

论著·基础研究

基于生物信息学方法的儿童急性不明确谱系白血病的潜在治疗靶基因筛选

吴任燕¹, 郭晓琳¹, 洪登礼¹, 陈磊²

1. 上海交通大学基础医学院病理生理学系, 细胞分化与凋亡教育部重点实验室, 上海 200025; 2. 上海交通大学医学院上海市免疫学研究所, 上海 200025

[摘要] **目的**·利用生物信息学分析方法探索儿童急性不明确谱系白血病 (acute leukemia of ambiguous lineage, ALAL) 致病通路、生存相关的独特基因表达谱及枢纽基因。**方法**·从 TARGET (Therapeutically Applicable Research to Generate Effective Treatment) 和 GEO (Gene Expression Omnibus) 数据库下载患者和健康人的表达数据, 利用 limma、clusterProfiler、survival 等生信工具对基因表达量进行差异分析、功能分析及生存分析, 最后通过构建蛋白-蛋白相互作用网络筛选得到与 ALAL 发病及生存相关的枢纽基因。**结果**·将 ALAL 组和健康对照组进行比较, 共筛选得到 4 053 个显著差异基因 (均 $P < 0.05$), 其中表达上调基因 1 844 个, 表达下调基因 2 209 个。利用 GO (Gene Ontology) 和 KEGG (Kyoto Encyclopedia of Genes and Genomes) 富集分析发现, 上调基因参与细胞周期及剪接, 下调基因参与免疫调节。通过与其他类型白血病的表达谱进行比较, 发现 ALAL 中生存相关基因呈现独特的表达模式。蛋白质-蛋白质相互作用网络显示生存基因网络的核心基因为 *CXCL8* (C-X-C motif chemokine ligand 8) 和 *LMNA* (lamin A/C)。**结论**·ALAL 的发病机制与细胞周期以及免疫相关, ALAL 预后不良可能与其生存基因的独特表达谱有关; *CXCL8* 和 *LMNA* 在 ALAL 中发挥重要作用, 可能作为潜在的治疗靶点; 这些结果对 ALAL 的机制研究和临床治疗具有提示作用。

[关键词] 生物信息学; 数据挖掘; 白血病; 生存分析

[DOI] 10.3969/j.issn.1674-8115.2021.03.006 **[中图分类号]** R733.73 **[文献标志码]** A

Identification of potential therapeutic target genes in pediatric acute leukemia of ambiguous lineage based on bioinformatics analysis

WU Ren-yan¹, GUO Xiao-lin¹, HONG Deng-li¹, CHEN Lei²

1. Key Library of Cell Differentiation and Apoptosis of Ministry of Education, Department of Pathophysiology, Shanghai Jiao Tong University College of Basic Medical Sciences, Shanghai 200025, China; 2. Shanghai Institute of Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[Abstract] **Objective**·To explore the pathogenetic pathways, unique gene expression profiles related to survival and hub genes in children with acute leukemia of ambiguous lineage (ALAL) by bioinformatics analysis. **Methods**·From TARGET (Therapeutically Applicable Research to Generate Effective Treatment) and GEO (Gene Expression Omnibus), the expression data of patients and healthy individuals were downloaded. The differential analysis of gene expression, as well as its function and survival analysis, were performed by bioinformatics tools, such as limma, clusterProfiler and survival. Finally, the hub genes of ALAL were screened by constructing protein-protein interaction network (PPI). **Results**·Four thousand and fifty-three significant differentially expressed genes were identified in the differential analysis between ALAL group and control group (all $P < 0.05$), of which 1 844 were up-regulated genes and 2 209 were down-regulated genes. GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis indicated that up-regulated genes were related to cell cycle and splicing, while the down-regulated genes were associated with immunity. By comparing the expression profiles with those of other types of leukemia, a unique expression pattern of survival-related genes in ALAL were found. Finally, PPI showed that *CXCL8* (C-X-C motif chemokine ligand 8) and *LMNA* (lamin A/C) were the hub genes of the survival gene network. **Conclusion**·The pathogenesis of ALAL is related to cell cycle and immunity. The poor prognosis of ALAL may be related to the unique expression profile of survival-related genes. *CXCL8* and *LMNA* play an important role in ALAL, and may act as potential therapeutic targets. These results have implications for the mechanism research and clinical treatment of ALAL.

[Key words] bioinformatics; data mining; leukemia; survival analysis

急性不明确谱系白血病 (acute leukemia of ambiguous lineage, ALAL) 是一种罕见的白血病类型, 不能清楚地分为淋巴系和髓系, 其发病率不到 5%。根据

2016 年世界卫生组织 (World Health Organization, WHO) 分类标准^[1], ALAL 可分为多个亚型, 包括急性未分化白血病 (acute undifferentiated leukemia, AUL)、

[作者简介] 吴任燕 (1995—), 女, 硕士生; 电子信箱: skye12138@sjtu.edu.cn。

[通信作者] 陈磊, 电子信箱: lei.chen@sjtu.edu.cn。

[Corresponding Author] CHEN Lei, E-mail: lei.chen@sjtu.edu.cn.

混合表型急性白血病 (mixed phenotype acute leukemia, MPAL) 等。尽管大多数白血病已经得到了很好的研究和治疗,但由于 ALAL 的罕见性,目前普遍采用的是与急性髓系白血病 (acute myeloid leukemia, AML) 和急性淋巴细胞白血病 (acute lymphoblastic leukemia, ALL) 相同的治疗手段,无法对 ALAL 进行针对性治疗,使得 ALAL 相较于 AML/ALL 治疗效果较差^[2-3]。因此探究 ALAL 发生和生存相关的基因和信号通路,有助于对 ALAL 发病机制的针对性研究,可为 ALAL 的临床治疗提供参考。

目前已有的研究主要针对 ALAL 的存活率、免疫分型以及相关 microRNA 而开展。白血病生存分析^[4-5]显示,大多数 ALAL 患者首选 ALL 的治疗方案;免疫表型分析^[6]显示,早期 T 细胞前体淋巴细胞白血病和 T/M-MPAL 在免疫表型上相似, *NOTCH1* 可作为 T 细胞混合表型白血病的潜在治疗靶点;通过高通量测序, ALAL 可以根据具有谱系特性的 microRNA 表达特性被归类为 AML 或 ALL^[7]。这些研究大多探讨了 ALAL 与其他白血病的谱系关系,但有关 ALAL 独特的基因表达谱、生存基因以及相关通路尚不清楚。高通量测序技术自问世以来,极大地推动了对于疾病标志物的研究进展。通过对公共数据库数据的生物信息学分析,我们可以对疾病的发病机制及治疗靶标进行预测,为疾病的机制研究和药物研发提供思路。

本研究利用小儿 ALAL 患者和健康对照组的表达谱数据 (RNA-seq), 对 ALAL 中的差异表达基因 (differential expressed genes, DEGs) 进行筛选,利用基因本体数据库 (Gene Ontology, GO) 和京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 富集分析,探讨 ALAL 发病的生物学过程和相关信号通路;结合生存分析与蛋白互作网络,我们发现了在 ALAL 中独特的生存基因表达谱,鉴定了与 ALAL 生存相关的枢纽基因。本研究可为 ALAL 的临床诊断和治疗提供新的思路。

1 材料与方法

1.1 RNA-seq 数据获得

基因表达数据来自于儿童 ALAL 患者和儿童健康对照者的血液或骨髓的 RNA-seq 数据。患者的基因表达数据是从 UCSC xena 数据门户 GDC 下载 TARGET-ALL-P3 数据集,该数据集由有效治疗应用研究 (Therapeutically Applicable Research to Generate Effective Treatments, TARGET, <https://ocg.cancer.gov/programs/TARGET>) 计划生成。健康对照者的基因表达数据 (GSE111459) 从基

因表达汇编数据库 (Gene Expression Omnibus, GEO) 下载。对异常样本进行过滤后,得到 21 个健康儿童样本和 69 个初诊儿童 ALAL 样本分别作为健康对照组和 ALAL 组。纳入诊断样本和复发样本的数据用于不同类型白血病基因表达量比较,以 TARGET-ALL-P2 的 532 个 ALL 样本为 ALL 组,以 TARGET-AML 的 187 个 AML 样本为 AML 组,以 TARGET-ALL-P3 中的 136 个 ALAL 患者样本为 ALAL 完全组。所有患者的数据均可从癌症基因组图谱 (The Cancer Genome Atlas, TCGA) 获得。

1.2 前期处理及 DEGs 筛选

所有样本的表达量数据均被标准化为每百万片段计数 (count per million, CPM) 值以进行相关性分析和去除低测序质量样本。低表达基因 (在所有样本中均 CPM<1.5) 和离群基因 ($<Q1-1.5\times IQR$ 或 $>Q3+1.5\times IQR$) 被移除。利用 Pearson 相关分析方法对样本表达谱相关性进行分析。R 语言 limma 软件包^[8]用于鉴别 ALAL 标本与健康对照组之间的差异表达基因。采用 Benjamini 和 Hochberg 修正法对多次试验进行修正,得到修正后的 *P* 值 (adjusted *P*-value)。为消除假阳性结果,DEGs 的筛选标准为 $|\log_2\text{Fold Change}| (|\log_2FC|) > 2$ 且修正后的 $P < 0.001$ 。

1.3 DEGs 的功能分析

为验证数据结果的可靠性,利用 R 语言 DOSE 软件包^[9]进行 DEGs 的疾病本体论 (Disease Ontology, DO)^[10]分析。为研究 ALAL 中的相关通路,使用 clusterProfiler^[11]的默认参数对上调和下调的 DEGs 分别进行 GO 和 KEGG 富集分析。显著性筛选标准为 $P < 0.05$ 。

1.4 生存相关基因鉴定及其与 AML/ALL 的表达比较

69 例 ALAL 患者中有 48 个病例有总生存时间记录,并选择用于生存相关基因的鉴定。根据每个基因在所有样本中的平均表达量,48 个样本被分为基因的高表达组和低表达组。生存分析采用 R 语言的 survival 和 survminer 软件包进行, Cox 回归模型用于鉴定生存相关基因;以 $P < 0.05$,在表达下调的基因中相关系数 > 0 或在表达上调的差异基因中相关系数 < 0 ,被认为是合理的 ALAL 生存相关基因。生存分析是针对与正常样本的差异基因进行,所得到的生存相关基因可能与小儿 ALAL 的致病相关,但这些生存相关基因的具体作用仍有待进一步的机制研究。

1.5 蛋白质-蛋白质相互作用网络构建及枢纽基因筛选

将生存相关基因导入 GeneMANIA^[12]构建蛋白质-蛋

白质相互作用网络 (protein-protein interaction network, PPI), 该网络中总共包含 22 个 ALAL 生存相关基因以及 100 个与这些基因具有功能或表达相关性的间接相关基因。利用 Cytoscape^[13] 绘制 PPI 网络图谱, 并利用 Cytohubba^[14] 计算网络图中每一个基因的最大群体中心度 (maximal clique centrality, MCC), 得分值最高的前 10 个基因, 被认为是 PPI 网络中的枢纽基因。

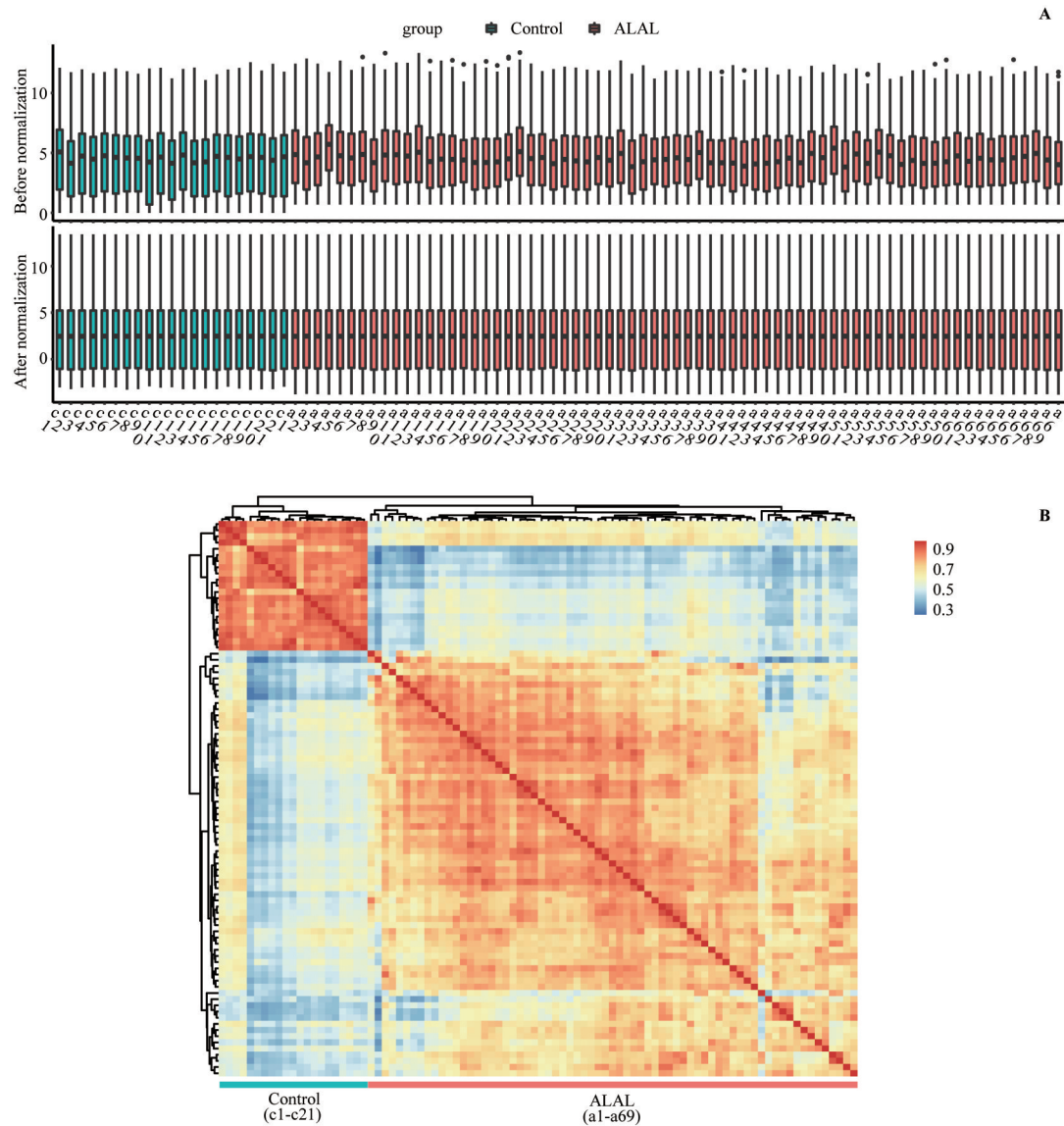
1.6 统计学方法

生存相关基因在不同类白血病及正常人中的表达量以 $\bar{x}\pm s$ 表示, 采用 R 语言 stats 4.0.2 软件包对基因表达量进行统计分析, 利用 Wilcoxon 秩和检验对基因表达的相关性进行评估。 $P<0.05$ 表示差异具有统计学意义。

2 结果

2.1 数据质控

由于测序技术的特性, 我们需要对不同来源的测序数据进行质控, 以提高最后分析结果的准确性。在去除低质量样本和离群基因后, 在原始数据中筛选出总共 90 个样本的 19 213 个基因的表达值进行后续分析, 其中包括 21 个健康对照组和 69 个儿童 ALAL 组的样本。标准化前后的表达值如图 1A 所示。在 ALAL 组和对照组中的基因表达量显示出较高的内部样本相关性 (图 1B), 证明筛选得到的数据内部表达谱具有均质性, 可以用于后续分析。样本编号与原始数据编号的对应关系见附表 1。



Note: A. Expression values before and after normalization; columns represent sample ID and rows represent gene expression values; the sample ID corresponding to the original material is in supplementary table 1. B. Correlation of gene expression profiles in all samples; color scale bar represents the correlation coefficient.

图1 样本数据质控

Fig 1 Quality control of sample data

表1 生存相关基因在不同类白血病患者及健康对照者中的表达值(log₂CPM)

Tab 1 The log₂CPM value of survival-related genes in the control and different leukemia types

Gene	Control group (N=21)	ALAL complete group (N=136)	AML group (N=187)	ALL group (N=532)
<i>CXCL8</i>	1.356±0.574	4.054±1.972 ^{①⑧⑩}	6.265±1.578 ^④	2.935±1.578 ^⑥
<i>SCHIP1</i>	0.092±0.138	1.606±1.312 ^{①⑧⑩}	0.912±0.939 ^④	1.175±1.288 ^⑥
<i>SLC1A2</i>	0.145±0.180	1.525±0.917 ^{①⑧⑩}	0.459±0.445 ^④	1.017±0.810 ^⑥
<i>CYGB</i>	0.269±0.253	1.566±1.370 ^{①⑩}	0.946±0.916 ^④	2.144±2.090 ^⑥
<i>GREM1</i>	0.158±0.269	1.154±1.434 ^{①⑩}	0.579±0.622 ^④	1.046±1.200 ^⑥
<i>PER1</i>	3.024±0.237	4.868±1.143 ^{①⑩}	5.807±0.697 ^④	5.075±0.985 ^⑥
<i>SMARCA5-ASI</i>	0.591±0.519	1.522±0.501 ^{①⑨⑩}	0.331±0.251	1.277±0.818 ^⑥
<i>LMNA</i>	2.208±0.444	3.760±1.355 ^{①⑧⑩}	4.993±1.025 ^④	3.225±1.196 ^⑥
<i>TMEM52B</i>	0.062±0.124	0.760±0.848 ^{①⑧⑩}	1.043±0.734 ^④	0.185±0.274 ^⑥
<i>RNU6-1208P</i>	0	0.581±0.408 ^{①⑩⑩}	0.153±0.187 ^④	0.495±0.416 ^⑥
<i>RPL23AP23</i>	0.643±0.416	1.987±0.924 ^{①⑩⑩}	0.293±0.277 ^④	1.700±1.086 ^⑥
<i>PDE6H</i>	0.073±0.120	0.678±0.669 ^{①⑧⑩}	0.472±0.538 ^④	0.172±0.225 ^⑥
<i>CCNL1</i>	5.680±0.177	6.811±0.683 ^{①⑩}	6.358±0.964 ^④	6.764±0.776 ^⑥
<i>LILRA4</i>	2.196±0.528	0.743±0.621 ^{①⑧⑩}	0.489±0.526 ^④	0.603±0.644 ^⑥
<i>BACH2</i>	5.090±0.433	3.240±1.276 ^{①⑧⑩}	2.044±1.017 ^④	4.010±1.469 ^⑥
<i>RHOXF2B</i>	0.573±0.491	0.075±0.052 ^{②⑧⑩}	0.049±0.172 ^⑤	0.010±0.031 ^⑥
<i>LEF1</i>	5.597±0.503	3.692±1.471 ^{①⑧⑩}	2.627±1.312 ^④	5.798±0.902 ^⑦
<i>SLC7A8</i>	1.664±0.440	0.729±0.521 ^{①⑧⑩}	1.128±0.818 ^④	0.505±0.506 ^⑥
<i>ABCA11P</i>	1.756±0.450	0.698±0.477 ^{①②⑩}	0.163±0.166 ^④	0.533±0.361 ^⑥
<i>AMIGO3</i>	1.116±0.721	0.364±0.186 ^{①⑧⑩}	0.108±0.083 ^④	0.262±0.175 ^⑥
<i>RPSAP41</i>	0.495±0.376	0.188±0.135 ^{①⑧⑩}	0.046±0.050 ^④	0.144±0.151 ^⑥
<i>ZNF853</i>	2.311±0.355	1.074±0.597 ^{①⑧⑩}	0.923±0.623 ^④	1.541±0.638 ^⑥
<i>LMOD1</i>	0.737±0.640	0.378±0.409 ^{②⑧⑩}	0.105±0.227 ^④	0.294±0.372 ^⑥
<i>BTN1A1</i>	0.781±0.358	0.349±0.251 ^{①⑧⑩}	0.087±0.175 ^④	0.322±0.418 ^⑥
<i>LINC01355</i>	3.194±0.326	1.946±1.140 ^{①⑧⑩}	0.393±0.396 ^④	1.099±0.791 ^⑥
<i>ZNF135</i>	2.485±0.434	1.566±1.173 ^{①⑧⑩}	0.559±0.524 ^④	2.008±1.203
<i>ST6GALNAC4P1</i>	0.780±0.471	0.397±0.257 ^{①⑩⑩}	0.169±0.159 ^④	0.349±0.279 ^⑥
<i>ZNF582-ASI</i>	1.500±0.319	0.723±0.351 ^{①⑩}	0.554±0.311 ^④	0.688±0.356 ^⑥
<i>PABPC1P7</i>	0.863±0.308	0.503±0.336 ^{①⑧⑩}	0.089±0.093 ^④	0.338±0.267 ^⑥
<i>GBP6</i>	0.573±0.394	0.347±0.299 ^{③⑧⑩}	0.060±0.094 ^④	0.153±0.188 ^⑥
<i>GOLGA6L9</i>	2.353±0.300	1.310±0.630 ^{①⑩}	0.791±0.544 ^④	1.400±0.736 ^⑥

Note: ^① $P=0.000$, ^② $P=0.003$, ^③ $P=0.002$, compared with the control; ^④ $P=0.000$, ^⑤ $P=0.004$, compared with the control; ^⑥ $P=0.000$, ^⑦ $P=0.017$, compared with the control; ^⑧ $P=0.000$, ^⑨ $P=0.050$, ^⑩ $P=0.009$, ^⑪ $P=0.002$, ^⑫ $P=0.001$, compared with the ALL; ^⑬ $P=0.000$, ^⑭ $P=0.008$, ^⑮ $P=0.007$, compared with the AML. *CXCL8*—C-X-C motif chemokine ligand 8; *SCHIP1*—schwannomin interacting protein 1; *SLC1A2*—solute carrier family 1 member 2; *CYGB*—cytoglobin; *GREM1*—gremlin 1, DAN family BMP antagonist; *PER1*—period circadian regulator 1; *SMARCA5-ASI*—SMARCA5 antisense RNA 1; *LMNA*—lamin A/C; *TMEM52B*—transmembrane protein 52B; *RNU6-1208P*—RNA, U6 small nuclear 1208, pseudogene; *RPL23AP23*—ribosomal protein L23a pseudogene 23; *PDE6H*—phosphodiesterase 6H; *CCNL1*—cyclin L1; *LILRA4*—leukocyte immunoglobulin like receptor A4; *BACH2*—BTB domain and CNC homolog 2; *RHOXF2B*—RhoX homeobox family member 2B; *LEF1*—lymphoid enhancer binding factor 1; *SLC7A8*—solute carrier family 7 member 8; *ABCA11P*—ATP binding cassette subfamily A member 11, pseudogene; *AMIGO3*—adhesion molecule with Ig like domain 3; *RPSAP41*—ribosomal protein SA pseudogene 41; *ZNF853*—zinc finger protein 853; *LMOD1*—leiomodin 1; *BTN1A1*—butyrophilin subfamily 1 member A1; *LINC01355*—long intergenic non-protein coding RNA 1355; *ZNF135*—zinc finger protein 135; *ST6GALNAC4P1*—ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 4 pseudogene 1; *ZNF582-ASI*—ZNF582 divergent transcript; *PABPC1P7*—poly(A) binding protein cytoplasmic 1 pseudogene 7; *GBP6*—guanylate binding protein family member 6; *GOLGA6L9*—golgin A6 family like 9.

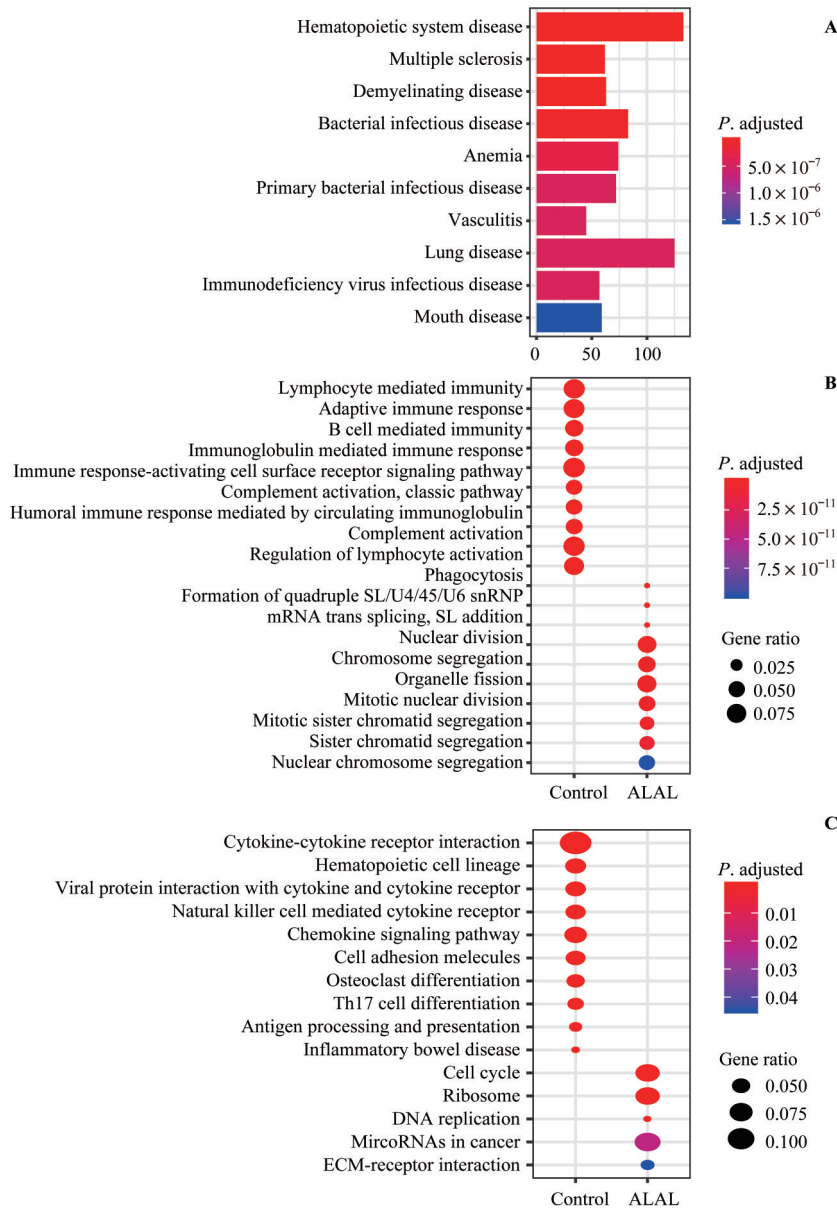
2.2 DEGs 筛选与功能分析

为进一步探究儿童ALAL的异常基因表达谱,利用limma软件包对ALAL组和健康对照组的基因表达谱进行差异分析;选择 $|\log_2FC|>2$ 、校正后的 $P<0.001$ 为条件进行基因筛选,去除假阳性基因,共鉴定得到4 053个

基因,其中上调基因1 844个,下调基因2 209个。基于所有DEGs的DO分析表明造血系统疾病是最显著富集的疾病(图2A),与我们的数据特征相符,进一步证明我们筛选使用的样本数据和差异分析的合理性。对上调和下调的差异基因分别进行GO和KEGG富集分

析, 找寻与 ALAL 发病相关的信号通路, 各组显著富集的前 10 个生物学过程和信号通路如图 2B 及图 2C 所示。ALAL 组中与免疫相关的生物学进程、造血细胞谱系和

细胞黏附分子相关的信号通路明显下调, 而细胞周期、有丝分裂以及剪接相关的生物学进程在 ALAL 组中明显富集。



Note: A. Disease ontology enriched in DEGs; columns presenting gene count. B. Biological progresses enriched in the control and ALAL samples; adaptive immune response—adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains. C. KEGG pathways enriched in the control and ALAL samples. snRNP—small nuclear ribonucleoprotein; SL—trans-splicing; Th17—interleukin-17-producing CD4⁺ effector T cell; ECM—extracellular matrix.

图2 差异基因的功能分析

Fig 2 Functional analysis of DEGs

2.3 生存相关基因筛选及表达谱的比较

利用 ALAL 患者的总体生存期资料, 我们从 4 053 个差异表达基因中鉴定出 31 个 ALAL 生存相关基因。其中 13 个基因在 ALAL 组中表达量增加, 与生存呈负相关, 被认为是负生存相关基因; 18 个基因在 ALAL 组中表达降低, 与生存呈正相关, 被认为是正生存相关基因。

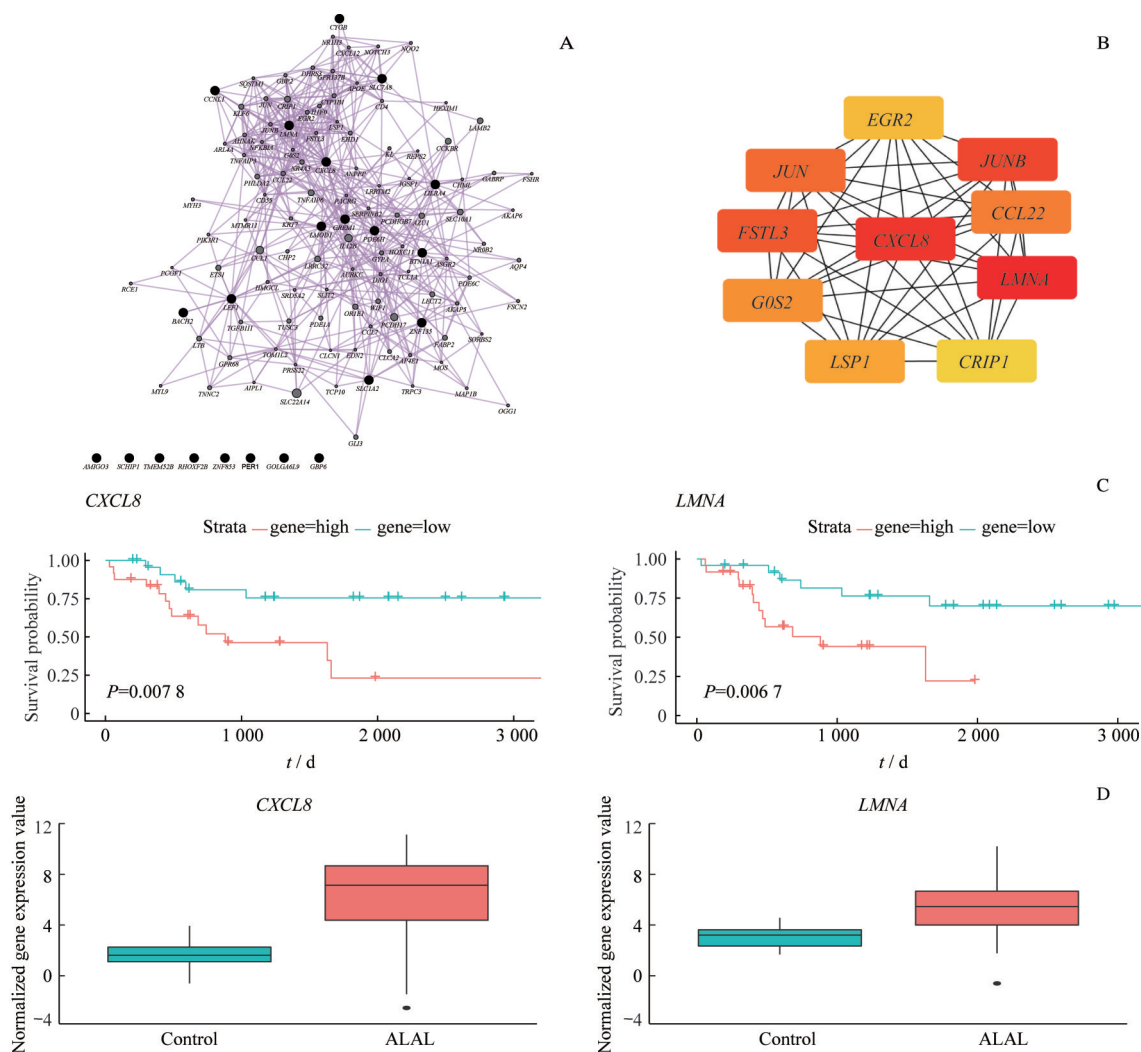
此外, 为了研究这些基因是特异在 ALAL 中表达变化还是白血病的普遍差异基因, 我们在 ALL 组以及 AML 组的表达谱中对这些基因进行了比较分析。结果表明, 与 ALAL 生存相关的基因, 在 ALL 组以及 AML 组中呈现出同样的差异表达趋势, 相对于健康组, 在 ALAL 完全组中表达上调的基因同样在 ALL 组和 AML 组中表达上

调,但这些基因的表达量在各类型白血病之间仍然存在显著差异(表1)。这些基因中也有一些“偏向性”表达特例,*SMARCA5-AS1*在AML组中的表达与对照组相比差异不具有统计学意义($P=0.151$),但在ALAL完全组和ALL组中表达较对照组有明显增加(均 $P=0.000$)。*ZNF135*在ALAL完全组和AML组中的表达值均低于对照组,且差异具有统计学意义(均 $P=0.000$),但在ALL患者中表达量无明显变化($P=0.073$)。这些数据为研究ALAL的独特生存特征及治疗方案的选择提供了线索。

2.4 PPI网络构建以及枢纽基因识别

枢纽基因是在PPI网络中具有紧密的相互联系,发挥核心作用的一群基因。为了进一步筛选出在ALAL患者

生存中发挥关键作用的枢纽基因,我们将所有生存相关基因导入到GeneMANIA中,构建了PPI网络(图3A)。GeneMANIA中的GO富集分析显示细胞趋化性和白细胞迁移相关的生物学过程在这个相互作用网络中显著富集,提示ALAL患者的生存与细胞趋化性及白细胞迁移相关。在这个PPI网络中MCC得分最高的前10个基因被认为是这个PPI网络中的枢纽基因(图3B)。在这10个基因中,只有*CXCL8*和*LMNA*与ALAL的生存相关且在ALAL组中表达量发生变化,被鉴定为ALAL的生存相关枢纽基因。在ALAL组和健康对照组中*CXCL8*和*LMNA*的生存曲线和表达谱如图3C和图3D所示,*CXCL8*和*LMNA*均为负生存相关基因,在ALAL组中表达上调,伴随ALAL患者总生存期下降。



Note: A. PPI of survival-related genes; black nodes present survival-related genes. B. Top 10 hub genes in PPI network; color presents the interaction density, from yellow to red, indicating that the genes have more interaction with others. C. Survival curves of *CXCL8* and *LMNA* in the ALAL group. D. Expression of *CXCL8* and *LMNA* in the control and ALAL groups. *EGR2*—early growth response 2; *JUN*—jun proto-oncogene, AP-1 transcription factor subunit; *FSTL3*—follistatin like 3; *JUNB*—junb proto-oncogene, AP-1 transcription factor subunit; *CCL22*—C-C motif chemokine ligand 22; *G0S2*—G0/G1 switch 2; *LSP1*—lymphocyte specific protein 1; *CRIP1*—cysteine rich protein 1.

图3 ALAL生存相关枢纽基因鉴定

Fig 3 Survival-related hub genes identification and its character.

3 讨论

本研究通过生物信息学分析的方法,将基因表达数据与临床生存数据结合,对ALAL的生存相关枢纽基因进行筛选。为使得研究结果更加准确,我们首先对公共数据库的数据进行了筛选,得到21个健康样本(GSE111459)和69个ALAL样本(TARGET-ALL-P3)的共19 213个基因表达数据。基因表达谱的差异分析显示在ALAL组中有1 844个基因表达上调,2 209个基因表达下调,揭示了ALAL的异常基因表达谱。疾病本体分析的结果显示造血系统疾病在差异基因中高度富集,进一步验证了我们的数据以及差异分析的可靠性。GO以及KEGG富集分析表明在ALAL组中上调表达的基因主要富集细胞周期以及有丝分裂的相关进程,而下调的差异基因在造血细胞谱系相关通路和免疫相关进程中富集。白血病的主要特征为造血干细胞的恶性增殖和分化障碍,因此细胞周期及有丝分裂进程的上调,造血细胞谱系相关进程的下调,符合白血病的基本特性。但目前尚未有研究指出ALAL中免疫相关进程的下调,因此本研究找到的通路可能为ALAL的免疫学治疗提供理论依据。

结合临床生存记录,我们从ALAL的差异基因中筛选出了31个与生存相关的基因。由于目前ALAL缺乏明确的靶向治疗方案,ALAL的治疗普遍是采用ALL或者AML的治疗方案进行,但这些治疗往往带来较差的预后效果^[2-3],因此我们将ALAL的生存相关基因表达谱与ALL/AML进行比较,探究特异在ALAL中异常表达的生存基因。结果显示大多数基因在ALAL、AML和ALL中具有共同的上下调趋势,这些ALAL的生存相关基因,在AML/ALL中同样差异表达,具有白血病致病的普遍性。但这些基因在各个白血病类型之间的表达谱存在明显差异,ALAL的生存相关基因表达量大多数介于ALL和AML之间,表现出与ALAL生存相关的特殊基因表达谱。此外我们发现,在这些基因中存在一些明显的“偏向”调节的情况,如*SMARCA5-AS1*在ALAL和ALL中表达增加,但在AML中没有改变;*ZNF135*在ALAL和

AML中表达下调,但在ALL中没有改变。这些数据可能为ALAL治疗方案的选择提供借鉴作用。

此外,通过PPI网络图谱的构建,我们进一步筛选出*CXCL8*和*LMNA*作为ALAL生存相关的枢纽基因。*CXCL8*又称*IL-8*,是CXC趋化因子家族的成员。*CXCL8*可从白细胞和非白细胞的体细胞中释放,在炎症中经常上调表达,并在白细胞诱导迁移和释放反应中发挥作用^[15]。近期的研究表明,白血病细胞广泛释放*CXCL*和*CCL*趋化因子,尤其是高水平地释放*CXCL8*^[16];*CXCL8*与AML的复发有关,*CXCL8*的敲除导致G0/G1细胞周期阻滞、凋亡和细胞外调节蛋白激酶1/2信号通路失活^[17];*CXCL8*可与*CXCR1/2*相互作用,调节细胞增殖和分化,介导肿瘤的发生和发展^[18-19]。然而,*CXCL8*在ALAL中是否发挥作用尚不清楚。在本研究中,*CXCL8*在ALAL中表达上调;并在ALAL患者的生存中充当枢纽基因,ALAL患者中富集与细胞周期相关的信号通路,*CXCR1/2*在ALAL中表达下调,结合已有报道,我们推测*CXCL8*可能通过参与细胞周期的进程介导了ALAL的发生和发展,而并非通过与*CXCR1/2*相互作用参与ALAL的发生。*LMNA*编码核包膜蛋白lamin A和lamin C,主要参与染色质组织、核组装和端粒动力学,与涉及凋亡和存活的Caspase级联途径有关。*LMNA*突变引起多种疾病^[20-21],但在白血病中对*LMNA*的研究非常少。此前,有报道称*LMNA*在活性T细胞中被诱导表达^[22],但*LMNA*在造血细胞中的表达情况尚不明确。通过高通量测序数据,我们在对照组和ALAL组中均检测到*LMNA*,并且*LMNA*在ALAL患者中的表达明显高于对照组。我们的结果首次提出*LMNA*在ALAL的发生和生存中发挥作用。

综上所述,我们对儿童ALAL病例的独特基因表达谱进行了研究,确定了参与ALAL发生的信号通路及生物学进程,并将ALAL生存相关基因的表达谱与ALL和AML的表达谱进行了比较,为研究ALAL的独特生存特征和现有治疗方案的选择提供了线索。基于生存期分析找到了ALAL生存相关枢纽基因,可能作为潜在的ALAL治疗靶点,但这些枢纽基因的功能仍有待进一步验证。

参·考·文·献

- [1] Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia[J]. Blood, 2016, 127(20): 2391-2405.
- [2] Slany RK. The molecular mechanics of mixed lineage leukemia[J]. Oncogene, 2016, 35(40): 5215-5223.
- [3] Guru Murthy GS, Dhakal I, Lee JY, et al. Acute leukemia of ambiguous lineage in elderly patients-analysis of survival using surveillance epidemiology and end results-medicare database[J]. Clin Lymphoma Myeloma Leuk, 2017, 17(2): 100-107.
- [4] Hrusak O, de Haas V, Stancikova J, et al. International cooperative study identifies treatment strategy in childhood ambiguous lineage leukemia[J]. Blood, 2018, 132(3): 264-276.
- [5] Lee HG, Baek HJ, Kim HS, et al. Biphenotypic acute leukemia or acute leukemia of ambiguous lineage in childhood: clinical characteristics and

- outcome[J]. *Blood Res*, 2019, 54(1): 63-73.
- [6] Noronha EP, Marques LVC, Andrade FG, et al. T-lymphoid/myeloid mixed phenotype acute leukemia and early T-cell precursor lymphoblastic leukemia similarities with *NOTCH1* mutation as a good prognostic factor[J]. *Cancer Manag Res*, 2019, 11: 3933-3943.
- [7] de Leeuw DC, van den Ancker W, Denkers F, et al. MicroRNA profiling can classify acute leukemias of ambiguous lineage as either acute myeloid leukemia or acute lymphoid leukemia[J]. *Clin Cancer Res*, 2013, 19(8): 2187-2196.
- [8] Ritchie ME, Phipson B, Wu D, et al. Limma Powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Res*, 2015, 43(7): e47.
- [9] Yu G, Wang LG, Yan GR, et al. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis[J]. *Bioinformatics*, 2015, 31(4): 608-609.
- [10] Schriml LM, Arze C, Nadendla S, et al. Disease Ontology: a backbone for disease semantic integration[J]. *Nucleic Acids Res*, 2012, 40(database issue): D940-D946.
- [11] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5): 284-287.
- [12] Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function[J]. *Nucleic Acids Res*, 2010, 38(web server issue): W214-W220.
- [13] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. *Genome Res*, 2003, 13(11): 2498-2504.
- [14] Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. *BMC Syst Biol*, 2014, 8(Suppl 4): S11.
- [15] Baggiolini M, Dewald B, Moser B. Human chemokines: an update [J]. *Annu Rev Immunol*, 1997, 15: 675-705.
- [16] Kittang AO, Hatfield K, Sand K, et al. The chemokine network in acute myelogenous leukemia: molecular mechanisms involved in leukemogenesis and therapeutic implications[J]. *Curr Top Microbiol Immunol*, 2010, 341: 149-172.
- [17] Li Y, Cheng J, Li Y, et al. CXCL8 is associated with the recurrence of patients with acute myeloid leukemia and cell proliferation in leukemia cell lines[J]. *Biochem Biophys Res Commun*, 2018, 499(3): 524-530.
- [18] Liu Q, Li A, Tian Y, et al. The CXCL8-CXCR1/2 pathways in cancer[J]. *Cytokine Growth Factor Rev*, 2016, 31: 61-71.
- [19] Kamohara H, Takahashi M, Ishiko T, et al. Induction of interleukin-8 (CXCL-8) by tumor necrosis factor- α and leukemia inhibitory factor in pancreatic carcinoma cells: impact of CXCL-8 as an autocrine growth factor[J]. *Int J Oncol*, 2007, 31(3): 627-632.
- [20] Bonne G, Di Barletta MR, Varnous S, et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy[J]. *Nat Genet*, 1999, 21(3): 285-288.
- [21] Raffaele Di Barletta M, Ricci E, Galluzzi G, et al. Different mutations in the LMNA gene cause autosomal dominant and autosomal recessive Emery-Dreifuss muscular dystrophy[J]. *Am J Hum Genet*, 2000, 66(4): 1407-1412.
- [22] Rocha-Perugini V, González-Granado JM. Nuclear envelope lamin-A as a coordinator of T cell activation[J]. *Nucleus*, 2014, 5(5): 396-401.

[收稿日期] 2020-05-12

[本文编辑] 张慧俊

学术快讯

上海交通大学基础医学院余健秀研究组发现 m⁶A 修饰阅读蛋白 YTHDF2 功能调控的新分子机制

2021年2月12日,上海交通大学基础医学院余健秀研究组在 *Nucleic Acids Research* 杂志上在线发表论文“SUMOylation of YTHDF2 promotes mRNA degradation and cancer progression by increasing its binding affinity with m⁶A-modified mRNAs”。该研究揭示了 SUMO 化修饰是 m⁶A 修饰阅读蛋白 YTHDF2 功能调控的一个新的分子机制,并阐明了其在肿瘤发生中的重要作用。该研究组一直致力于与肿瘤相关的蛋白质修饰和 RNA 修饰功能机制的研究,近年来取得了多项创新性成果。