



中国科学院北京基因组研究所
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES



国家基因组科学数据中心

国家基因组科学数据中心冬季培训

甲基化资源平台功能介绍

MethBank

ScMethBank

EWAS Open Platform

汇报人：杨 飞

日 期：2022年12月1日

国家生物信息中心

China National Center for Bioinformation



National Genomics Data Center



1

Why 背景与问题

2

What 数据处理流程

3

What 数据库信息统计

4

How 数据库使用教程

环境因子通过表观组与基因组的互作影响机体表型

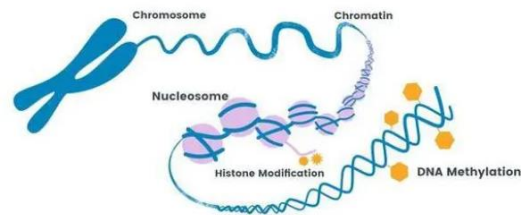
环境因子



表观组

基因组

调控基因



影响表型

发育、衰老、疾病等



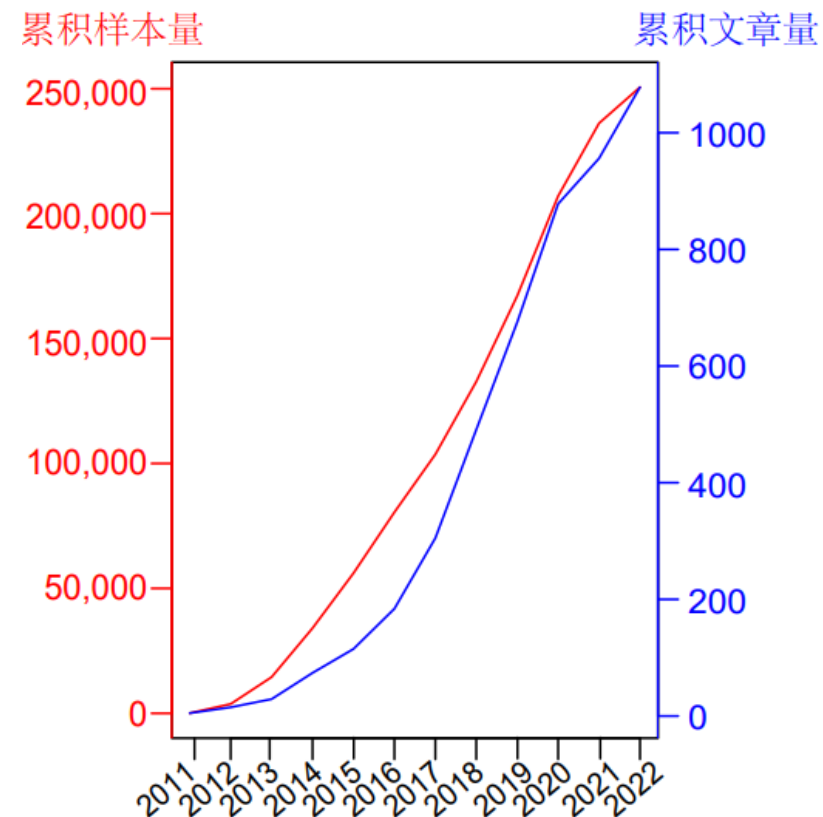
DNA甲基化
测序数据类型

全基因组甲基化
(WGBS)

芯片甲基化
(27K、450K、850K)

单细胞甲基化
(scBS-seq...)

.....

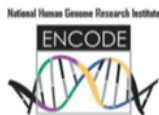


数据来源:

GEO(<https://www.ncbi.nlm.nih.gov/geo>)

EWAS Atlas(<https://bigd.big.ac.cn/ewas>)

公共数据难以充分挖掘使用



Tissue
sample.type
Source.name
sample.tissue
cell.type
Histology
tissue.type
...

1 散在分布于各平台 “数据孤岛”

2 元数据录入不规范、不统一

3 数据质量参差不齐



Steve Horvath创建表观时钟
花费近两年时间整合处理数据

https://en.wikipedia.org/wiki/Epigenetic_clock#cite_note-Biome2013-12

许多关联知识并没有明确

tissue

age

gender

ancestry

BMI

...

1 表观模式是多性状综合的结果

2 小样本结果可能存在不一致性



1. 目前公共数据库中积累了多少甲基化数据？能否提供高效的数据检索下载？
2. 各个实验项目公开的数据如何进行整合分析？
3. 如何快速获得感兴趣的生物背景下的特征甲基化基因或者特征甲基化位点？
4. 如何获得多模态的参考甲基化图谱？

**为了解决这些问题，我们开发了
集数据、知识及工具为一体的DNA甲基化资源平台。**

全基因组甲基化

MethBank

(Nucleic acids research, 2015, 2018, 2022)

单细胞甲基化

scMethBank

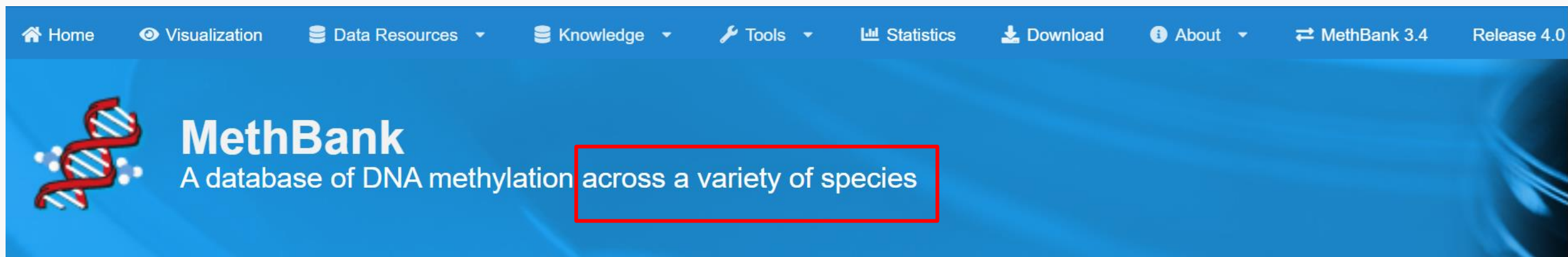
(Nucleic acids research, 2021)

芯片甲基化

EWAS Open Platform

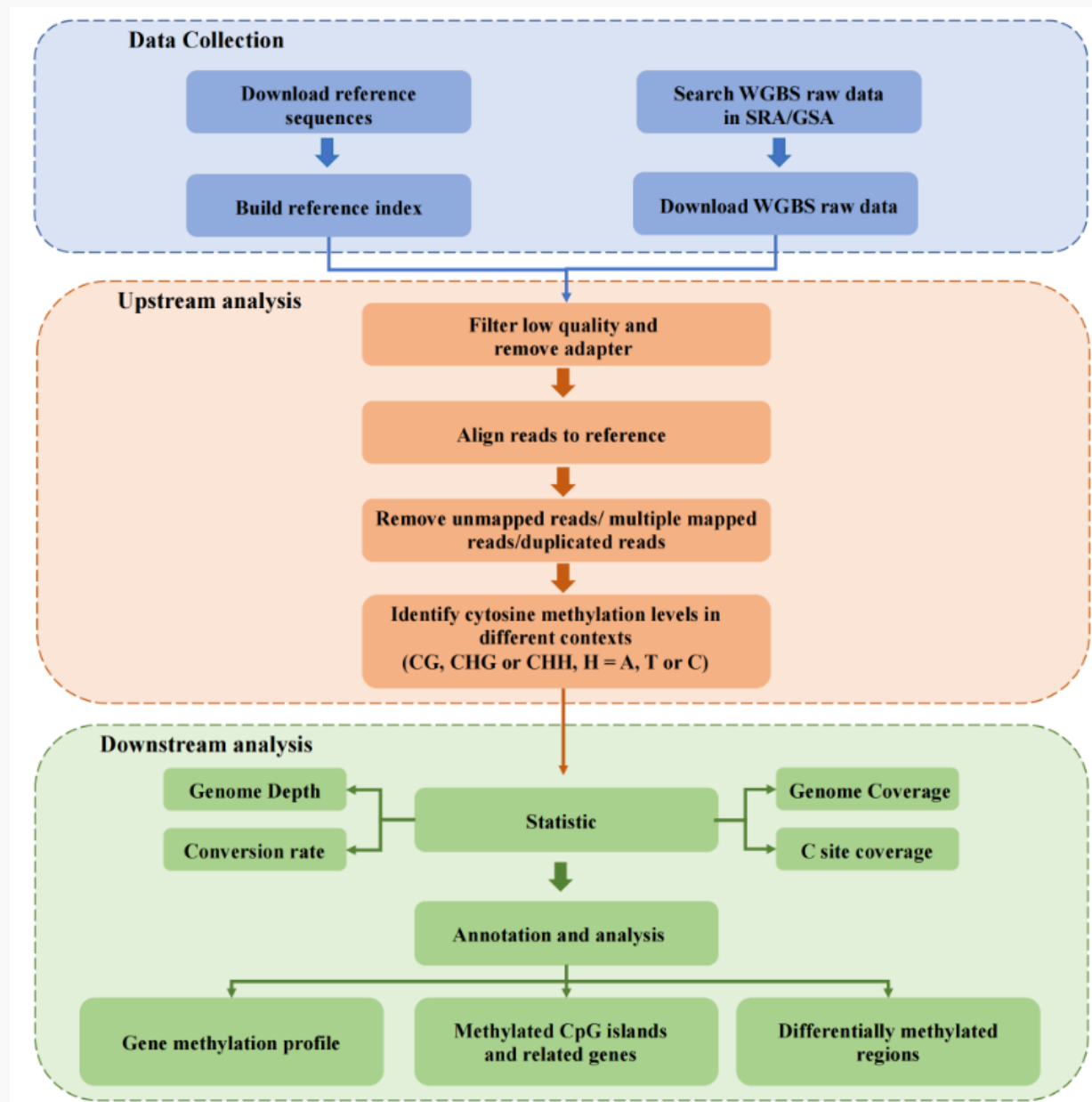
(Nucleic acids research, 2019, 2020, 2022)

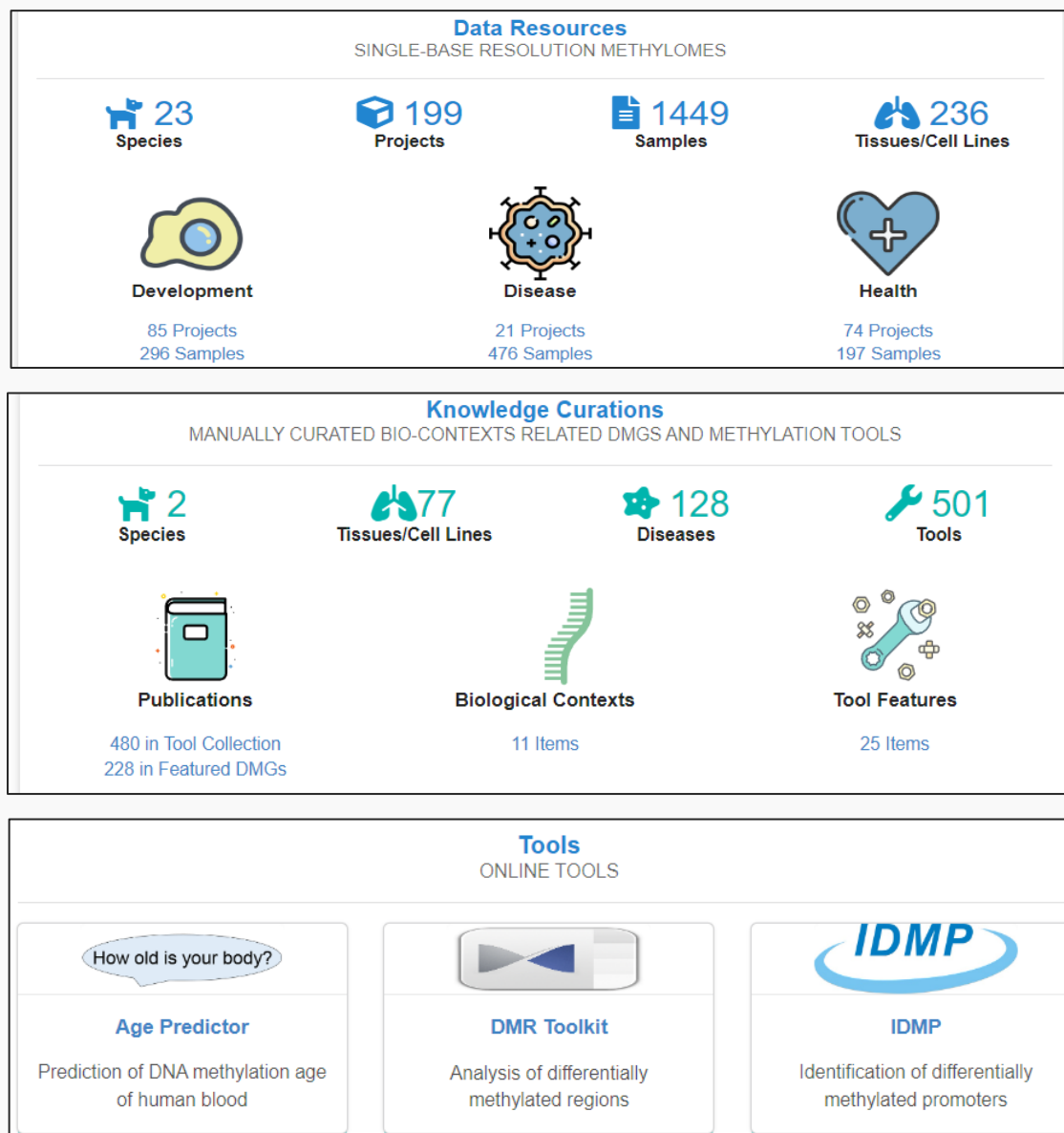
— 跨物种的全基因组甲基化数据库 MethBank



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

MethBank 数据处理流程





数据

知识

工具



The screenshot shows the MethBank 3.4 homepage. The top navigation bar includes links for Home, Visualization, Data Resources, Knowledge, Tools, Statistics, Download, About, and MethBank 3.4. The Knowledge dropdown menu is open, showing options like Projects, Samples, Genes, CpG Islands, DMRs, Publications, Featured DMGs, Tool Collection, Age Predictor, IDMP, and DMR Toolkit. The Tools dropdown menu is also open, showing Documentation, Release Notes, and Contact Us. The 'Explore data' search bar is highlighted, showing a search query: '282353; Spleen; embryonic development; 28472656; Bismark; tp53'. Below the search bar, the 'Data Resources' section is highlighted, showing counts for Species (23), Projects (199), and Samples (1449). The 'Knowledge Curations' section is also highlighted, showing counts for Species (2), Tissues/Cell Lines (77), Diseases (128), and Tools (501). The bottom section features icons for Development, Disease, Health, Publications, Biological Contexts, and Tool Features, each with associated project and sample counts.

Navigation Bar:

- Home
- Visualization
- Data Resources
- Knowledge
 - Projects
 - Samples
 - Genes
 - CpG Islands
 - DMRs
 - Publications
 - Featured DMGs
 - Tool Collection
- Tools
 - Age Predictor
 - IDMP
 - DMR Toolkit
- Statistics
- Download
- About
 - Documentation
 - Release Notes
 - Contact Us
- MethBank 3.4

Release 4.0 (2022-09)

Explore data

Select Item: Enter query... Search

e.g., 282353; Spleen; embryonic development; 28472656; Bismark; tp53

Data Resources
SINGLE-BASE RESOLUTION METHYLOMES

- Species: 23
- Projects: 199
- Samples: 1449

Knowledge Curations
MANUALLY CURATED BIO-CONTEXTS RELATED DMGS AND METHYLATION TOOLS

- Species: 2
- Tissues/Cell Lines: 77
- Diseases: 128
- Tools: 501

Development
85 Projects
296 Samples

Disease
21 Projects
476 Samples

Health
74 Projects
197 Samples

Publications
480 in Tool Collection
228 in Featured DMGs

Biological Contexts
11 Items

Tool Features
25 Items

1.1 如何快速找到感兴趣的样本数据？

例如：寻找含有以下信息的项目及样本

物种： 小鼠

研究内容： 发育

如何快速找到感兴趣的样本数据？方法一

Home Visualization Data Resources Knowledge Tools Statistics Download About MethBank 3.4 Release 4.0 (2022-09)

 **MethBank**
A database of DNA methylation across a variety of species

 **Explore data**

Select Item development **Search**

e.g., Homo sapiens; human; PRJNA574550; GSM1282353; Spleen; embryonic development; 28472656; Bismark; tp53

Data Resources
SINGLE-BASE RESOLUTION METHYLOMES

Species	Projects	Samples	Tissues/Cell Lines
23	199	1449	236

Development
85 Projects
296 Samples

Disease
21 Projects
476 Samples

Health
74 Projects
197 Samples

Knowledge Curations
MANUALLY CURATED BIO-CONTEXTS RELATED DMGS AND METHYLATION TOOLS

Species	Tissues/Cell Lines	Diseases	Tools
2	77	128	501

Publications
480 in Tool Collection
228 in Featured DMGs

Biological Contexts
11 Items

Tool Features
25 Items

Home / Search

Search results for: **development**

Select Item development Search

Projects (88) Samples (6) Publications (173) Tool Collection (2) Featured DMGs (12)

Projects

Show 5 entries

Search: Mus musculus

Project ID	Title	Overall Design	Species	Tissue/Cell Line	Sample Number	Related Biological Context
GSE101000	Embryonic to postnatal mouse developmental time course...		Mus musculus	• Hindbrain	2	Health, Development
GSE101023	Embryonic to postnatal mouse developmental time course...		Mus musculus	• Limb	2	Health, Development
GSE101302	Embryonic to postnatal mouse developmental time course...		Mus musculus	• Midbrain	2	Health, Development
GSE101322	Embryonic to postnatal mouse developmental time course...		Mus musculus	• Heart	2	Health, Development
GSE101360	Embryonic to postnatal mouse developmental time course...		Mus musculus	• Forebrain	2	Health, Development

Showing 1 to 5 of 73 entries (filtered from 88 total entries)

Previous 1 2 3 4 5 ... 15 Next



如何快速找到感兴趣的样本数据？方法二

The screenshot shows the MethBank 3.4 website. The top navigation bar includes links for Home, Visualization, Data Resources, Knowledge, Tools, Statistics, Download, About, and MethBank 3.4. The 'Data Resources' dropdown menu is open, showing options: Projects, Samples, Genes, CpG Islands, DMRs, and Publications. The 'Projects' option is highlighted with a red box. Below the navigation bar, there is a search bar with the text 'Explore data' and a 'Search' button. Below the search bar, there is a text input field with the placeholder 'Enter query...' and a 'Select Item' dropdown. Below the input field, there is a list of example queries: 'e.g., Homo sapiens; human; PRJNA574550; GSM1282353; Spleen; embryonic development; 28472656; Bismark; tp53'. The main content area is divided into two sections: 'Data Resources' and 'Knowledge Curations'. The 'Data Resources' section is titled 'SINGLE-BASE RESOLUTION METHYLOMES' and contains three sub-sections: 'Species' (23), 'Projects' (199), and 'Samples' (1449). The 'Knowledge Curations' section is titled 'MANUALLY CURATED BIO-CONTEXTS RELATED DMGS AND METHYLATION TOOLS' and contains three sub-sections: 'Species' (2), 'Tissues/Cell Lines' (77), and 'Diseases' (128). The 'Tools' section (501) is also present. The 'Data Resources' section also includes 'Development' (85 Projects, 296 Samples), 'Disease' (21 Projects, 476 Samples), and 'Health' (74 Projects, 197 Samples). The 'Knowledge Curations' section also includes 'Publications' (480 in Tool Collection, 228 in Featured DMGs), 'Biological Contexts' (11 Items), and 'Tool Features' (25 Items).

Data Resources
SINGLE-BASE RESOLUTION METHYLOMES

- Species: 23
- Projects: 199
- Samples: 1449
- Tissues/Cell Lines: 236

Knowledge Curations
MANUALLY CURATED BIO-CONTEXTS RELATED DMGS AND METHYLATION TOOLS

- Species: 2
- Tissues/Cell Lines: 77
- Diseases: 128
- Tools: 501

Data Resources Details:

- Development:** 85 Projects, 296 Samples
- Disease:** 21 Projects, 476 Samples
- Health:** 74 Projects, 197 Samples

Knowledge Curations Details:

- Publications:** 480 in Tool Collection, 228 in Featured DMGs
- Biological Contexts:** 11 Items
- Tool Features:** 25 Items

Home / Projects

View All

Species

Animal Tissue

Human Disease

Search...

Search

Cardiovascular system disease

Endocrine system disease

Gastrointestinal system disease

Hematopoietic system disease

Immune system disease

Integumentary system disease

Musculoskeletal system disease

Nervous system disease

Reproductive system disease

Respiratory system disease

Thoracic disease

Display columns

Select AllSelect Default

Basic Information:

☐ Data Resource

☐ BioProject ID

☒ Project ID

☒ Title

☐ Summary

☒ Overall Design

☒ Species

☒ Tissue/Cell Line

☐ Cell Type

☐ Healthy Condition

☐ Development Stage

☐ Disease State

☒ Sample Number

☐ Submitter

☐ PMID

☐ Submission Date

☐ Last Update Date

☒ Related Biological Context

Show High Quality Datasets

Download Table

Show5entries

Search:

Project ID	Title	Overall Design	Species	Tissue/Cell Line	Sample Number	Related Biolo Context
Search Project ID	Search Title	Search Overall Design	Search Species	Search Tissue/Cell Line	min sample# max sample#	Search Relat
GSE139817	Chromatin dynamics reveal circadian...	Whole-genome bisulfite sequencing...	Homo sapiens	hPSC Cell Pancreatic Islet Cell	7	Differentiat
GSE42410	Plants regenerated from tissue...	Whole genome methylation maps of...	Oryza sativa	Leaf Callus	18	Genotype
GSE146620	The dynamic methylome of the broiler...	Whole genome methylation analysis...	Gallus gallus	Breast Ileum Jejunum More...	51	Health Age
GSE128859	Genome-wide analysis of DNA methylation...	Whole genome bisulfite sequencing...	Zea mays	Seedling root Seedling leaf Seedling shoot	6	Genotype Tissue Com
GSE138598	Whole genome bisulfite sequencing...	Whole genome bisulfite sequencing...	Homo sapiens	Semen	17	Disease

Note: The project containing several species is shown in multiple rows.

Showing 1 to 5 of 205 entries

Previous

1

2

3

4

5

...

41

Next

1.2 如何筛选高质量的感兴趣内容相关的样本？

例如：筛选含有以下信息的样本

物种： 小鼠

研究内容： 胚胎发育

Mapping Rate: $\geq 80\%$

Mapped Read Depth: ≥ 20



[Home](#) / [Samples](#)

▼ Display columns

Select All

Select Default

Basic Information:

☒ Sample Name

☐ Data Resource

☒ Description

☐ BioProject ID

☒ Project ID

☒ Sample ID

☐ Experiment ID

☐ Status

☐ Submission Date

☐ Last Update Date

☐ Donor ID

Sample Characteristic:

☒ Species

☒ Tissue/Cell Line

☐ Cell Type

☐ Source Name

☐ Strain

☐ Breed

☐ Cultivar

☐ Sex

☐ Age

Biological Condition:

☐ Healthy Condition

☐ Disease State

☐ Development Stage

☐ Genotype

☐ Treatment

☐ Knockout

☐ Resistant Phenotype

Protocol:

☐ Growth Protocol

☐ Treatment Protocol

☐ Extraction Protocol

☐ Construction Protocol

☐ Library Strategy

☐ Library Source

☐ Library Selection

☐ Layout

☐ Platform

Assessing Quality:

☒ Quality Assessment ?

☐ Mapping Rate

☐ Uniquely Mapping Rate

☐ Genome Coverage

☐ Cytosine Coverage

☐ Conversion Rate

☐ Mapped Read Depth

Analysis:

☐ Genome Version for Mapping

☐ Genome Version for Annotation

Action:

☒ View in JBrowse

☒ CpG Island

▼ Filter by Quality Assessment ?

Mapping Rate

0.8

max mapping rate

Uniquely Mapping Rate

min uniquely mapping rate

max uniquely mapping rate

Genome Coverage

min genome coverage

max genome coverage

Cytosine Coverage

min cytosine coverage

max cytosine coverage

Conversion Rate

min conversion rate

max conversion rate

Mapped Read Depth

20

max mapped read depth

您可以自定义质控范围进行样本的筛选

MethBank / Data Resources / Samples

Show	10	entries	Download Table		Search:		
Sample Name	Description	Project ID	Sample ID	Species	Tissue/Cell Line	Quality Assessment	View In
Search Sample Name	Search Description	Search Project ID	Search Sample ID	Mus musculus	Search Tissue/Cell Line	-	-
Mus_Healthy_FacialProminence_E10.5_PooledMaleAndFemale_Rep1	Healthy E10.5 facial prominence...	GSE101478	GSM2704163	Mus musculus	Face		🔗
Mus_Healthy_FacialProminence_E10.5_PooledMaleAndFemale_Rep2	Healthy E10.5 facial prominence...	GSE101478	GSM2704164	Mus musculus	Mapped Read Depth: 40.3865473204483		🔗
Mus_Healthy_FacialProminence_E11.5_PooledMaleAndFemale_Rep1	Healthy E11.5 facial prominence...	GSE83041	GSM2192489	Mus musculus	Face		🔗
Mus_Healthy_FacialProminence_E11.5_PooledMaleAndFemale_Rep2	Healthy E11.5 facial prominence...	GSE83041	GSM2192490	Mus musculus	Face		🔗
Mus_Healthy_FacialProminence_E12.5_PooledMaleAndFemale_Rep1	Healthy E12.5 facial prominence...	GSE82828	GSM2192025	Mus musculus	Face		🔗
Mus_Healthy_FacialProminence_E12.5_PooledMaleAndFemale_Rep2	Healthy E12.5 facial prominence...	GSE82828	GSM2192027	Mus musculus	Face		🔗
Mus_Healthy_FacialProminence_E13.5_PooledMaleAndFemale_Rep1	Healthy E13.5 facial prominence...	GSE82812	GSM2191988	Mus musculus	Face		🔗
Mus_Healthy_FacialProminence_E13.5_PooledMaleAndFemale_Rep2	Healthy E13.5 facial prominence...	GSE82812	GSM2191990	Mus musculus	Face		🔗

1.3 如何查看感兴趣的基因的甲基化谱？

例如： 物种： 小鼠

基因： Narf 、 Tbx2

不同发育时期



Home / Genes

Search

Species: Mus musculus

Gene ID: ENSMUSG000000000056 [Narf] x ENSMUSG000000000093 [Tbx2] x ENSMUSG000000000103 [Zfy2] x

Gene Symbol: Search Gene Symbol...

Position: Chr Search Chromosome... From To

Description:

View Reset

Show 10 entries

Gene ID	Gene Symbol	Chromosome	Start Position	End Position	Strand	Description
ENSMUSG000000000056	Narf	chr11	121237253	121255856	+	nuclear prelamins A recognition factor
ENSMUSG000000000093	Tbx2	chr11	85832551	85841948	+	T-box 2
ENSMUSG000000000103	Zfy2	chrY	2106175	2170409	-	zinc finger protein 2, Y-linked

Showing 1 to 3 of 3 entries

Previous 1 Next



Mus musculus house mouse Taxonomy ID: 10090 ENSMUSG000000000056

Gene Information

Gene ID	Gene Symbol	Chromosome	Start Position	End Position	Strand	Description
ENSMUSG000000000056	Narf	chr11	121237253	121255856	+	nuclear prelamin A recognition factor

Gene Methylation Profiles

Project

Select Project

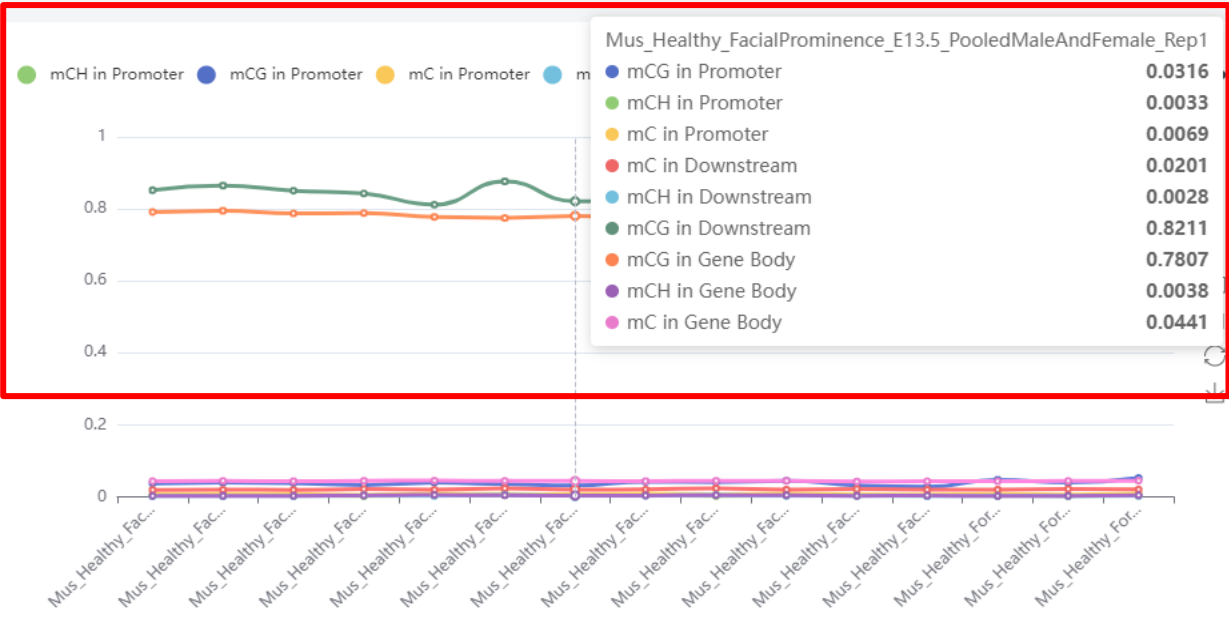
Samples

Select Samples

Display 5 records Showing 1 to 5 of 153

Sample	CG			CH			F
	Promoter	Gene Body	Downstream	Promoter	Gene Body	Downstream	
Mus_2-cell	0.0119560422527714	0.626662547264011	0.7530317301585	NA	NA	NA	N
Mus_4-cell	0.0113017350837755	0.579317763275653	0.64535563530525	NA	NA	NA	N
Mus_E13.5f	0.0087554366295551	0.0344120454778789	0.0489086552657125	NA	NA	NA	N
Mus_E13.5m	0.00729374456549388	0.0505642757119329	0.0815433348328125	NA	NA	NA	N
Mus_E6.5	0.0267457897688959	0.687097162894663	0.646527412672875	NA	NA	NA	N

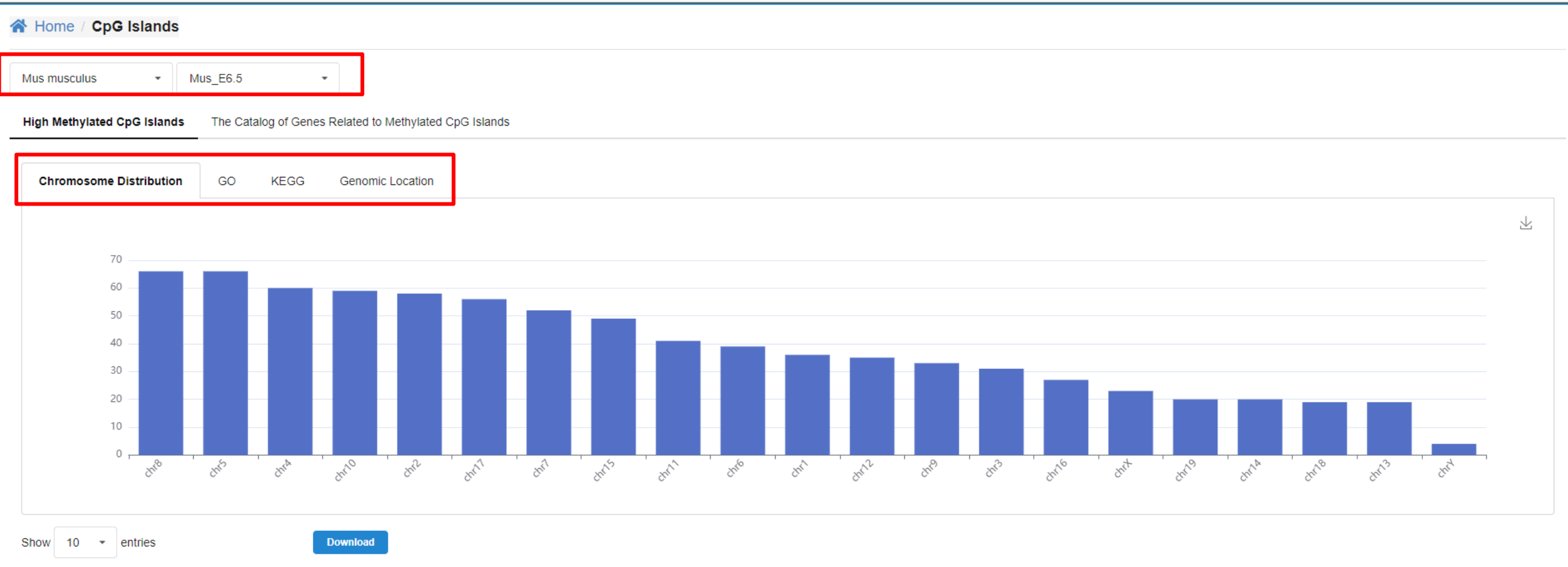
Previous12345...31Next



1.4 如何查看感兴趣样本的高甲基化的CpG岛特征?

例如： 物种： 小鼠
时期： E6.5

Home / CpG Islands								
Mus musculus Mus_E6.5								
High Methylated CpG Islands The Catalog of Genes Related to Methylated CpG Islands								
Show 10 entries Download								
Chromosome	Strand	Start Position	End Position	CpG Length	CpG Number	Average Methylation Level	Genomic Location	Gene ID
Search Chromosome	Search Strand	position >	position <	min cpg length max cpg length	min cpg number max cpg number	min max	Search Genomic Location	Search Gene ID
chr1	+	1040116	1044925	4809	347	67.38	Exon (ENST00000379370.7/375790, exon 4 of 36)	ENSG00000188157
chr1	+	1162663	1165204	2541	180	67.17	Distal Intergenic	ENSG00000207730
chr1	+	1048826	1049049	223	19	74.71	Exon (ENST00000379370.7/375790, exon 24 of 36)	ENSG00000188157
chr1	+	1045688	1046020	332	26	95.19	Exon (ENST00000379370.7/375790, exon 15 of 36)	ENSG00000188157



1.5 如何获得感兴趣的样本间的差异甲基化区域 (DMR) ?

例如： 物种： 小鼠
时期： E6.5 VS E7.5

Species *

Mus musculus

Project *

PRJNA244322 [Programming and inheritance of parental DNA methylomes in mammals]

Group *

Mus_E6.5VSMus_E7.5

P-Value *

0.01

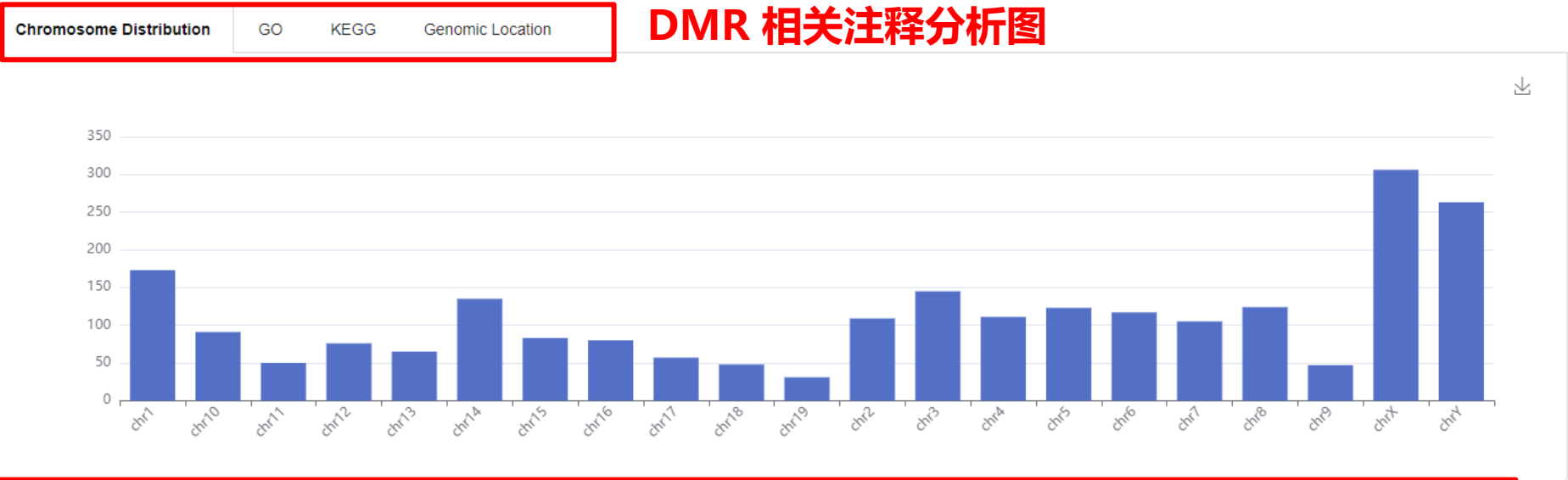
Context *

CG

View

Reset

输入特定的样本比较组

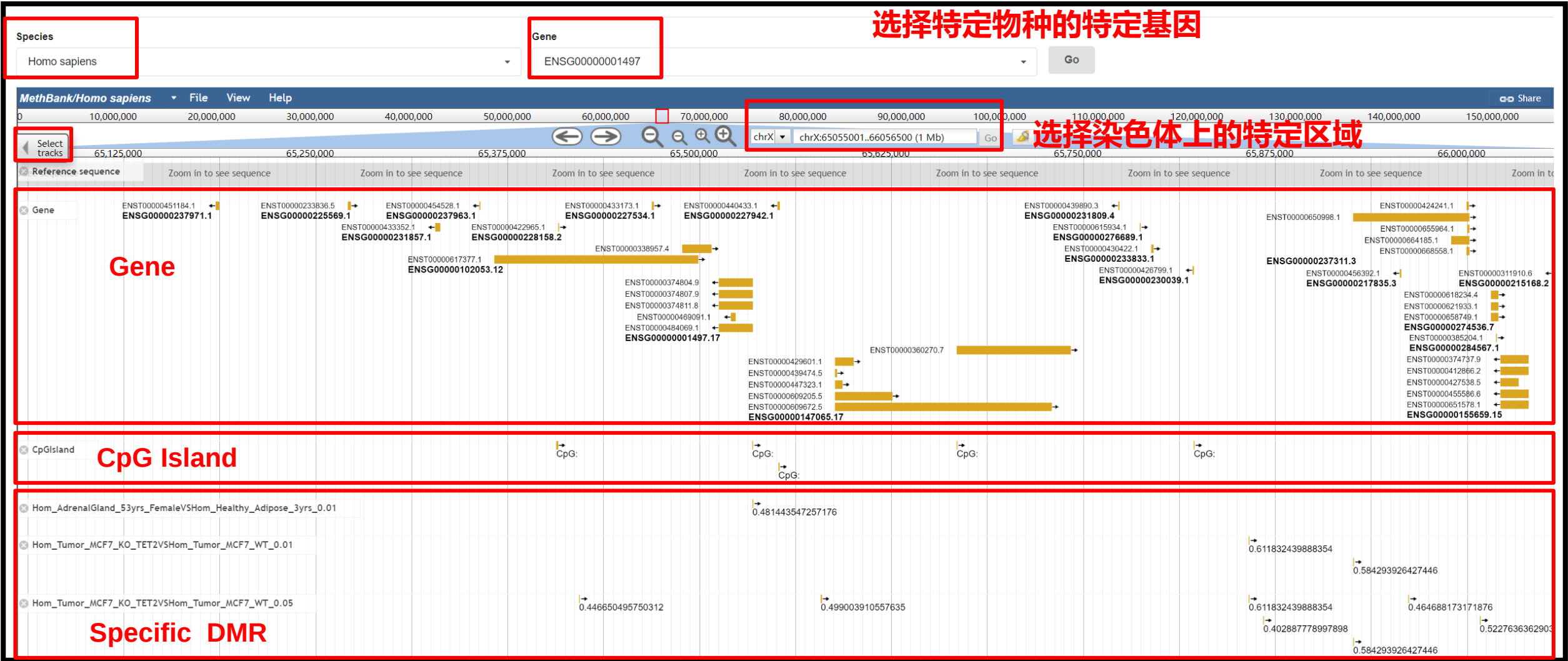


Show 10 entries

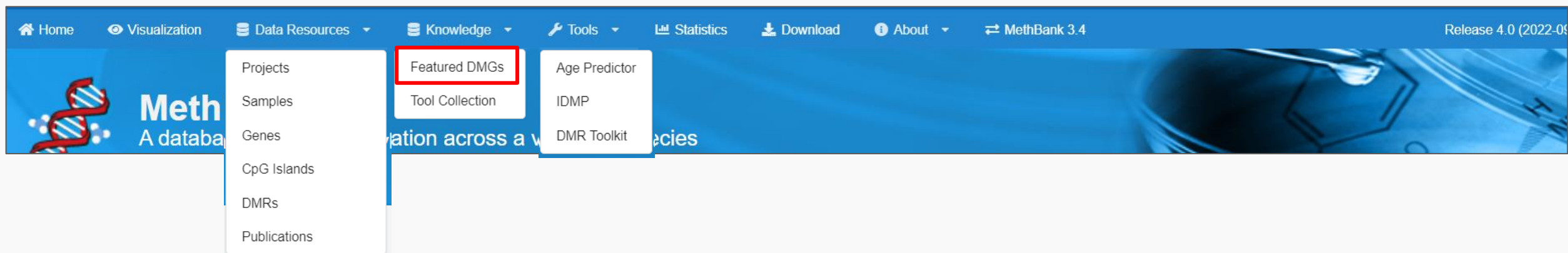
Download

DMR 表

Gene ID	Gene Symbol	Chromosome	Start Position	End Position	Length	Annotation
Search Gene ID	Search Gene Symbol	Search Chromosome	Position >	Position <	Min Length# Max Length#	Search Annotation
ENSMUSG00000000078	Klf6	chr13	6062328	6062597	269	Distal Intergenic
ENSMUSG000000000103	Zfy2	chrY	2288435	2288526	91	Distal Intergenic
ENSMUSG000000000103	Zfy2	chrY	2232008	2232689	681	Distal Intergenic



1.6 现在已经报道了哪些基因的甲基化与小鼠的衰老相关？



► Display columns

Select All

Select Default

☒ Epi Factor

☒ Species

☐ PMID

☐ Title

☐ Journal

☐ Publication Year

☒ Library

☒ Tissue/Cell Line

☐ Sampling Site

☐ Cell Type

☒ Disease

☐ Method

☒ Compare Group

☒ Condition

☒ Featured DMG

☒ Global Methylation Level

☐ Global Methylation Entropy

☐ Non-coding

☐ Enrichment Function

Show

50

 entries

Search:

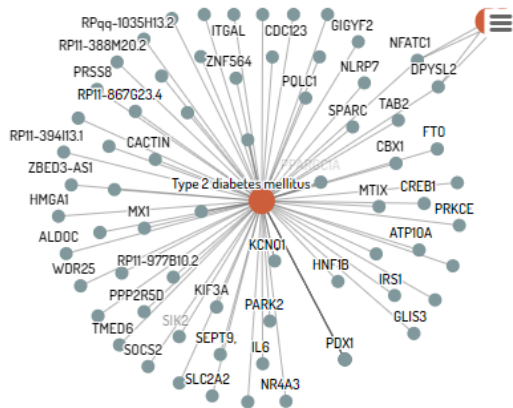


[Home](#) / [Bio-contexts Curation](#)

Disease Network

Search Disease

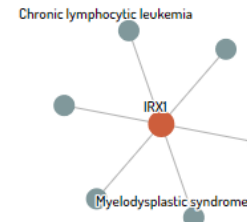
Search Disease...



Gene Network

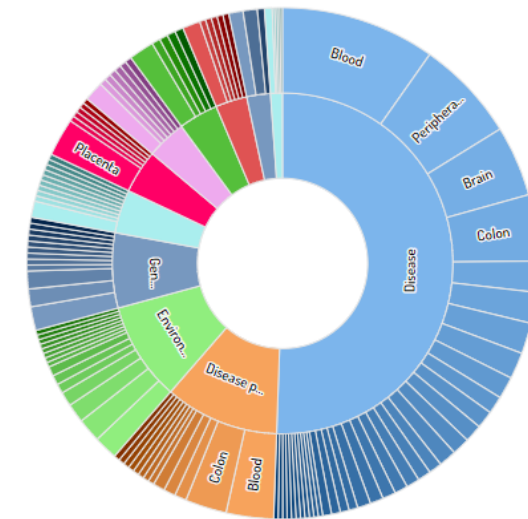
Search gene

Search Gene...

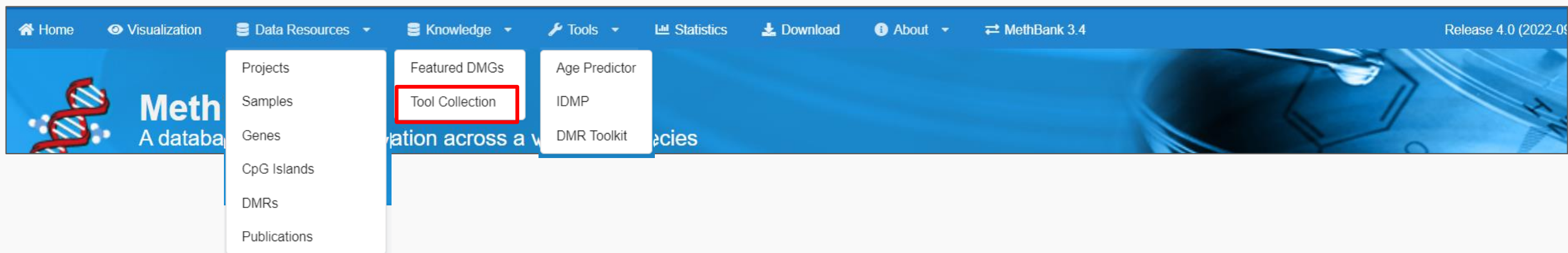


- View in full screen
- Print chart
- Download PNG image
- Download JPEG image
- Download PDF document
- Download SVG vector image

Tissue Sunburst



1.7 现存有哪些针对BS-seq数据的DMR分析工具?



[Home](#) / Tool Collection

Filter

Name

Description

Software Type

Category

Country/Region

Operation System

Filter

Reset

Show 10 entries

工具集合列表

Search: BS-seq

	Accession	Name	Short Description	Software Type	Category	Operation System	Interface	Description
+	MBT000032	Bi-Seq	Processing and analyzing bisulfite sequencing data	• Package/Module	• DNA Methylation • BS-Seq • Analysis	• Unix/Linux • Mac OS • Windows	• Command line interface	The BiSeq package provides useful...
-	MBT000040	Bisulfighter	Accurate detection of methylated cytosines and differentially methylated regions	• Package/Module	• DNA Methylation • BS-Seq • Analysis • More...	• Unix/Linux	• Command line interface	Bisulfighter combines the LAST...

Download	http://epigenome.cbrc.jp/bisulfighter
Documentation	https://github.com/yutaka-saito/Bisulfighter/issues
Stability	Stable
Computer skills	Advanced
Maintained	Yes
Programming	Python
Restrictions to use	None
Country & Region	Japan
Host Institution	Computational Biology Research Center
Publication	Bisulfighter: accurate detection of methylated cytosines and differentially methylated regions.

选定工具的具体信息介绍

Saito Y, Tsuji J,
Mituyama T.Nucleic acids research, Volume:42,
Issue:6, 201431 citation(s) (from EuropePMC,
2022-11-25)

筛选特定条件的工具



[Home](#) / Publications

Data Resources (78)

Tool Collection (480)

Featured DMGs (228)

Show 5 entries

Search:

Defiant: (DMRs: easy, fast, identification and ANnoTation) identifies differentially Methylated regions from iron-deficient rat hippocampus.

Condon DE, Tran PV, Lien YC, Schug J, Georgieff MK, Simmons RA, Won KJ.

BMC bioinformatics, 2018-02-05 DOI: [10.1186/s12859-018-2037-1](#), PMID: [29402210](#)

Citations:10 (Sourced from EuropePMC: 2022-11-25)

[GSE98064](#) [Rattus norvegicus](#) [Hippocampus](#)

[MBT000096](#)

[Abstract](#) ▶

Guide Positioning Sequencing identifies aberrant DNA methylation patterns that alter cell identity and tumor-immune surveillance networks.

Li J, Li Y, Li W, Luo H, Xi Y, Dong S, Gao M, Xu P, Zhang B, Liang Y, Zou Q, Hu X, Peng L, Zou D, Wang T, Yang H, Jiang C, Peng S, Wu F, Yu W.

Genome research, 2019-01-22 DOI: [10.1101/gr.240606.118](#), PMID: [30670627](#)

Citations:9 (Sourced from EuropePMC: 2022-11-25)

[GSE92328](#) [Homo sapiens](#) [Liver](#)

[MBT000158](#)

[Abstract](#) ▶



Note: HTTP download speed may be slow, it is highly recommended that you download the dataset using a dedicated FTP tool (such as FileZilla Client).
FTP: <ftp://download.cncb.ac.cn/methbank/MethBank4.0>

Single-base Resolution Methylomes (SRMs)

Show 10 entries

多物种的单碱基精度甲基化组

支持直接下载或FTP下载

Search:

Species	Tissues/Cell Lines	Methylomes		Gene methylation profiles	Methylated CpG islands (mCpGIs) and related genes	Differentially Methylated Regions (DMRs)
		WIG	BED			
<i>Ailuropoda_melanoleuca</i>	3	29G	18G	2.7M	30M	91M
<i>Bos_taurus</i>	17	790G	479G	81M	55M	592M
<i>Brassica_napus</i>	3	29G	17G	54M	41M	47M
<i>Canis_lupus_familiaris</i>	2	93G	59G	12M	18M	33M
<i>Danio_rerio</i>	16	166G	103G	65M	23M	2.1G
<i>Gallus_gallus</i>	5	240G	147G	48M	12M	98M
<i>Glycine_max</i>	18	46G	52G	122M	946K	57M
<i>Gorilla_gorilla</i>	2	36G	22G	4.6M	2.0M	13M
<i>Homo_sapiens</i>	92	4.7T	2.9T	1.8G	557M	7.8G
<i>Macaca_fascicularis</i>	2	31G	19G	3.4M	770K	-

Note: The old workflow data of 40 Human SRMs can be downloaded at https://download.cncb.ac.cn/methbank/MethBank3.0/Human/SRM/Homo_sapiens

Showing 1 to 10 of 23 entries

Previous

1

2

3

Next

Human Healthy Consensus Reference Methylomes (CRMs)

Show 10 entries

人的多组织的参比甲基化组

Search:

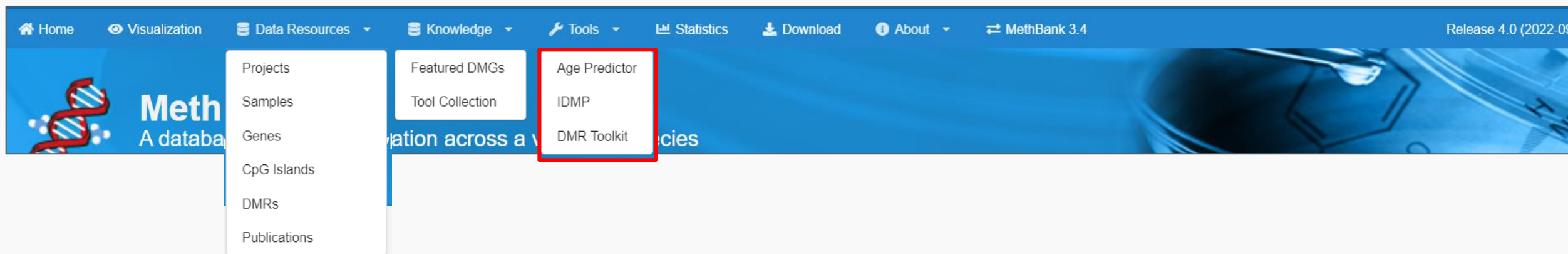
Tissue	Female(450k)	Male(450k)	No Sex Message(450K)	Female(850k)	Male(850k)	No Sex Message(850K)
abdominal_adipose_tissue	89	53				



1.8 如何使用**甲基化年龄预测**工具？

如何使用**差异甲基化启动子区域分析**工具？

如何使用**差异甲基化区域识别分析**工具？



[Home](#) / [Age Predictor](#)

年龄预测在线分析工具

Age Predictor is an online tool to estimate methylation age of the sample by three machine learning methods (SVM, Random Forest and Elastic Net). Users can input the idat file, processed data or sample ID of GEO datasets to obtain the prediction results.

Using idat Files

Using Processed Data

Using NCBI Sample ID

支持三种格式的输入文件

Age Predictor accepts compressed raw data files (.gz format), which contain two files (one for green signal intensity and another for red signal intensity). The Age Predictor will process the raw data using standard Illumina pipeline and return a predicted age.

[7970376047_R03C01_Grn.idat.gz](#)

[7970376047_R03C01_Red.idat.gz](#)

File1 (gz format):

选择文件 未选择任何文件

File2 (gz format):

选择文件 未选择任何文件

Submit

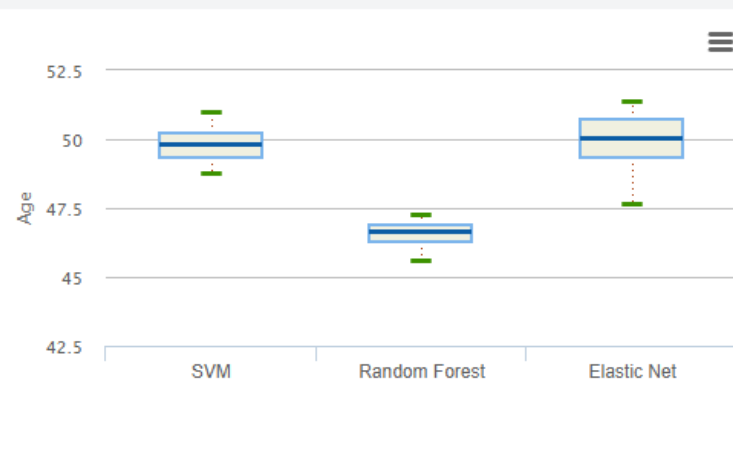
Age Predictor Result

The adopted methods and their corresponding predicted ages are listed below:

SVM: 50 yr [46.36,50.95]

Random Forest: 47 yr [45.59,47.26]

Elastic Net: 50 yr [47.66,51.35]



差异甲基化启动子区域分析工具

[Home](#) / IDMP

IDMP

IDMP requires working Perl and R, therefore it is a requirement that they are installed on your computer. IDMP will assume that R software is in the working path unless its path is specified manually.

1. Prerequisite software and packages**2. Download and install****3. Usage and options summary**

3.1 Typical command

3.2 Command parameters

3.3 Input file format

4. Result Analysis[Back to top](#)

具体操作流程

1. Prerequisite software and packages

IDMP requires a working of **Perl** and **R**, therefore it is a requirement that they are installed on your computer. IDMP will assume that R software is in the working path unless its path is specified manually.

2. Download and install

Download: [IDMP_v1.0.tar.gz](#)

Install:

```
tar -xzf IDMP_v1.0.tar.gz
```

3. Usage and options summary**3.1 Typical command**

```
IDMP_v1.0 --perlDir script --bedFile1 testData/sample1.bed --bedFile2 testData/sample2.bed --gene testData/annotation_gene.gff3 --seqType C:CH:CG --diffLev 0.1:0.1:0.2 --upRegion 1500 --pValue 0.01 --fdrCorr y --RPath ../../R-2.14.2/bin --outDir IDMPresult
```

```
IDMP_v1.0 --perlDir script --bedFile1 testData/sample1.bed --bedFile2 testData/sample2.bed --gene testData/annotation_gene.gff3 --seqType CH:CG --diffLev 0.1:0.2 --upRegion 1500 --pValue 0.01 --fdrCorr y --outDir IDMPresult
```

```
IDMP_v1.0 --perlDir script --bedFile1 testData/sample1.bed --bedFile2 testData/sample2.bed --gene testData/annotation_gene.gff3 --seqType C --diffLev 0.1 --upRegion 1500 --pValue 0.01 --fdrCorr y --outDir IDMPresult
```



[Home](#) / [DMR Toolkit](#)

差异甲基化区域识别分析工具

DMR Toolkit

DMR toolkit is a pipeline package for DMRs identification, annotation and enrichment for multiple species. Step-by-step example of its use is provided in [Documentation](#).

1. Prerequisite software and packages

2. Download and install

3. Usage and options summary

4. Result Analysis

3.1 Recommended file structure

3.2 Typical command

3.3 Command parameters

3.4 Input file format

4.1 Output of DMRs identification

4.2 Output of annotation

4.3 Output of GO enrichment

4.4 Output of KEGG enrichment

1. Prerequisite software and packages

DMR toolkit requires the installation of [R \(version 4.0.5\)](#) and [bedtools \(version 2.28\)](#) to run locally. DMR toolkit will work properly only when both R and bedtools software are set up correctly and added to the working path.

Identification of differentially methylated regions (DMRs)

- [DSS 2.42.0](#)
- [getopt 1.20.3](#)
- [bsseq 1.30.0](#)

Downstream Analysis (Annotation, GO & KEGG enrichment)

- [clusterProfiler 3.18.1](#)
- [tidyverse 1.3.1](#)
- [data.table 1.14.2](#)
- [enrichplot 1.10.2](#)
- [topGO 2.42.0](#)
- [KEGG.db 1.0](#)
- [DO.db 2.9](#)
- [DOSE 3.16.0](#)
- [Rgraphviz 2.34.0](#)
- [ChIPseeker 1.26.2](#)

When using clusterprofiler package to do KEGG analysis for some samples, the package may report errors. You can generate your KEGG.db package with the following steps:

具体操作流程

二 单细胞甲基化数据库 scMethBank



Home

Browse ▾

Visualize ▾

Tools ▾

Download

Documentation



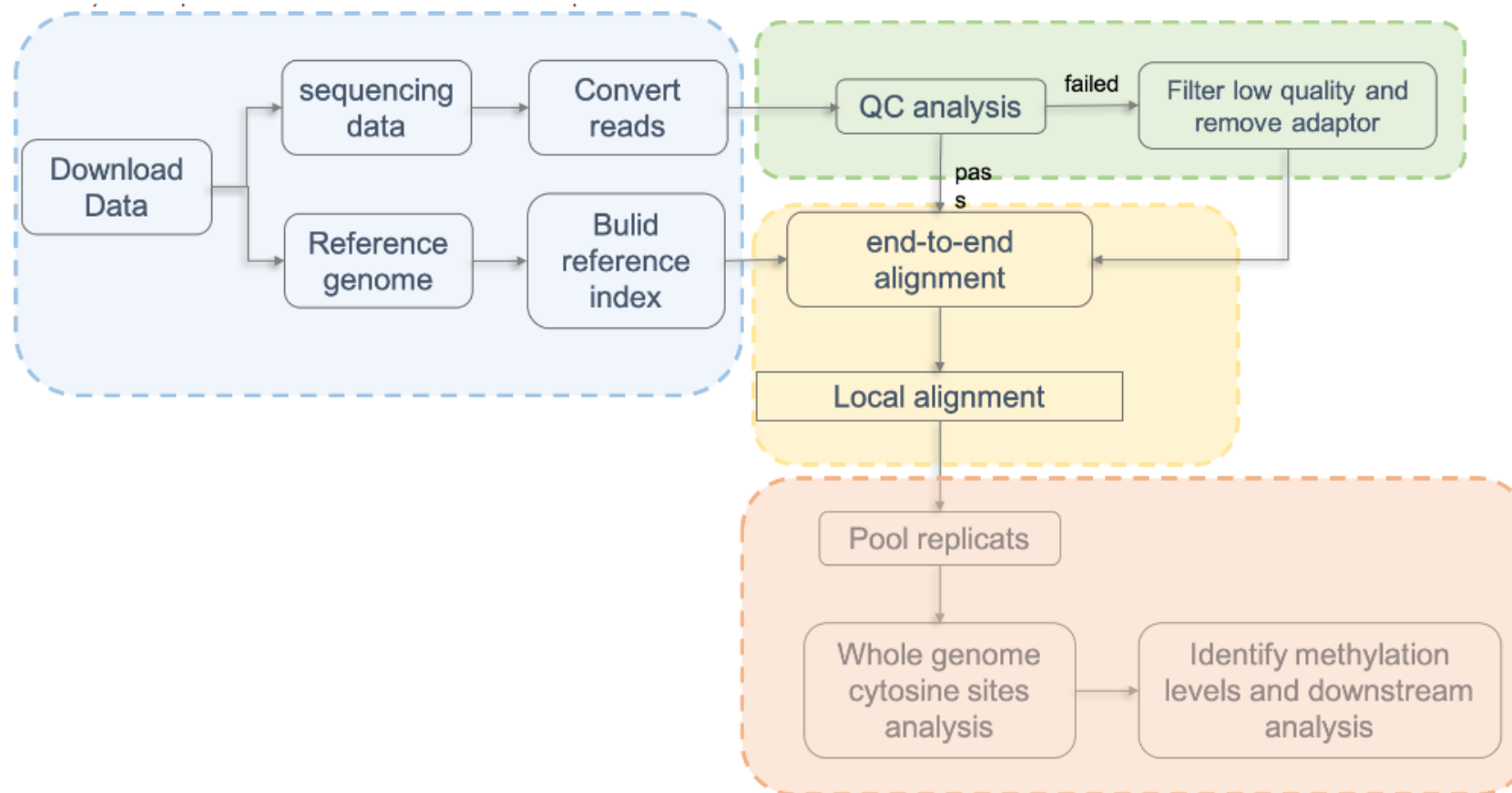
scMethBank

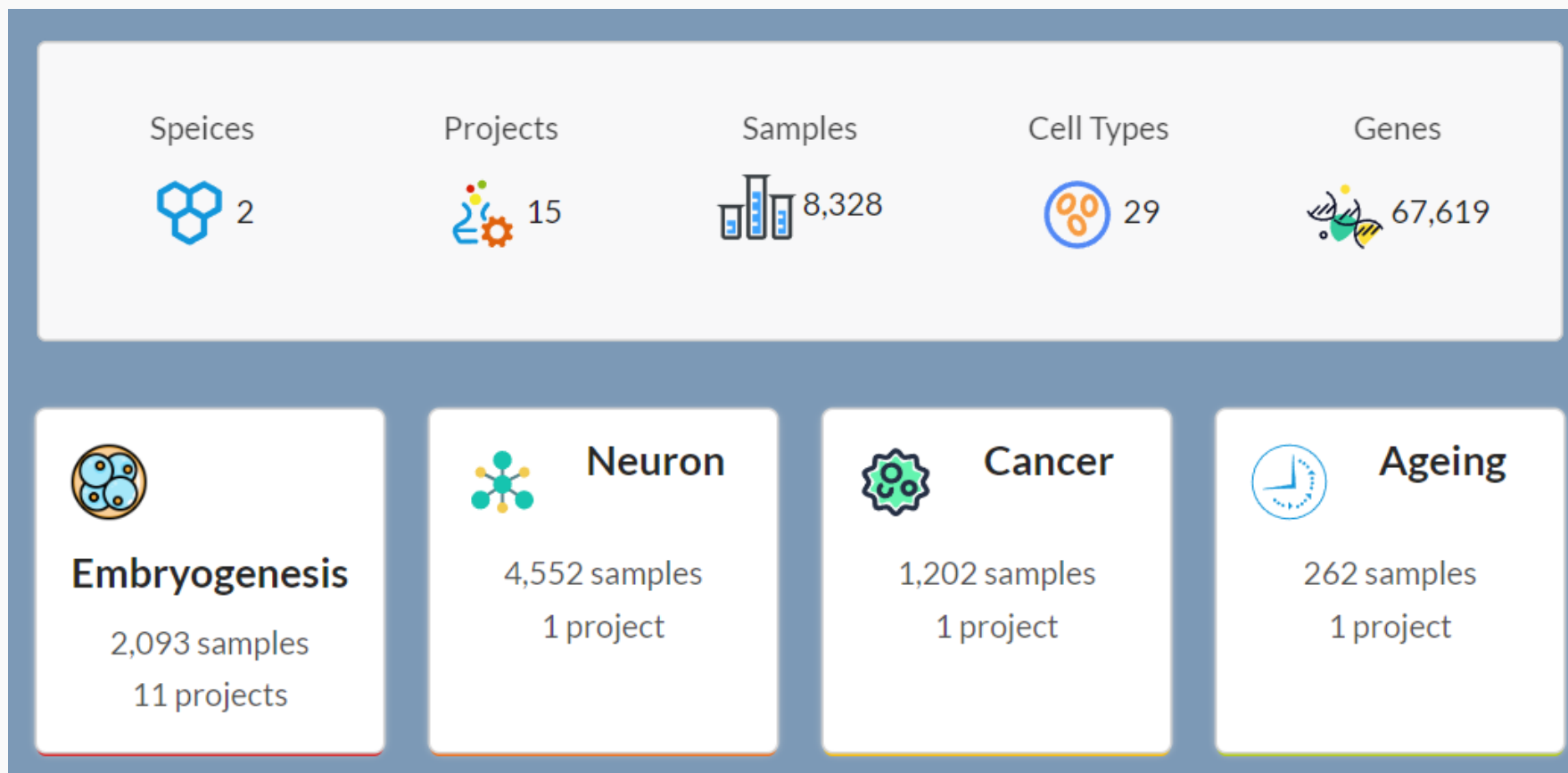
A database of single-cell methylation maps for human and mouse.



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







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

[Home](#) > [Browse](#) > [Projects](#)

PMID	Title	Year	Journal	Datasets	Strategy	Sample Count	Cell Groups (#sample)	View Dataset	Link to scRNA-seq database
30560925	Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes	2019	Cell Res.	GSE114822	iscCOOL-seq	766	ESC(40), MII oocyte(19), oocyte(707)	View Dataset	GEN
31554804	Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells	2019	Nat Commun.	GSE121436	scBS-seq	260	Muscle stem cell(260)	View Dataset	Single Cell Portal
-	Decoding aging and maturation related transcriptome and methylome dynamics in mouse oocytes by single cell parallel sequencing technology	2019	-	GSE119906	scM&T-seq	24	MII oocyte(12)	View Dataset	GEN
31847873	A KHDC3L Mutation Resulting in Recurrent Hydatidiform Mole Causes Genome-Wide DNA Methylation Loss in Oocytes and Persistent Imprinting Defects Post-Fertilisation	2019	Genome Med	GSE122872	scBS-seq	5	oocyte(5)	View Dataset	GEN
30078558	High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations	2018	Stem Cell Reports.	GSE89545	scPBAL	219	Hematopoietic Progenitor Cell(110), Hematopoietic Stem Cell(109)	View Dataset	Single Cell Portal



2.1 如何快速找到感兴趣的样本数据？

例如：寻找含有以下信息的样本

物种： 小鼠

性别： 雄性

细胞类型： 胚胎干细胞


[Home](#)
[Browse](#)
[Visualize](#)
[Tools](#)
[Download](#)
[Documentation](#)
[Home](#) > [Browse](#) > [Samples](#)

筛选特定特征下的单细胞甲基化测序的样本

Filter

Series

Tissue

Organism

Mus musculus

Development Stage

Source

ESC

Age

Cell Type

Sex

Male

Clear all

Advanced

Organism: Mus musculus;

Sex: Male;

Source: ESC;

🛒 (0)

Download Metadata

Download Data

数据下载

☐	Sample Accession	Series	Organism	Source	Tissue	Cell Type	Development Stage	Strain	Disease	Age	Sex	Race	Other Fields	#More Details
☐	GSM1370535 🔗	GSE56879	Mus musculus	ESC	Embryo	ESC	-	C57BL/6	-	-	-		-	+
☐	GSM1370536 🔗	GSE56879	Mus musculus	ESC	Embryo	ESC	-	C57BL/6	-	-	-		-	+
☐	GSM1370537 🔗	GSE56879	Mus musculus	ESC	Embryo	ESC	-	C57BL/6	-	-	-		-	+
☐	GSM1370538 🔗	GSE56879	Mus musculus	ESC	Embryo	ESC	-	C57BL/6	-	-	-		-	+
☐	GSM1370539 🔗	GSE56879	Mus musculus	ESC	Embryo	ESC	-	C57BL/6	-	-	-		-	+
☐	GSM1370540 🔗	GSE56879	Mus musculus	ESC	Embryo	ESC	-	C57BL/6	-	-	-		-	+



2.2 如何获得感兴趣的基因在不同组织或不同发育阶段中的甲基化情况？

例如：Pax9（广泛地参与脊椎和无脊椎动物胚胎发育过程中众多器官的形成）


[Home](#)
[Browse](#)
[Visualize](#)
[Tools](#)
[Download](#)
[Documentation](#)

[Home](#) > [Browse](#) > [Human Gene](#)

human

▼

Filters

Gene ID

▼

Gene Symbol

PAX9

▼

Position

chr --select--

▼

from

to

Go!

Description

▼

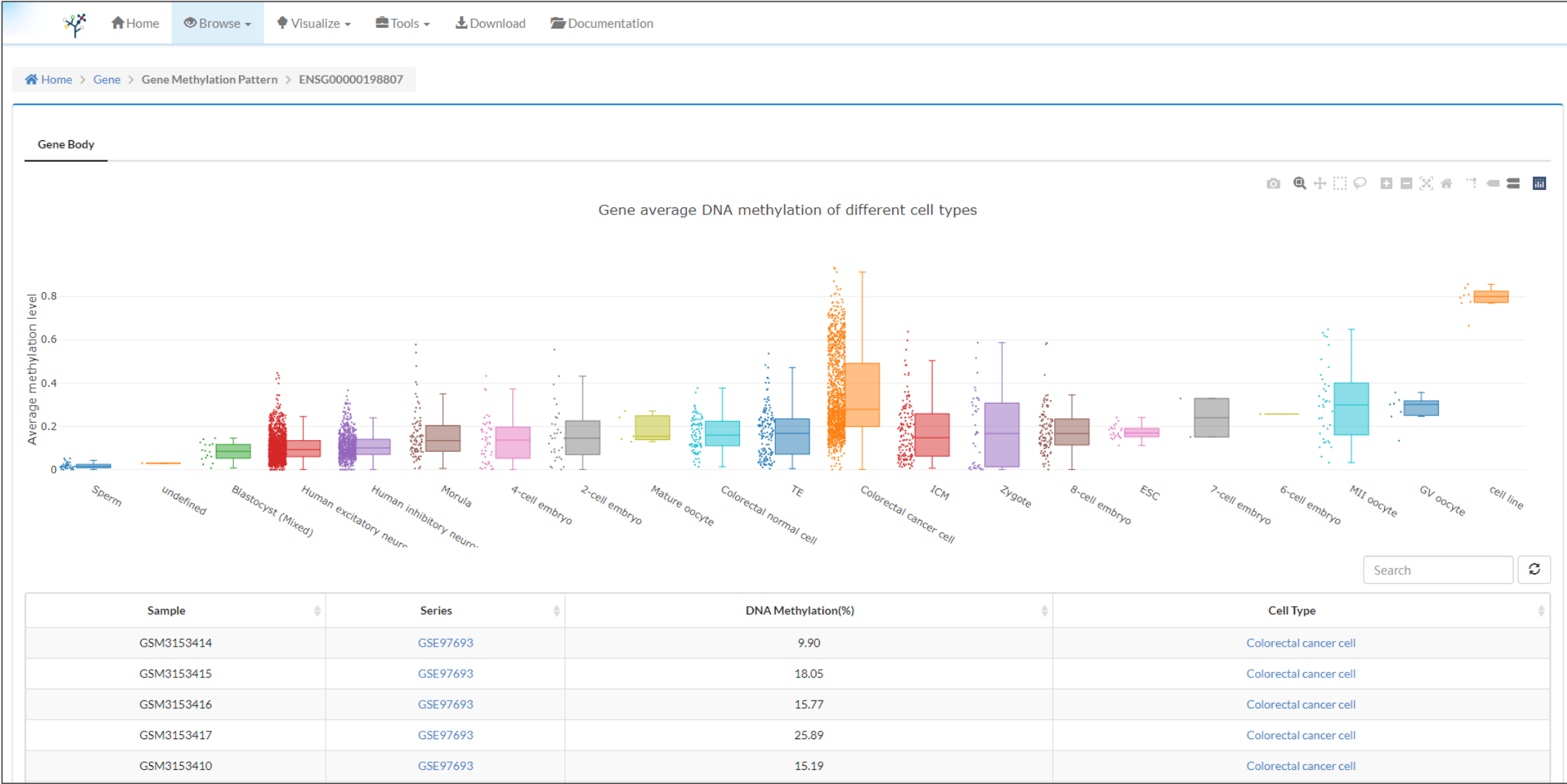
Gene ID	Gene symbol	Chr	Start	End	Strand	Description	Sample Pattern
ENSG00000198807	PAX9	14	36657568	36679362	+	paired box 9	

Showing 1 to 1 of 1 rows

选择基因或特定区域

点击跳转到基因结果页面





2.3 如何获得两种细胞类型间的差异甲基化?

例如：小鼠的E4.5和E5.5的差异甲基化区域 (DMR)



Home > Browse > DMR

DMRs

A visual overview of the DMRs (Differentially Methylated Regions) between different two cell groups.

Select one dataset and cell groups, then click the submit button to generate a DMR overview of your selection.

Selection of Datasets:

GSE116165-- [Genome-Scale Oscillations in DNA Methylation d...]

you can select one dataset in database

Selection of Cell Groups:

E4.5 VS E5.5

you can select any two cell groups

Submit

Reload

Organism: Mouse

Project: GSE116165

Cell Groups: E4.5 - E5.5

Chromosome Distribution

GO

KEGG

Genomic Location

Chromosome distribution graph

Download DMR Table

DMR 表

Search

Chr	Position	P Value	Differential Methylation Value	Annotation	Related Gene Ensemble ID	Related Gene Symbol	Methylation distr
chr10	58231802-58232100	5.3e-30	0.463	Promoter (<=1kb)	ENSMUSG00000075046	Dux	+

选择数据集和两种细胞，
点击查看具体的DMR结果。

National Genomics Data Center

51

输入
DMR

分析
参数
选择

Differentially methylated regions (DMRs) enrichment and pathway analysis

Organism ☒ Human ☐ Mouse

选择物种

Input File ?

选择文件 未选择任何文件

DMR List

输入 DMR list

#Example
chr1 6793377 6793562
chr1 9000992 9001174
chr1 9431014 9431130
chr1 11656042 11656229
chr1 14268679 14268831
chr1 21815768 21815839

Analysis

☒ DMR Annotation

☒ Go Analysis

☒ KEGG Analysis

Reference version

参考基因组版本

hg38

TSS upstream

2000

TSS上下游长度

TSS downstream

2000

pValue cutoff

0.05

p值阈值

padj cutoff

0.1

Run

This input has been submitted before. Job ID: 86b8a3407ca71820a836bfb2d30736fb copy

Job will take a few seconds, Thanks for waiting! Or

You can query the results by this job ID later

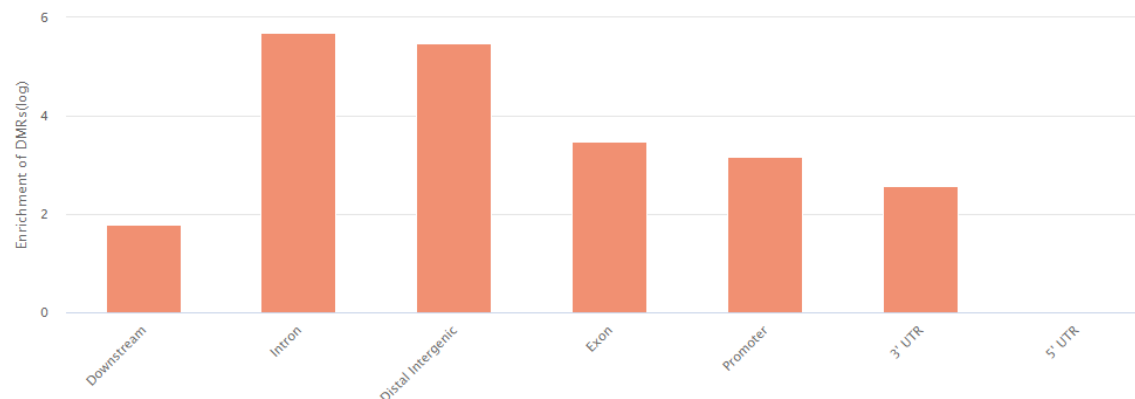
Annotation ✓

GO ✓

KEGG ✓

☒ log transformation ☐ absolute value

Genomic Element Location Enrichment



Chromosome	Start	End	Annotation	Related Gene	Ensemble ID
chr1	1239844	1240086	Downstream (2-3kb)	C1QTNF12	388581
chr1	2278181	2278347	Intron (ENST00000378536.5/6497, intron 1 of 6)	SKI	6497
chr1	4022604	4022803	Distal Intergenic	LINC01345	105376672
chr1	5072993	5073049	Distal Intergenic	AJAP1	55966
chr1	5306938	5307146	Exon (ENST00000641871.1/ENST00000641871.1, exon 1 of 4)	MIR4689	100616421



Gene functional enrichment and pathway analysis

Organism ☒ Human ☐ Mouse

Gene List **输入Gene List** #Example

ENSG00000000457
ENSG00000002330
ENSG00000002834
ENSG00000001561
ENSG00000001630

Analysis

☒ Go Analysis

☒ KEGG Analysis

pValue cutoff

0.05

padj cutoff

0.1 **p值阈值**

Run

This input has been submitted before. Job ID: a4fb5cd7f6bf30e34b425cbb491672ce copy
Job will take a few seconds, Tanks for waiting! Or
You can query the results by this job ID later

GO✓

KEGG✓

Gene Ontology Enrichment

Ontology Type	Term	p Value	FDR
MF	protein kinase B binding(GO:0043422)	2.79e-3	8.01e-3
MF	nucleotide diphosphatase activity(GO:0004551)	4.31e-3	8.01e-3
MF	cysteine-type endopeptidase activator activity involved in apoptotic process(GO:0008656)	4.31e-3	8.01e-3

2.4 如何进一步直观查看DMR内的差异情况？

例如：针对于小鼠的E4.5和E5.5的差异甲基化区域（DMR）

Chr	Position	P Value	Differential Methylation Value	Annotation	Related Gene Ensemble ID	Related Gene Symbol	Methylation distr
chr10	58231802-58232100	5.3e-30	0.463	 Promoter (<=1kb)	ENSMUSG00000075046	Dux	+

[Home](#) > [Visualization](#) > [Lollipop](#)

Methylation Lollipop

A visual overview of the DNA methylation pattern of region with lollipop plot.



Step1
Sample Selection



Step2
Lollipop Visualization

Select Gene

Select Region

Chromosome

10

Start

58231802

End

58232100

Run

Reload

选择特定的基因或区域

选择特定的样本（至少5个） E4.5 E5.5

Samples selected:

You have selected 6 samples with 1 series and 1 Organism and 1 Cell_type and 2 Development_stage

Sample Accession	Series	Organism	Source	Cell Type	Development Stage	Strain	Disease	Age	Sex
GSM3211748	GSE116165	Mus musculus	Embryo	Embryonic Cell	E4.5	C57Bl/6Babr			
GSM3211749	GSE116165	Mus musculus	Embryo	Embryonic Cell	E4.5	C57Bl/6Babr			
GSM3211750	GSE116165	Mus musculus	Embryo	Embryonic Cell	E4.5	C57Bl/6Babr			
GSM3211905	GSE116165	Mus musculus	Embryo	Embryonic Cell	E5.5	C57Bl/6Babr			
GSM3211906	GSE116165	Mus musculus	Embryo	Embryonic Cell	E5.5	C57Bl/6Babr			

Showing 1 to 5 of 6 rows

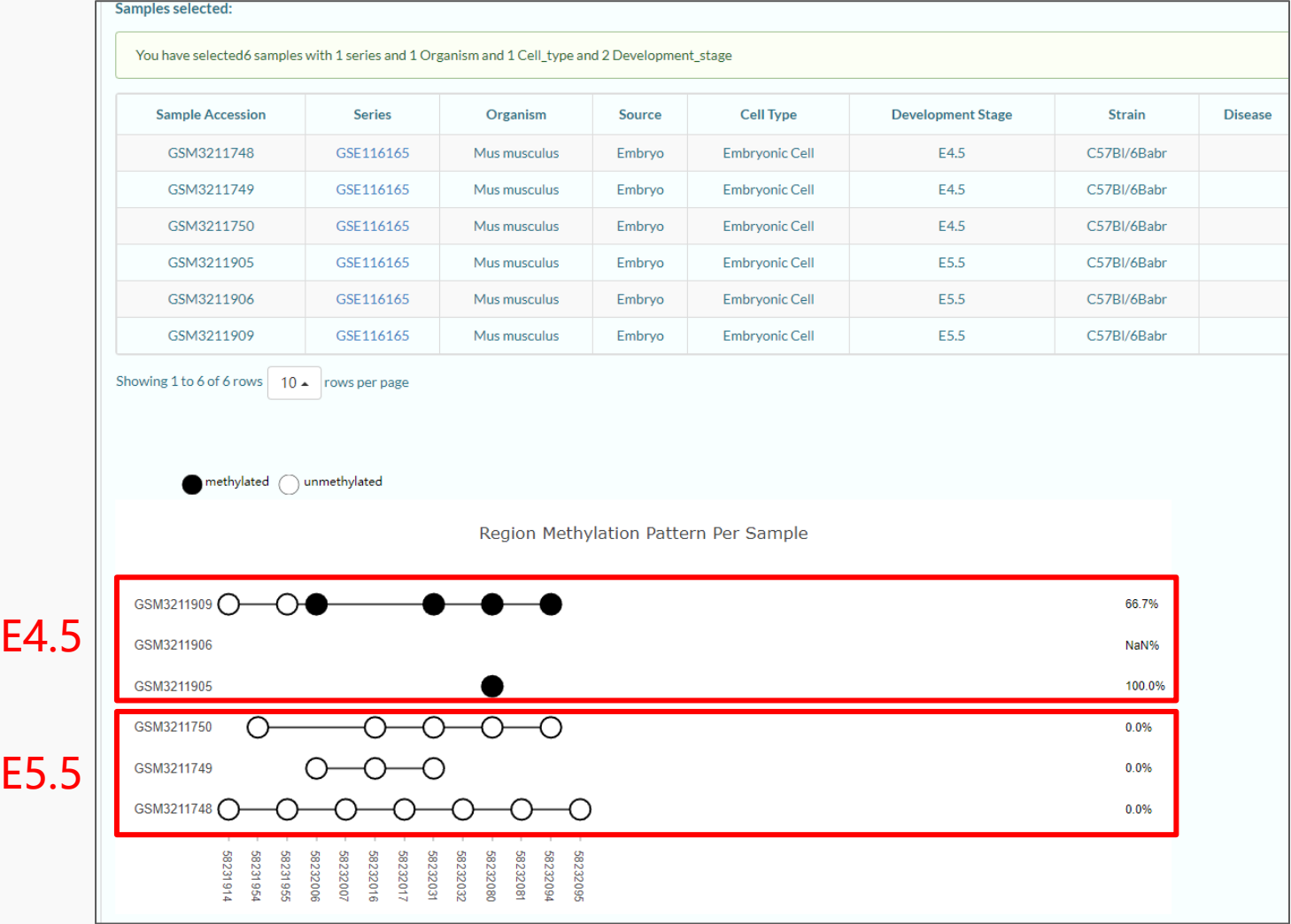
5

rows per page

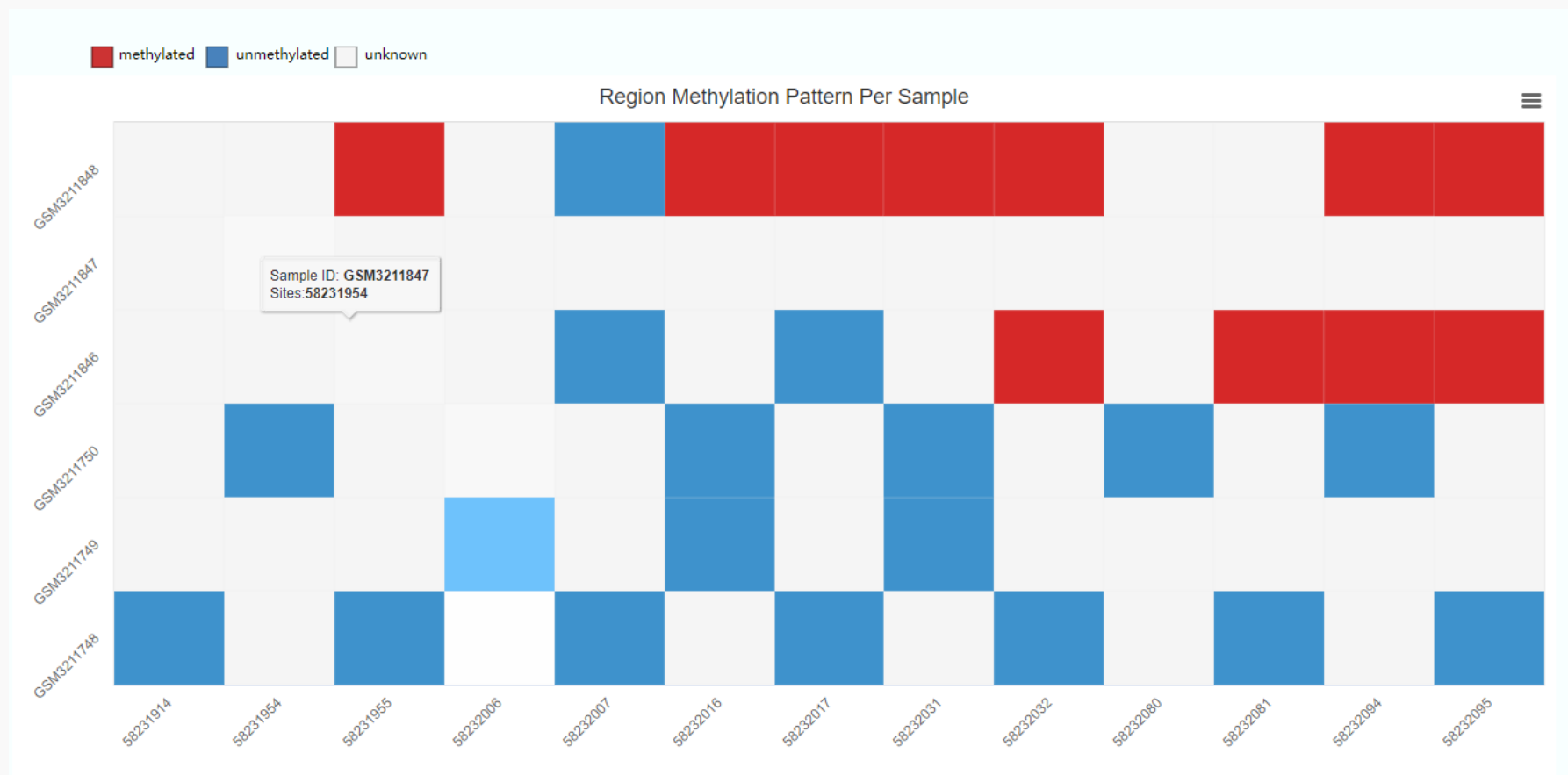
< 1 2 >

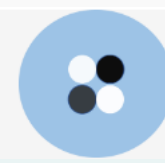


棒棒糖图：直观展示小鼠E4.5和E5.5在chr10 特定DMR上的具体差异情况。



热图展示小鼠E4.5和E5.5在chr10 特定DMR上的具体差异情况。





1

Please input your file :

未选择任何文件

[#Example File](#)

点击查看示例文件

Or you can paste the text to here:

```
position 13 1341 13414 134143 1341439 134143991 234143906 334143823 434144036
Oocyte1 NA 0.0113 0.0291 0.9108 0.0652 0.9222 0.519 0.9562 0.0253
ESC1 0.0151 0.0574 0.0492 0.0251 0.0957 0.0568 0.0396 NA 0.0252
Oocyte2 0.0143 0.0514 0.0727 0.0764 0.0748 0.0686 0.0219 0.0739 0.0207
ESC2 0.0148 0.053 0.9456 0.0162 0.0846 0.9675 0.0219 0.076 0.0185
Oocyte3 0.9627 0.9622 0.0763 0.9563 0.9868 0.0854 0.9306 0.9405 0.9851
```

Methylation file must be a tab separated file with the following structure:

- 1.Methylation levels must be indicated in percentage (from 0 to 1).
- 2.The first row must indicate the position of the CpGs (positions can be unsorted).
- 3.Each row contains the methylation levels for a sample.
- 4.First column indicates the sample identifier. Each middle column contains the methylation levels for a CpG.
- 5.Mind that the maximum number of both CpGs and samples to be plotted is 100.

2

Methylation threshold value:

You can customize the threshold or use our default parameters.

Default parameter means CpG sites with greater than 0.9 DNA methylation level were considered methylated, whereas CpG sites with less than 0.1 were considered unmethylated.

Methylation(<): 0.1

Unmethylation(>): 0.9

3

Plot Setting:

You can customize the plot parameters or use our default parameters.

Dot type parameter determines whether the dots are arranged continuously or discontinuously according to the location information of the site, Dot size determines the size of the drawing point.

Dot type:

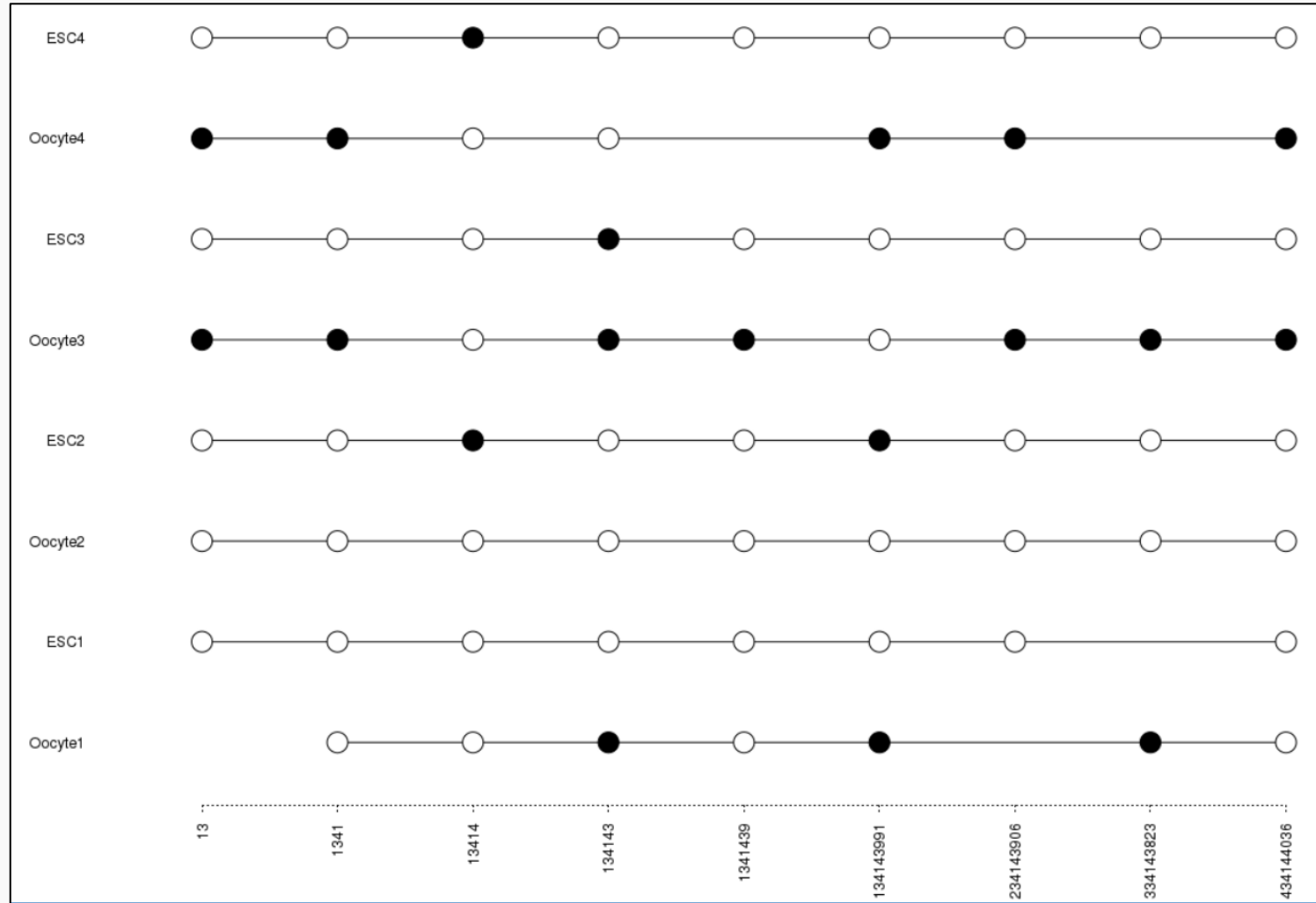
Non Proportional ▾

Dot size:

Dot Size: 3

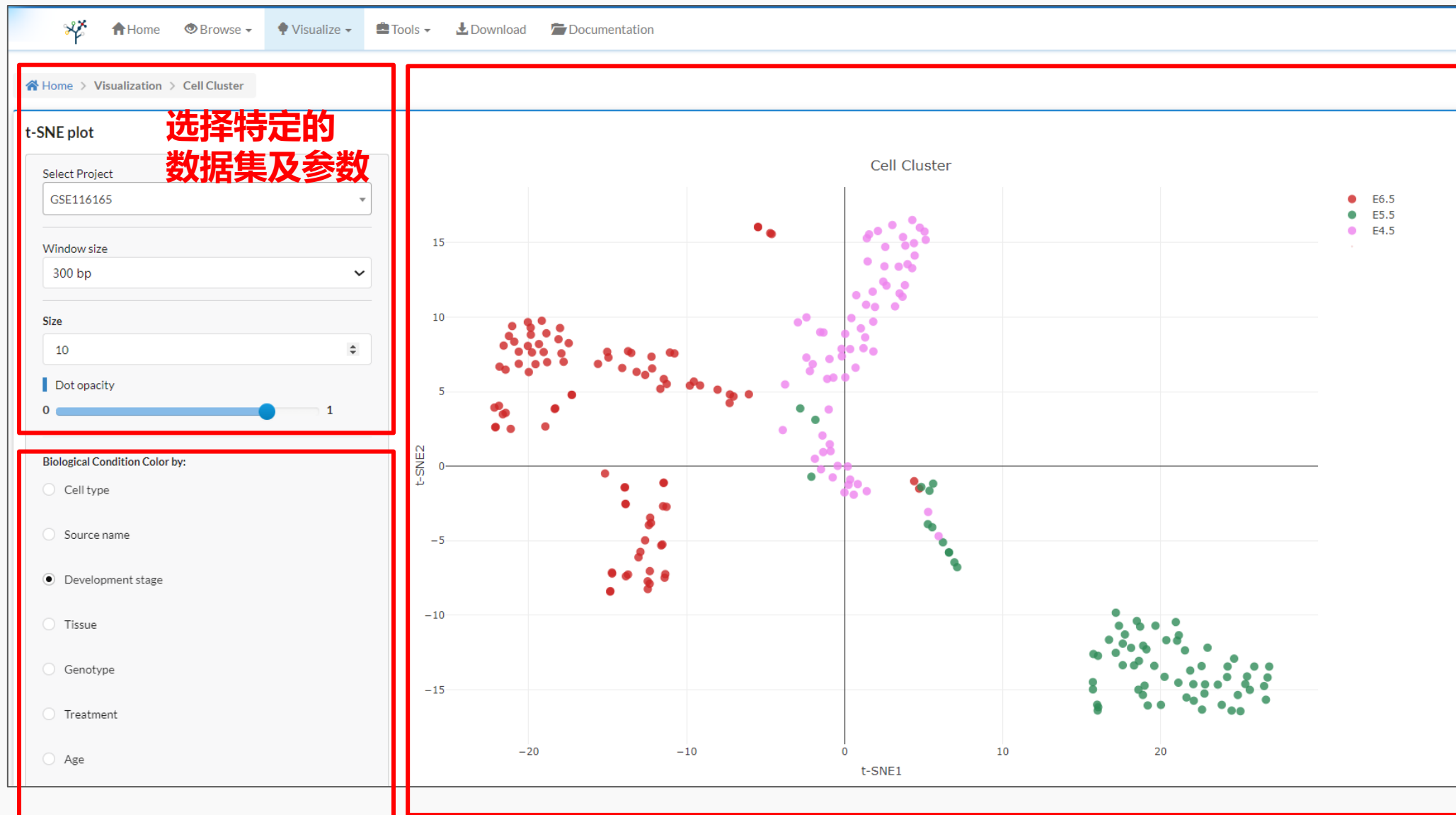


棒棒糖图 展示所有样本中每个CpG位点上的甲基化状态



2.5 如何查看细胞分群结果?





Download FTP:

The FTP url is convenient for advanced users to download BED files by Sample Accessions.

`ftp://download.big.ac.cn/scmethmap/singlebed/{SampleAccession}.bed.gz`

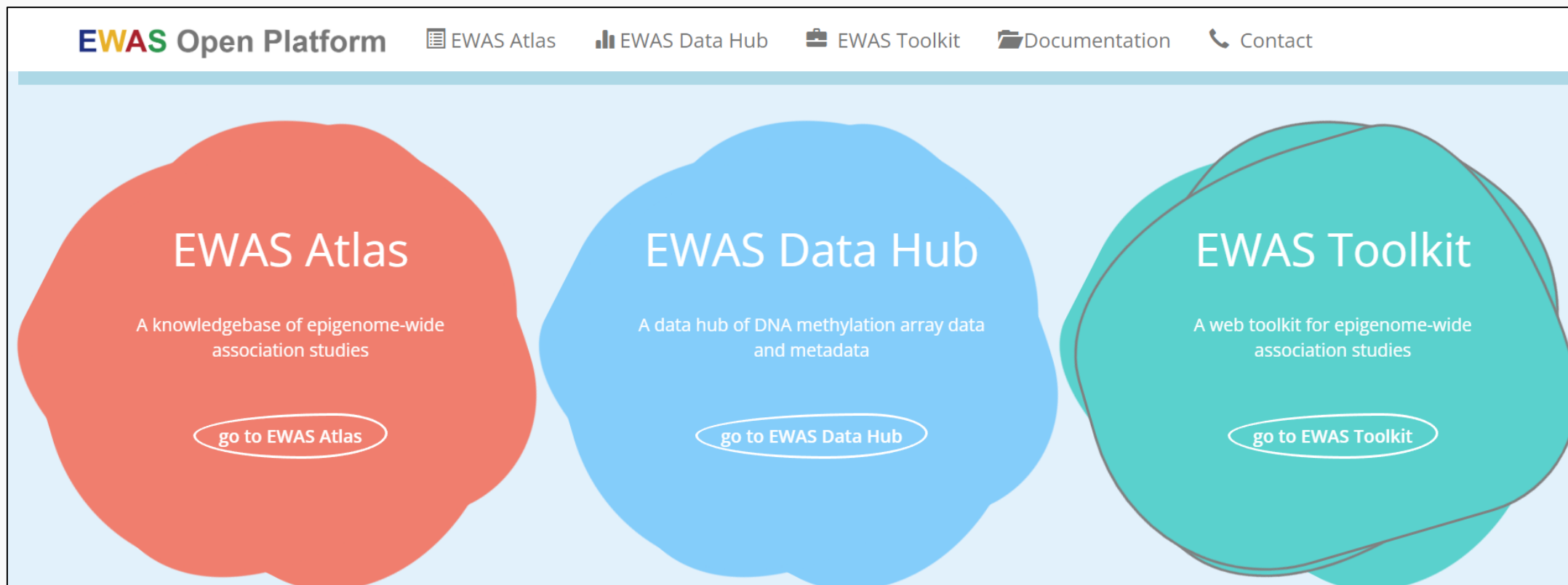
Example: <ftp://download.big.ac.cn/scmethmap/singlebed/GSM1370522.bed.gz>

Batch Download:

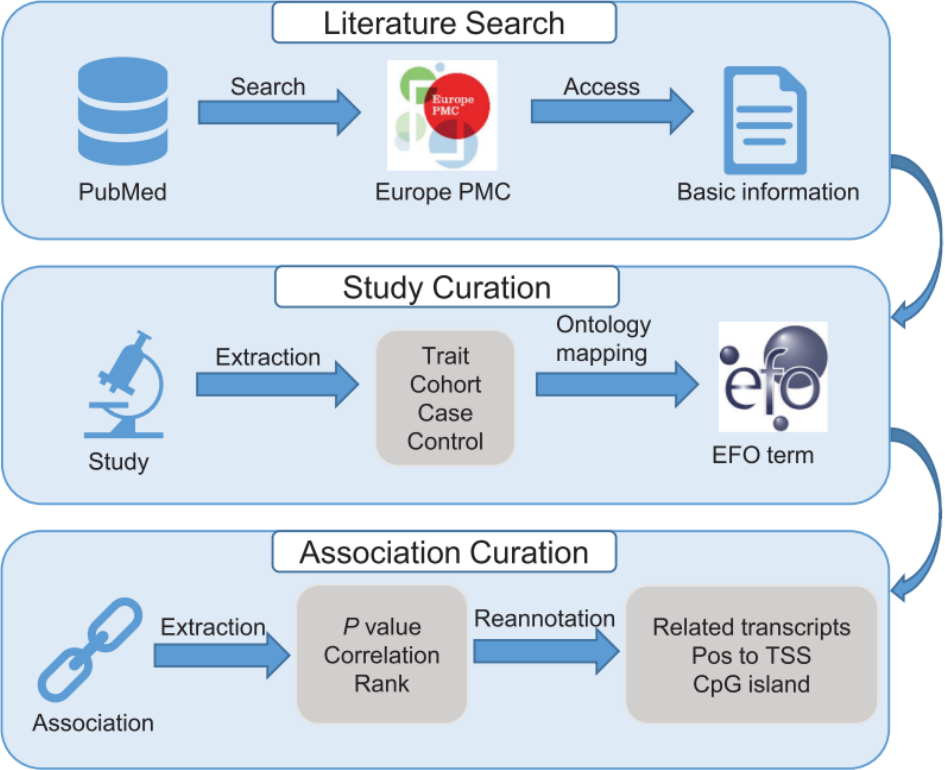
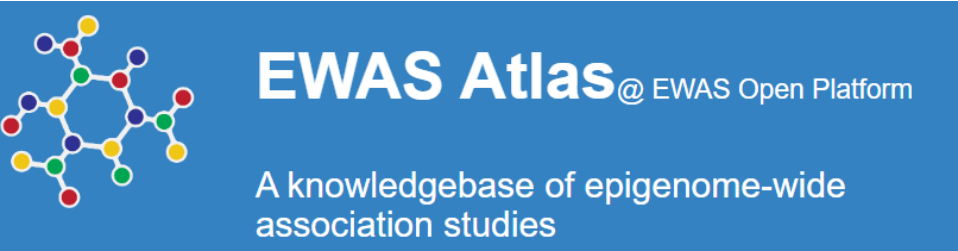
Cell type 	Organism 	SampleNumber 	Download (.tar.gz)	File Size
oocyte	Mus musculus	707	oocyte_Mouse.tar.gz	71 GB
oocyte	Homo sapiens	5	oocyte_Human.tar.gz	1.2 GB
MII oocyte	Mus musculus	71	MII_oocyte_Mouse.tar.gz	14 GB
MII oocyte	Homo sapiens	45	MII_oocyte_Human.tar.gz	8.5 GB
4 cell embryo	Mus musculus	36	4_cell_embryo_Mouse.tar.gz	22 GB
4 cell embryo	Homo sapiens	48	4_cell_embryo_Human.tar.gz	17 GB
GV oocyte	Mus musculus	12	GV_oocyte_Mouse.tar.gz	1.7 GB
GV oocyte	Homo sapiens	8	GV_oocyte_Human.tar.gz	1.6 GB
ESC	Mus musculus	261	ESC_Mouse.tar.gz	57 GB
ESC	Homo sapiens	22	ESC_Human.tar.gz	14 GB
TE	Homo sapiens	140	TE_Human.tar.gz	101 GB
Carcinoma cell	Homo sapiens	578	CRC_Human.tar.gz	157 GB
ICM	Homo sapiens	146	ICM_Human.tar.gz	102 GB
8 cell embryo	Homo sapiens	116	8_cell_embryo_Human.tar.gz	48 GB
Morula	Homo sapiens	88	Morula_Human.tar.gz	30 GB
embryo	Mus musculus	238	embryo_Mouse.tar.gz	27 GB
Blastocyst	Homo sapiens	28	Blastocyst_Human.tar.gz	17 GB



三 芯片甲基化资源库 EWAS Open Platform



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



Associations

643,805

Traits

728

Cohorts

3,510

Tissues/Cells

199

Studies

1,599

Publications

1002

EWAS Atlas
EWAS Open Platform

Browse
EWAS Toolkit
Downloads
Statistics
API
Help
EWAS Data Hub

Filter
Hide

Trait(s)

Gene symbol

Promoter/Body

Probe ID

P value rank
from to Go!

Clear all
Hide

Position
chr --select--
from to Go!

CpG island

Trait type

Ontology

Traits (727)
Probes (301524)
Genes (36041)
Studies (1599)
Publications (1002)

Correlation: Hypermethylation Hypomethylation Not report
CpG Island: Island Shore Shelf Open sea
+ Show detail information or genome browser
Click on a [hyperlink](#) will use it as a search condition

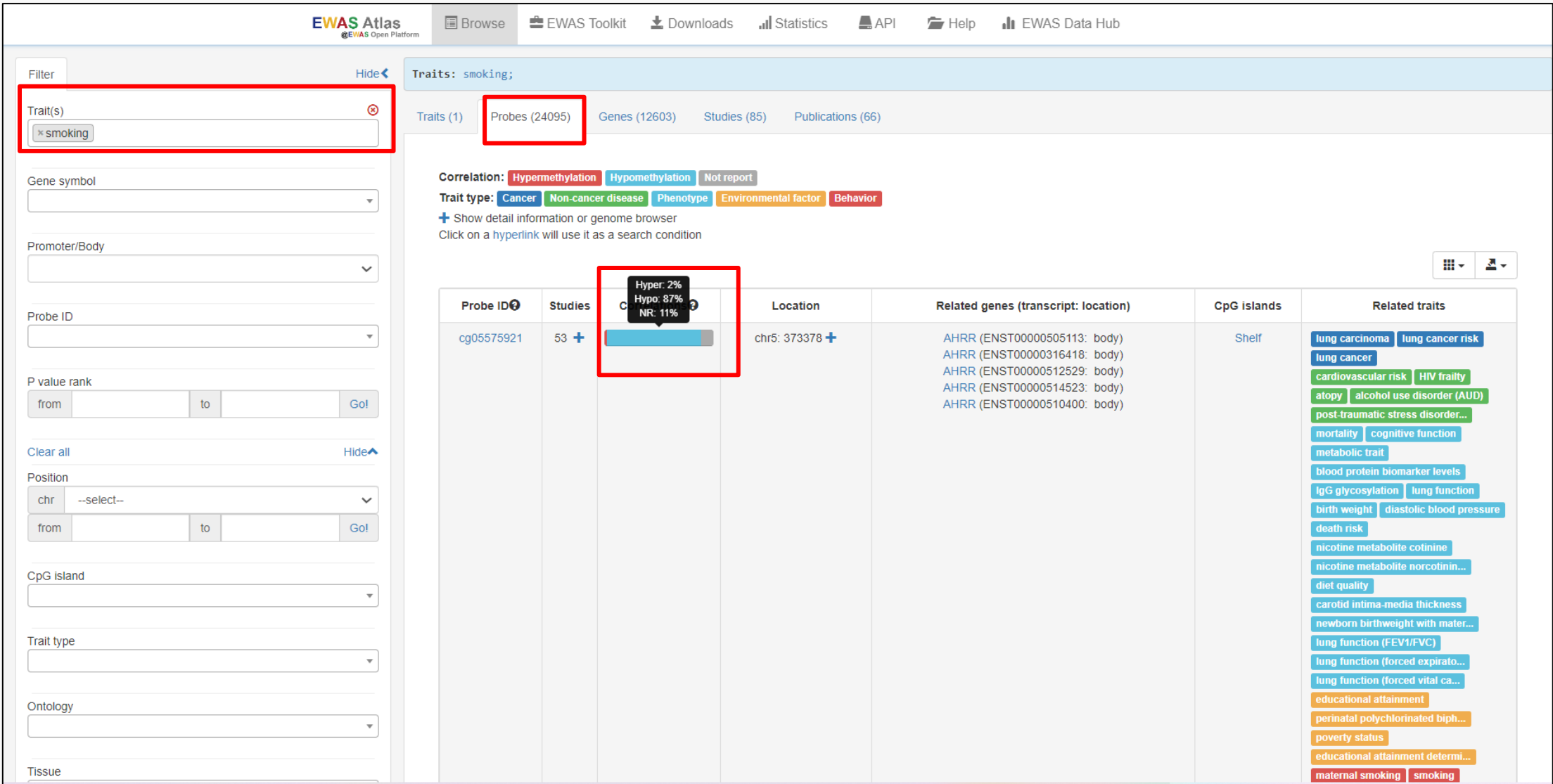
Trait	Type	#Publications	#Studies	#Associations	Correlations?	CpG Island
1,6-hexamethylene diisocyanate exposure	environmental factor	1	2	32		
17q12 deletion	phenotype	1	1	21		
abdominal obesity	phenotype	1	1	14		
absolute fat free mass (FFM, kg)	phenotype	1	1	250		
absolute fat mass (FM, kg)	phenotype	1	1	250		
acute coronary syndrome	non-cancer disease	1	1	9		
acute mania	non-cancer disease	1	1	1		
acute myelocytic leukemia (AML)	cancer	1	2	28		
acute respiratory distress syndrome (ARDS)	non-cancer disease	1	1	4		
adiposity	phenotype	1	3	6		

Showing 1 to 10 of 727 rows
10 rows per page

1 2 3 4 5 ... 73

3.1 如何查看感兴趣的trait与甲基化的研究现状？

例如：吸烟会引发哪些位点的甲基化变化呢？哪些文章报道了呢？结果是否稳定呢？



吸烟引发哪些基因相关的位点的甲基化变化呢?

吸烟与AHRR基因甲基化相关, 还有哪些Trait与AHRR甲基化相关呢?

EWAS Atlas
Browse EWAS Toolkit Downloads Statistics API Help EWAS Data Hub

Filter Traits: smoking; Hide

Trait(s) * smoking

Gene symbol

Promoter/Body

Probe ID

P value rank from to Go!

Clear all Hide

Position chr --select-- from to Go!

CpG island

Trait type

Ontology

Traits (1) Probes (24095) Genes (12603) Studies (85) Publications (66)

Promoter/Body: Promoter Body
Correlation: Hypermethylation Hypomethylation Not report
Trait type: Cancer Non-cancer disease Phenotype Environmental factor Behavior
+ Show detail information or genome browser
+ Jump to other related website
Click on hyperlink will use it as a search condition

Gene ID	Location	Expression	#Association	Promoter/Body	Correlations	#Related Traits	Most Related traits (#Associations)
AHRR ENSG00000063438	(+) chr5:304291-438406	+	584			75 +	smoking (584) maternal smoking (111) smoking cessation (25) HIV frailty (16) educational attainment (16)
AC068134.5 ENSG00000204121	(-) chr2:233280528-233284526	+	204			36 +	smoking (204) alcohol consumption (15) smoking cessation (10) aging (7) educational attainment (7)
VAR5 ENSG00000204394	(-) chr6:31745295-31763730	+	127			41 +	smoking (127) aging (20) SETD1B-related syndrome (9) Werner syndrome (7) asthma (6)
GFI1 ENSG00000162676	(-) chr1:92940319-92952433	+	118			64 +	smoking (118) maternal smoking (72)

该基因在不同组织中的表达情况。

https://ngdc.cncb.ac.cn/ewas/browse#study

不同研究的结果存在差异，应该怎么寻找原因呢？
查看85个吸烟相关的研究，使用的是什么样本

Filter

Trait(s)

* smoking

Gene symbol

Promoter/Body

Probe ID

P value rank

from

to

Go!

Clear all

Advanced

Traits: smoking;

Traits (1) Probes (24095) Genes (12603) **Studies (85)** Publications (66)

+ Show detail information or genome browser

Jump to other related website

Click on [hyperlink](#) will use it as a search condition

Grid

Print

Study ID	Trait	Ontology	#Associations	#Cohorts	Source	Case	Control
ES00080	smoking	smoking behavior smoking status measurement	5	6 +	peripheral blood	smoking status	
ES00081	smoking	smoking behavior smoking status measurement	15	2 +	whole blood	current smokers	never smokers
ES00082	smoking	smoking behavior pack-years measurement	2	2 -	whole blood	highest pack-years quartile	lowest pack-years quartile

Details

Stage	Platform	Tissue	Sample size	Age(min)	Age(max)	Age(mean)	Age(sd)	Male(%)	Ancestries	Cohort name	Description
Initial	450K	whole blood	1085	45	65	57.3	8.1	0.54	European	ICGN	
Replication	450K	whole blood	369	-	53	47.5	7.1	0.36	Not reported	EOCOPD	

Showing 1 to 2 of 2 rows

| [ES00084](#) | smoking | [smoking behavior](#) [smoking status measurement](#) | 29 | 1 + | [lymphocyte](#) | smoking status | |

National Genomics Data Center

69

EWAS Atlas
EWAS Open Platform

Browse

EWAS Toolkit

Downloads

Statistics

API

Help

EWAS Data Hub

Filter

Hide

Trait(s)
* smoking

Gene symbol

Promoter/Body

Probe ID

P value rank
from to Go!

Clear all

Hide

Position
chr --select--
from to Go!

CpG island

Trait type

Traits: smoking;

Traits (1) Probes (24095) Genes (12603) Studies (85) Publications (66)

[Jump to other related website](#)
Click on hyperlink will use it as a search condition

PMID

Title

Year

Journal

Citation

Datasets

21457905

Tobacco-smoking-related differential DNA methylation: 27K discovery and replication.

2011

Am J Hum Genet

328

22492999

Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome.

2012

Hum Mol Genet

153

23070629

Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers.

2012

Epigenetics

74

23175441

Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking.

2013

Hum Mol Genet

201

23657504

Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans.

2013

Hum Genet

78

23691101

Tobacco smoking leads to extensive genome-wide changes in DNA methylation.

2013

PLoS One

291

24334605

Smoke-related DNA methylation changes in the etiology of human disease.

2014

Hum Mol Genet

75

24485148

Differences in smoking associated DNA methylation patterns in South Asians and Europeans.

2014

Clin Epigenetics

86

24559495

The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women.

2014

BMC Genomics

102

GSE53045

24704585

CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the Sister Study.

2014

Environ Health Perspect

54

Showing 1 to 10 of 66 rows 10 rows per page

<

1

2

3

4

5

6

7

>



3.2 如何快速查找限定条件下trait与甲基化的关联关系？

例如：全血中，与aging相关的位于promoter区域的CpG island甲基化

Filter

Trait(s)

× aging

Gene symbol

Promoter/Body

promoter

Probe ID

P value rank

from

to

Go!

Clear all

Hide

Position

chr --select--

from to Go!

CpG island

Island

Trait type

Ontology

Tissue

whole blood

Published year

PMID

Study ID

Traits: aging;
CpG island: Island;
Promoter/Body: promoter;
Tissue: blood;

Traits (1) Probes (208) Genes (202) Studies (3) Publications (3)

Correlation: **Hypermethylation** Hypomethylation Not report

Trait type: **Cancer** Non-cancer disease Phenotype Environmental factor Behavior

+ Show detail information or genome browser

Click on a [hyperlink](#) will use it as a search condition

Probe ID	Studies	Correlations	Location	Related genes (transcript: location)
cg22920873	2 +		chr7: 139025153 +	C7orf55 (ENST00000468383: promoter) C7orf55 (ENST00000481123: promoter) C7orf55 (ENST00000482181: promoter) C7orf55 (ENST00000488886: promoter) C7orf55 (ENST00000297534: promoter) LUC7L2 (ENST00000541170: promoter) C7orf55-LUC7L2 (ENST00000541515: promoter)



EWAS Data Hub

EWAS Open Platform

[Probes & Genes](#)
[Samples](#)
[EWAS Toolkit](#)
[Download](#)
[Statistics](#)
[Help](#)
[EWAS Atlas](#)



EWAS Data Hub@ EWAS Open Platform

A data hub of DNA methylation array data and metadata

Example: cg16867657; DNMT1; 850K; age (year); brain - cerebellum;

Browse

 Probes

 Genes

 Samples

Reference DNA Methylation Profiles (cg16867657)

 Basic Information

 Aging

 Cancers

 Tissue

 Ancestry Categories

 Other Diseases

 Sex

 BMI

 Public EWAS

Other Features

 Normalization

 Download

 Documentation

Resource Overview

133,672 Samples	1,099 Tissues/Cells
612 Diseases	257 Fields

Recent Updates

2022-06-15

New 17820 samples were added

2021-06-15

New 218 cell lines were added

2021-06-10

New 40508 samples were added

2019-06-27

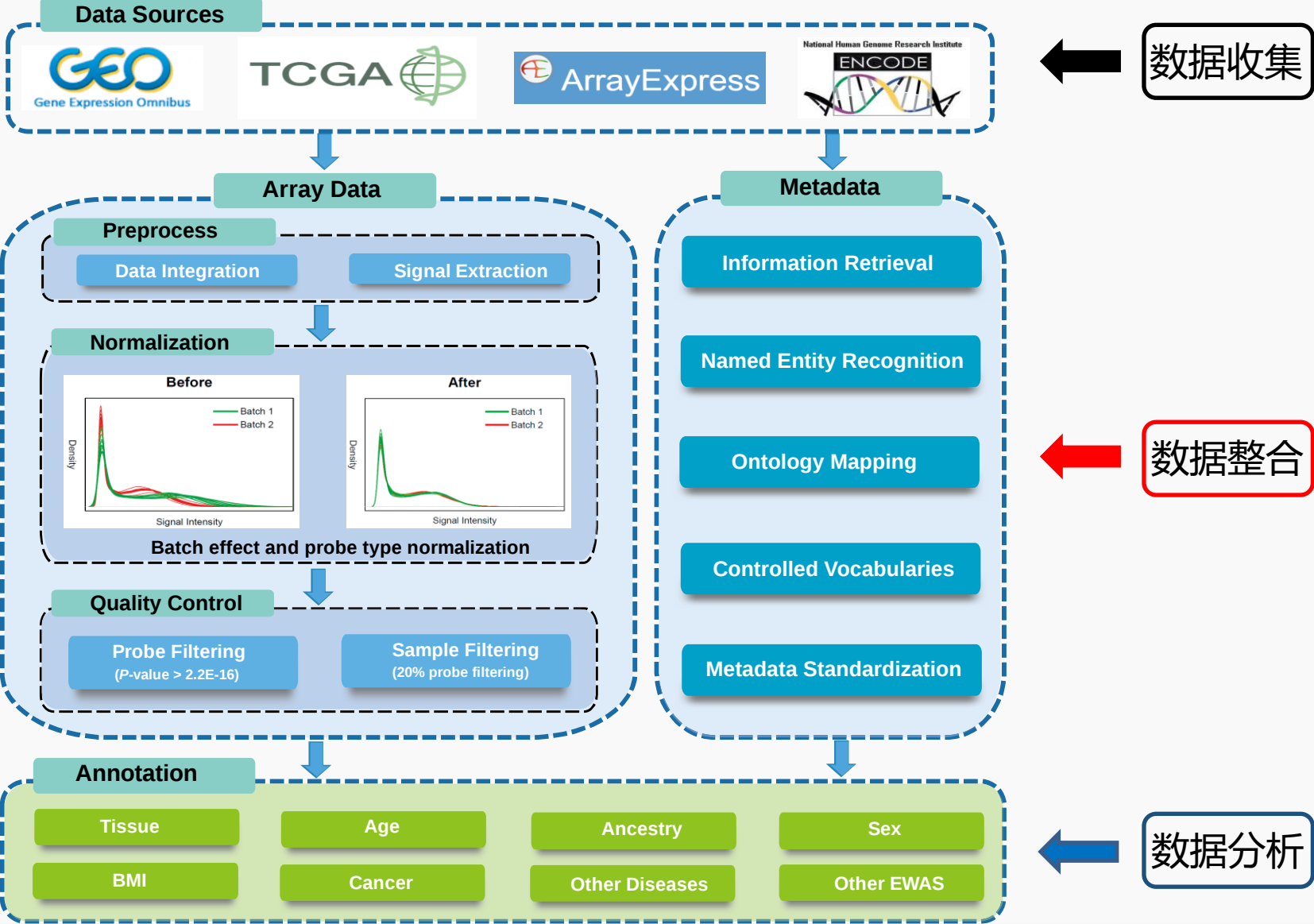
Three new 850K projects were added

2019-06-08

EWAS Data Hub version 1 was released

Please Cite

- EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res* 2022. [PMID=34718752]
- EWAS Data Hub: a resource of DNA methylation array



Gaussian Mixture Quantile Normalization (GMQN)

Methods

To remove the batch effects and other unwanted noise, we develop Gaussian Mixture Quantile Normalization (**GMQN**), a reference based method that removes unwanted technical variations at signal intensity level. GMQN adjusts batch effects as well as bias associated with type II probe values in 450k and EPIC/850K studies. The principle behind this method is that the signal intensity of each channel displays a Gaussian mixture distribution. The first component is the background signal which has a mean slightly greater than 0. The second component is the signal from probes which have been hybridized to input DNA successfully. Variance of the second component is much larger than the first component because the degrees of hybridization are different among probes.

The object of GMQN is to rescale the signal intensity to make the two Gaussian component from different array have the same mean and variance. There are four steps to perform GMQN.

1. Fitting of a two-state Gaussian mixture model to the median values of each type I probe signal intensity from a large single study (GEO project id: [GSE105018](#)). The mean and variance of two components are used as reference for rescaling type I probes.
2. Fitting of a two-state Gaussian mixture model to the input type I probe signal intensity.
3. For type I probes from each component of input data, transform their probabilities to quantiles using the inverse of the cumulative Gaussian distribution with mean and variance estimated from the corresponding reference component.
4. Calculating beta value and normalizing type II probes on the basis of type I probes using BMIQ.

Citation

GMQN: A Reference-Based Method for Correcting Batch Effects and Probe Bias in HumanMethylation BeadChip. *Front.Genet.* 2022. [PMID=35069703]



133,672

Samples



257

Fields



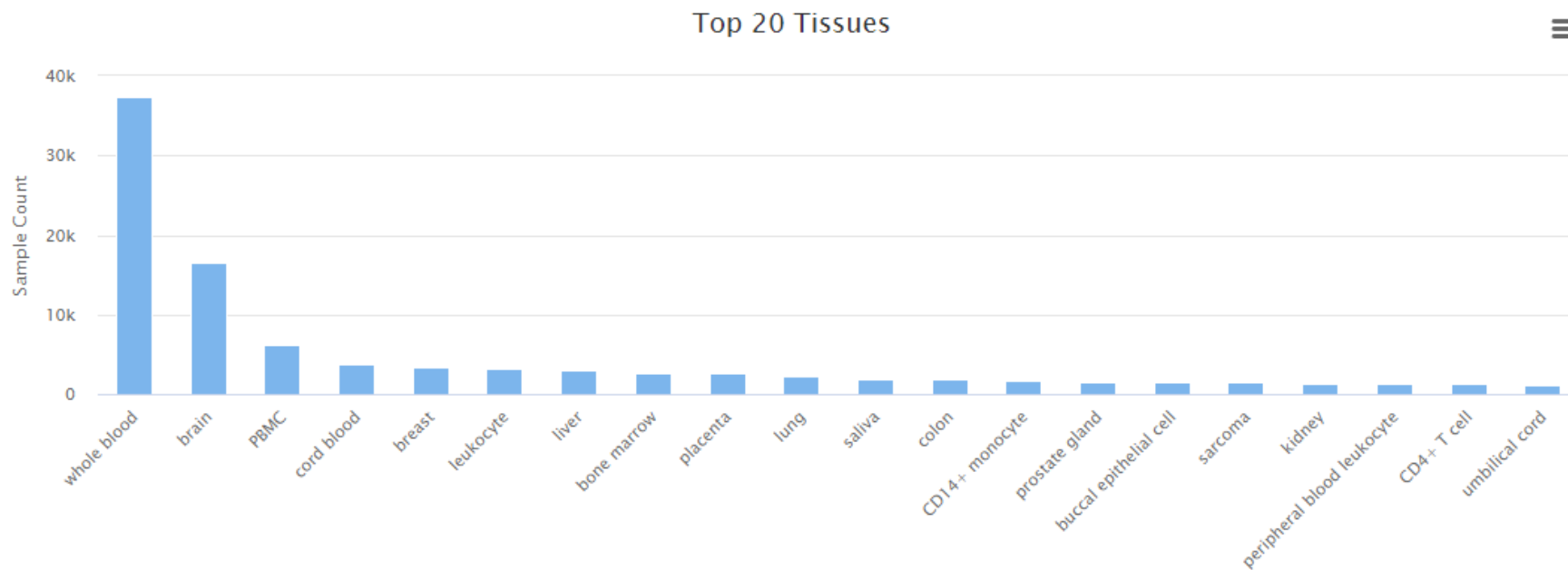
1,099

Tissues/Cells



612

Diseases



3.3 如何快速找到感兴趣的样本数据？

例如：寻找含有以下信息的样本

年龄：30-80岁

性别：男

组织：肺

疾病：肺癌

Advanced

Example: cg16867657; DNMT1; 850K; age (year); brain - cerebellum;

age (year): 30.4160001 to 80.7953973; AND sex: M; AND tissue: lung; AND disease: lung cancer;

检索结果

高级检索

(4)

Download Metadata

Download Data

检索条件

下载

Click on a [hyperlink](#) will use it as a search condition

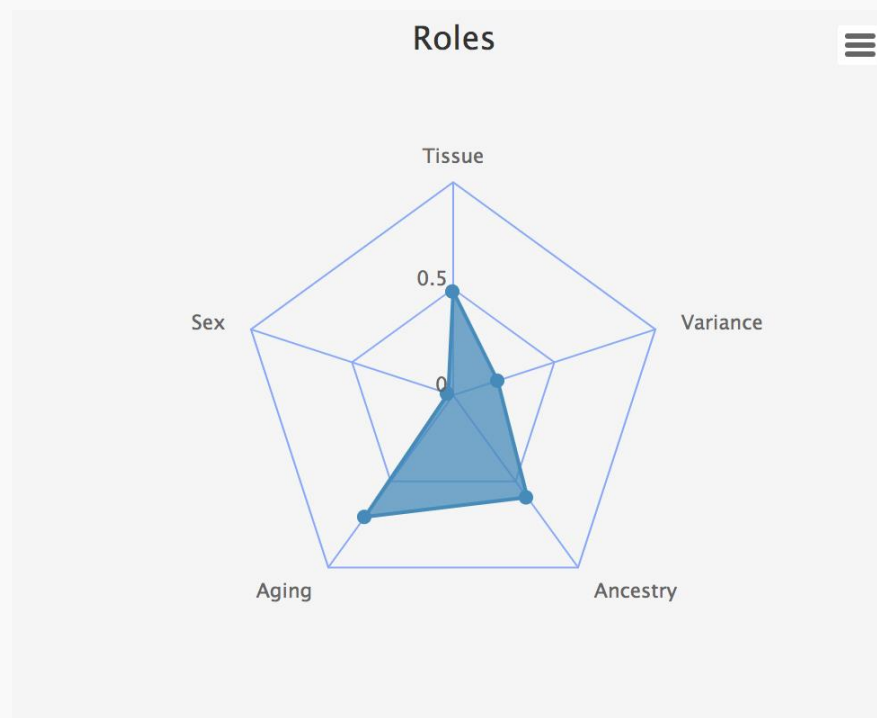
☐	Sample ID	Project ID	Tissue	Sample Type	Platform	Other Fields
☒	GSM1281077	GSE53051	lung	disease tissue	450K	age (year) 50 disease lung cancer sex M
☒	GSM1281113	GSE53051	lung	disease tissue	450K	age (year) 57 disease lung cancer sex M
☐	GSM1281147	GSE53051	lung	disease tissue	450K	age (year) 78 disease lung cancer sex M
☐	GSM1281202	GSE53051	lung	disease tissue	450K	age (year) 67 disease lung cancer sex M
☐	GSM1281266	GSE53051	lung	disease tissue	450K	age (year) 74 disease lung cancer sex M
☒	GSM2482857	GSE94785	lung	disease tissue	450K	treatment asbestos unexposed age (year) 70 disease lung cancer sex M pack years 38 paired id(disease) 1003
☒	GSM2482858	GSE94785	lung	disease tissue	450K	treatment asbestos unexposed age (year) 64 disease lung cancer sex M pack years 66 paired id(disease) 1004
☐	ENCSR050XGE	ENCODE	artery - tibial	control	850K	age (year) 37 sex M

Add a filter

Search

来源: <https://bigd.big.ac.cn/ewas/datahub/repository>

3.4 如何快速筛选出特异的甲基化位点和基因？



Filter

Hide

Probe ID

Gene symbol

Position

chr

--select--

from

to

Go!

Tissue-specific hypermethylation score

0

1

Tissue-specific hypomethylation score

0

1

Correlation with age

-1

1

Methylation level difference between sex (F-M)

-1

1

Probes (77)

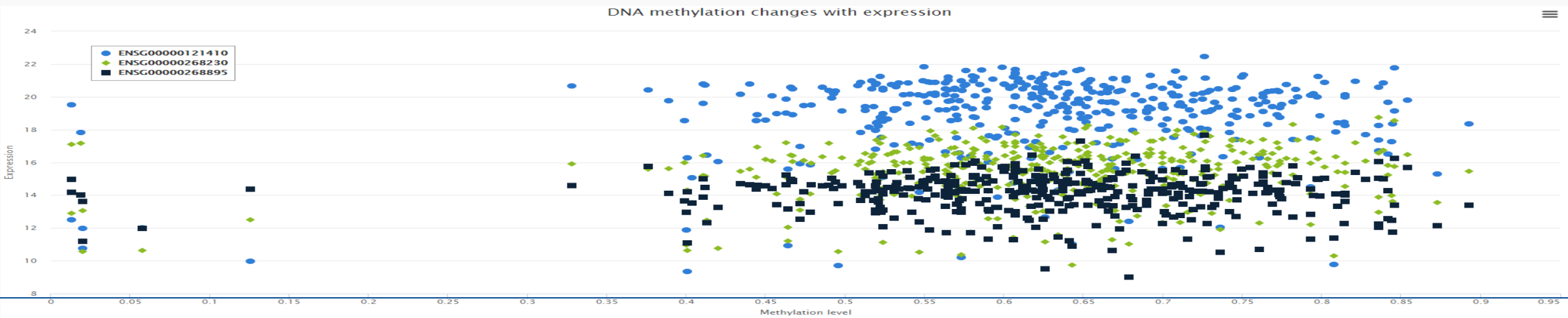
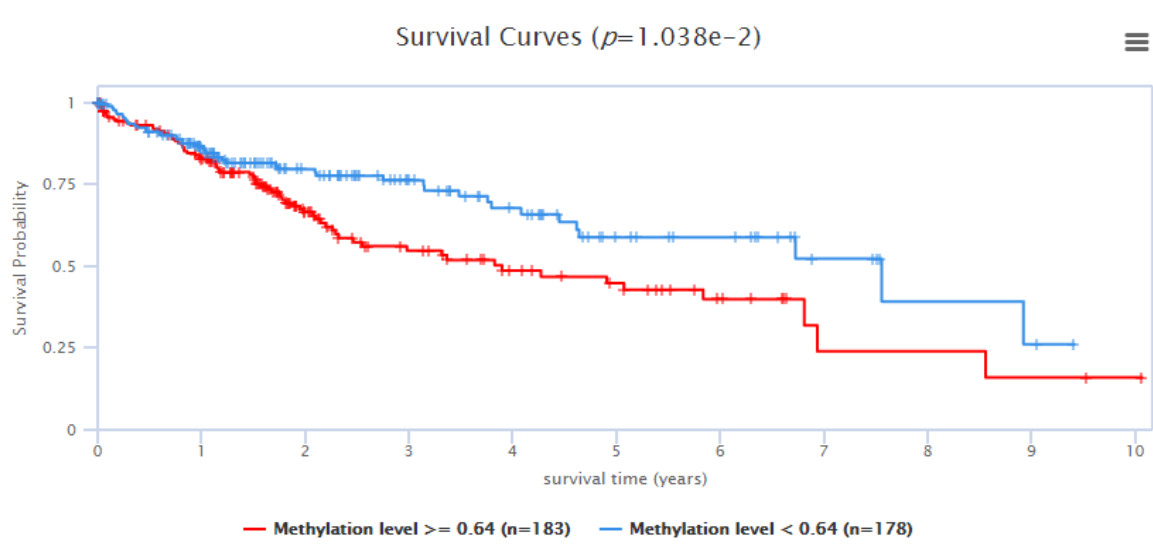
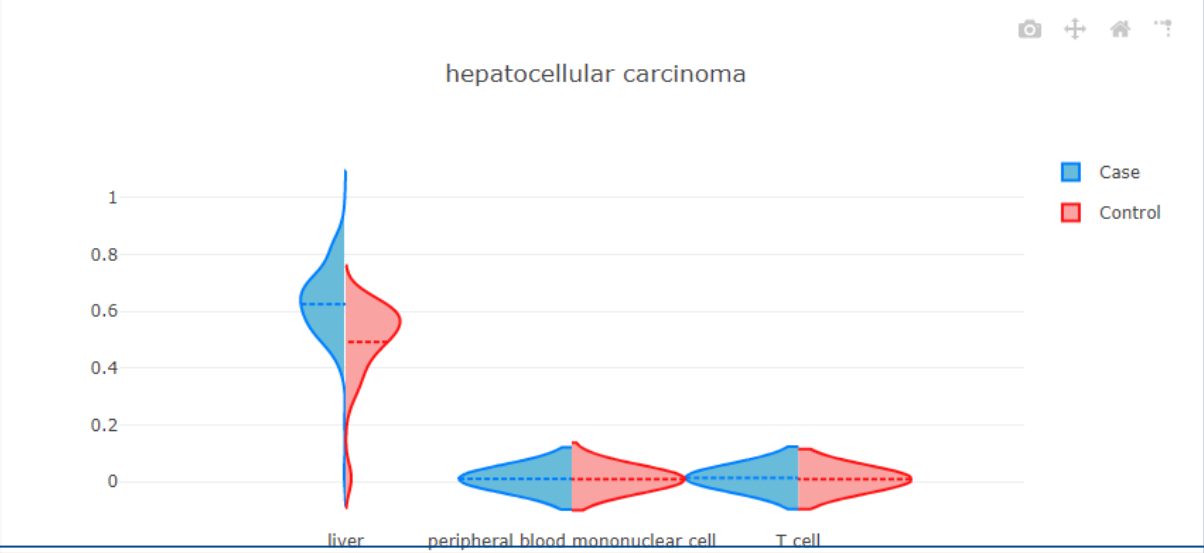
Genes (68)

Probe ID	Basic Information	TS (Hyper)	TS (Hypo)	Cor(age)	Sex difference score	Ancestry (Hyper)	Ancestry (Hypo)
ch.18.734816F	Location:chr18 37716523 CpG Island: Open sea Related Gene:	0.950272	0.0056558	0.266682	-0.0126306	0.567678	0.00236195
ch.4.179413987F	Location:chr4 179176993 CpG Island: Open sea Related Gene:	0.950584	0.0143543	0.120079	-0.00485827	0.7816	0.00954494
ch.4.27504789F	Location:chr4 27895691 CpG Island: Open sea Related Gene:	0.950596	0.010786	0.172793	-0.00399807	0.361049	0.00260969
cg09425247	Location:chr10 52178225 CpG Island: Shore Related Gene: SGMS1	0.950885	0.0142073	-0.0636153	-0.00349724	0.252067	0.00172077
cg22286978	Location:chr19 58858806 CpG Island: Island Related Gene: A1BG CTD-2619J13.8 A1BG-AS1	0.951478	0.0381678	0.0677108	-0.00994882	0.158318	0.00274152
ch.1.194862R	Location:chr1 5228961 CpG Island: Open sea Related Gene:	0.951522	0.00740018	0.231407	-0.00751737	0.581766	0.00318644
cg13341720	Location:chr19 6740888 CpG Island: Shelf Related Gene: TRIP10	0.951768	0.0126817	0.033887	0.000428855	0.150361	0.00177123

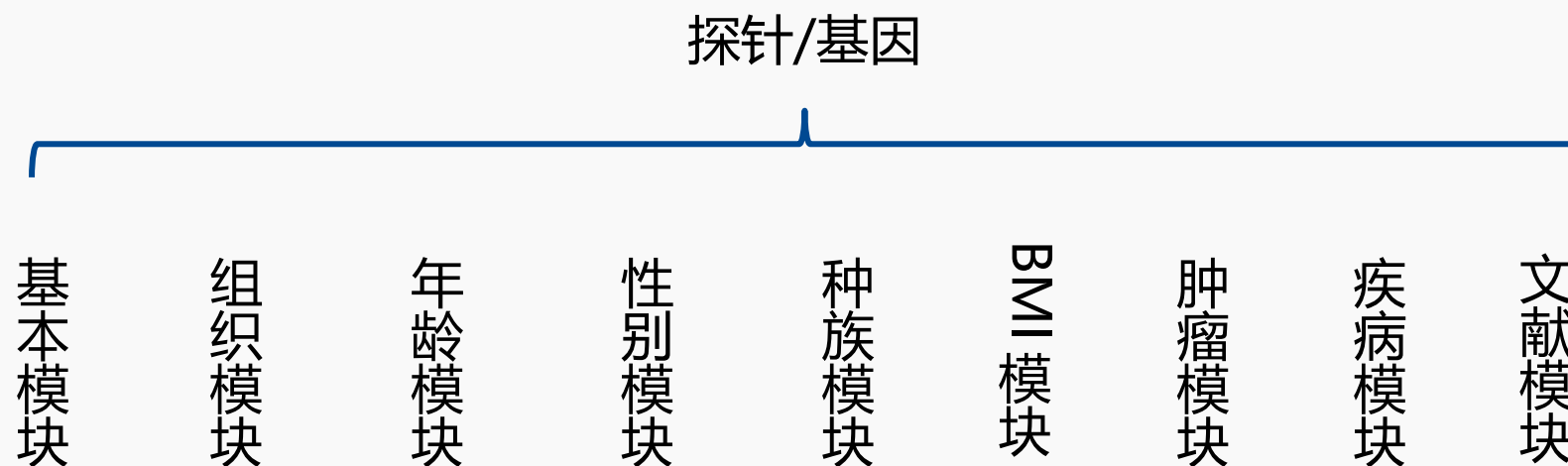




Cancer name: hepatocellular carcinoma



如何查看自己关注的甲基化位点和基因的多模态甲基化图谱？

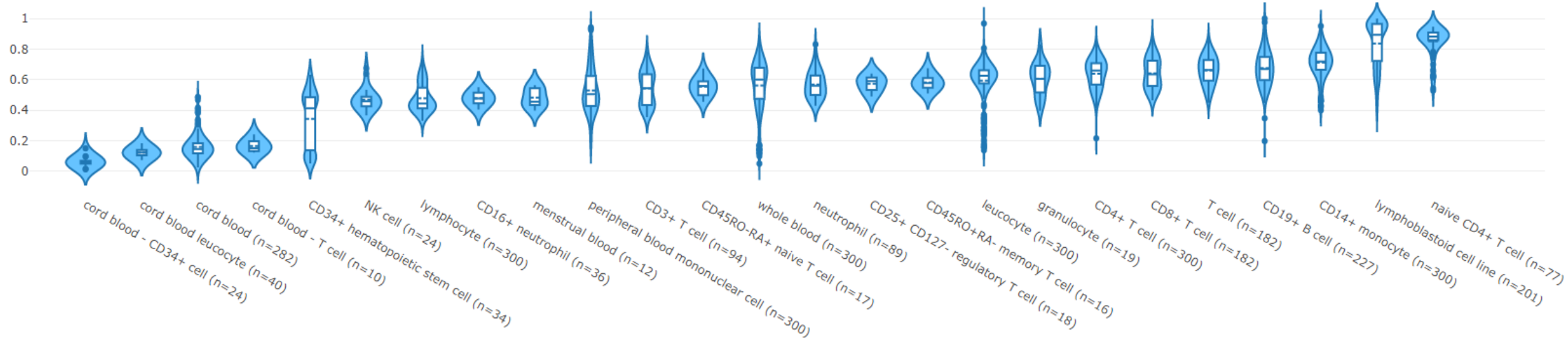


cg16867657

Basic Tissue Sex Age Ancestry Category BMI Cancer Other diseases Public EWAS

Tissues Brain parts Blood cell types

DNA methylation profiles of different blood cell types

来源: <https://bigd.big.ac.cn/ewas/datahub/probe/cg16867657>

cg16867657

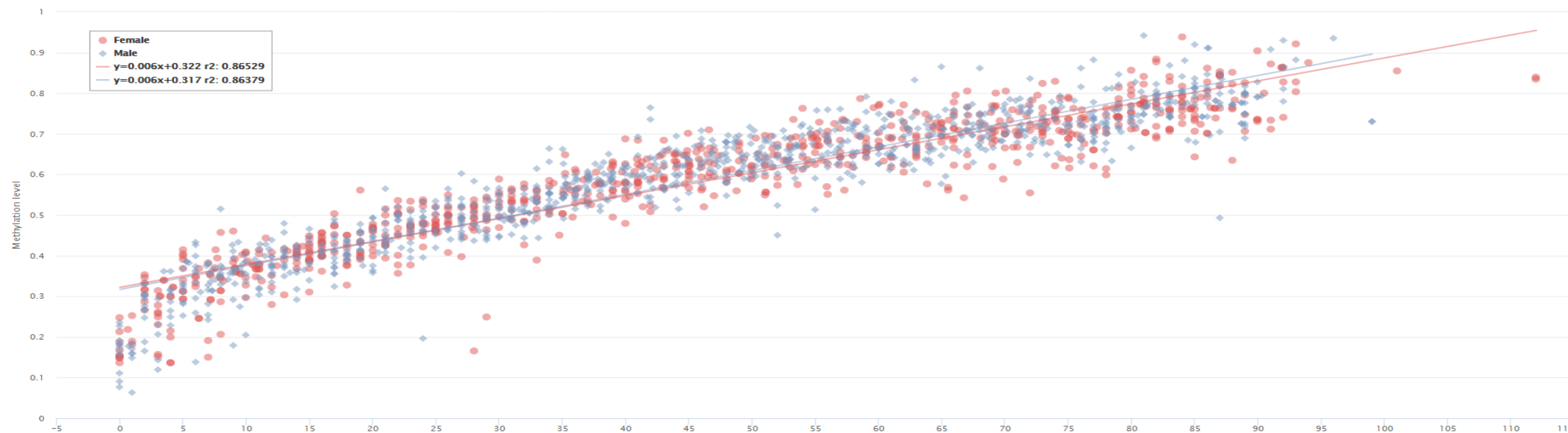
Basic Tissue Sex **Age** Ancestry Category BMI Cancer Other diseases Public EWAS



breast buccal epithelium CD14+ monocyte CD4+ T cell CD8+ T cell kidney leukocyte liver lymphocyte mucosa PBMC saliva thyroid brain - cerebellum brain - dorsolateral prefrontal cortex brain - entorhinal cortex brain - frontal cortex brain - occipital lobe **whole blood** brain - frontal lobe brain - superior temporal gyrus

DNA methylation changes with age

年龄模块



来源: <https://bigd.big.ac.cn/ewas/datahub/probe/cg16867657>

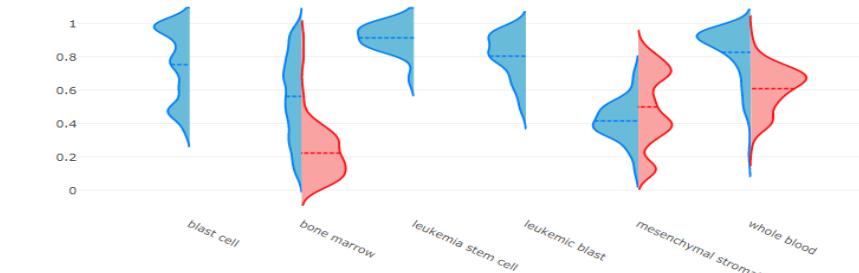
cg16867657

Basic Tissue Sex Age Ancestry Category BMI **Cancer** Other diseases Public EWAS

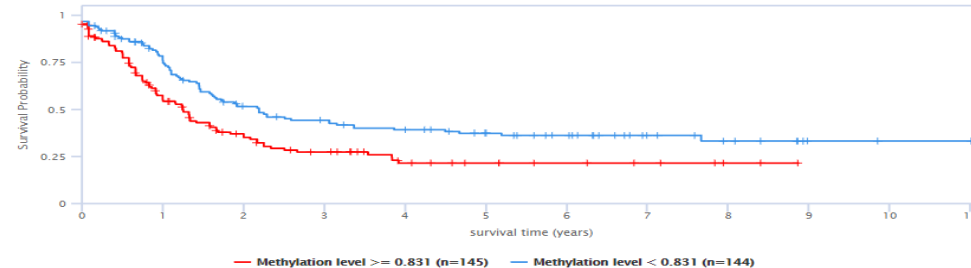


Cancer name: acute myeloid leukemia

acute myeloid leukemia

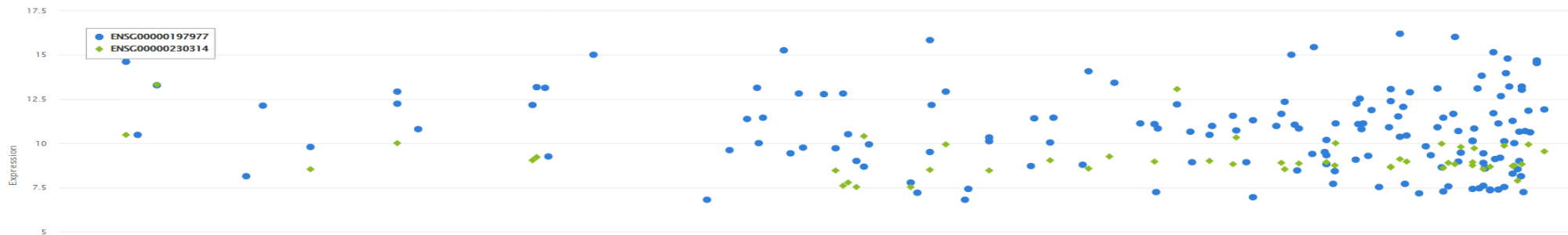


Survival Curves ($p=1.047e-3$)



肿瘤模块

DNA methylation changes with expression



来源: <https://bigd.big.ac.cn/ewas/datahub/probe/cg16867657>



**EWAS Toolkit** @ EWAS Open Platform
a web toolkit for epigenome-wide association study

Enrichment & Annotation

Network Visualization

Input File ⓘ

选择文件 未选择任何文件

Input Probe ID:

#example: vitamin B12 supplementation related DMP (PMID:29135286)
Background: ☒ 450K ☐ EPIC/850K ☐ Others

Go!

#Example File

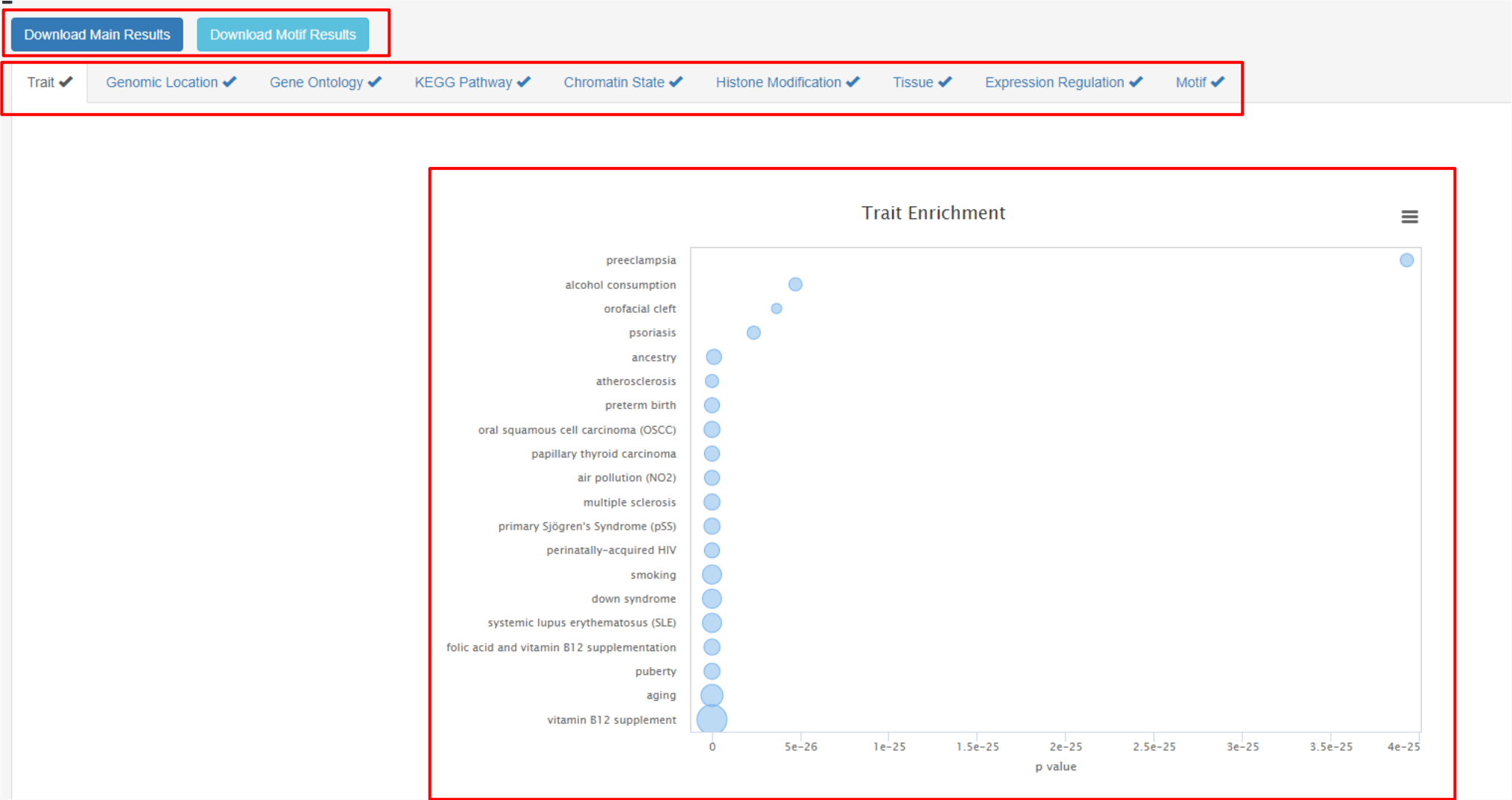
Clear Input

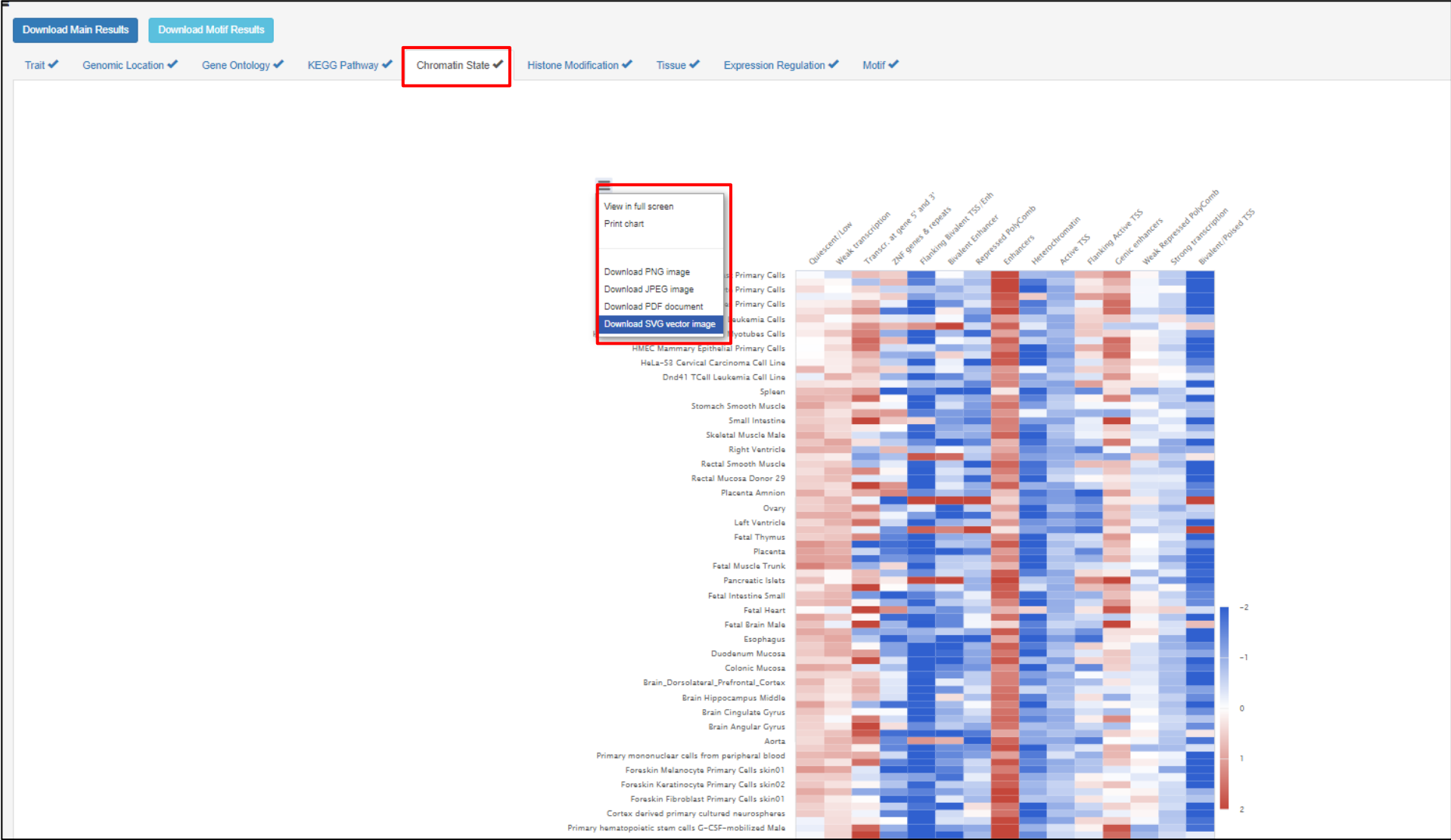
Trait From EWAS Atlas:

Go!

Input Job ID:

Go!

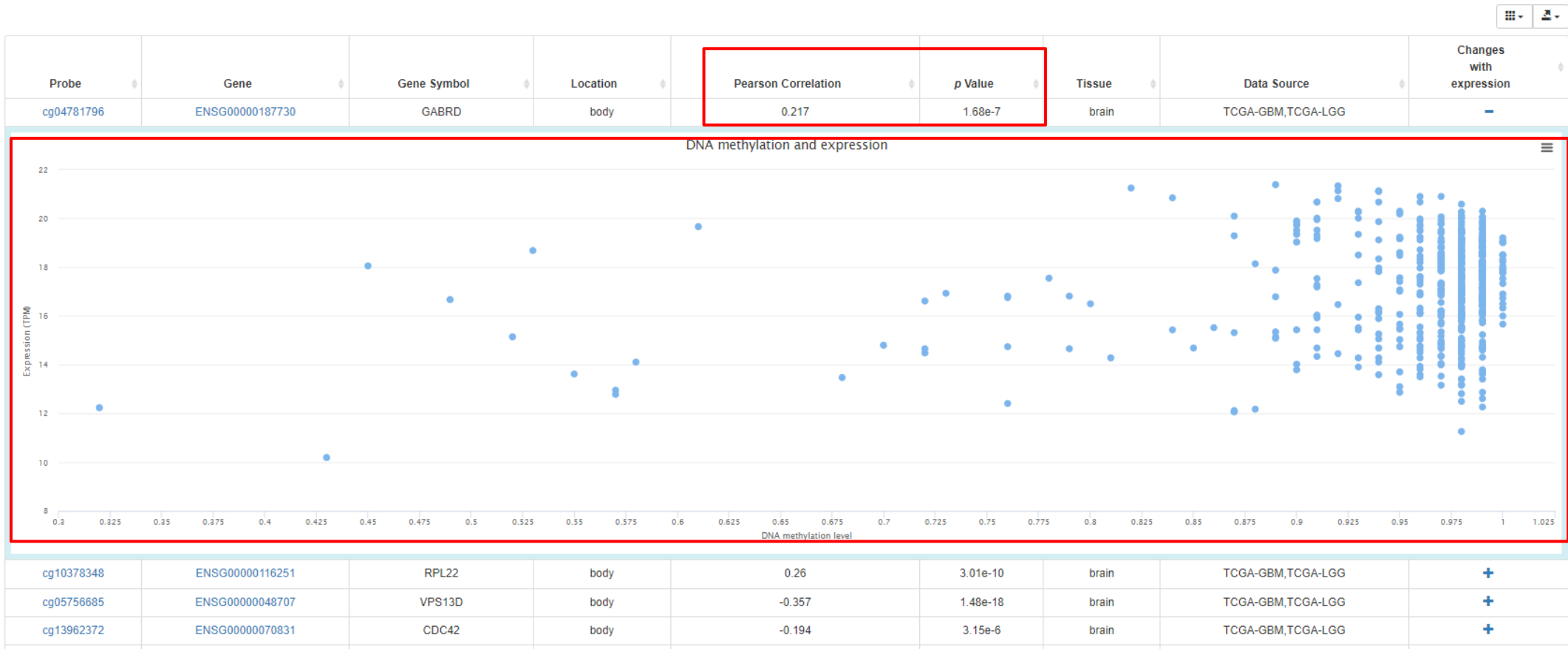




Download Main Results

Download Motif Results

Trait ☒ Genomic Location ☒ Gene Ontology ☒ KEGG Pathway ☒ Chromatin State ☒ Histone Modification ☒ Tissue ☒ Expression Regulation ☒ Motif ☒



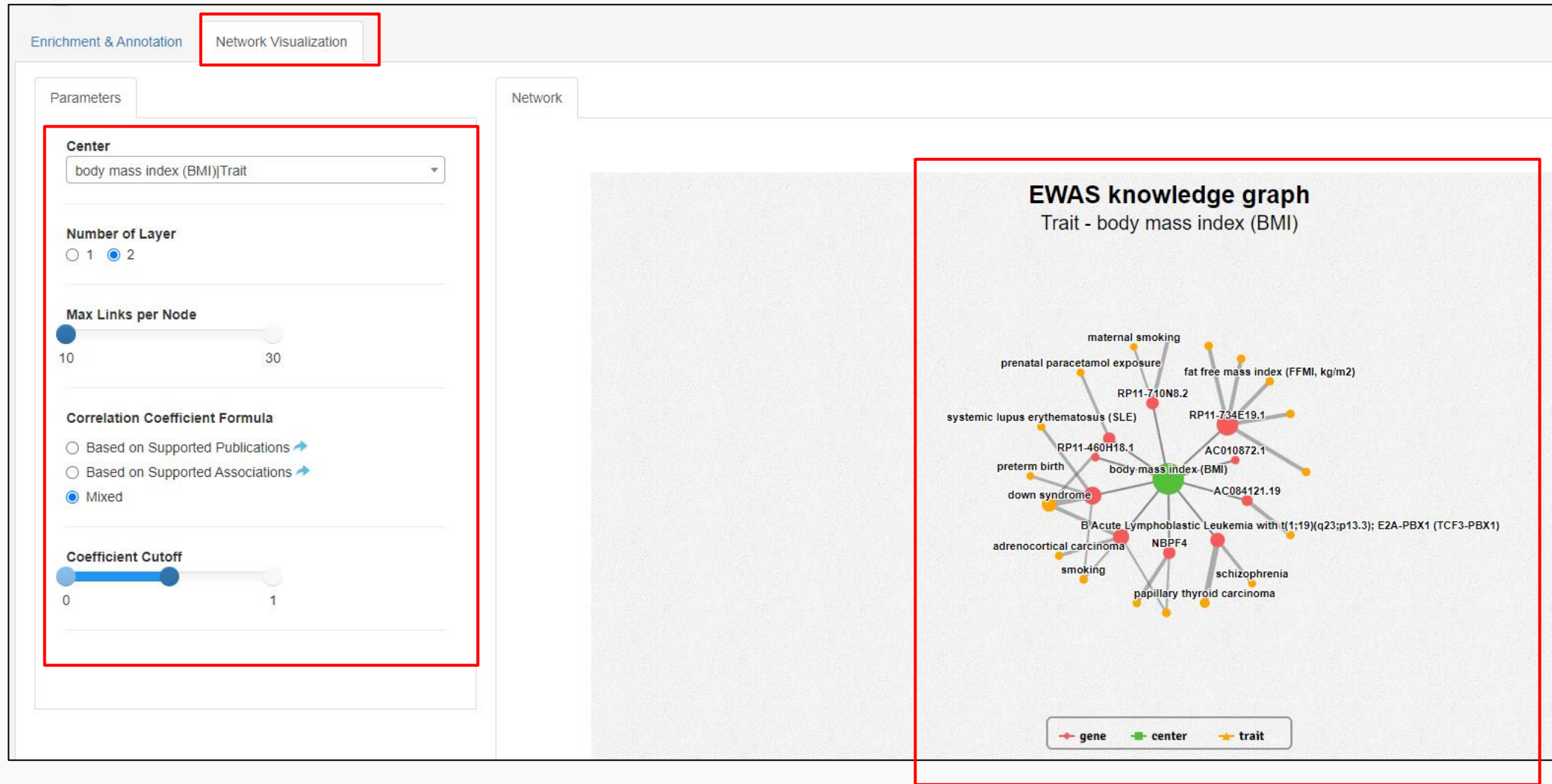
Download Main ResultsDownload Motif Results

Trait ✓Genomic Location ✓Gene Ontology ✓KEGG Pathway ✓Chromatin State ✓Histone Modification ✓Tissue ✓Expression Regulation ✓Motif ✓

Name	Motif	p value	Targets Sequences with Motif	Background Sequences with Motif
EHF(ETS)		1e-9	0.226	
NFY(CCAAT)		1e-7	0.141	
ELF5(ETS)		1e-6	0.134	
ELF3(ETS)		1e-6	0.127	0.070
Elf4(ETS)		1e-6	0.166	0.101
RUNX2(Runt)		1e-5	0.124	0.073
PU.1(ETS)		1e-4	0.083	0.043
EBF2(EBF)		1e-4	0.138	0.085
SpiB(ETS)		1e-4	0.051	0.022
NF1(CTF)		1e-4	0.070	0.035

Showing 1 to 10 of 48 rows10 rows per page

12345



Contact

使用甲基化资源平台时，您有任何的想法、建议、意见或疑问，都随时欢迎与我们邮件联系！

MethBank

methbank@big.ac.cn

scMethBank

zongwenting2018m@big.ac.cn
lirj@big.ac.cn

EWAS Open Platform

ewas-user@big.ac.cn



生物数据分析工程师

岗位要求

- (1) 诚信踏实、责任心强，具有较好的沟通能力，富有团队合作精神。
- (2) 获得或即将获得生物信息学、基因组学等相关专业硕士及以上学历。
- (3) 掌握生信数据分析流程操作，熟悉Linux操作系统，有Linux系统脚本编程能力，精通Python、Perl、R、C/C++、Java等至少一种编程语言。
- (4) 优先考虑具有算法开发经验、表观组学数据分析经验的申请者。

IT开发工程师

岗位要求

- (1) 诚信踏实、责任心强，具有较好的沟通能力，富有团队合作精神。
- (2) 获得或即将获得计算机、软件工程等相关专业硕士及以上学历，特别优秀者可适当放宽。
- (3) 熟练掌握Java语言开发，至少精通SpringCloud、SpringBoot、VUE等一种微服务开发框架。
- (4) 具有主观能动性，至少具备MySQL数据库开发经验，能独立开展数据库开发，解决开发中的技术问题，以及后期系统维护工作。



DNA甲基化资源平台招聘

投递简历邮箱

lirj@big.ac.cn

招聘电话

010-84097638

微 信



鱼加猫

中国大陆





Thanks

欢迎您使用甲基化资源平台，并提出宝贵意见和建议！