Advanced Genetics PDMSA：一款基于DNA甲基化水平的泛癌生存分析网络工具

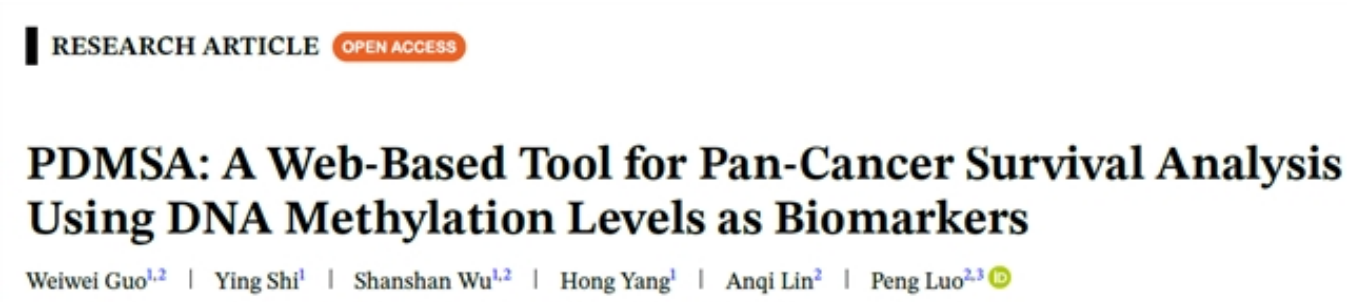
作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/41895.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

Advanced Genetics PDMSA：一款基于DNA甲基化水平的泛癌生存分析网络工具。

大家好，今天与大家分享我们课题组近期在 Advanced Genetics 上发表的一篇文章，题目为 PDMSA: A web-based tool for pan-cancer survival analysis using DNA methylation levels as biomarkers。


RESEARCH ARTICLE OPEN ACCESS

PDMSA: A Web-Based Tool for Pan-Cancer Survival Analysis Using DNA Methylation Levels as Biomarkers

Weiwei Guo[1,2] | Ying Shi[1] | Shanshan Wu[1,2] | Hong Yang[1] | Anqi Lin[2] | Peng Luo[2,3]

摘要

DNA甲基化水平与肿瘤的发生、发展及治疗效果密切相关。准确分析DNA甲基化水平与肿瘤预后之间的关系，有助于深入探究肿瘤发生机制，优化临床决策，从而提升癌症患者的生存率。然而，目前用于分析肿瘤甲基化水平与生存预后的网络工具仍存在显著局限性。为此，我们开发了一款基于Shiny框架的网络工具——PDMSA（Pan-cancer DNA Methylation Survival Analysis），整合了来自TCGA和GEO数据库的DNA甲基化数据及临床信息。PDMSA目前涵盖30个TCGA数据集和15个GEO数据集，共计16,205,211条记录，覆盖39种癌症类型、45个数据集、19,909个基因和8,369个样本。该工具支持Kaplan-Meier生存分析与Cox回归分析，并提供两种分组方式（中位数分组与最佳截断值分组），用户可自定义可视化结果。PDMSA旨在为生物医学研究者提供一个友好、全面的分析平台，助力肿瘤精准医学的发展。

访问地址：robinl-lab.com/PDMSA

方法概述

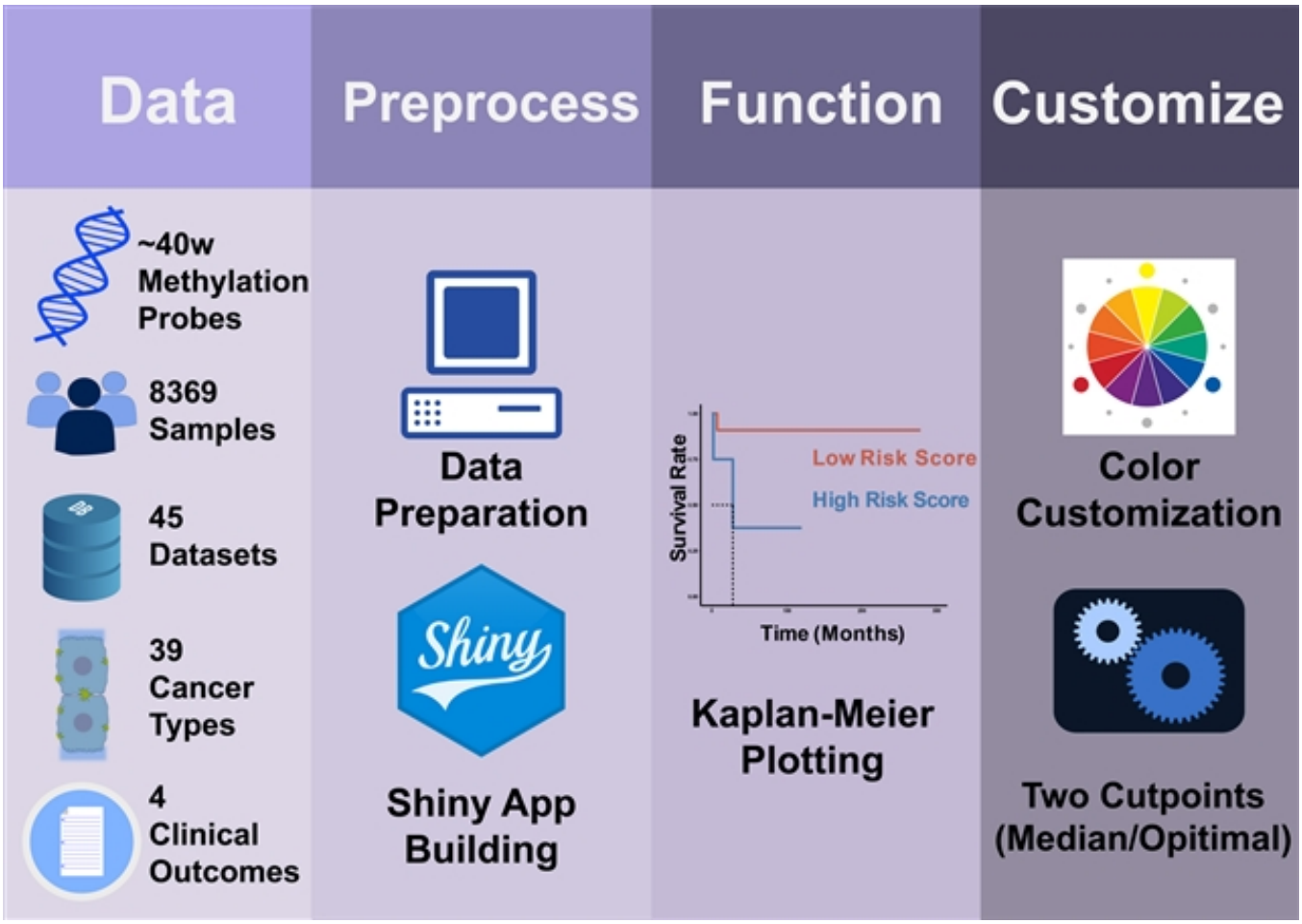
PDMSA整合了TCGA和GEO中基于Illumina 450K或EPIC平台的甲基化数据，筛选标准包括：

- 样本量 ≥ 20；
- 包含完整的生存信息；
- 数据经过标准化处理。

生存分析采用Kaplan-Meier曲线和Cox回归模型，支持中位数分组与最佳截断值分组，并可根据性别、年龄、种族、病理分期等因素进行多变量校正。此外，PDMSA还引入了Benjamini-Hochberg FDR校正，以控制多重检验带来的假阳性风险。

核心功能模块详解

PDMSA的用户界面围绕六大功能模块构建，每个模块都针对甲基化生存分析的一个特定环节进行了优化设计，为研究者提供从初步探索到深度挖掘的全流程支持。



GENE OVERVIEW：跨癌种甲基化位点全景扫描

为帮助用户直观比较同一基因在不同癌症类型中的甲基化位点分布差异，我们设计了Gene Overview模块。该模块基于预计算的多癌种生存分析结果，生成跨癌种甲基化位点分布棒棒糖图。

用户可首先根据风险类型（高风险，HR > 1；或低风险，HR < 1）筛选基因。系统会根据筛选条

件，动态更新基因列表，并展示每个基因在所有癌症中具有显著预后意义的探针总数。

选择目标基因后，PDMSA调用Ensembl数据库进行基因组坐标注释。图表以转录起始位点（红色方块）和基因体（黑色方块）为锚点，X轴代表探针的精确基因组位置。

用户可以通过交互式参数设置，自定义显示最显著的Top
N个探针。系统会根据探针在染色体上的分布情况，自动调整图表布局。该模块支持PDF格式。

PDMSA

HOME | GENE OVERVIEW | KM PLOT! | COX FOREST! | FDR CORRECTION | DATA | ABOUT

STEP1: Select Data to Be Processed

Do you want to explore high-risk or low-risk genes?

- (•) High HR (>1)
- ( ) Low HR (<1)

Select a gene symbol: PRDM16

Search to find more genes, or you can find the complete list of available genes in the Data tab.

Number of Top Probes: 30

ANALYZE!

Gene: PRDM16

Risk type: HR>1

Total Probes: 253

Important Notice

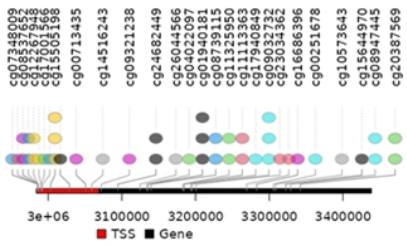
For screening analyses across multiple genes or cancer types, users are advised to perform FDR correction on candidate biomarkers during subsequent validation stages.

STEP2: View, Customize and Download Your Plot

Note: This feature is still under development. If you encounter any issues, please feel free to contact us via the ABOUT page.

Gene: **PRDM16**
Risk type: **High-risk (HR > 1)**
Top probes shown: **30** out of **253** available probes
Cancer types covered: **Multiple cancer types**
Datasets analyzed: **TCGA + GEO**
P-value range: **p < 0.05**

DOWNLOAD PDF



PLOT INFORMATION >>

Plot Interpretation:
1. Each lollipop represents a CpG probe in the **PRDM16** gene
2. The position on the x-axis shows the genomic location
3. The height of the lollipop indicates how many cancer types show significant association
4. The red block represents the Transcription Start Site (TSS)
5. The black block represents the gene body
6. Probes are sorted by p-value (smallest to largest)

Data Summary:
Total probes analyzed: 253
Probes displayed: 30
Cancer types: Multiple cancer types
Datasets: TCGA + GEO
P-value range: p < 0.05

KM PLOT：核心生存分析与可视化定制

在获得甲基化位点全景扫描后，用户可使用此模块探究单个甲基化位点与生存结局的关联。

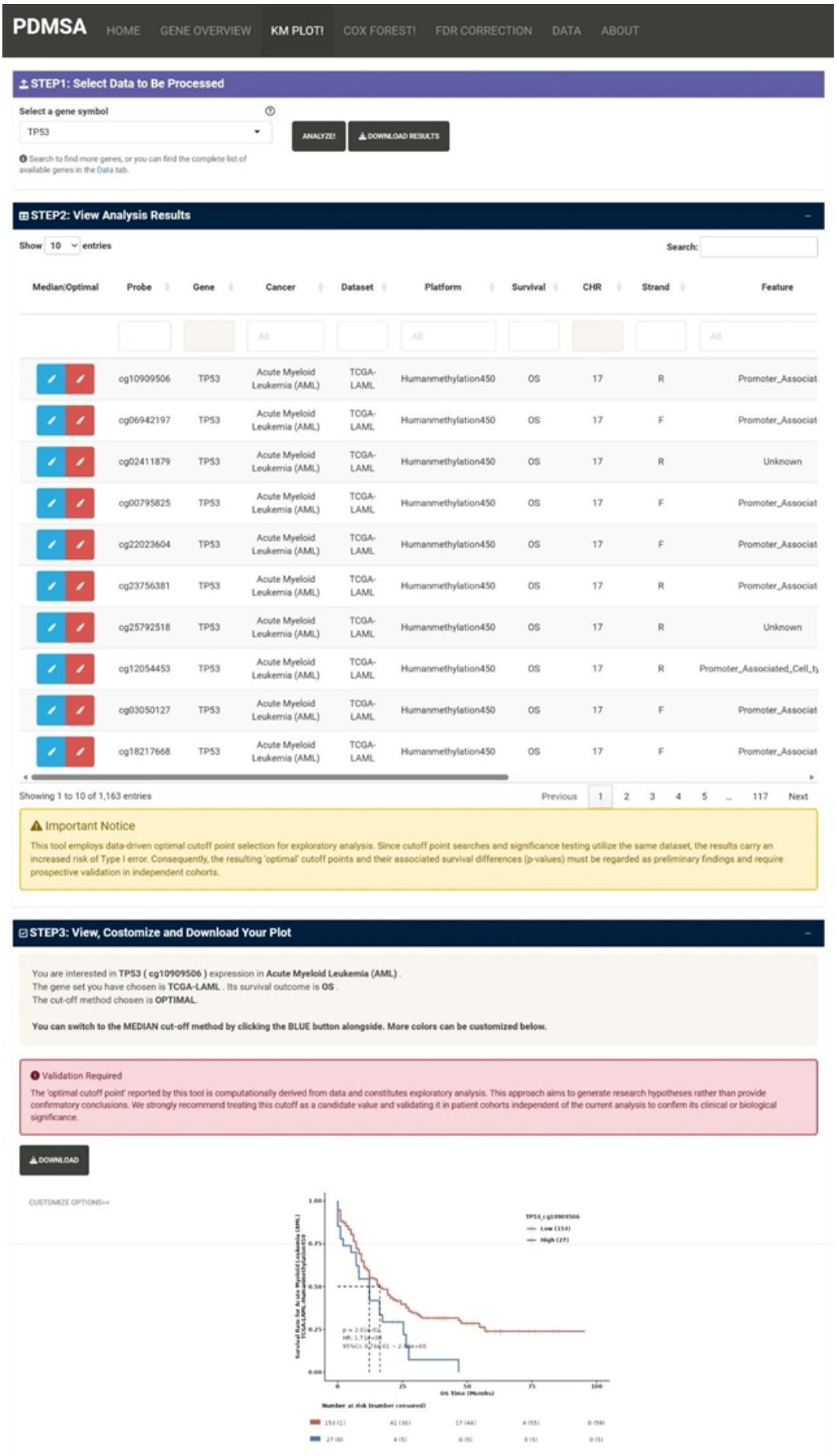
用户可以在预设的基因和对应的甲基化探针中进行选择。PDMSA支持多种生存终点，包括总生存期（OS）、无进展生存期（PFS）、无远处转移生存期（DMFS）和无转移生存期（MFS）。

本模块提供两种将连续甲基化水平转化为分组变量的方法：

1. 中位数分组：快速、稳定，适合进行大规模的初步筛查。

2. 最佳截断值分组：通过统计检验寻找能最大程度区分生存差异的阈值，更精准地识别预后不同的患者群体。（注：此方法为数据驱动，结果需视为探索性发现，并在独立队列中验证。）

用户可以自由定制Kaplan-Meier曲线的颜色，使图表符合发表要求。所有生存分析结果表格（包括探针ID、基因名、癌症类型、HR、置信区间、P值等）均可一键保存为CSV、Excel或PDF格式，结果图也可直接导出为PDF。


PDMSA
HOME
GENE OVERVIEW
KM PLOT!
COX FOREST!
FDR CORRECTION
DATA
ABOUT
STEP1: Select Data to Be Processed
Select a gene symbol
TP53
ANALYZE!
DOWNLOAD RESULTS
Search to find more genes, or you can find the complete list of available genes in the Data tab.
STEP2: View Analysis Results
Show 10 entries
Search:
Median/Optimal
Probe
Gene
Cancer
Dataset
Platform
Survival
CHR
Strand
Feature
cg10909506 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Promoter_Associat
cg06942197 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg02411879 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Unknown
cg00795825 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg22023604 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg23756381 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Promoter_Associat
cg25792518 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Unknown
cg12054453 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Promoter_Associated_Cell_t
cg03050127 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg18217668 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
Showing 1 to 10 of 1,163 entries
Previous 1 2 3 4 5 ... 117 Next
Important Notice
This tool employs data-driven optimal cutoff point selection for exploratory analysis. Since cutoff point searches and significance testing utilize the same dataset, the results carry an increased risk of Type I error. Consequently, the resulting 'optimal' cutoff points and their associated survival differences (p-values) must be regarded as preliminary findings and require prospective validation in independent cohorts.
STEP3: View, Costomize and Download Your Plot
You are interested in TP53 (cg10909506) expression in Acute Myeloid Leukemia (AML) .
The gene set you have chosen is TCGA-LAML . Its survival outcome is OS .
The cut-off method chosen is OPTIMAL.
You can switch to the MEDIAN cut-off method by clicking the BLUE button alongside. More colors can be customized below.
Validation Required
The 'optimal cutoff point' reported by this tool is computationally derived from data and constitutes exploratory analysis. This approach aims to generate research hypotheses rather than provide confirmatory conclusions. We strongly recommend treating this cutoff as a candidate value and validating it in patient cohorts independent of the current analysis to confirm its clinical or biological significance.
DOWNLOAD
CUSTOMIZE OPTIONS>>

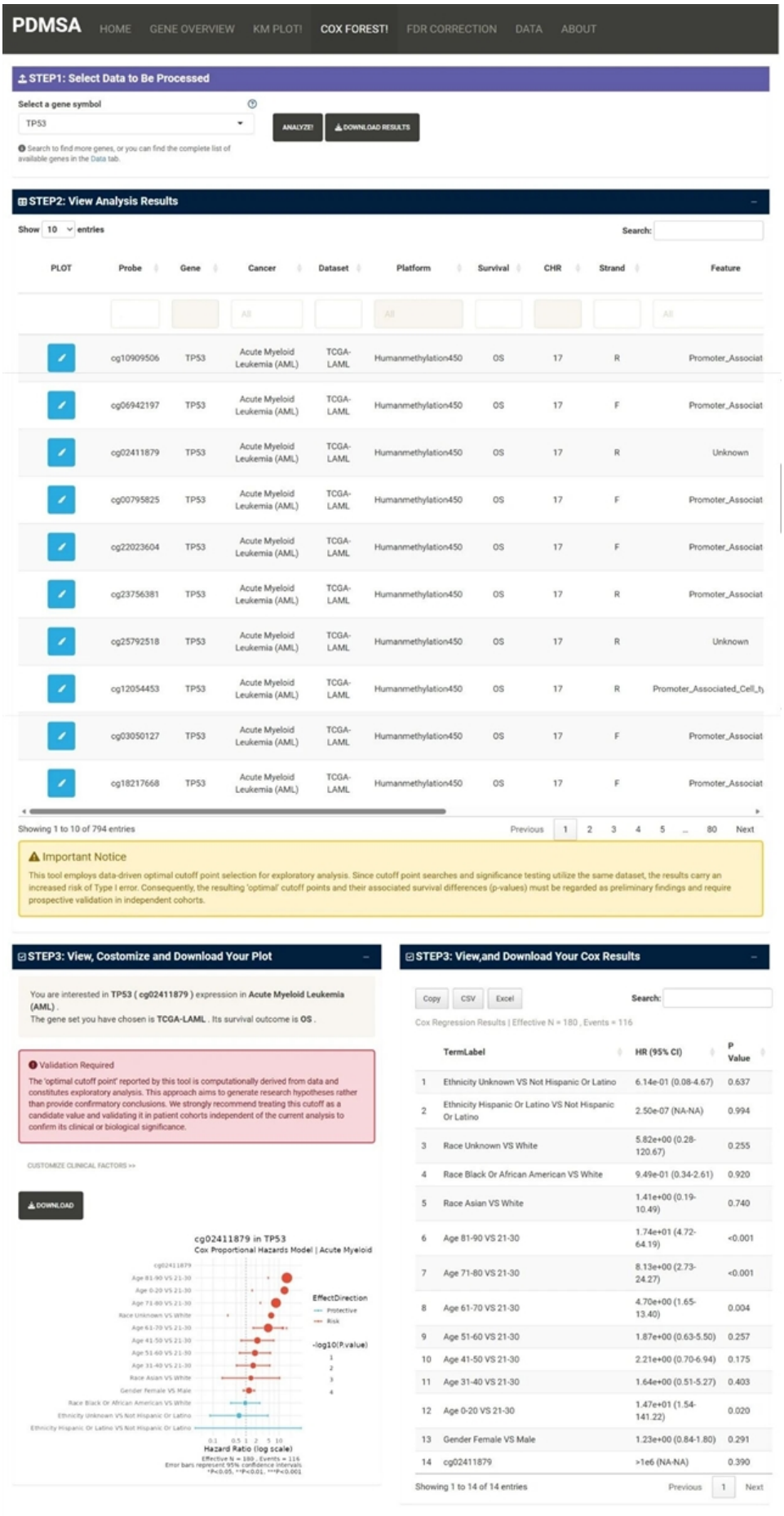
COX FOREST ：多变量回归与混杂因素校正

单变量生存分析无法排除其他临床因素的干扰。为此，PDMSA提供了多变量Cox回归分析模块，并以森林图形式直观展示结果。

在CUSTOMIZE CLINICAL FACTORS面板中，用户可以从系统提取的核心临床变量（包括性别、年龄、种族、民族和病理分期）中，自由选择2-5个变量纳入多变量模型，以校正潜在的混杂因素。

系统会自动对分类变量进行因子化编码（如性别以男性为参照，分期以I期为参照），并将临床信息缺失的样本归为Unknown组一并纳入分析，确保模型的稳健性。

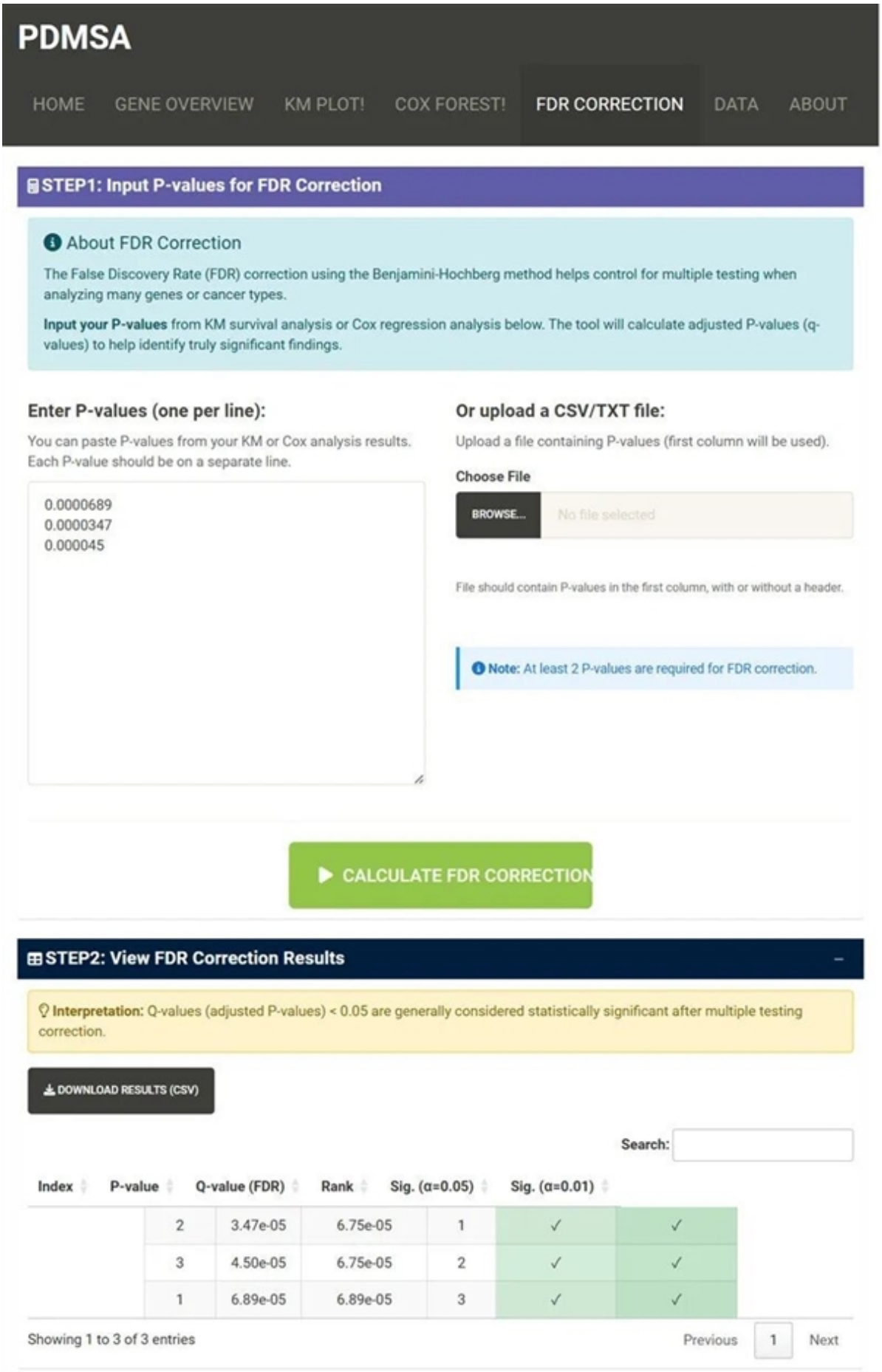
分析结果以森林图呈现，清晰展示每个变量（包括目标甲基化探针）的风险比（HR）及其95%置信区间和P值，帮助用户评估甲基化水平是否为独立的预后因素。


PDMSA
HOME
GENE OVERVIEW
KM PLOT!
COX FOREST!
FDR CORRECTION
DATA
ABOUT
STEP1: Select Data to Be Processed
Select a gene symbol
TP53
ANALYZE!
DOWNLOAD RESULTS
Search to find more genes, or you can find the complete list of available genes in the Data tab.
STEP2: View Analysis Results
Show 10 entries
Search:
PLOT
Probe
Gene
Cancer
Dataset
Platform
Survival
CHR
Strand
Feature
cg10909506 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Promoter_Associat
cg06942197 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg02411879 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Unknown
cg00795825 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg22023604 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg23756381 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Promoter_Associat
cg25792518 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Unknown
cg12054453 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 R Promoter_Associated_Cell_t
cg03050127 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
cg18217668 TP53 Acute Myeloid Leukemia (AML) TCGA-LAML Humanmethylation450 OS 17 F Promoter_Associat
Showing 1 to 10 of 794 entries
Previous 1 2 3 4 5 ... 80 Next
Important Notice
This tool employs data-driven optimal cutoff point selection for exploratory analysis. Since cutoff point searches and significance testing utilize the same dataset, the results carry an increased risk of Type I error. Consequently, the resulting 'optimal' cutoff points and their associated survival differences (p-values) must be regarded as preliminary findings and require prospective validation in independent cohorts.
STEP3: View, Costomize and Download Your Plot
You are interested in TP53 (cg02411879) expression in Acute Myeloid Leukemia (AML) .
The gene set you have chosen is TCGA-LAML . Its survival outcome is OS .
Validation Required
The 'optimal cutoff point' reported by this tool is computationally derived from data and constitutes exploratory analysis. This approach aims to generate research hypotheses rather than provide confirmatory conclusions. We strongly recommend treating this cutoff as a candidate value and validating it in patient cohorts independent of the current analysis to confirm its clinical or biological significance.
CUSTOMIZE CLINICAL FACTORS >>
DOWNLOAD
cg02411879 in TP53
Cox Proportional Hazards Model | Acute Myeloid
EffectDirection
Protective
Risk
-log10(P.value)
Hazard Ratio (log scale)
Effective N = 180 , Events = 116
Error bars represent 95% confidence intervals
*P<0.05, **P<0.01, ***P<0.001
STEP3: View,and Download Your Cox Results
Copy
CSV
Excel
Search:
Cox Regression Results | Effective N = 180 , Events = 116
TermLabel
HR (95% CI)
P Value
1 Ethnicity Unknown VS Not Hispanic Or Latino 6.14e-01 (0.08-4.67) 0.637
2 Ethnicity Hispanic Or Latino VS Not Hispanic Or Latino 2.50e-07 (NA-NA) 0.994
3 Race Unknown VS White 5.82e+00 (0.28-120.67) 0.255
4 Race Black Or African American VS White 9.49e-01 (0.34-2.61) 0.920
5 Race Asian VS White 1.41e+00 (0.19-10.49) 0.740
6 Age 81-90 VS 21-30 1.74e+01 (4.72-64.19) <0.001
7 Age 71-80 VS 21-30 8.13e+00 (2.73-24.27) <0.001
8 Age 61-70 VS 21-30 4.70e+00 (1.65-13.40) 0.004
9 Age 51-60 VS 21-30 1.87e+00 (0.63-5.50) 0.257
10 Age 41-50 VS 21-30 2.21e+00 (0.70-6.94) 0.175
11 Age 31-40 VS 21-30 1.64e+00 (0.51-5.27) 0.403
12 Age 0-20 VS 21-30 1.47e+01 (1.54-141.22) 0.020
13 Gender Female VS Male 1.23e+00 (0.84-1.80) 0.291
14 cg02411879 >1e6 (NA-NA) 0.390
Showing 1 to 14 of 14 entries
Previous 1 Next

FDR CORRECTION：多重假设检验的统计学保障

在高通量的甲基化数据分析中，进行数千次统计检验会带来假阳性风险升高的问题。为此，PDMSA专门集成了FDR校正模块，以增强研究结果的可靠性。该模块基于经典的Benjamini-Hochberg方法，对原始P值进行校正，计算调整后的q值。

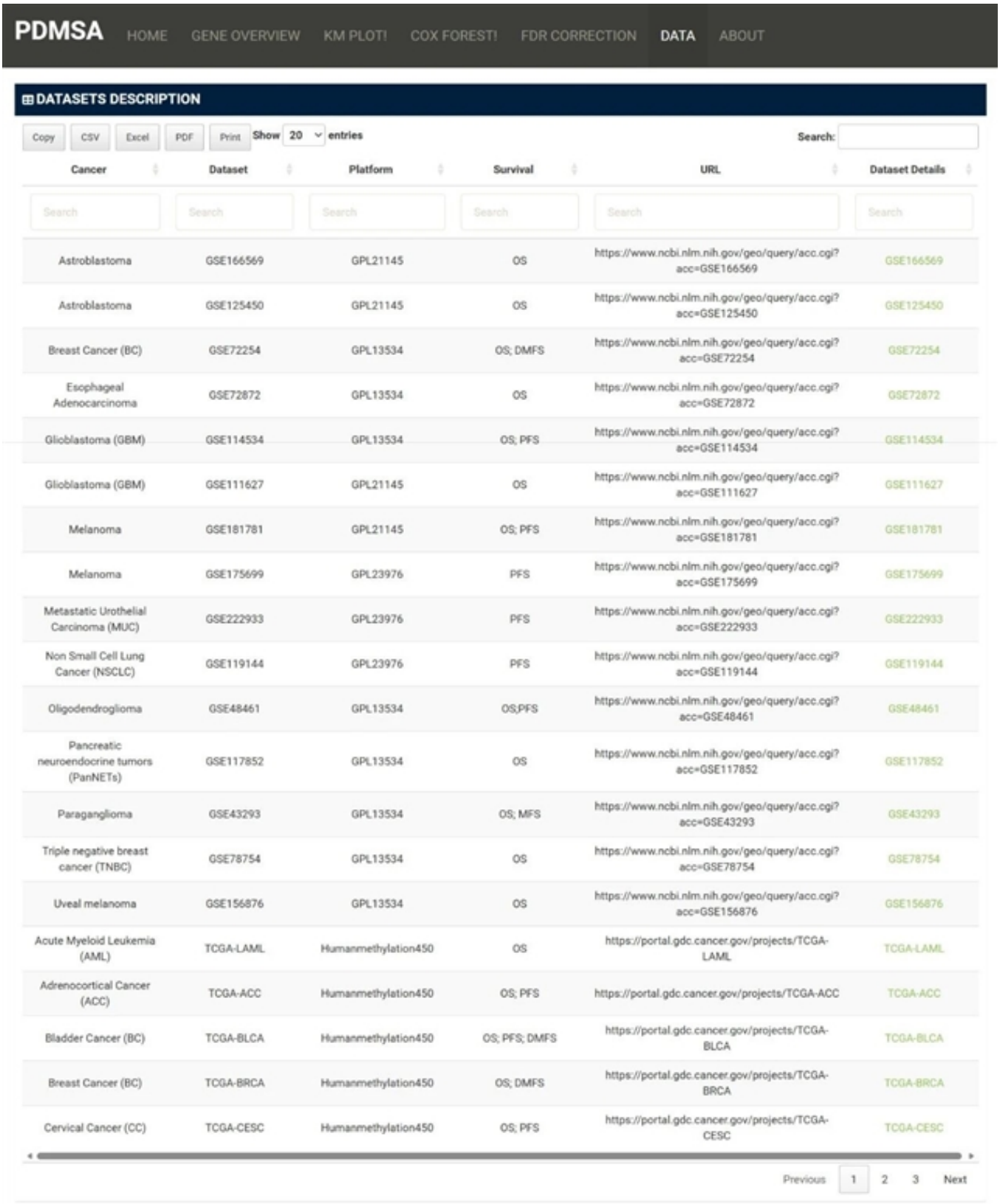
校正结果会清晰标注在 $\alpha = 0.05$ 和 $\alpha = 0.01$ 水平下的显著性，帮助研究者从海量数据中准确区分出统计学上更为稳健的信号，减少假发现，为后续的验证研究提供更可靠的候选列表。


PDMSA
HOME
GENE OVERVIEW
KM PLOT!
COX FOREST!
FDR CORRECTION
DATA
ABOUT
STEP1: Input P-values for FDR Correction
About FDR Correction
The False Discovery Rate (FDR) correction using the Benjamini-Hochberg method helps control for multiple testing when analyzing many genes or cancer types.
Input your P-values from KM survival analysis or Cox regression analysis below. The tool will calculate adjusted P-values (q-values) to help identify truly significant findings.
Enter P-values (one per line):
You can paste P-values from your KM or Cox analysis results. Each P-value should be on a separate line.
0.0000689
0.0000347
0.000045
Or upload a CSV/TXT file:
Upload a file containing P-values (first column will be used).
Choose File
BROWSE...
No file selected
File should contain P-values in the first column, with or without a header.
Note: At least 2 P-values are required for FDR correction.
CALCULATE FDR CORRECTION
STEP2: View FDR Correction Results
Interpretation: Q-values (adjusted P-values) < 0.05 are generally considered statistically significant after multiple testing correction.
DOWNLOAD RESULTS (CSV)
Search:
Index
P-value
Q-value (FDR)
Rank
Sig. (α=0.05)
Sig. (α=0.01)
2 3.47e-05 6.75e-05 1 ✓ ✓
3 4.50e-05 6.75e-05 2 ✓ ✓
1 6.89e-05 6.89e-05 3 ✓ ✓
Showing 1 to 3 of 3 entries
Previous
1
Next

DATA：数据来源透明可溯

PDMSA的所有分析都建立在公开、透明的数据基础上。在DATA选项卡中，我们以交互式表格的形式，提供了所有纳入数据集的详细信息。

用户可以查询每个数据集的癌症类型、测序平台、样本量、可用生存信息，以及指向原始数据库（TCGA或GEO）的链接。这不仅方便用户了解数据的广度与深度，也便于在需要时进一步探索原始数据的更多细节，保证了研究的可追溯性。



ABOUT：用户支持与持续更新

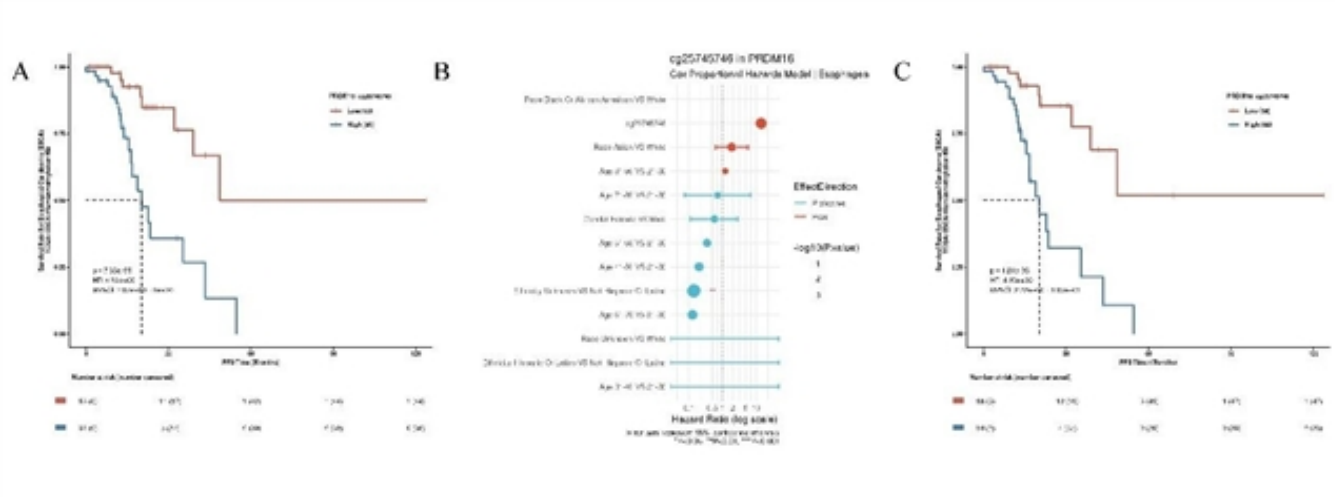
为降低使用门槛，PDMSA在ABOUT模块中提供了详尽的使用教程视频和常见问题解答。同时，该模块也是用户与开发团队沟通的桥梁。用户可以通过邮箱或评论系统提交反馈、建议或报告问

题。我们承诺将认真考虑每一条反馈，并持续优化和更新PDMSA，在网站上实时公布版本更新日志，确保工具的透明性和持续改进。

案例分析：PRDM16基因在食管癌中的甲基化预后价值

我们通过对PDMSA数据库中显著相关的甲基化探针进行统计分析，筛选出PRDM16作为潜在甲基化调控的预后标志物。该基因在多种癌症中与不良预后相关，其甲基化位点主要富集于启动子区域。

在食管癌（ESCA）队列中，PRDM16探针cg25745746的高甲基化与患者较差的无进展生存期显著相关（HR = 4.65, p = 2.66×10-4）。多变量Cox回归分析进一步证实其独立预后价值。与中位数分组相比，最佳截断值分组能更显著地识别出预后较差的患者群体（p = 1.21×10-5 vs 7.35×10-5）。



与现有工具对比

特性 PDMSA MethSurv MethHC 2.0 UALCAN LinkedOmics 数据来源 TCGA + GEO TCGA TCGA + GEO TCGA TCGA 数据集数量 45 25 33 33 32 癌症类型 39 25 33 33 32 样本量 8,369 7,358 27,190 >10,000 11,158 生存终点 4种 1种 1种 1种 1种 最佳截断值 √ √ × × × Cox回归 √ × × × √ 颜色自定义 √ × × × × 多变量分析 √ √ × √ ×

讨论与展望

PDMSA作为一款专注于DNA甲基化与生存预后的分析平台，具有数据覆盖面广、分析方法多样、用户界面友好等优势。尽管如此，目前仍存在以下局限性：

- 数据来源仅限于TCGA与GEO，尚未整合EWAS Data Hub等其他数据库；
- 暂不支持用户上传自定义数据；
- 尚未实现比例风险假定的自动化检验。

未来，我们将持续更新数据集，拓展分析功能，探索与免疫治疗、代谢重编程等新兴领域的交叉研究，助力肿瘤精准医学的发展。

引用信息

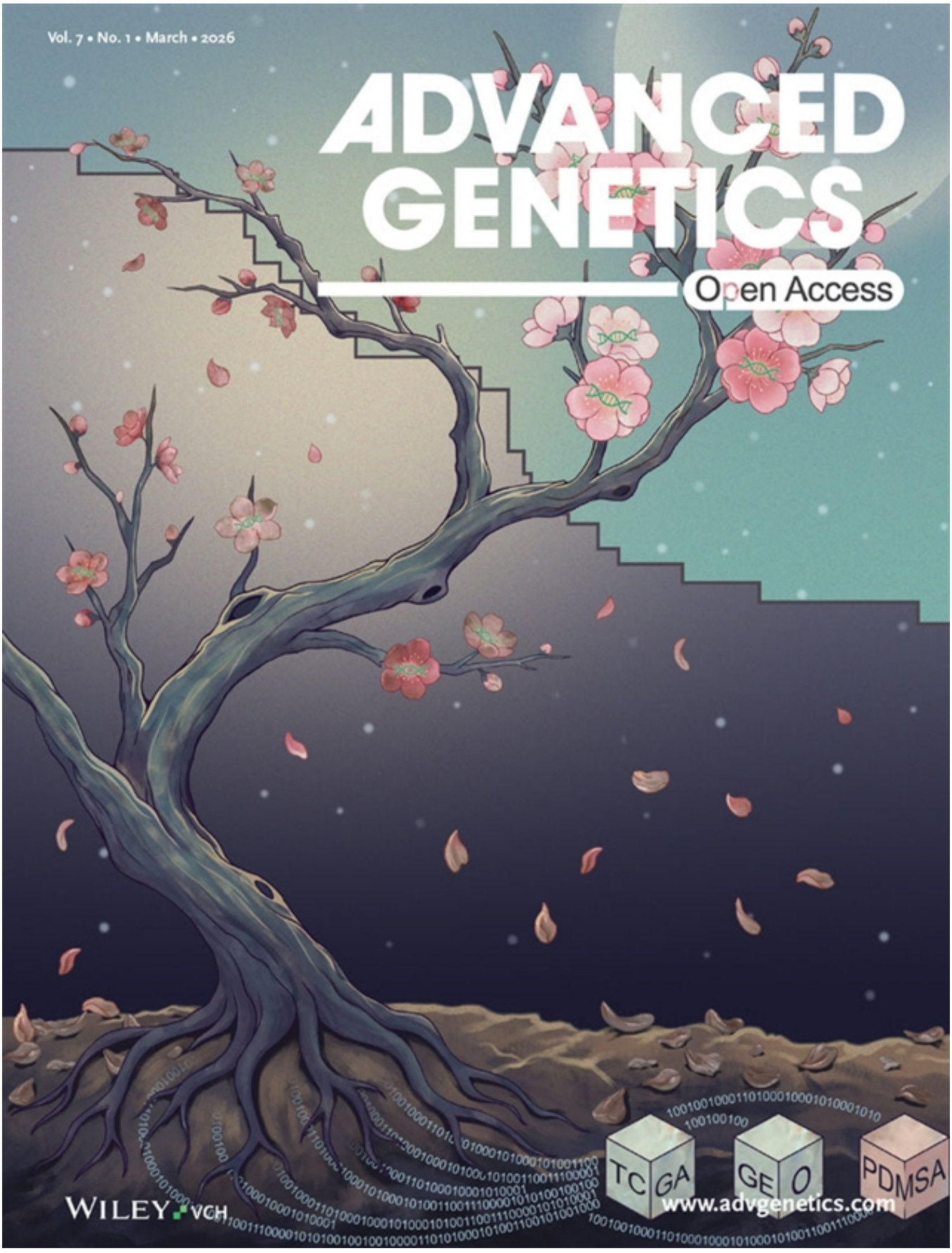
Guo, W., Shi, Y., Wu, S., Yang, H., Lin, A., Luo, P. (2026). PDMSA: A web-based tool for pan-cancer survival analysis using DNA methylation levels as biomarkers. Advanced Genetics, e202500069. https://doi.org/10.1002/ggn2.202500069

访问地址

robinl-lab.com/PDMSA

欢迎使用并反馈！

期刊简介


Vol. 7 • No. 1 • March • 2026
ADVANCED GENETICS
Open Access
WILEY VCH
TC GA
GE O
PDMSA
www.advgenetics.com

Advanced Genetics是隶属于享有盛誉的 Advanced Portfolio 的一本广泛领域期刊，专注于基因相关领域的前沿研究。

本刊致力于覆盖广泛的遗传学主题，包括人类、动物、植物和微生物遗传学，基因组学与表观基因组学，以及将遗传学、组学、生物信息学和计算生物学作为研究工具，用于发现与创新的相关研究方向。期刊欢迎基础研究、应用研究、临床研究和转化研究等不同类型的投稿，旨在应对重要的科学与社会挑战，这些挑战包括提升粮食安全、保护生物多样性、推动人类健康与福祉，以及促进有助于 One Health 成果产生实质影响的全球性倡议。

本刊同样欢迎与遗传学研究共同体相关的创新性软件工具、方法或数据资源的投稿。这类成果以及其他所有原创研究，均可通过本刊灵活设置的研究论文类型进行投稿。

来源：Advanced Genetics

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发