



中国科学院北京基因组研究所（国家生物信息中心）

BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES / CHINA NATIONAL CENTER FOR BIOINFORMATION

单细胞多维组学数据库资源 介绍与在线分析平台实践



主讲人

曾滢瑶



日期

2025.5.9

目录

CONTENTS

01

NGDC数据资源体系

02

基因表达图谱数据库

03

癌症单细胞表达图谱数据库

04

空间转录组综合资源存储库

05

单细胞全基因组甲基化数据库

06

细胞分类学知识库

07

单细胞全转录组关联知识库

01

NGDC数据资源体系

国家基因组科学数据中心建设历程

1999年6月

郝柏林《建议尽快
组建国家级的生物
医学信息中心》

2014年11月

中科院学部《关于中国
“国家生物医学信息中心”
的调研报告与建议》

2016年2月

北京基因组所“生
命与健康大数据中
心”正式成立

2018年8月

国家生物信息
中心建设方案

2019年8月

国家生物信息
中心建设方案

2019年11月

批复加挂
“国家生物信息中心”

2017年10月

国家生物信息
中心建设方案

2019年6月

科技部、财政部批复成立
“国家基因组科学数据中心”

1999

.....

2014

2015

2016

2017

2018

2019

2020

建议尽快组建国家级的生物医学信息中心

郝柏林

1999年6月10日

从细菌到人类，众多物种的基因和蛋白质数据正在以科学史上从未有过的高速度增长。目前已测定出二十多种细菌，以及比细菌更高等的一些物种如酵母菌和线虫的完全基因组。人类基因组，即一个典型的“人”的全部基因，也将提前在2003年完全测定。现在，基因数据的总量每14个月翻一番，每个月至少测出150种蛋白质的结构。增长速度本身也在加速。

国家生物信息中心筹建搁浅三年！

郝柏林

2003年1月1日

徐冠华同志：

这封信只提两件事：

一、国家生物信息中心筹建工作搁浅三年多，情况太不正常，实在不应再拖下去。国家生物信息中心首先是艰苦的服务，而不是值得哄抢的肥肉。凡是拿了人民的钱作出的数据，都应当提交国家生物信息中心，按情况、分层次

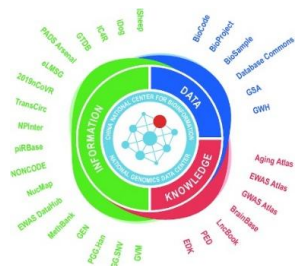
关于中国“国家生物医学信息中心”
的调研报告与建议

2014年

中科院学部咨询项目

《生物信息学的发展与共享平台的建设》项目组

调研项目组院士：陈润生、赵国屏、强伯勤、郝柏林、
杨焕明、贺福初、李衍达、康乐、张春霆



4

02

基因表达图谱数据库

Gene Expression Nebulas (GEN)

基因表达图谱数据库GEN



NGDC / CNCB

NCBI + EBI

对标



Gene Expression Nebulas

A data portal of transcriptomic profiles across multiple species



Gene Expression Omnibus



Expression Atlas



Gene Expression Nebulas

A data portal of transcriptomic profiles analyzed by a unified pipeline across multiple species

[Home](#) [Datasets](#) [Genes](#) [Metadata](#) [Tools](#) [Statistics](#) [Documentation](#) [Download](#) [Contact](#)

Gene Expression Nebulas (GEN) is a data portal of transcriptomic profiles under various conditions derived entirely from bulk and single-cell RNA-Seq data analysis in multiple species.

GEN Dataset

e.g., GEND000097; Homo sapiens; Lung; SARS-CoV-2 infection; TP53

41

SPECIES

577

DATASETS

59,339

SAMPLES

19,254,189

CELLS

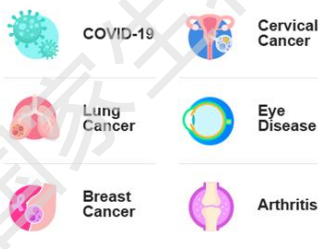
549

PUBLICATIONS

Release Note

- 2024-12-30 A new class of data - Binary Alignment/Map (BAM) data were added to FTP for downstream analysis of personalization.
- 2024-12-20 Gene expression profiles of 7 *Oryza sativa* datasets and 1 cancer-related dataset were added.
- 2023-11-13 Gene expression profiles of 34 cancer-related datasets were added.
- 2023-09-30 Gene expression profiles of 28 *Oryza sativa* (Geng/Japonica Group) related datasets were added.

Featured Projects



Homo sapiens
Human
330 Datasets
8899 Bulk Samples
9469443 Cells



Mus musculus
House mouse
37 Datasets
585 Bulk Samples
1436995 Cells



Macaca fascicularis
Crab-eating macaque
16 Datasets
472 Bulk Samples
89492 Cells



Macaca mulatta
Rhesus monkey
11 Datasets
248 Bulk Samples
41141 Cells



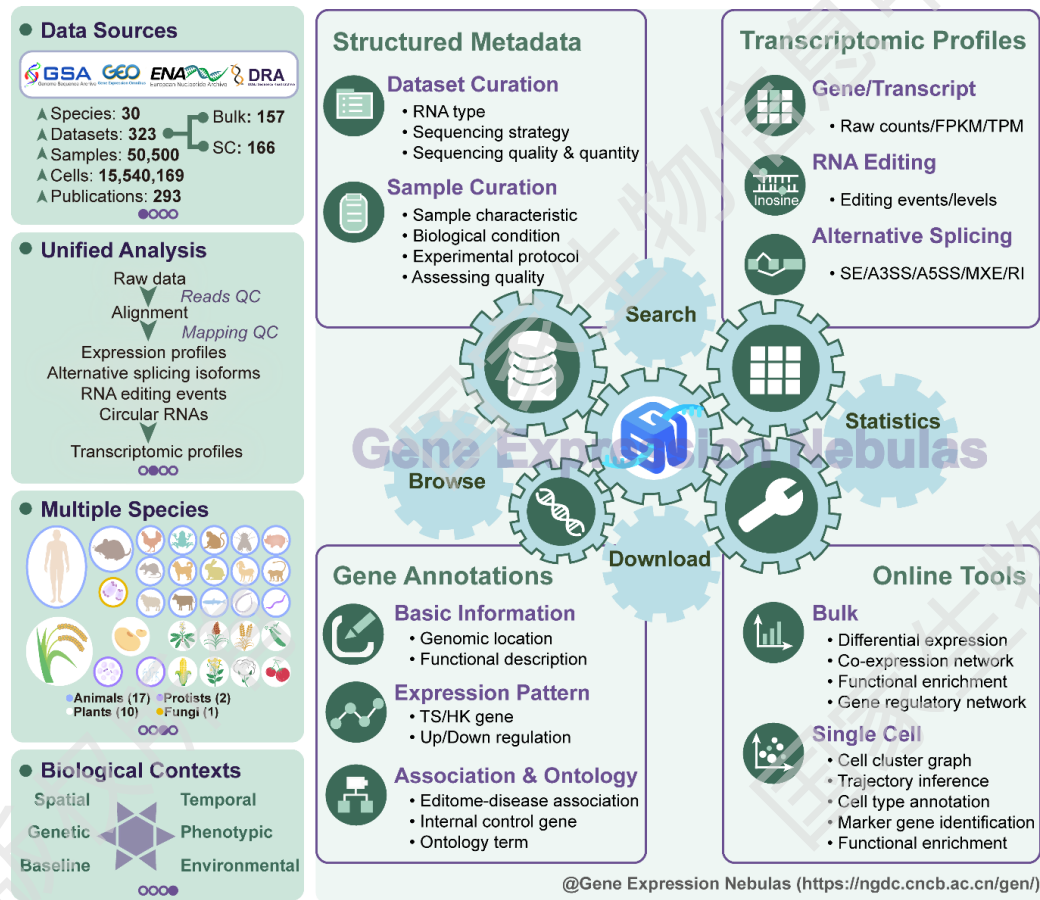
Callithrix jacchus
white-tufted-ear marmoset
8 Datasets
73 Bulk Samples
196 Cells

<https://ngdc.cncb.ac.cn/gen/>, *Nucleic Acids Res.* 2022

问题与挑战：我国缺少RNA组学数据纵深挖掘整合资源平台，公共RNA组学数据挖掘维度单一，其所蕴含的丰富基因表达调控信息未充分释放

GEN主要特点：

- 建立结构化**审编模型和标准化数据处理流程**
- 转录图谱：基因/转录本表达、RNA选择性剪接、环形RNA表达、RNA编辑
- 兼容**组织**和**单细胞**水平的表达数据
- 丰富的基因注释信息
- 方便快捷的在线、离线分析工具



数据统计:

- 整合**577**套高质量转录组数据集，包含**41**个物种的**59,339**个样本和**19,254,189**个细胞
- 涵盖**94**种人类疾病、**141**种组织、**130**种细胞或细胞系类型

访问情况:

- 截止2024年12月，GEN数据库获得**158,178**次访问，**90,927**次下载，被下载数据量总计**2,119.82GB**

建设意义:

- 初步解决海量异构RNA组学数据挖掘整合问题，弥补RNA组学纵深挖掘研究中的资源空白



Browse Datasets

View All

快速预览

▼ Biological Context (469)

生物学场景

Search

- Baseline (469)
 - No (313)
 - Yes (156)
- Environmental (106)
- Genetic (60)
- Phenotypic (286)
- Spatial (469)
- Temporal (122)

▶ Animal Tissue (261)

组织

▼ Human Disease (239)

人类疾病

Search...

Search

- Cardiovascular system disease (1)
- Coronavirus infectious disease (4)
- Endocrine system disease (8)
- Gastrointestinal system disease (12)
- Hematopoietic system disease (10)
- Immune system disease (6)
- Integumentary system disease (12)
- Musculoskeletal system disease (13)
- Nervous system disease (17)
- Reproductive system disease (4)
- Respiratory system disease (145)
- Thoracic disease (7)
 - Breast cancer (7)
- Urinary system disease (5)

▶ Transcriptomic Profile (469)

转录图谱

▼ Display columns

Basic Information ☒ Dataset ID ☒ BioProject ID ☐ Project ID ☒ Dataset Name ☒ Species ☒ Strategy ☒ Sample Number ☐ Healthy Condition ☐ Tissue ☐ Cell Type
☐ Cell Line ☐ Development Stage ☐ Case Detail ☐ Control Detail ☐ Cell Number

Biological Context? ☒ Baseline ☒ Genetic ☒ Phenotypic ☒ Environmental ☒ Spatial ☒ Temporal

Quality and Quantity ☒ RNA Type ☒ Median Mapping Quality ☒ Median Coverage ☒ Max Sequencing Length ☒ Max Replicate Number

元信息展示选项

Show 10 entries

Search:

Download	Dataset ID	Dataset Name	Species	Strategy	Sample Number	Tissue
-	Search Dataset ID	Search Dataset Name	Search Species	scRNA	min sample# max sample#	Search Tissue
	GEND000004	Single cell RNA sequ...	Macaca fascicularis	• scRNA 10x Genomics	2	• Fetal Ovary
	GEND000028	A single-cell RNA-se...	Schistosoma mansoni	• scRNA 10x Genomics	9	-
	GEND000029	Single cell analysis...	Homo sapiens	• scRNA Smart-seq2	1446	• Human Liver Parenchymal Tissue
	GEND000030	Single-cell omics re...	Homo sapiens	• scRNA Smart-seq2	221	-
	GEND000031	Spatial and single-c...	Homo sapiens	• scRNA Smart-seq2	190	• Primary Papillary Dermis • Primary Reticular Dermis
	GEND000032	Single cell RNA sequ...	Homo sapiens	• scRNA Smart-seq2	30	• Bone Marrow

输入scRNA筛选单细胞数据集



Home > Datasets > GEND000004

Basic Information

Samples

Transcriptomic Profiles

Visualization

数据集相关信息

GEND000004: Single cell RNA sequencing reveals fetal ovary development in Macaca fascicularis

BioProject ID: PRJNA629482 Original Project ID: GSE149629

Strategy: scRNA 10x Genomics

Species: Macaca fascicularis

Tissue: Fetal Ovary

Healthy Condition: Normal

Cell Type: Germ Cell

Cell Number: 22737

Biological Context: Baseline: No Genetic: - Phenotypic: - Environmental: - Spatial: Tissue Temporal: -

Quality and Quantity: RNA Type: Poly(A)+ RNA Median Mapping Quality: 0.9505 Median Coverage: 19.30 Max Sequencing Length: 300 #Max Replicate: 1

Show 10 entries

Search:

GEN Sample ID	BioSample ID	Tissue	Cell Type	Cell Line	Disease	Case Detail
Search GEN Sample ID	Search BioSample ID	Search Tissue	Search Cell Type	Search Cell Line	Search Disease	Search Case Detail
GENS00000069	SAMN03070042	Brain	iPSC-derived Neural Progenitor	iPSC line	Autism Spectrum Disorder (ASD)	Autism spectrum disorder with macrocephaly
GENS00000068	SAMN03070041	Brain	iPSC-derived Neural Progenitor	iPSC line	Autism Spectrum Disorder (ASD)	Autism spectrum disorder with macrocephaly



Transcriptomic Profiles

Note: **GENDX:** Gene Expression Profiles **GENDE:** RNA Editing Profiles

ID: GENDX000004 [Download](#)

Reference Genome: Macaca_fascicularis_5.0.101

Genome Annotation: Macaca_fascicularis_5.0.101

Data Processing:

1. Data generation: Cellranger

- ◆ 点击**Transcriptomic Profiles**，可以显示该数据集分析所使用的参考基因组和注释信息
- ◆ 点击**Visualization**，可设置不同的参数进行可视化，展示数据集中细胞降维聚类情况

GEND000004

Single cell RNA sequencing reveals fetal ovary development in Macaca fascicularis

GENS00014204

Main Parameters

Projection

tSNE

Color cells by

- ☐ Metadata
☐ Annotation
☒ Resolution

0.8

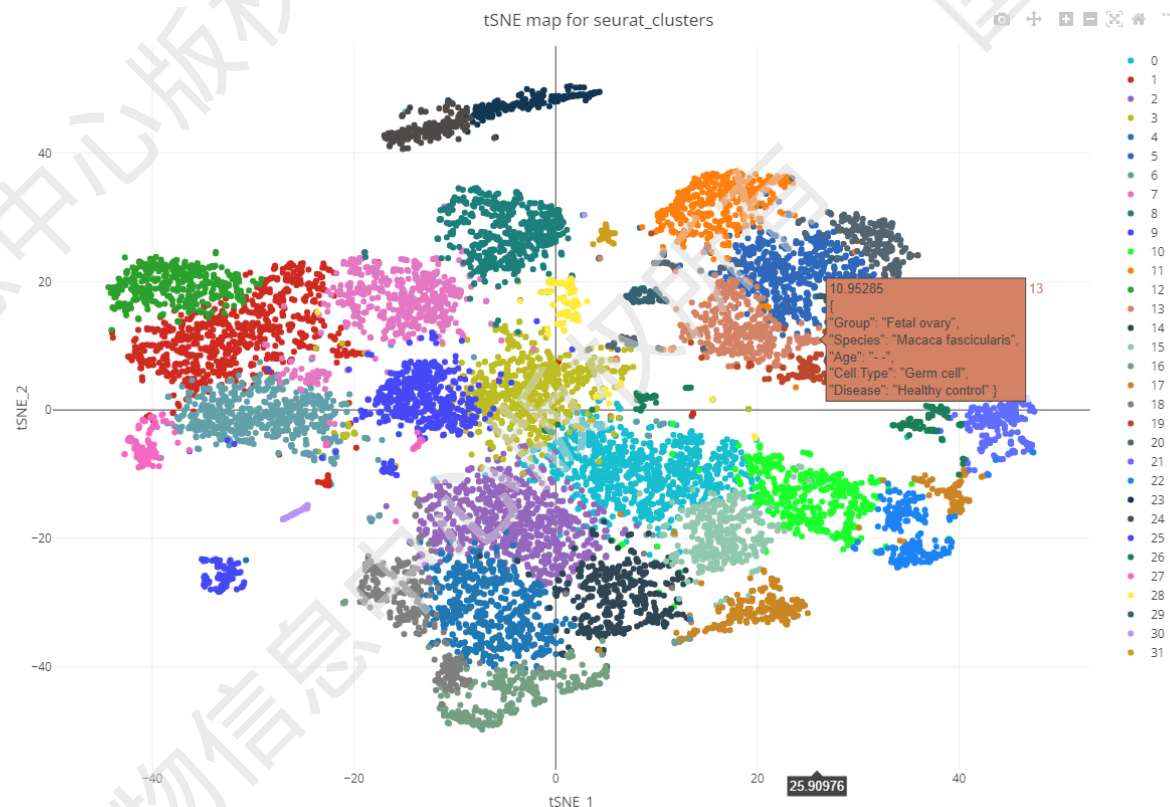
Additional Parameters

Point Size

0 10 20

Point opacity

0 0.5 1



单细胞数据在线分析工具



Home Datasets Genes Metadata Tools Statistics Documentation Download Contact

Home > Tools > Online Analysis

Online Analysis

GEN Toolkit

Tools

GEN is equipped with a series of online tools in aid of further downstream analysis of both bulk and single-cell RNA-seq data and visualization of scRNA-seq data analysis results. All the analysis results can be downloaded in CSV and/or Excel formats and visualized images can be exported to PNG or PDF.

Bulk RNA-seq Data Analysis

Single-cell RNA-seq Data Analysis

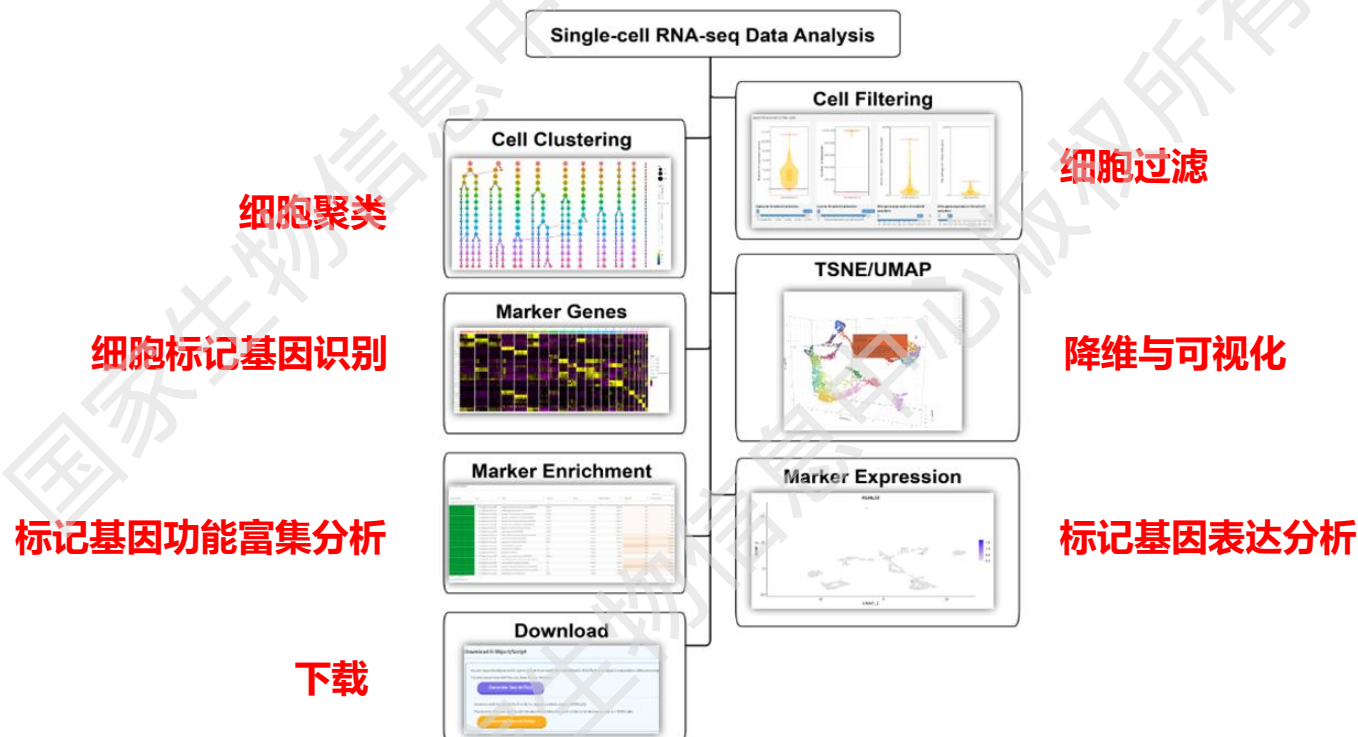
Visualization of scRNA-seq Data Analysis Results

[Single-cell RNA-seq Data Analysis](#)

2. 点击并跳转到单细胞数据在线分析工具

For scRNA-seq data, users can perform multiple analyses including quality control, data normalization, scaling and regression, dimensional reduction, graph-based clustering, identification of marker genes for cell clusters, gene enrichment analysis for cell markers, cell trajectory inference, and cell type annotation. To start 'Single-cell RNA-seq Data Analysis', users should select single-cell RNA-seq dataset of interest first. network inference. To start 'Bulk RNA-seq Data Analysis', users should select bulk RNA-seq dataset of interest first.

1. 选择单细胞数据分析



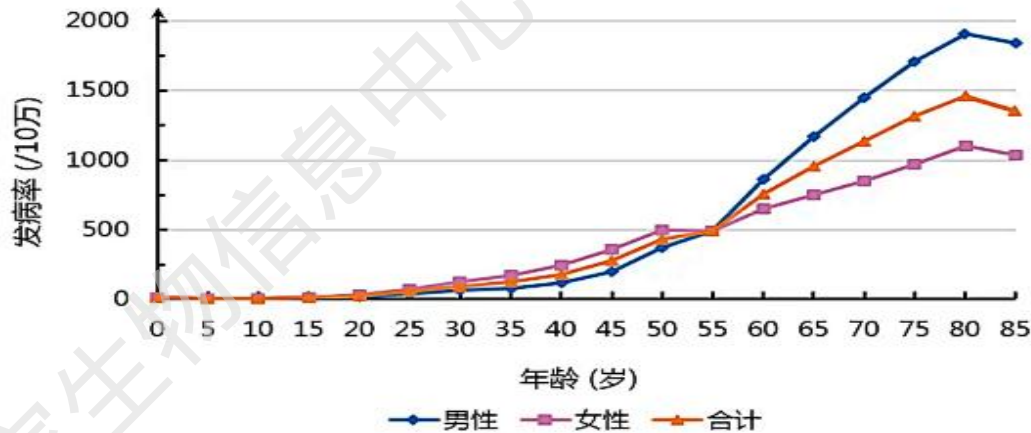
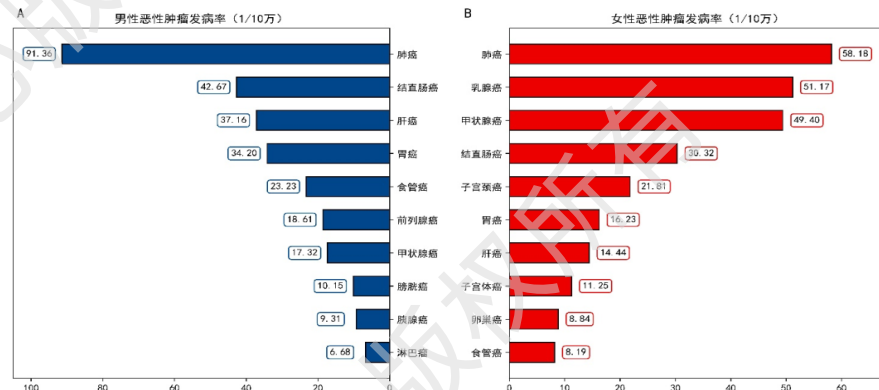
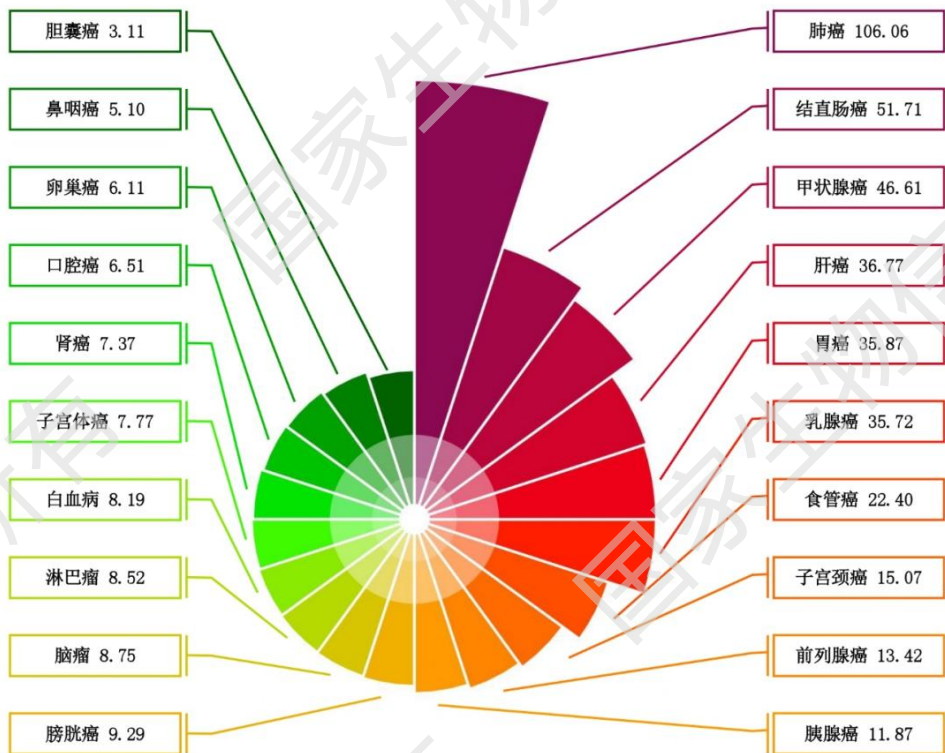
03

癌症单细胞表达图谱数据库

Cancer Single-cell Expression Map (CancerSCEM)

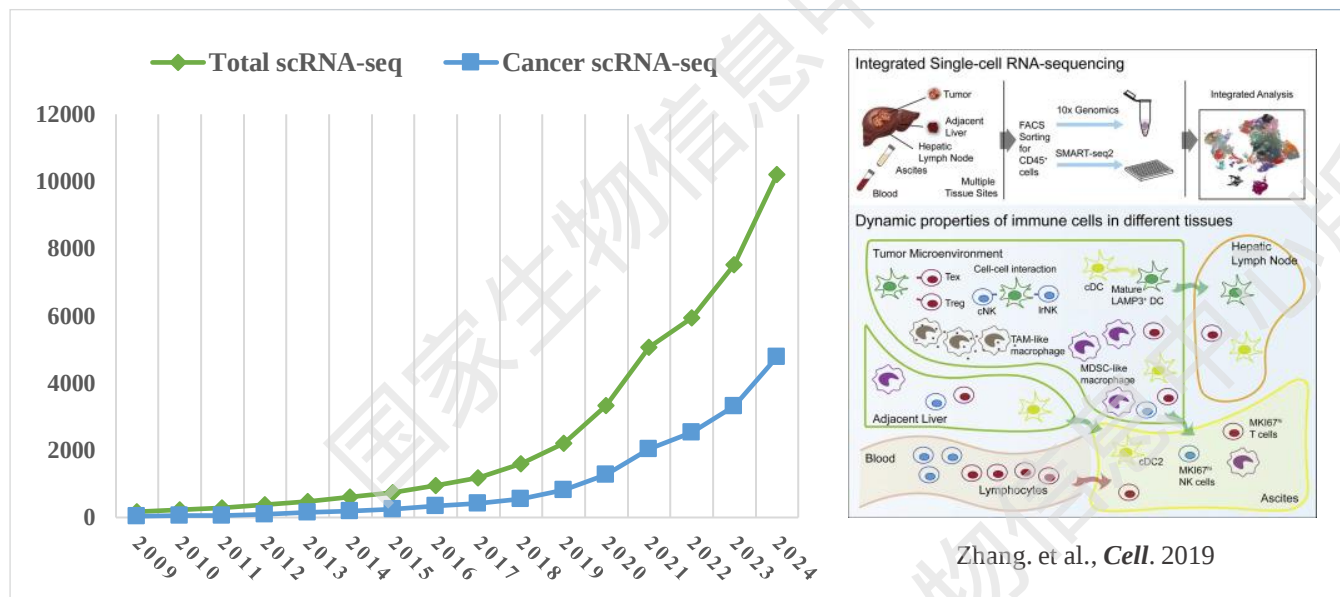
癌症发病率和死亡率逐年攀升

- 2024年全球最新癌症报告显示，中国 2022 年癌症新增病例占全球**24%**，死亡病例占**26.7%**，均位居第一
- 在新发恶性肿瘤中，肺癌位居首位，其次是结直肠癌（51.71万例）、甲状腺癌（46.61万例）、肝癌等
- 继续推动**癌症筛查、早诊早治和规范化治疗**，将有助于减轻肿瘤负担，显著改善癌症患者的预后

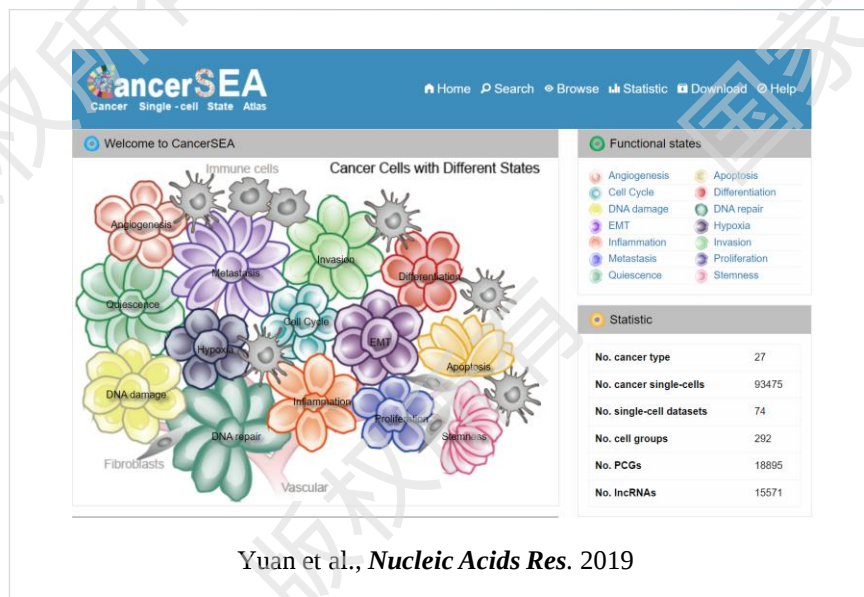


单细胞转录组数据海量激增

单细胞转录组学研究广泛开展



癌症单细胞数据库资源—CancerSEA



问题与挑战

- scRNA-seq研究呈现逐年快速增长趋势，癌症scRNA-seq数据爆炸式增长，尚缺乏标准化处理的、功能完善的综合性数据库
- CancerSEA是第一个旨在以单细胞水平全面解码肿瘤细胞不同功能状态的数据库，TME解析较少

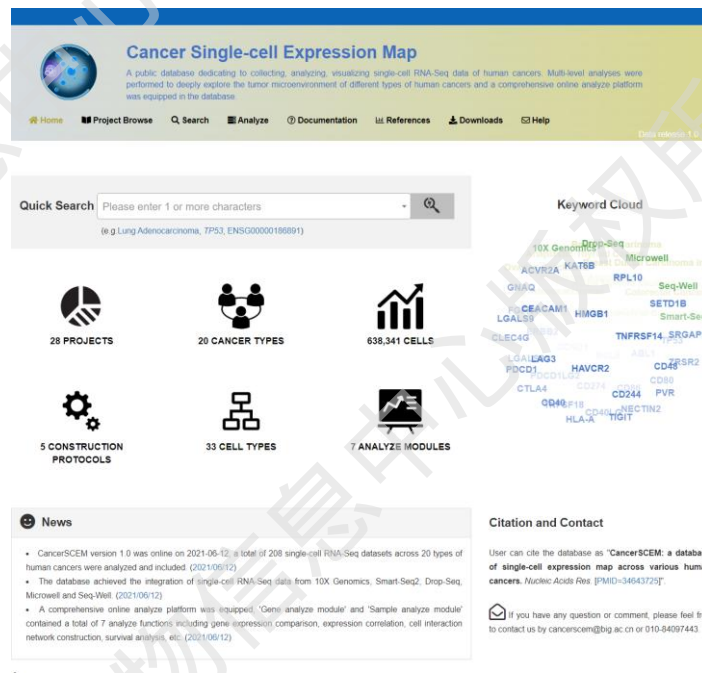
癌症单细胞表达图谱数据库CancerSCEM



- 围绕NGDC精准健康数据资源体系的建设和发展需求，从头建设了癌症单细胞表达图谱数据库

CancerSCEM

- CancerSCEM的建立将有效促进癌症单细胞转录组数据的二次利用与开放共享，不断推动癌症相关知识的发现



	Version 1.0 (2022)	Version 2.0 (2025)
Projects	28	127
Datasets	208	1,466
Cancer types	20	74
Cells	638,341	7,320,070
Cell types	33	80

CancerSCEM多维度数据浏览



国家生物信息中心

Cancer Single-cell Expression Map

Home / Project Browse

Project **Clinical**

Project Country

- ☐ Australia (6)
- ☐ Belgium (68)
- ☐ Canada (36)
- ☐ China (400)
- ☐ France (12)
- ☐ Germany (16)
- ☐ Japan (39)
- ☐ Korea (28)
- ☐ Singapore (2)
- ☐ Switzerland (17)
- ☐ United Kingdom (61)
- ☐ United States (781)

Cancer Type

- ☐ Acral Melanoma (7)
- ☐ Actinic Keratosis (2)
- ☐ Acute Myeloid Leukemia (36)
- ☐ Aldosterone Producing Adenoma (2)
- ☐ Anaplastic Astrocytoma (1)
- ☐ Anaplastic Thyroid Cancer (5)
- ☐ Appendiceal Neoplasms (16)
- ☐ Basal Cell Carcinoma (27)
- ☐ Biliary Tract Cancer (4)
- ☐ Bladder Cancer (11)
- ☐ Breast Cancer (48)
- ☐ Breast Ductal Carcinoma in Situ (1)
- ☐ Cervical Cancer (7)
- ☐ Clear Cell Renal Cell Carcinoma (49)
- ☐ Colorectal Cancer (106)
- ☐ Cutaneous Melanoma (4)
- ☐ Cutaneous Squamous Cell Carcinoma (6)
- ☐ Distal Cholangiocarcinoma (3)

[More](#)

Construction Protocol

Home / Transcriptome analysis of MCC-00

Home / Metabolic profile of CRC-016-21-1A

Basic Metabolism Analysis

- Metabolic flux balance loss
Convergence of the total loss and four loss terms during the training of neural networks.

- Metabolic flux in 168 modules
A total of 168 metabolic modules were included in scFEA, the distribution of their calculated metabolic fluxes were visualized.

- Metabolic Comparative Analysis
Cell-type Log2FC of 168 modules

- Cell-type Log2FC of 34 pathways

Related information

- [SRA](#)
- [BioProject](#)
- [GEO Datasets](#)
- [PubMed](#)

therapy) followed by checkpoint inhibitors (anti-PD1 and anti-CTLA4)

and adjacent non-malignant colon tissue

of 9 colorectal tumors as well as to matched adjacent non-malignant

is and their tumor microenvironment.

ons for each tumor) and adjacent non-malignant tissue from the same

from the pathologist. Samples were rinsed with PBS, and each sample

id transferred to 10 mL digestion medium (DMEM (Gibco) with 2mg/mL

ere incubated for 15 minutes at 37°C with manual shaking every 5

down for 1 minute using pipettes of descending sizes (25 mL, 10 mL,

[ThermoFisher Scientific]) was added, samples were filtered using a

300 G at 4°C for 5 minutes, the supernatant was decanted and

buffer (Sigma) and transferred to a 2 mL DNA low bind tube. Following

(150 G, 4°C, 5 min) using a swing-out rotor. Samples were next

min (50 mg/mL, AM2616 (ThermoFisher Scientific)) and filtered over

low retention filter tips (Mettler-Toledo). 10 µL (9 µL of cell suspension

cell counter (Luna) to determine the concentration of live cells.

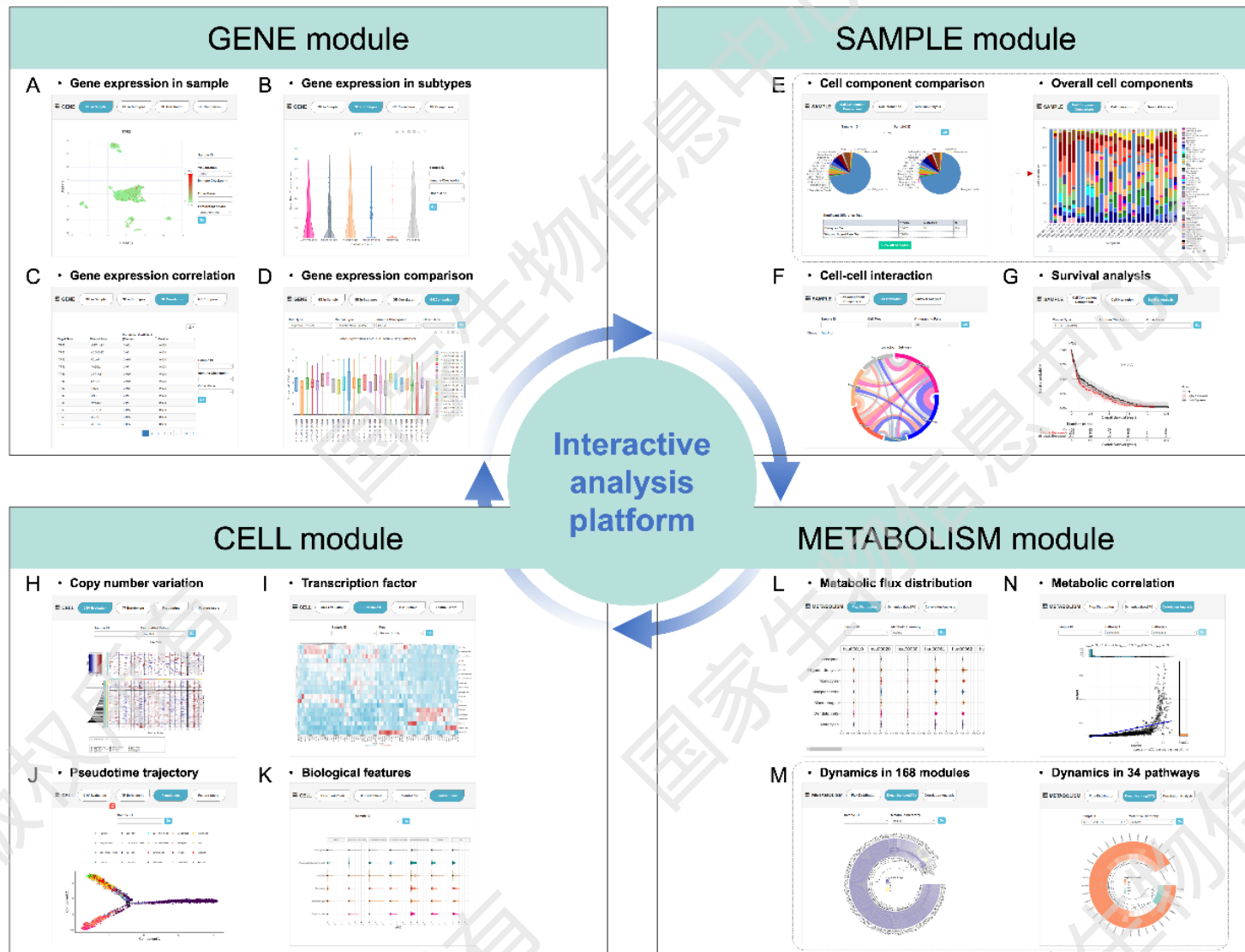
omics

eq 4000

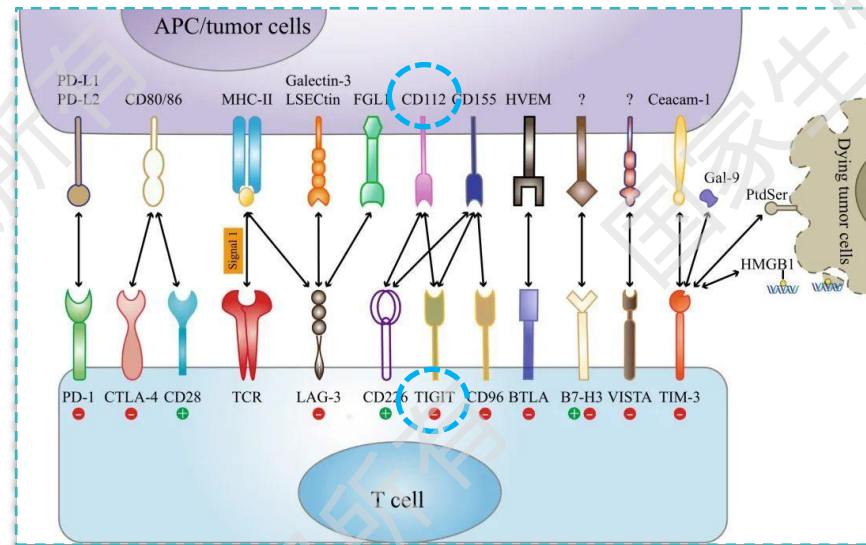
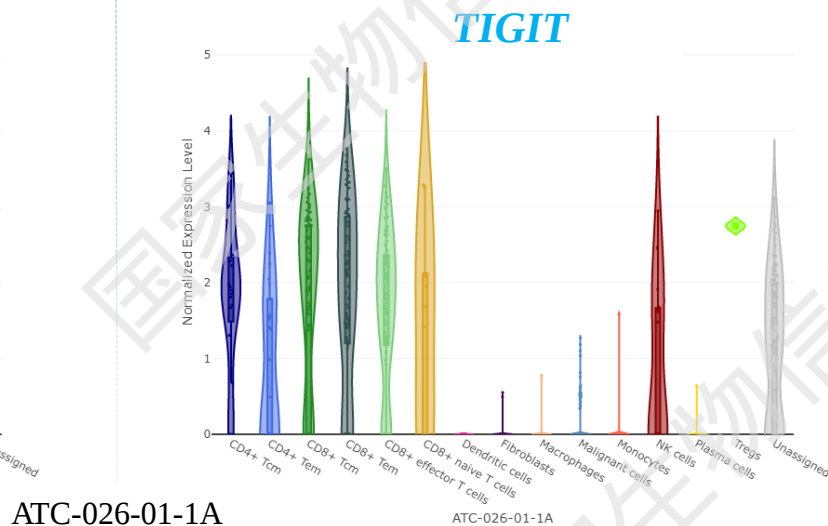
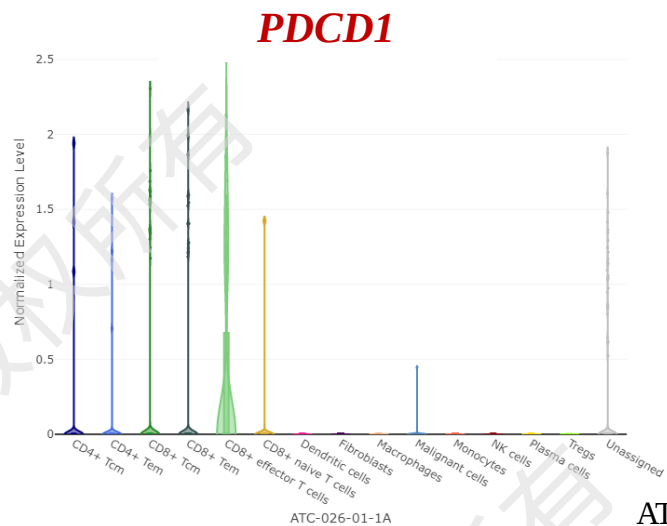
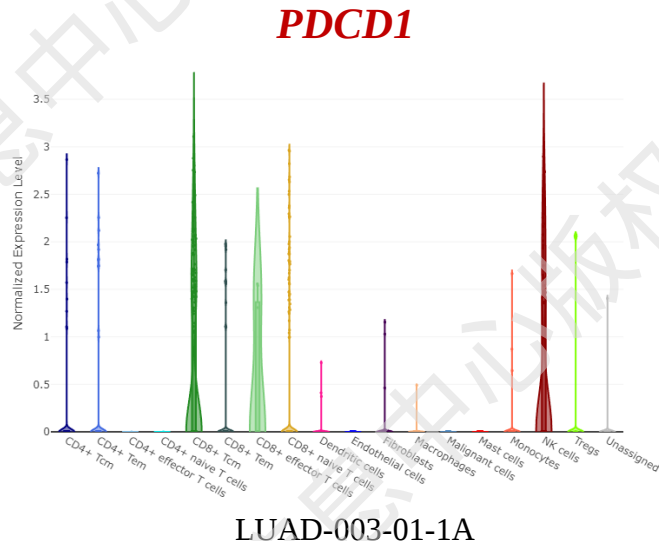
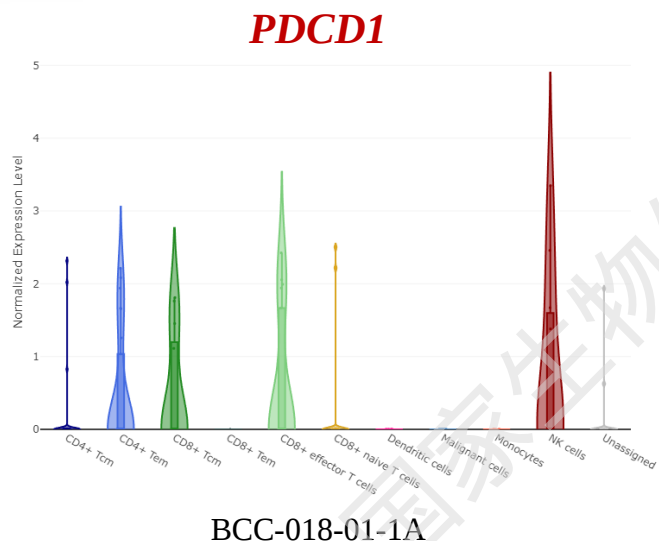
eq

D

the immune landscape of colorectal cancer. Nature Genetics, 2020.



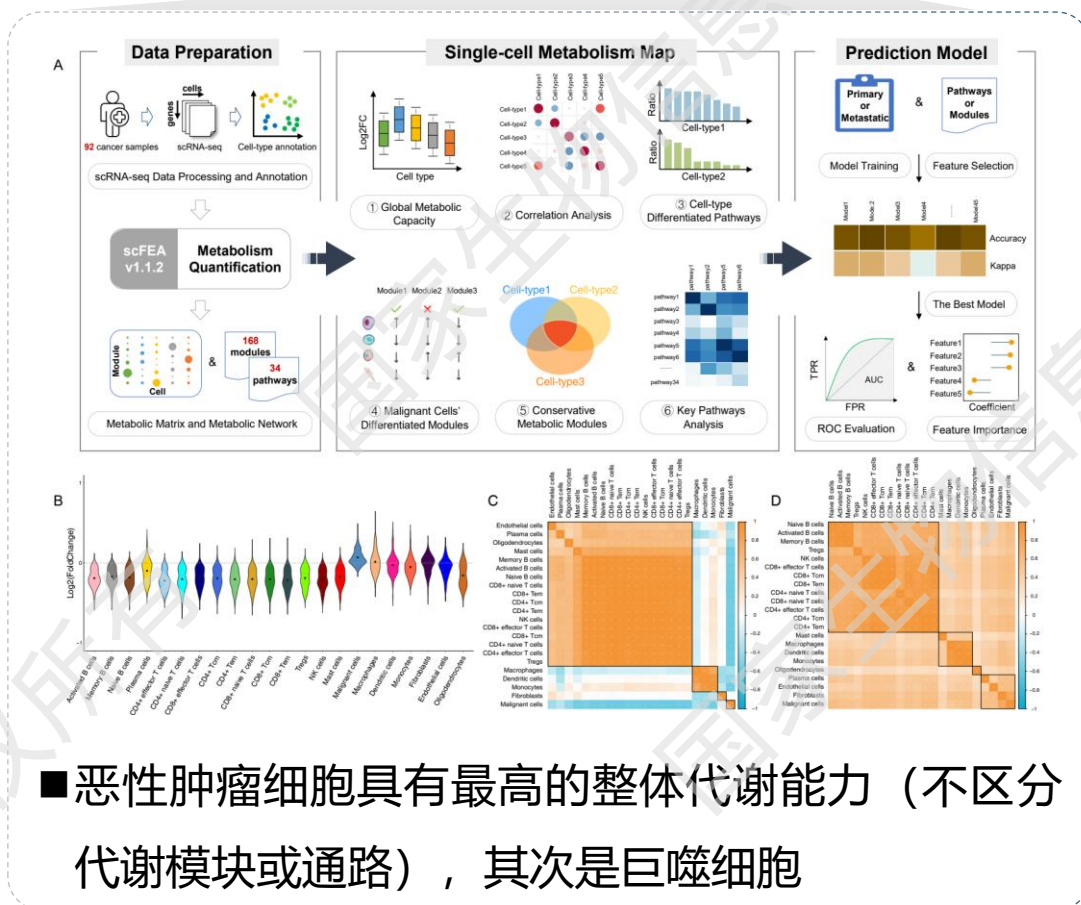
模块	功能
基因模块	1) 样本内基因表达分布
	2) 跨细胞类型基因表达分布
	3) 基因-基因表达相关性
	4) 单细胞与bulk水平表达
样本模块	1) 样本间细胞组成比较
	2) 细胞间通讯网络构建
	3) 基于TCGA的生存分析
细胞模块	1) 基因组拷贝数变异评估
	2) 转录因子富集分析
	3) 样本内拟时序构建
	4) 7项生物学特征打分
代谢模块	1) 原始代谢通量分布
	2) 细胞代谢动态变化追踪
	3) 代谢通路的两两相关性



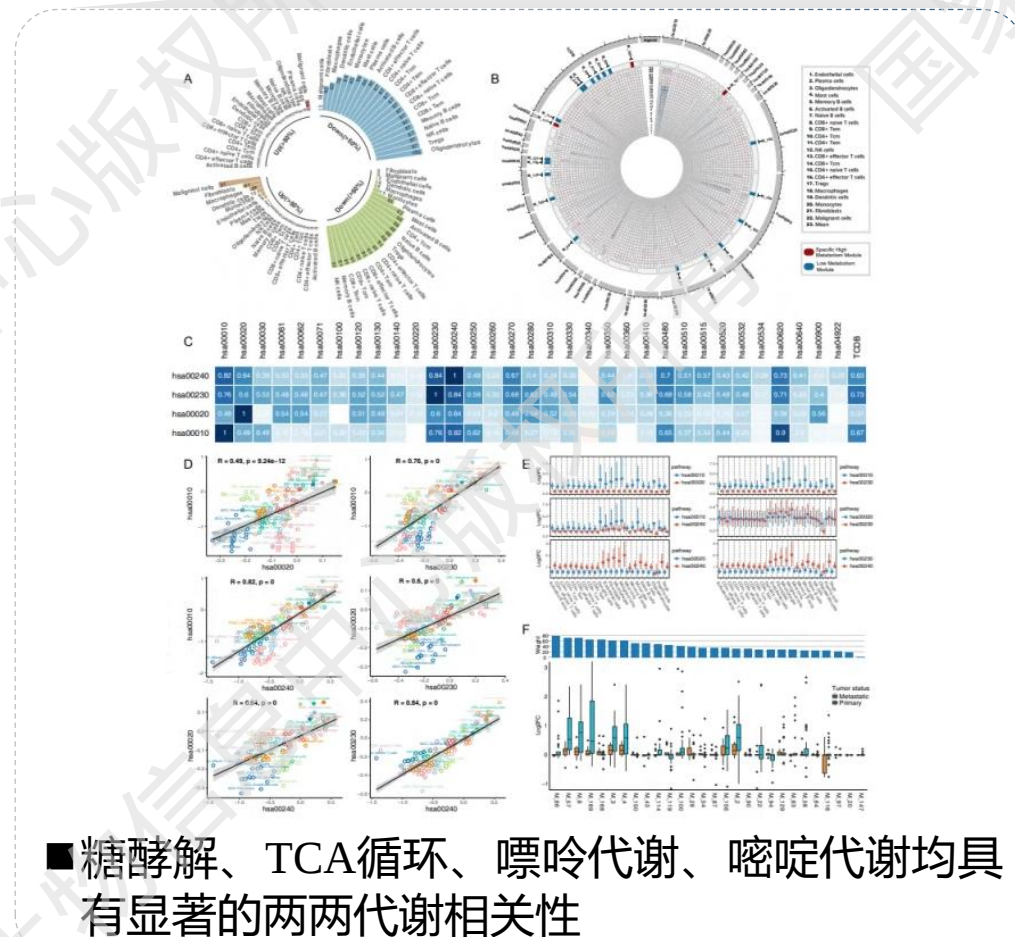
- 常用的免疫检查点分子**PDCD1**在基底细胞癌、肺腺癌、甲状腺未分化癌中的表达呈梯度下降趋势，暗示其在不同癌症类型中免疫治疗疗效的差异
- 对于甲状腺未分化癌，免疫检查点分子**TIGIT**在免疫细胞中的表达远远高于**PDCD1**，揭示其潜在的临床应用价值



不同细胞类型整体代谢能力评估

Shang Y et al., *Science Bulletin*. 2024

跨细胞类型代谢动态变化追踪



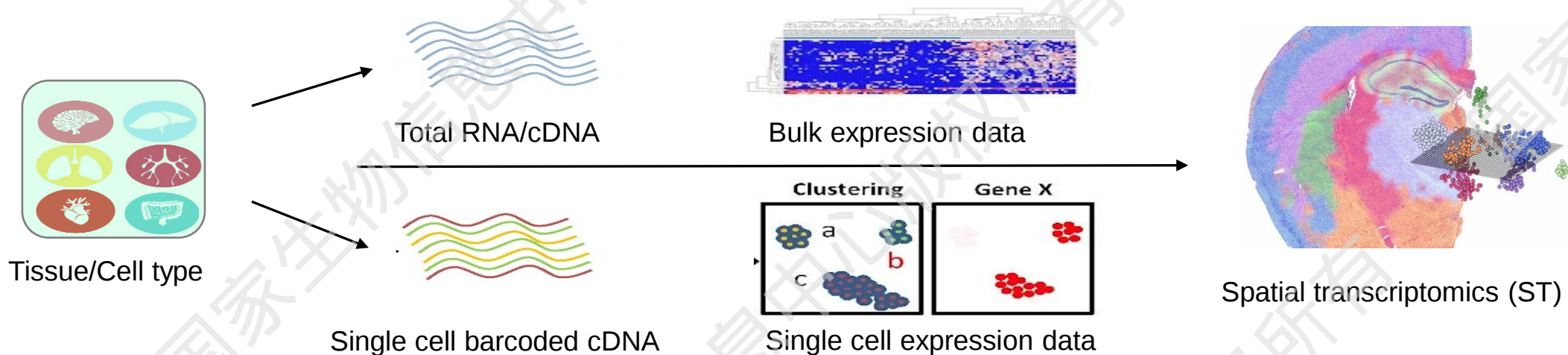
04

空间转录组综合资源存储库

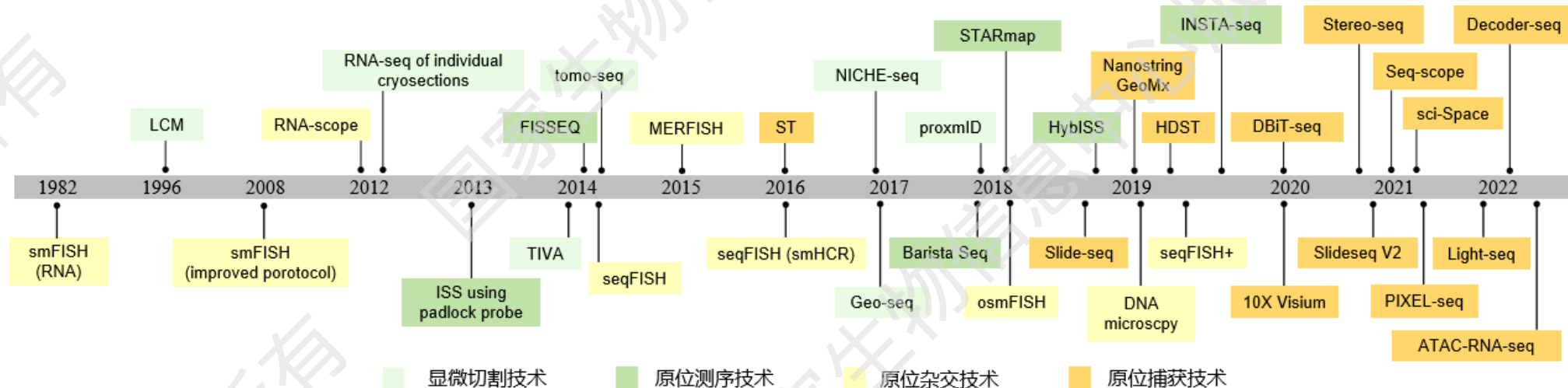
Comprehensive Repository of Spatial Transcriptomics (CROST)

空间转录组测序技术发展历程

- 细胞群体 (bulk) → 单细胞 → 空间转录组



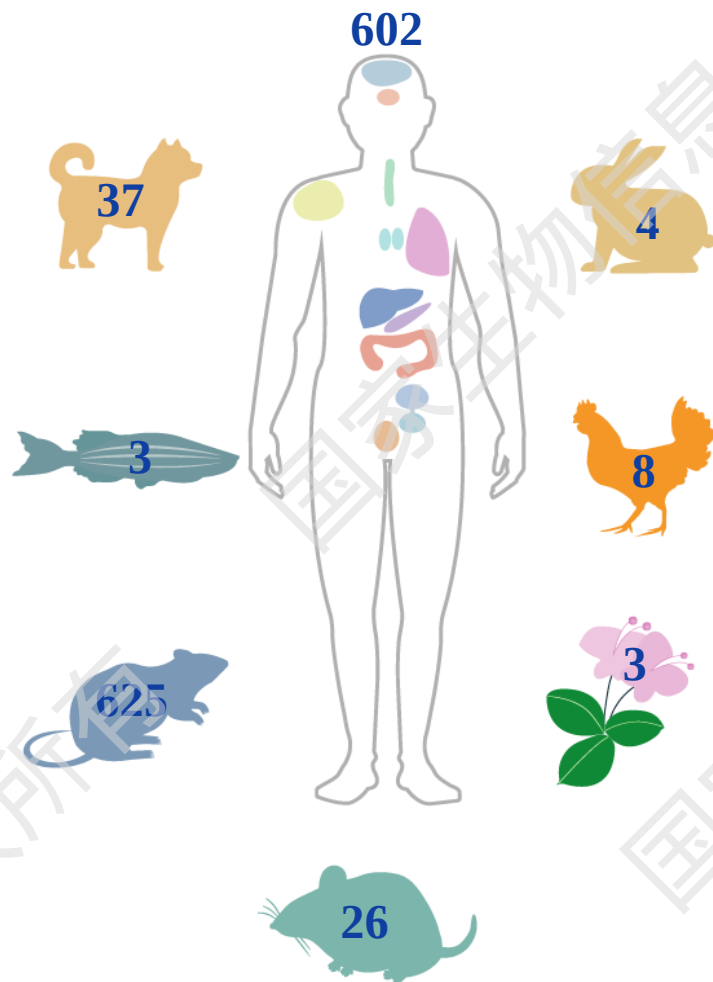
- 空间组学技术发展历程



空间转录组综合资源存储库CROST

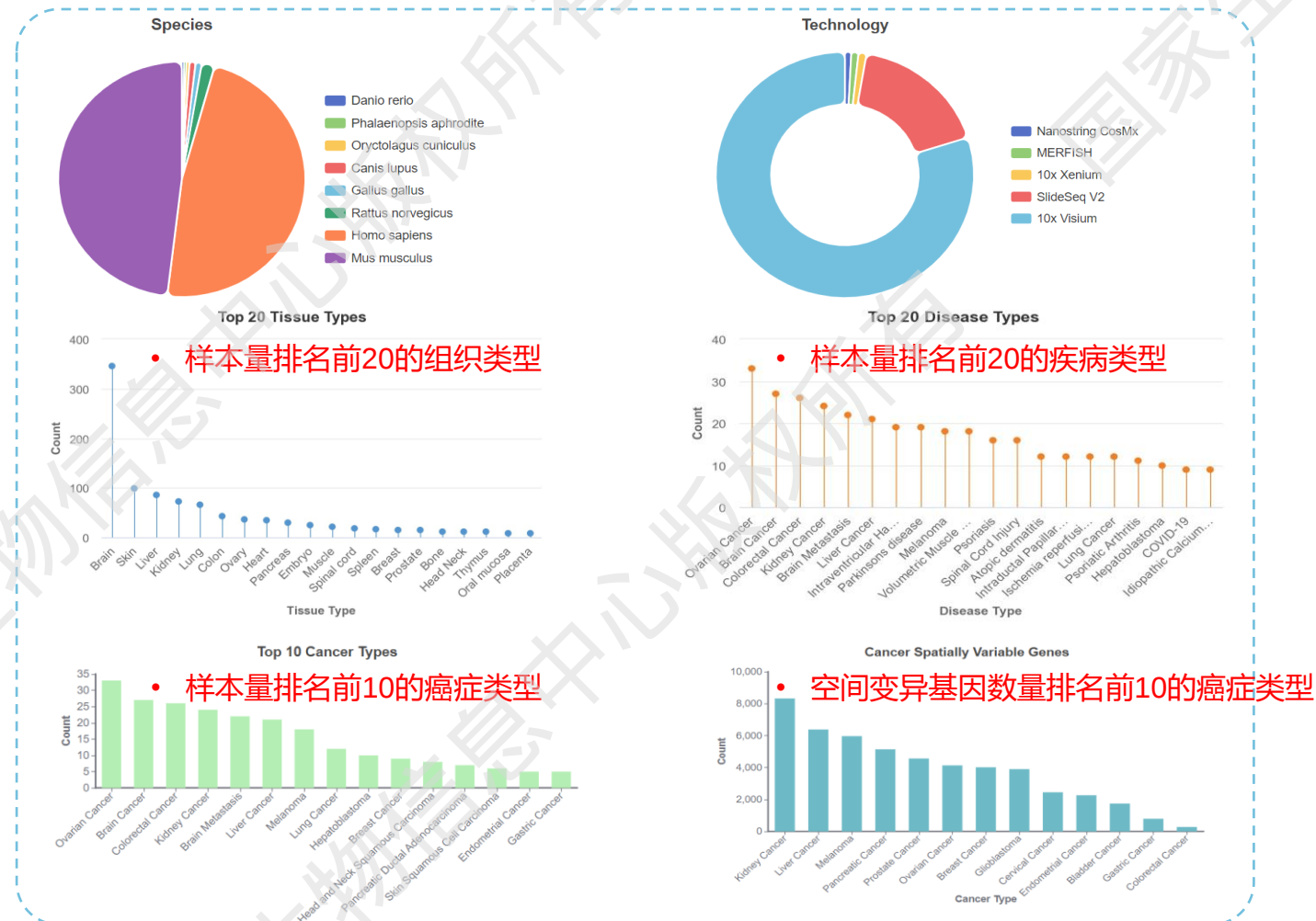


➤ 数据库多物种样本数统计



• 涵盖人类、小鼠、大鼠等多个物种

➤ 空间测序技术：10x Visium、SlideSeq V2、MERFISH等5种



CROST数据浏览模块



➤ 数据库浏览页面 – 数据集Dataset / 样本Sample / 文献Publication

• 提供数据筛选功能

• 筛选展示指定列

Technology	Dataset ID	Technology	Species	Data Resource	Project ID	Paired Project ID	PMID	Tissue/Cell Line	Sample Count	Paired Single Cell
10x Visium	VISDP000001	10x Visium	Mus musculus	CNCB	CRA000651	CRA002489	34341490	Liver	4	●
10x Visium	VISDP000002	10x Visium	Homo sapiens	NCBI	GSE130662	E-MTAB-7407	37157887	Liver	4	●
10x Visium	VISDP000009	10x Visium	Homo sapiens	NCBI	GSE144239	GSE144236	32579074	Skin	4	●
10x Visium	VISDP000010	10x Visium	Gallus gallus	NCBI	GSE146457	-	33741943	Heart	4	●
10x Visium	VISDP000011	10x Visium	Mus musculus	NCBI	GSE153424	GSE153424	35219624	Brain	8	●
10x Visium	VISDP000012	10x Visium	Mus musculus	NCBI	GSE153859	GSE108097	35837994	Brain	3	●
10x Visium	VISDP000013	10x Visium	Homo sapiens	NCBI	GSE158326	GSE158704	33406409	Colon, Small Intestine	8	●
10x Visium	VISDP000014	10x Visium	Mus musculus	NCBI	GSE158450	GSE158450	33489025	Brain	3	●
10x Visium	VISDP000015	10x Visium	Homo sapiens	NCBI	GSE159687	GSE142489	33458620	Prostate	2	●
10x Visium	VISDP000016	10x Visium	Danio rerio	NCBI	GSE159709	-	34725363	Skin	3	●
10x Visium	VISDP000017	10x Visium	Mus musculus	NCBI	GSE160137	-	33979190	Embryo	4	●
10x Visium	VISDP000018	10x Visium	Mus musculus	NCBI	GSE161318	GSE108097	34773081	Muscle	3	●
10x Visium	VISDP000019	10x Visium	Homo sapiens	NCBI	GSE162950	-	35416655	Embryo	7	●

Identification of HSC/MPP expansion units in fetal liver by single-cell spatiotemporal transcriptomics.

Suwei Gao, Qiang Shi, Yifan Zhang, Guixian Liang, Zhixin Kang, Baofeng Huang, Dongyuan Ma, Lu Wang, Jianwei Jiao, Xiangdong Fang, Cheng-Ran Xu, Longqi Liu, Xun Xu, Berthold Göttgens, Cheng Li, Feng Liu

Author Information

PMID: 34341490 DOI: 10.1038/s41422-021-00540-7

Abstract

Limited knowledge of cellular and molecular mechanisms underlying hematopoietic stem cell and multipotent progenitor (HSC/MPP) expansion within their native niche has impeded the application of stem cell-based therapies for hematological malignancies. Here, we constructed a spatiotemporal transcriptome map of mouse fetal liver (FL) as a platform for hypothesis generation and subsequent experimental validation of novel regulatory mechanisms. Single-cell transcriptomics revealed three transcriptionally heterogeneous HSC/MPP subsets, among which a CD93-enriched subset exhibited enhanced stem cell properties. Moreover, by employing integrative analysis of single-cell and spatial transcriptomics, we identified novel HSC/MPP 'pocket-like' units (HSC PLUS), composed of niche cells (hepatoblasts, stromal cells, endothelial cells, and macrophages) and enriched with growth factors. Unexpectedly, macrophages showed an 11-fold enrichment in the HSC PLUS. Functionally, macrophage-HSC/MPP co-culture assay and candidate molecule testing, respectively, validated the supportive role of macrophages and growth factors (MDK, PTN, and IGFBP5) in HSC/MPP expansion. Finally, cross-species analysis and functional validation showed conserved cell-cell interactions and expansion mechanisms but divergent transcriptome signatures between mouse and human FL HSCs/MPPs. Taken together, these results provide an essential resource for understanding HSC/MPP development in FL, and novel insight into functional HSC/MPP expansion ex vivo.

MeSH Term

- Animals
- Endothelial Cells
- Hematopoiesis
- Hematopoietic Stem Cells
- Liver
- Mice
- Transcriptome

Journal Article Research Support, Non-U.S. Gov't

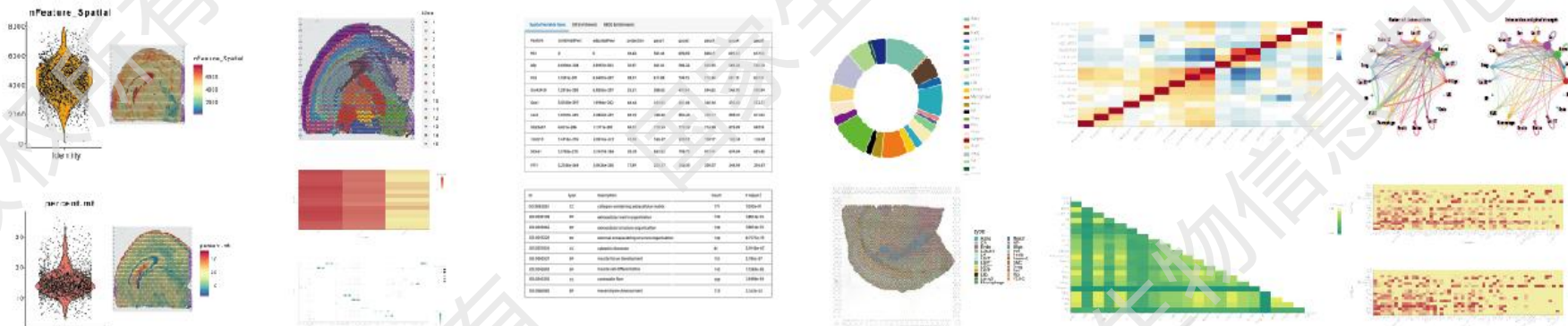


Word Cloud



• 抽提疾病、物种、基因等信息用于关键词

➤ 数据库浏览页面 – 单样本分析结果

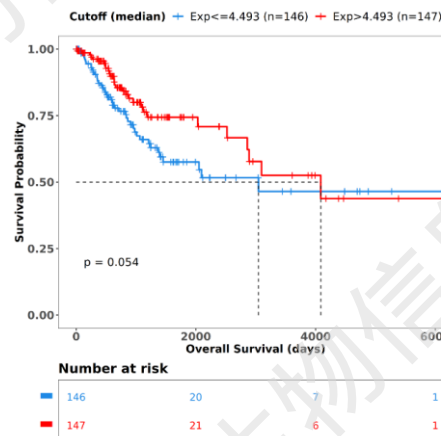
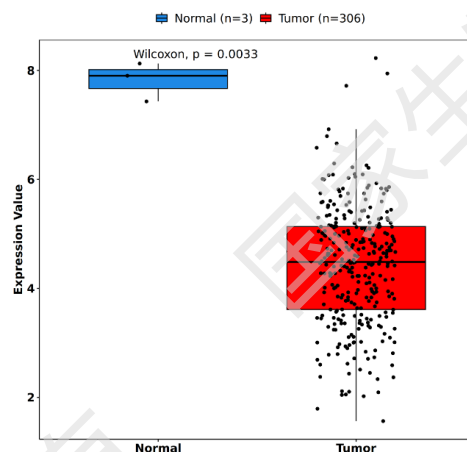
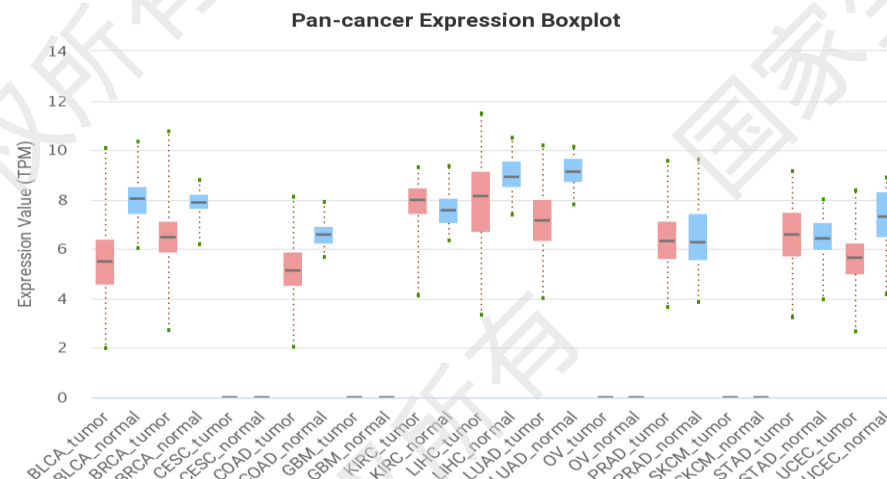
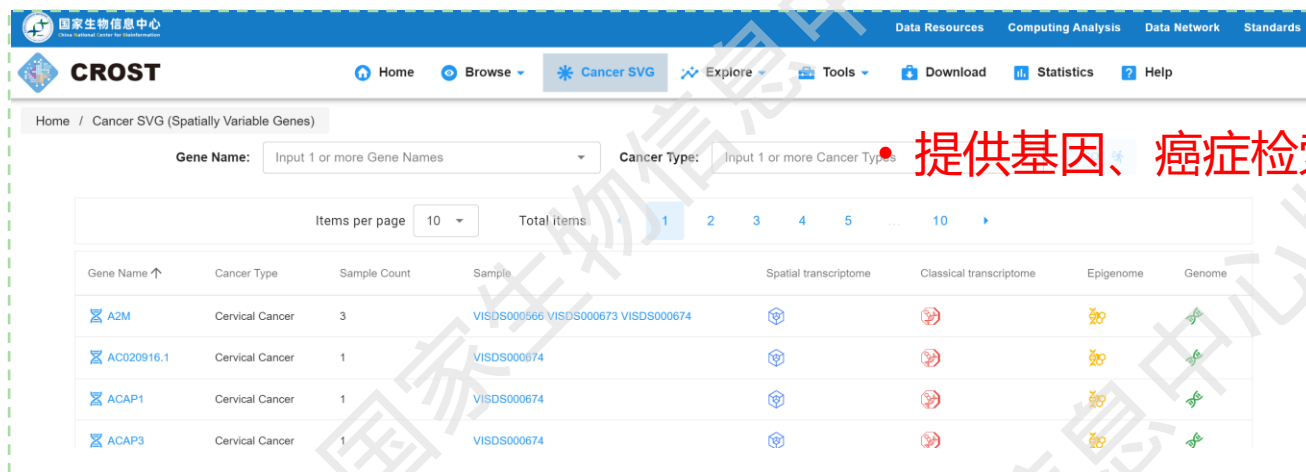


- 样本概述
- 维度 & 聚类分析
- 空间变异基因
- Spot解卷积
- 相关性 & 共定位分析
- 细胞通讯网络

CROST — Cancer SVG浏览



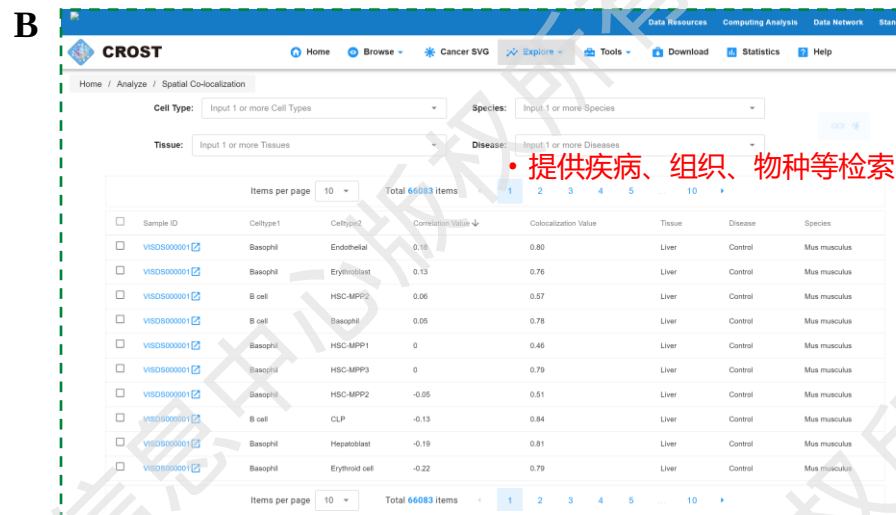
➤ 数据库功能模块介绍 – 癌症空间变异基因 (Cancer SVG)



- 识别SVG可为系统分析特定位置的细胞状态、推断细胞间通讯及确定生物体重要的表型和功能提供见解
- 模块展示肿瘤相关SVG的基本信息、泛癌RNA-seq表达、甲基化水平、泛癌基因组CNV以及发表文献列表

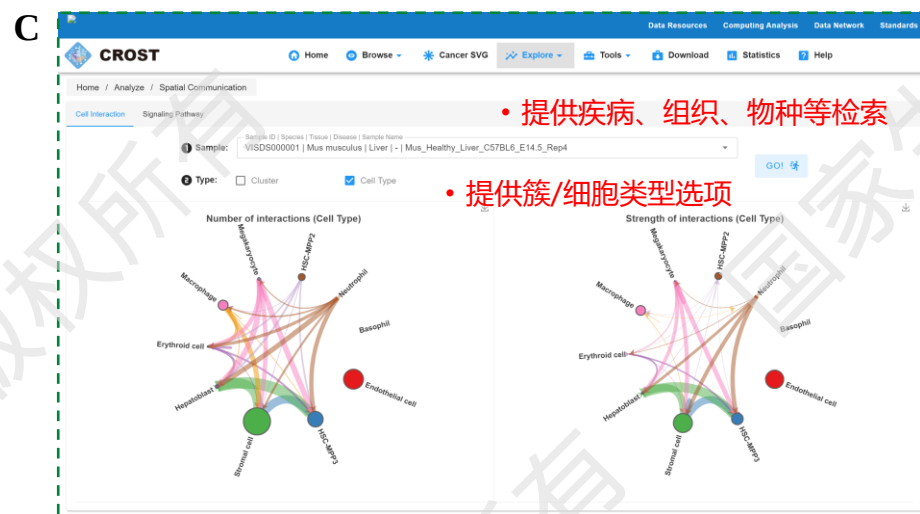


➤ 数据库功能模块介绍 – 数据探索 (Explore)



A. 空间转录组样本和配对单细胞转录组样本的交互式可视化

B. 细胞类型空间相关性和空间共定位数据查询



C. 空间聚类簇/细胞类型的全局通讯网络可视化分析

D. 空间聚类簇/细胞类型在特定信号通路的可视化分析

CROST在线分析工具



➤ 数据库功能模块介绍 – 在线分析工具

Toolkit --ssGSEA

gene_name	cluster1/ celltype1	cluster2/ celltype2	cluster3/ celltype3
gene1	0	0.18	0.40
gene2	0.33	4	7
gene3	0	2	8
gene4	5	0.8	1



- Cell State
- Gene Ontology
- TF Targets
- Chromosome Cytogenetic Bands
- KEGG Pathway
- Biological Process
- MicroRNA Targets



Toolkit --SpatialAP



Spatial Analysis Content



Home / Tools / ssGSEA

1 Upload a matrix file: [#Example File](#)

Note: Input 1 File (txt, csv, tsv). Support the maximum upload

Or, input the matrix: [#Example](#) [#Clear](#)

2 Species: ☐ Homo Sapiens ☒ Mus Musculus

3 Run: [Run Demo](#) [#Clear](#)

* This task is estimated to take approximately 1-2 minutes to complete. Once it has successfully finished, we will show you the results.

Query Job ID: [#Example](#) [#Clear](#)

- **ssGSEA**: 用于单样本的基因功能富集分析，涵盖细胞状态、转录因子、生物学过程等7类背景基因集，适用于空间转录组、单细胞转录组、经典转录组的分析

Home / Tools / Spatial Analysis Pipeline

1 Spatial File: ☒ CROST Data ☐ Upload Data

Sample ID | Species | Tissue | Disease | Sample Name

2 Single Cell File: ☒ CROST Data ☐ Upload Data

Species | Tissue | Disease

3 Species: ☒ Homo Sapiens ☐ Mus Musculus

4 Threshold Value: 20

Mitochondrion Percent: 0.001

Spatially Variable Gene q Value: 0.01

GO & KEGG q Value: 0.01

5 Email: name@email.com

6 Run: [Run Demo](#) [#Clear](#)

* This task is estimated to take approximately 2 hours to complete. Once it has successfully finished, we will notify you via email.

Query Job ID: [#Example](#) [#Clear](#)

- **SpatialAP**: 提供一站式的空间转录组学与单细胞转录组学数据分析功能，可满足空间转录组学分析的基本需求，包括空间变异基因分析、细胞类型注释、共定位与相关性分析、细胞通讯等

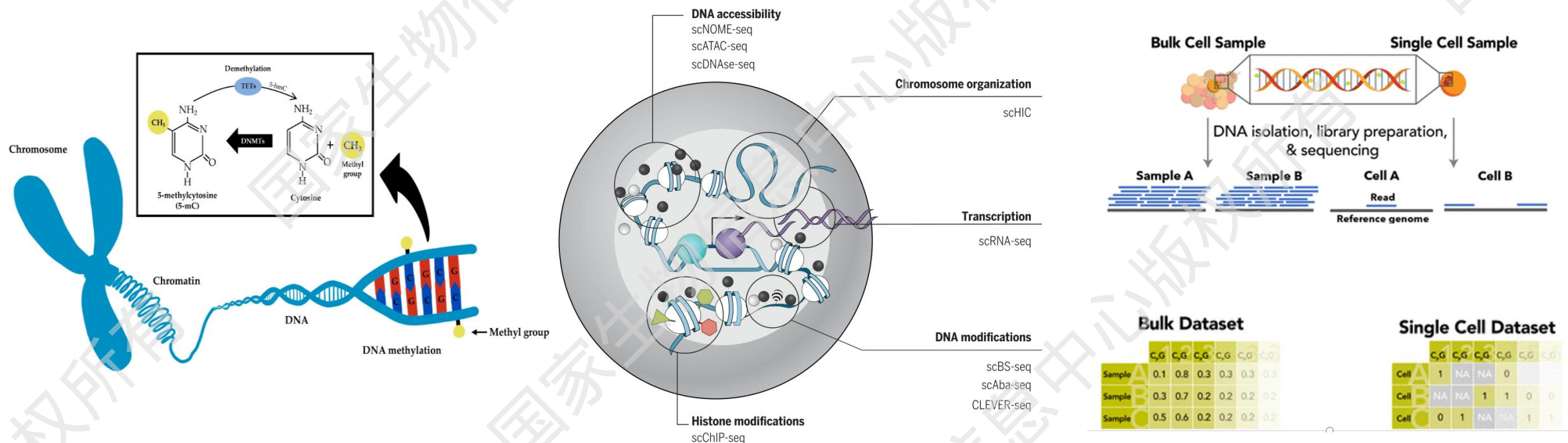
05

单细胞全基因组甲基化数据库

The single-cell Methylation Bank (scMethBank)

单细胞DNA甲基化测序

- DNA 甲基化 (DNA methylation)是最早被研究的重要的表观遗传修饰之一
- 单细胞DNA甲基化测序可以检测细胞水平单碱基的甲基化模式，在单个细胞中，甲基化状态只有两种



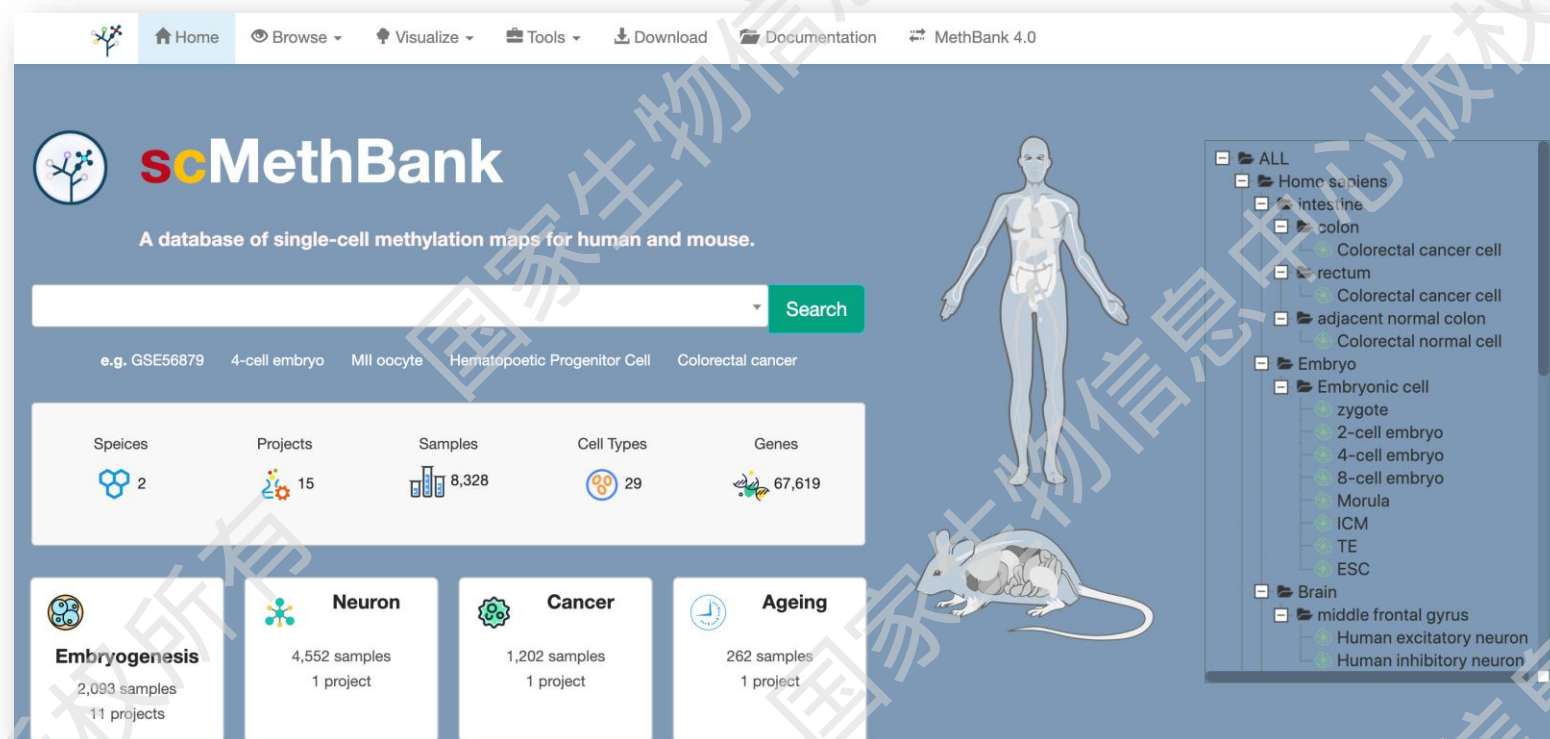
Valente et al., *Nanomaterials*. 2023

Kelsey G et al., *Science*. 2017

单细胞DNA甲基化数据库scMethBank



- 整合了多种类型和多个生物学场景下的高质量单核苷酸精度的单细胞DNA甲基化数据



15 个数据集

8,328 个单细胞甲基化图谱

29 种细胞类型

数据

实体组织细胞、早期胚胎细胞、癌细胞、生殖细胞... ..

6 类生物学场景

可视化

胚胎发育、细胞分化、疾病... ..

工具

- 单碱基精度甲基化图谱可视化浏览器
- 差异甲基化区域模式可视化
- 基于表观特征的细胞聚类
- 差异甲基化区域识别与富集分析
- 差异甲基化区域可视化工具

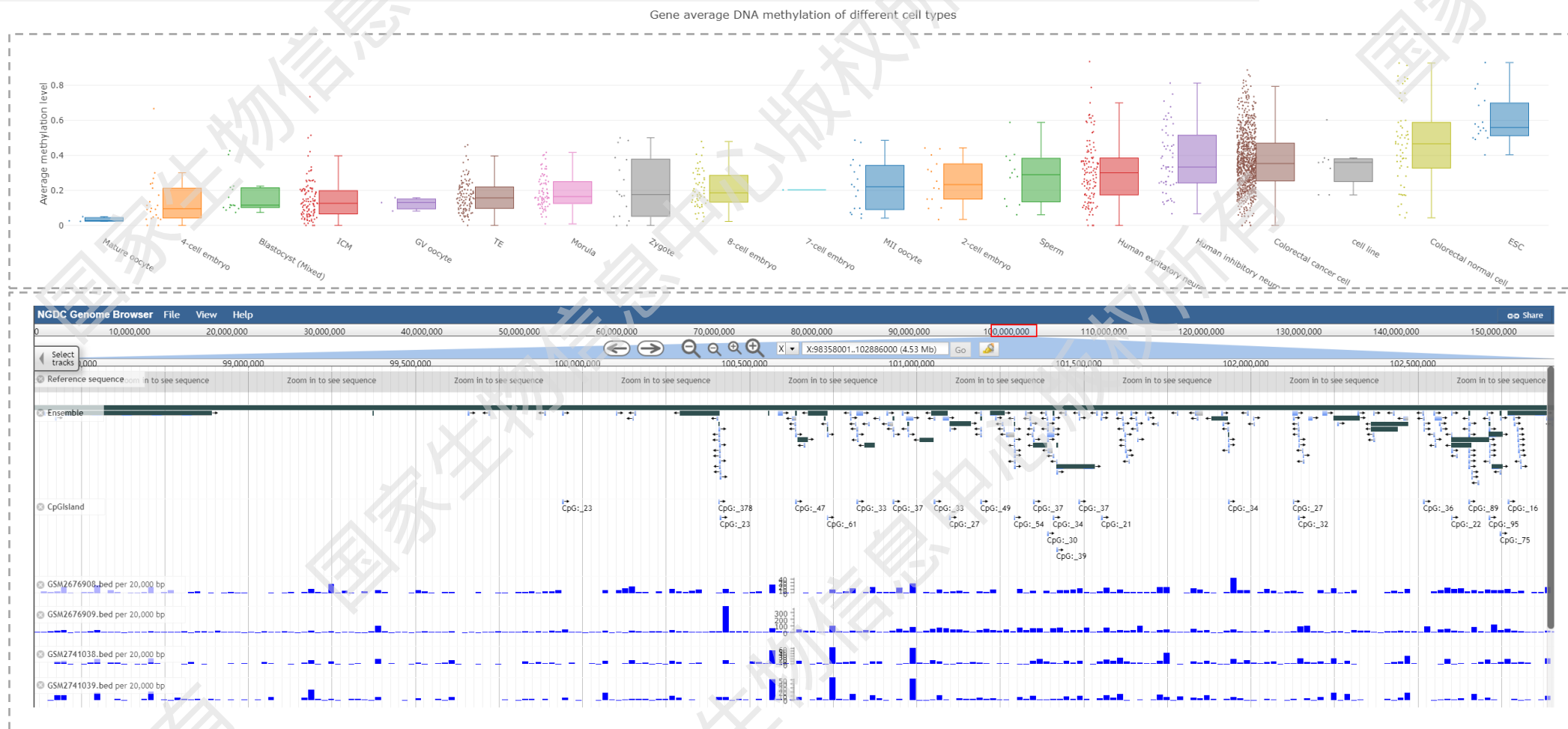
<https://ngdc.cncb.ac.cn/methbank/scm>, *Nucleic Acids Res.* 2022

基因在细胞中DNA甲基化水平查询



- 基因在不同细胞类型中呈现的甲基化模式反映细胞特异性及其调控模式

基因的平均
甲基化水平

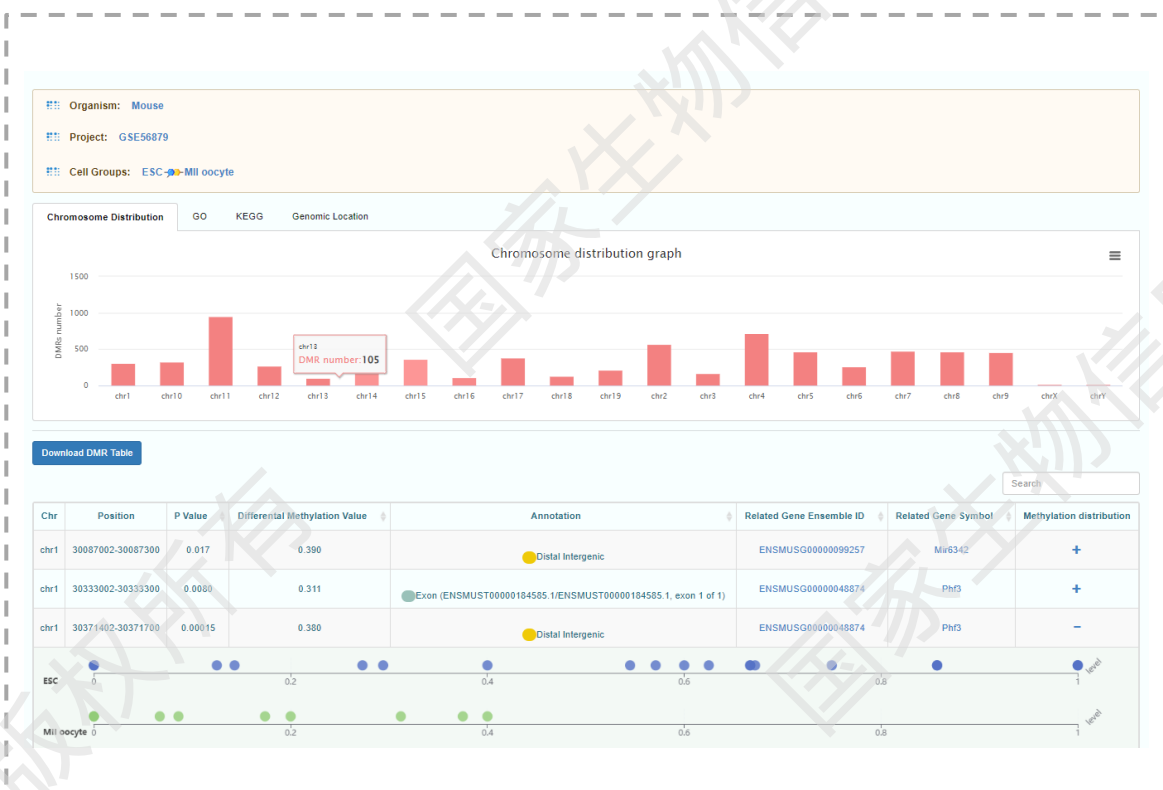


基因在基因组
浏览器中的甲
基化模式

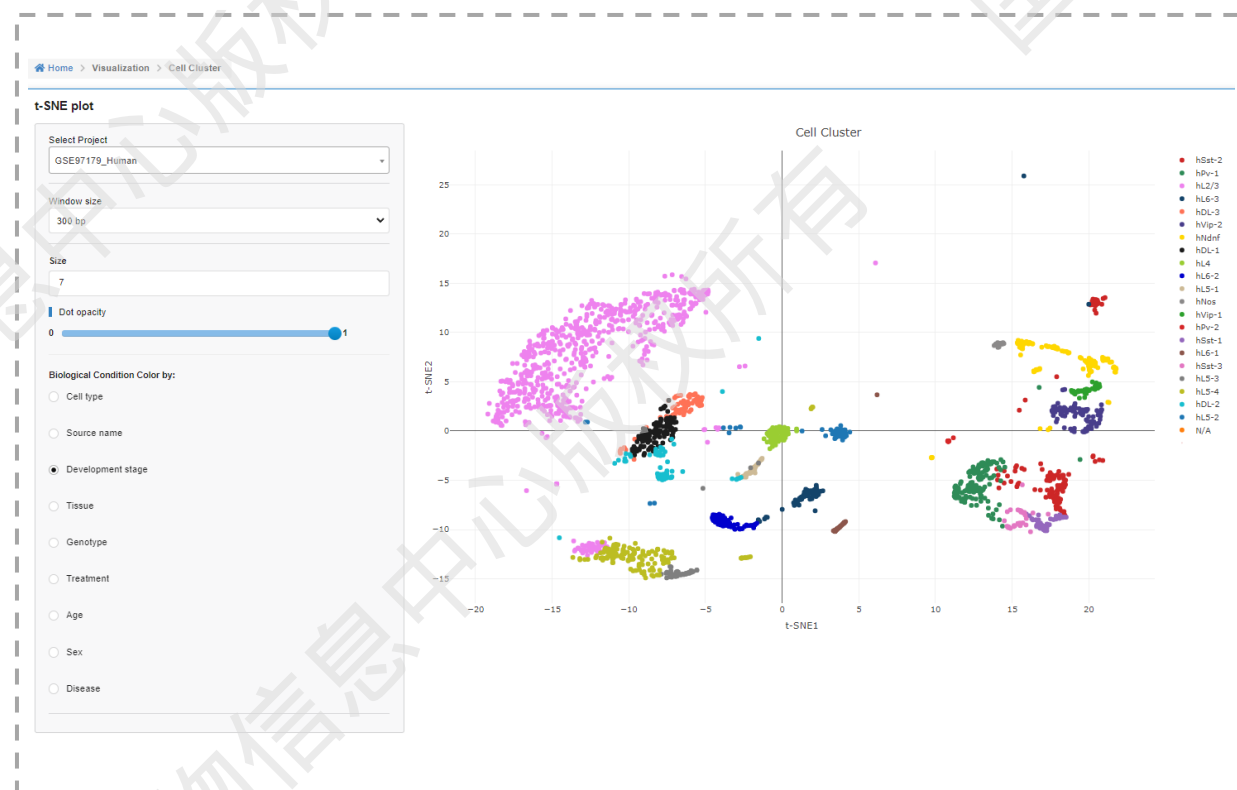
细胞类型特异的DNA甲基化图谱 (DMR分析)



- 提供细胞特异性的DNA甲基化图谱，不同细胞类型的单细胞甲基化谱可以与特定基因功能联系起来



细胞类型差异甲基化标记



单细胞图谱聚类

单核苷酸分辨率的DNA甲基化图谱



■ 单核苷酸分辨率查询DNA甲基化模式揭示了高精度的细胞异质性表现

- 优化的数据存储和索引方法
- 允许在任意基因组区域上快速定位和查询单个碱基的甲基化状态

- 基于基因和启动子
- 基于基因组区域

Select Gene

Select Region

Gene Name/Ensemble ID

NANOG

☒ Gene Body

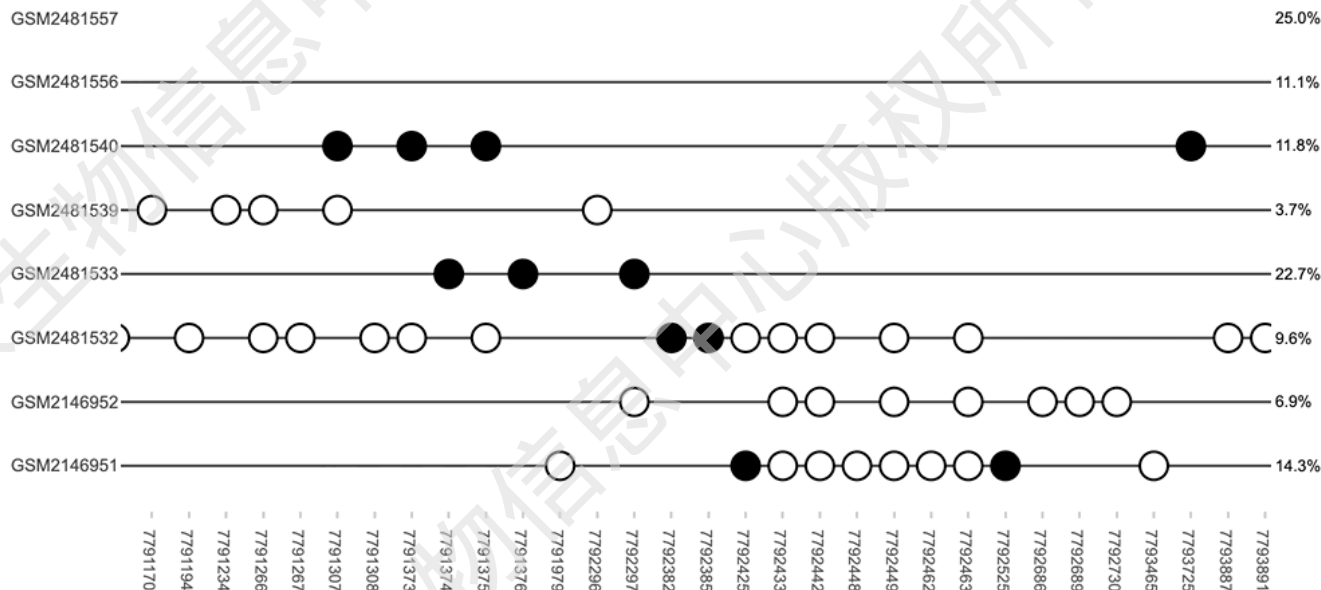
☐ Promoter

☐ Promoter & Body

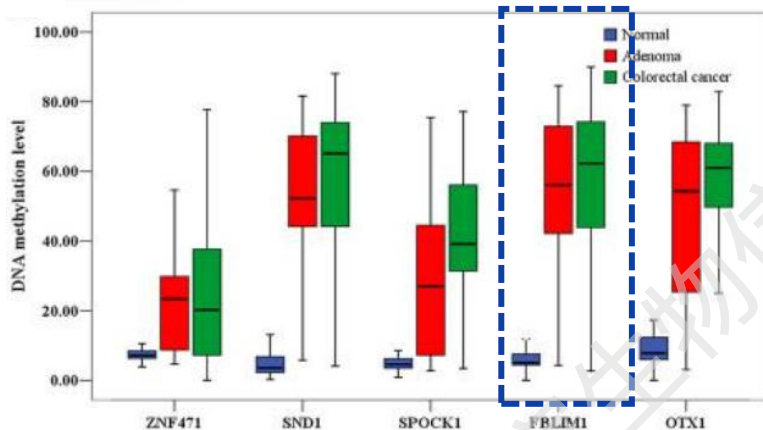
Run

Reload

Region Methylation Pattern Per Sample



单核苷酸精度的甲基化查询细胞类型间具体的甲基化模式差异



基因和调控区域的单碱基查询

Select Gene

Select Region

Gene Name/Ensemble ID

FBLIM1

☐ Gene Body

☐ Promoter

☒ Promoter & Body

TSS upstream:

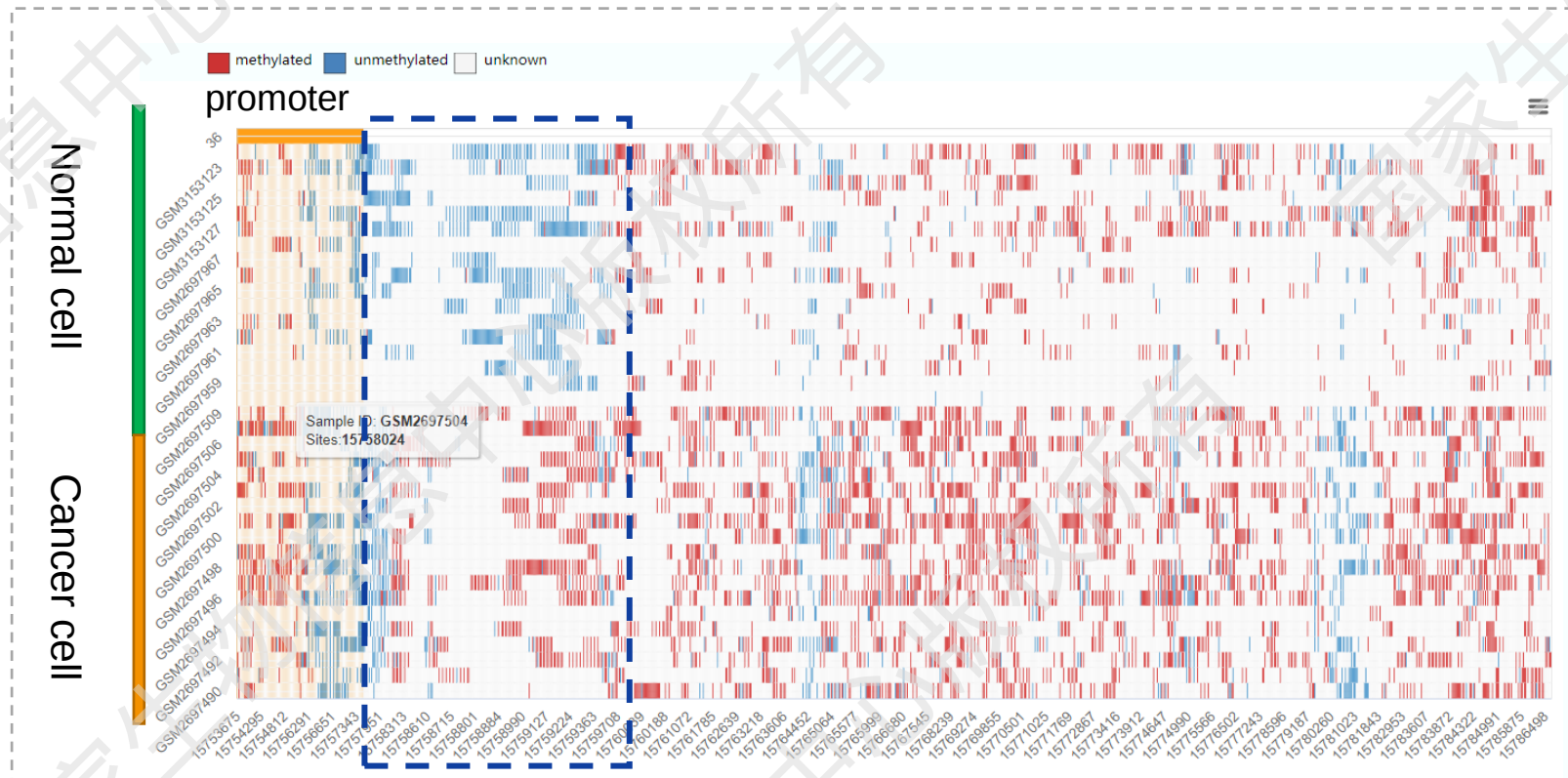
3000

TSS downstream:

1000

Run

Reload



- FBLIM1转录起始位点周围的胞嘧啶在癌症细胞和正常细胞之间存在明显的差异甲基化现象
- 基于scMethBank单碱基模式可以看到具体的差异甲基化模式发生在TSS下游而非整个基因区域

06

细胞分类学知识库

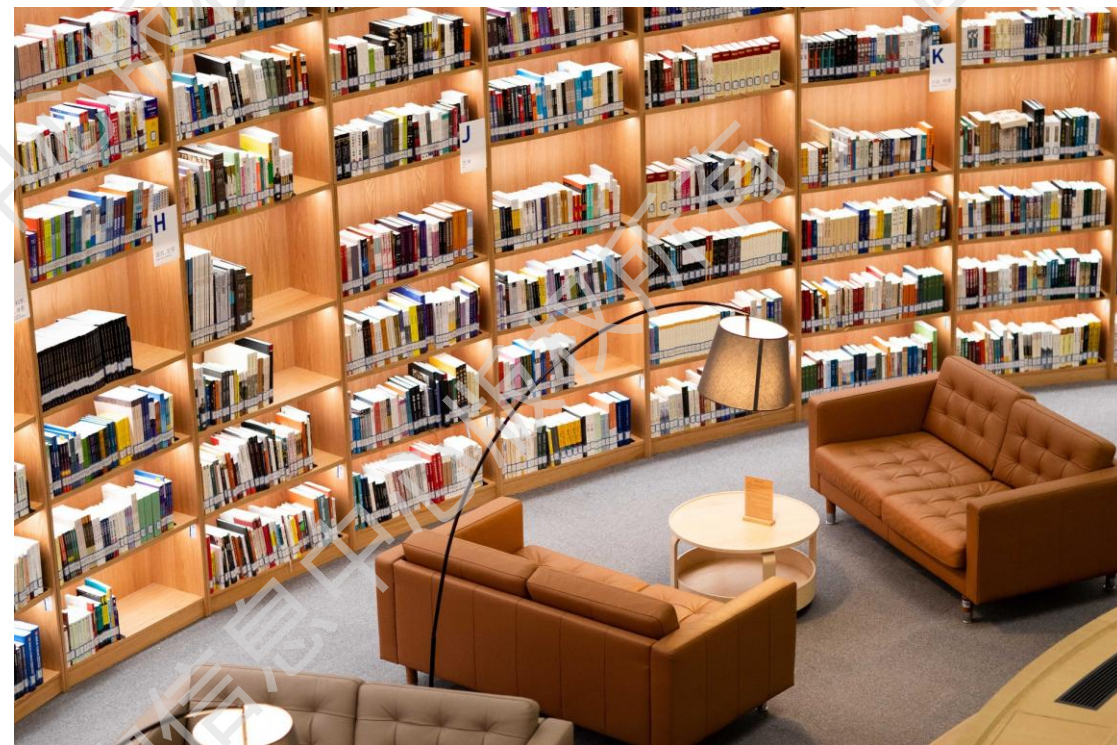
Cell Taxonomy

Cell Taxonomy: 解码细胞身份的生命“图书馆”

- 2015年后，单细胞技术让科学家每天发现数百万个“新细胞”，但如何区分它们？就像一堆混乱的“书本”，没有标准的分类



杂乱无章的“书本”



整齐分类的“书架”

当科学撞上“细胞身份危机”



Cell Taxonomy应运而生——解决从单细胞测序大爆炸到标准化分类的迫切需求

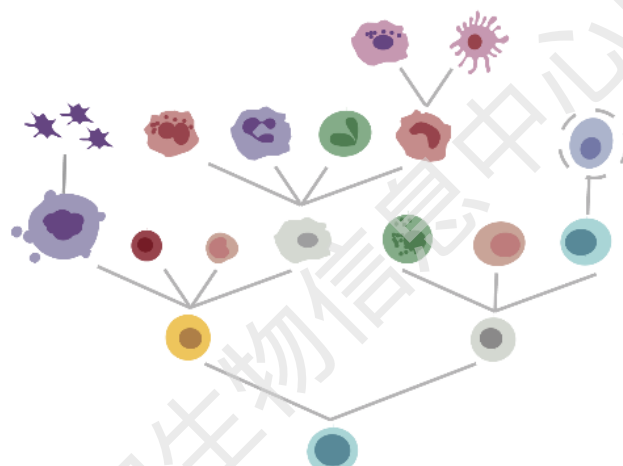
A curated repository of cell types and associated characteristics

Browse tree ?

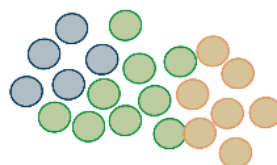
Search

- Cell
 - Native cell
 - Germ line cell
 - Ciliated cell
 - Circulating cell
 - Secretory cell
 - Contractile cell
 - Sensory receptor cell
 - Electrically active cell
 - Absorptive cell
 - Barrier cell
 - Motile cell
 - Anucleate cell
 - Eukaryotic cell
 - Structural cell
 - Stuff accumulating cell
 - Oxygen accumulating cell
 - Polyploid cell
 - Haploid cell
 - Diploid cell

3,143 cell types



146 scRNA-seq studies



26,613 cell markers

Cell surface markers

Drug targets

257 conditions

4,299 publications



387 tissues



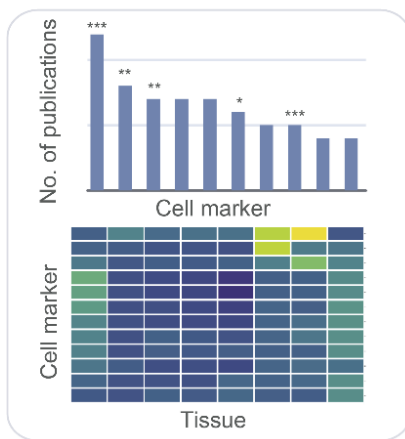
34 species

构建细胞世界的“宪法体系”

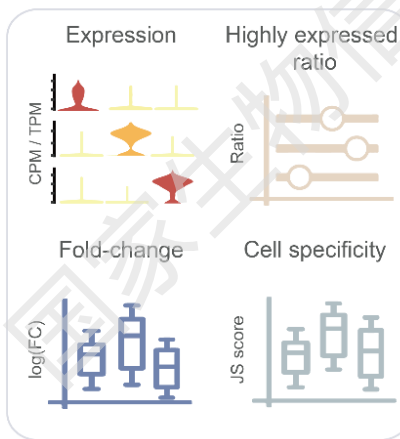


Cell Taxonomy能够从多方面系统表征细胞类型特征

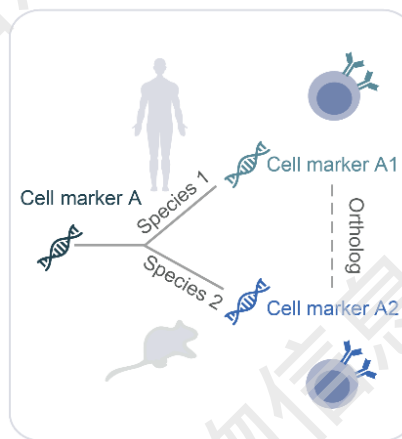
Supported publications



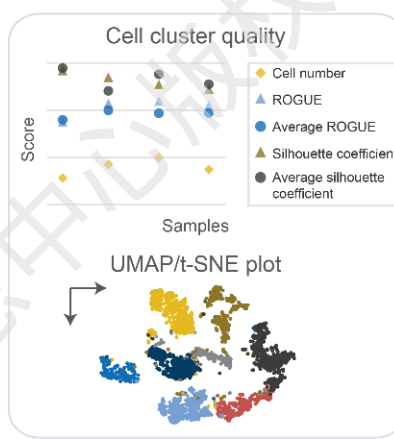
Cell marker enrichment



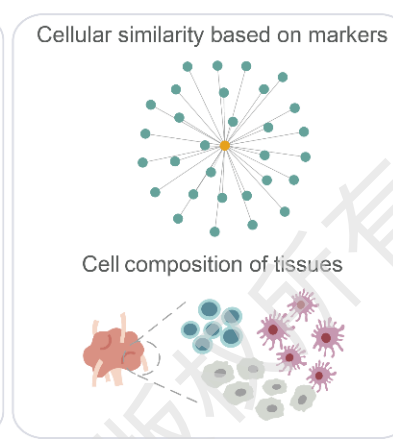
Cross-species comparison



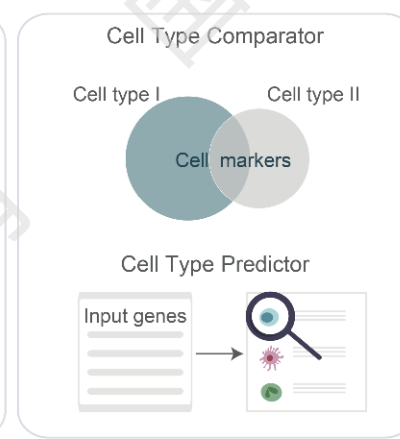
Cell cluster evaluation



Cellular characterization



Analysis tools



细胞标志物评估

细胞群质量评估

跨物种比较

细胞组分刻画

细胞相似性定量

在线分析工具

如果您发现一个细胞团簇，会如何为它命名？——现在有标准答案了。

Cell Taxonomy: 从"手工分类"到"智能导航"



□ 使用Cell Taxonomy, 输入3个Marker基因, 秒获候选细胞类型

研究生小王用了3周比对10篇文献标记细胞类型



Cell Type Predictor: predict cell types according to gene list

Input gene list Load from disk

Species
Homo sapiens

Example Run

Results

Cell type	Cell Taxonomy ID	Similarity score	Matched genes
Exhausted T cell	CT:00001844	1.00	CLTA4, CXCL13, HAVCR2, PDCD1, TIGIT
Exhausted CD4+ T cell	CT:00000644	0.800	CLTA4, CXCL13, PDCD1, TIGIT
Exhausted regulatory T cell	CT:00002612	0.800	CLTA4, HAVCR2, PDCD1, TIGIT
Exhausted CD4+ effector T-cell	CT:00002620	0.600	CLTA4, CXCL13, PDCD1
Megakaryocyte-erythroid progenitor cell	CT:00000054	0.400	CLTA4, PDCD1
T-helper 1 cell	CT:00000502	0.400	CLTA4, CXCL13
Regulatory T cell	CT:00000825	0.400	CLTA4, TIGIT
Tr1 cell	CT:00000907	0.400	CLTA4, PDCD1



□ 您是哪种细胞分类者？

“新手用户”

- 请分别找到含有最多cell marker的Top 10细胞类型、在不同细胞类型分布最广泛的Top 10的cell markers以及含有最多cell type和cell marker的tissue

“入门用户”

- 能够较好的理解和利用细胞类型、marker基因或组织等详情页面的说明及统计学图表

“进阶用户”

- 单细胞数据集细胞类型分类实战（推荐使用真实数据集，需要熟悉单细胞常规分析流程）

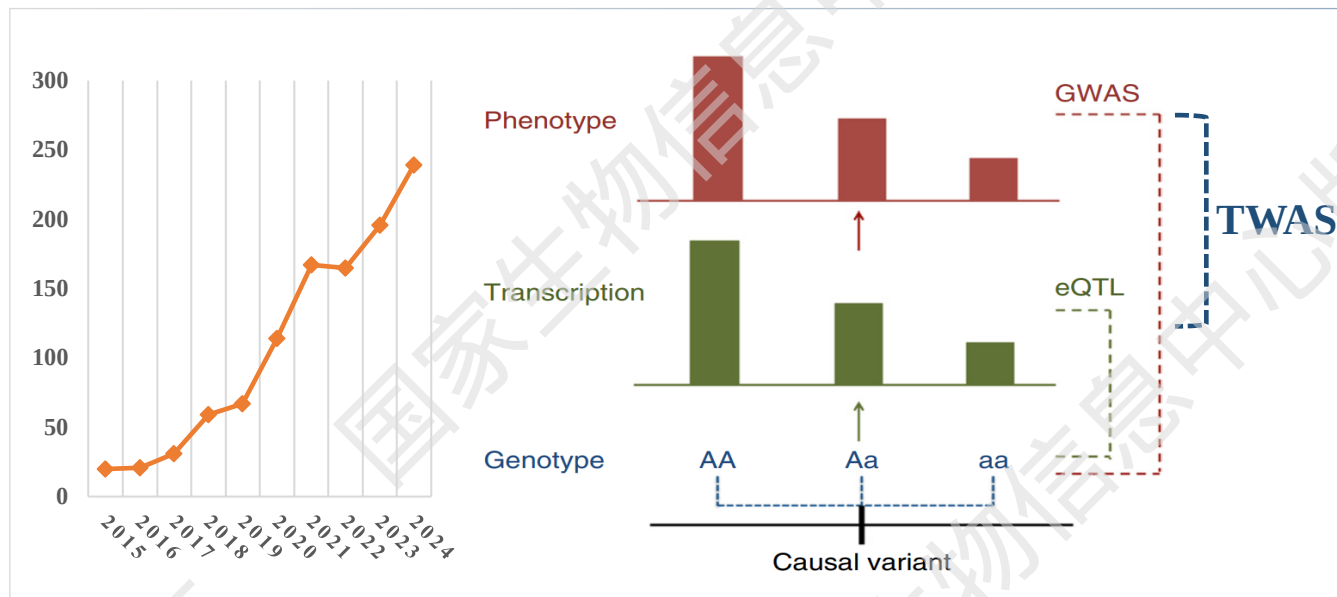
07

单细胞全转录组关联知识库

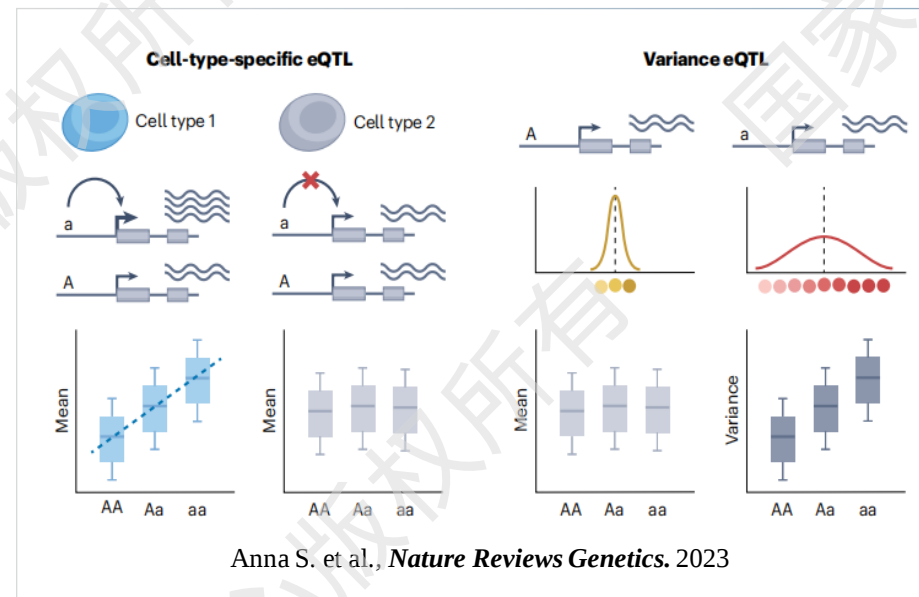
Single-cell TWAS Atlas (scTWAS Atlas)

全转录组关联研究逐渐兴起

全转录组学关联研究文献稳定增长



单细胞转录调控特异性



问题与挑战

- 全转录组关联研究 (TWAS) 广泛开展, 尚缺乏经人工审编的高质量TWAS关联图谱知识库
- 研究发现基因的转录调控具有细胞特异性, 然而目前尚缺乏单细胞层面的TWAS分析方法, 用于细胞特异性的TWAS关联基因的鉴定

单细胞全转录组关联知识库scTWAS Atlas



- **研究案例：** *ZNF703*等基因首次报道在乳腺上皮细胞和基质细胞中的表达与乳腺癌显著相关(Song et al., 2023); 在抑郁症研究中, *NEGR1*基因在不同细胞亚型中的表达对疾病风险的影响方向相反(Zeng et al., 2024)
- 团队开发的单细胞全转录组关联知识库scTWAS Atlas, **结合数据分析与文献审编**, 为科研人员提供高质量的单细胞转录组关联研究数据与分析工具

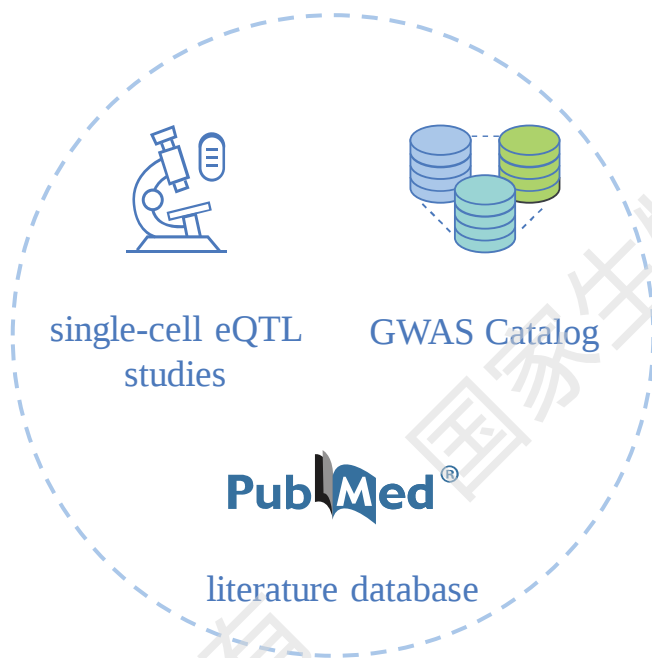


<https://ngdc.cncb.ac.cn/sctwas>, *Nucleic Acids Res.* 2025

scTWAS Atlas核心数据



● 研究策略



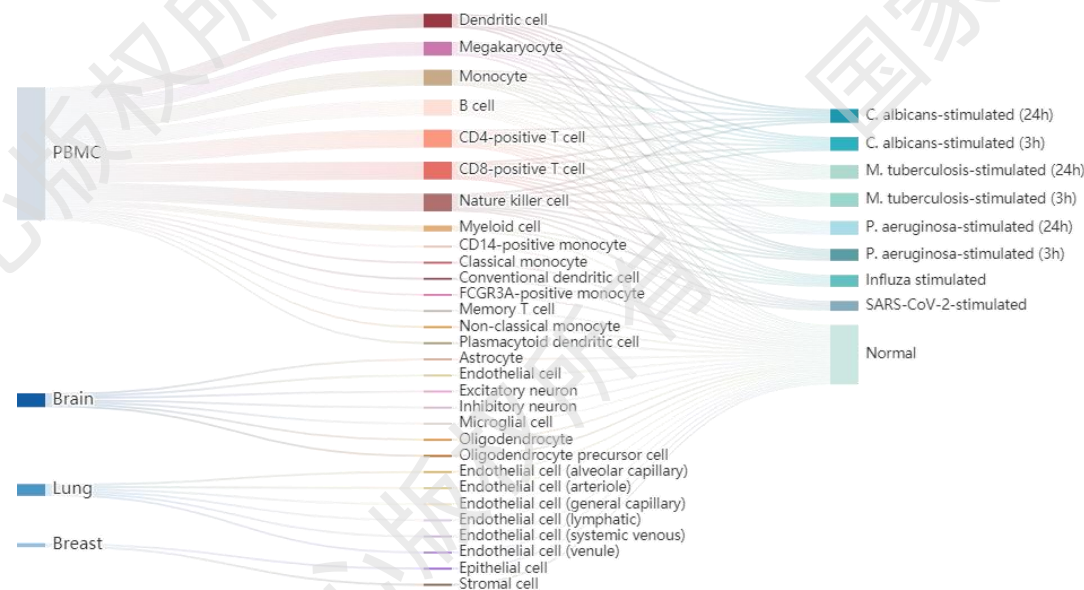
- 10个单细胞eQTL数据集
- 15个GWAS数据集
- 5篇单细胞TWAS研究

● 表型本体及分类



- 17种表型大类
- 34种性状表型

● 细胞类型本体



- 9种人类组织
- 30种细胞类型
- 9种细胞状态



➤ 如何查看性状/细胞类型/基因的scTWAS关联总体情况？

Home

Trait

Cell Type

Gene

Visualization

Analysis

Download

Documentation

Filters

Trait Type

Show all

Trait Label

Show all

GWAS Dataset Sample Size

☐ >100,000

☐ 10,000 - 100,000

Strategy

☐ Literature Curation

☐ Data Analysis

Trait Label	Trait Type	Associated Cell Types	#Associated Genes	GWAS Datasets
Alzheimer's Disease	Nervous System Diseases	B cell, CD14-positive monocyte, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Conventional dendritic cell, Dendritic cell, Endothelial cell (alveolar capillary), Endothelial cell (lymphatic), Endothelial cell (systemic venous), Endothelial cell (venule), Megakaryocyte, Memory T cell, Monocyte, Myeloid cell, Nature killer cell, Non-classical monocyte, Plasmacytoid dendritic cell	124	GCST90027158; Thompson et al. 2022 Nat Commun
Anemia	Hemic and Lymphatic Diseases	B cell, CD14-positive monocyte, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Conventional dendritic cell, Dendritic cell, Endothelial cell (alveolar capillary), Endothelial cell (arteriole), Endothelial cell (general capillary), Endothelial cell (lymphatic), Endothelial cell (systemic venous), Endothelial cell (venule), FCGR3A-positive monocyte, Megakaryocyte, Memory T cell, Monocyte, Myeloid cell, Nature killer cell	10610	GCST90077851
Arthritis	Musculoskeletal Diseases	B cell, CD4-positive T cell, CD8-positive T cell, Dendritic cell, Endothelial cell (lymphatic), Endothelial cell (systemic venous), Endothelial cell (venule), Megakaryocyte, Monocyte, Nature killer cell	52	GCST90080404
Asthma	Respiratory Tract Diseases	B cell, CD14-positive monocyte, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Conventional dendritic cell, Dendritic cell, Endothelial cell (alveolar capillary), Endothelial cell (arteriole), Endothelial cell (general capillary), Endothelial cell (lymphatic), Endothelial cell (systemic venous), Endothelial cell (venule), FCGR3A-positive monocyte, Megakaryocyte, Memory T cell, Monocyte, Myeloid cell, Nature killer cell, Non-classical monocyte, Plasmacytoid dendritic cell	2591	GCST007799; Thompson et al. 2022 Nat Commun
Bipolar Disorder	Mental Disorders	B cell, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Conventional dendritic cell, Nature killer cell, Non-classical monocyte, Plasmacytoid dendritic cell	50	Thompson et al. 2022 Nat Commun
Breast Cancer	Skin and Connective Tissue Diseases; Neoplasms	Epithelial cell, Stromal cell	11	Song et al. 2023 Nat Commun
Chronic Kidney Disease	Urogenital Diseases	B cell, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Nature killer cell, Non-classical monocyte, Plasmacytoid dendritic cell	10	Thompson et al. 2022 Nat Commun
Coronary Artery Disease	Cardiovascular Diseases	B cell, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Conventional dendritic cell, Nature killer cell, Plasmacytoid dendritic cell	13	Thompson et al. 2022 Nat Commun
COVID-19	Respiratory Tract Diseases	B cell, CD14-positive monocyte, CD4-positive T cell, CD8-positive T cell, Classical monocyte, Conventional dendritic cell, Dendritic cell, Endothelial cell (alveolar capillary), Endothelial cell (arteriole), Endothelial cell (general capillary), Endothelial cell (lymphatic), Endothelial cell (systemic venous), Endothelial cell (venule), FCGR3A-positive monocyte, Megakaryocyte, Memory T cell, Monocyte, Myeloid cell, Nature killer cell	8478	GCST011079

性状筛选



如何检索【银屑病】？

Quick search ① 快速检索

② 高级检索

Advanced search

Search by trait, cell type or gene (e.g. Heart Failure, CD4-positive T cell, DR1)

Home / Advanced Search

Trait:

Trait Label:

Arthritis

Psoriasis

Depression

Sarcoidosis

Cell Type:

Tissue Type:

Cell Type:

Gene:

Gene Symbol:

Gene Ensembl ID:

Reset

Search

结果1：【银屑病】的总体描述

Summary information	
Trait Label:	Psoriasis
MeSH ID:	D011565
Ontology ID:	NCIT:C3346
MeSH Type:	Skin and Connective Tissue Diseases
Synonyms:	Psoriasis; Pustulosis of Palms and Soles; Pustulosis Palmaris et Plantaris; Palmoplantar Pustulosis; Pustular Psoriasis of Palms and Soles
Description:	A common genetically determined, chronic, inflammatory skin disease characterized by rounded erythematous, dry, scaling patches. The lesions have a predilection for nails, scalp, genitalia, extensor surfaces, and the lumbosacral region. Accelerated epidermopoiesis is considered to be the fundamental pathologic feature in psoriasis.

结果2：【银屑病】的scTWAS关联

TWAS Associations												
Cell type		Cell condition										
Show all		Show all		Download								
Show 10 entries												
Trait	Cell Type	Cell Condition	Tissue	Strategy	Gene	Ensembl ID	Model	TWAS.Z	TWAS.P	TWAS.FDR	eQTL Dataset	GWAS Dataset
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000230074	ENSG00000230074	PRScs	5.90	3.72e-09	3.73e-07	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000232876	ENSG00000232876	PRScs	6.13	8.92e-10	9.78e-08	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000234902	ENSG00000234902	PRScs	3.86	1.11e-04	4.50e-03	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000245869	ENSG00000245869	PRScs	3.80	1.45e-04	5.54e-03	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000247121	ENSG00000247121	PRScs	-4.30	1.74e-05	9.01e-04	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000261505	ENSG00000261505	PRScs	-3.35	8.19e-04	2.40e-02	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000263080	ENSG00000263080	PRScs	-3.83	1.27e-04	5.00e-03	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000271858	ENSG00000271858	PRScs	-3.60	3.19e-04	1.09e-02	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000272182	ENSG00000272182	PRScs	-3.33	8.61e-04	2.50e-02	Oelen et al. 2022 Nat Commun	GCST005527
Psoriasis	B cell	C. albicans-stimulated (24h)	PBMC	Analysis	ENSG00000272221	ENSG00000272221	PRScs	-6.57	5.03e-11	6.44e-09	Oelen et al. 2022 Nat Commun	GCST005527

Showing 1 to 10 of 29,598 entries

Previous 1 2 3 4 5 ... 296





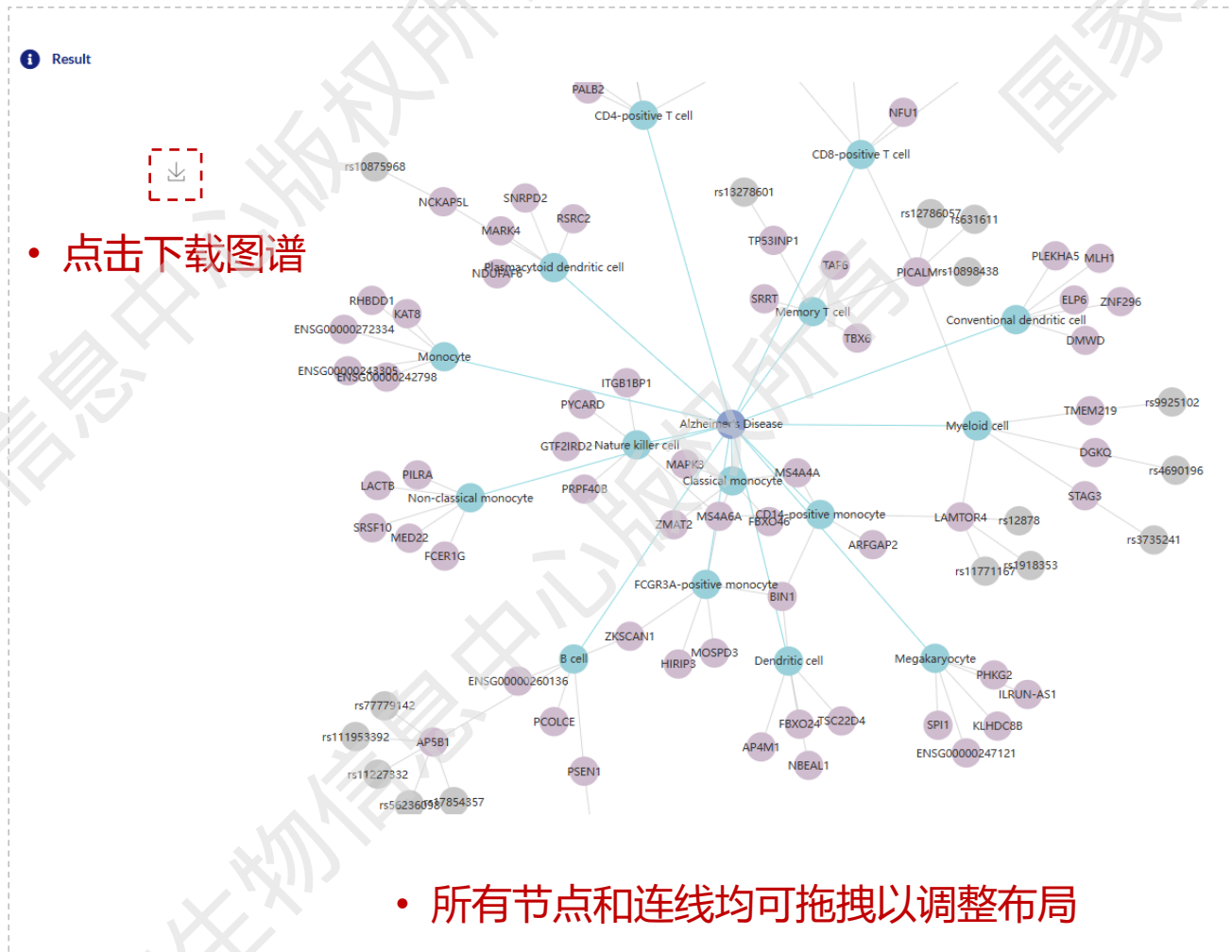
04 实时绘制和编辑知识图谱

02 选择感兴趣的中心节点

03 选择感兴趣的组织类型

Tissue

PBMC



总结



意义：单细胞多模态测序技术不断革新，相应的生物信息学分析工具也呈现蓬勃发展态势，有力支撑了研究人员**将原始测序数据转化成生物学或医学相关信息**



不足：现有的单细胞多组学数据库为疾病、植物育种、公共卫生等领域的研究提供了非常丰富的数据资源与信息支撑，**但在数据质量和数据库功能方面尚不完善**



展望：**亟需持续产出高质量单细胞测序数据以及开发高性能的大数据解析算法**，更好地服务于各领域基础科学研究及产业应用，同时有效应对美西方数据脱钩风险

致谢



Email: gen@big.ac.cn

cancerscem@big.ac.cn

crost@big.ac.cn

zongwenting2018m@big.ac.cn

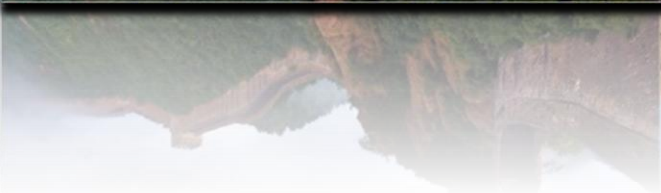
wangshiting@big.ac.cn

maijialin@big.ac.cn

**欢迎使用NGDC单细胞系列数据库
期待您的宝贵意见!**



感谢您的聆听!
欢迎批评指正!



01
GEN



<https://ngdc.cncb.ac.cn/gen/>

02
CancerSCEM



<https://ngdc.cncb.ac.cn/cancerscem/>

03
CROST



<https://ngdc.cncb.ac.cn/crost/>

04
scMethBank



<https://ngdc.cncb.ac.cn/methbank/scm/>

05
Cell Taxonomy



<https://ngdc.cncb.ac.cn/celltaxonomy/>

06
scTWAS Atlas



<https://ngdc.cncb.ac.cn/sctwas/>