



# 转录组学的分析与临床应用



# 学习框架

## 第一课



1. 了解基础知识框架

2. 收集相关资料，自主学习

## 第二课



3. 练习编程、绘图

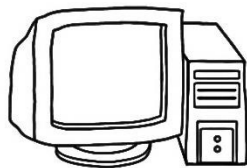
4. 将分析技能应用到研究中

5. 编写、发表生信算法、文章

生信自主学习的主要方式：读说明文档，读代码

# 计算机与编程

基本编程技能： R, Python ( 或 Perl), Linux



运算功能

存储功能

01101110101101010101001

01101110011001011011011

编译

编码

汇编语言： MOV  
C： int height=1 ;  
C++： class

dwadawxzcvas  
xwadaw+)dwad  
wwad&\*%(^\*(^

编译

编码

R, Python, Perl

Hello world



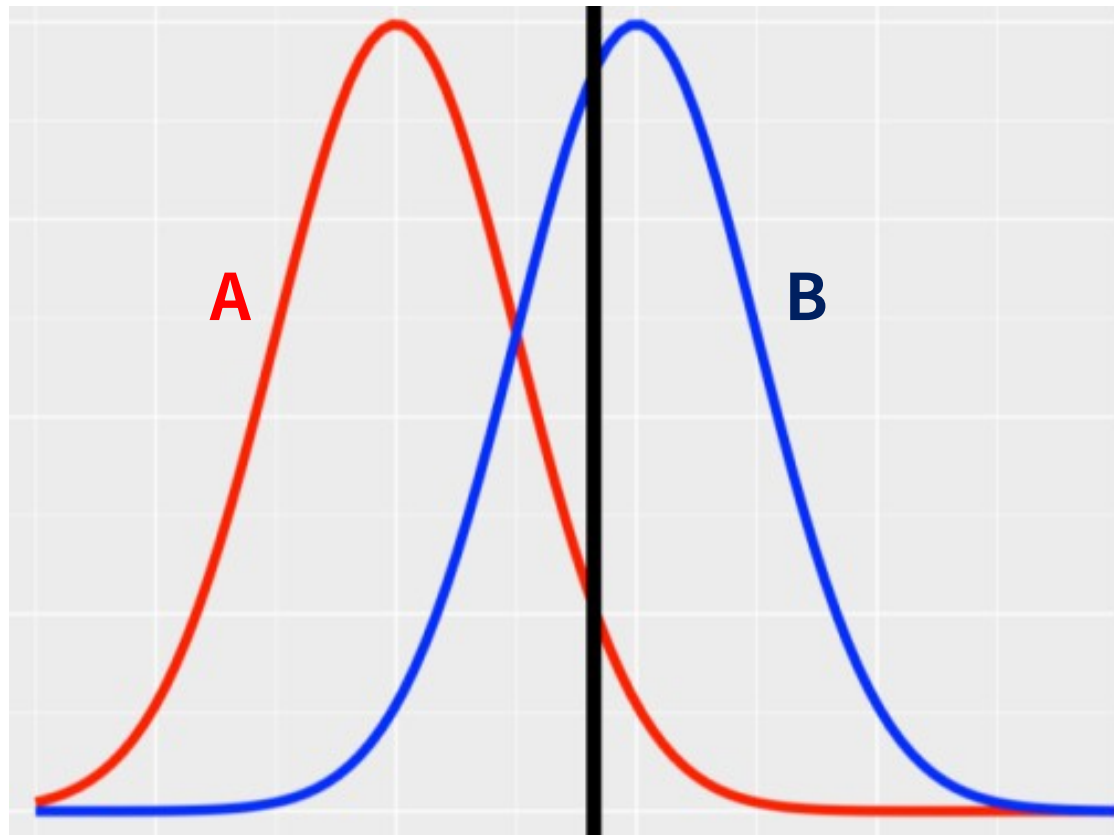
# 统计检验

到底来自 A 还是 B？

P 值为什么要矫正？

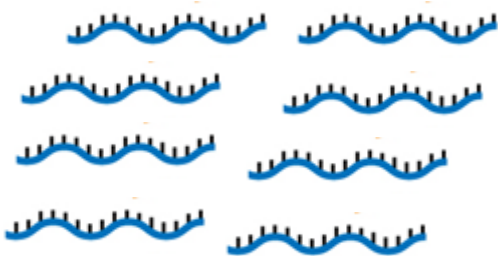
概率密度函数

统计量

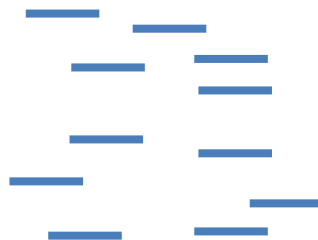


# 转录组测序

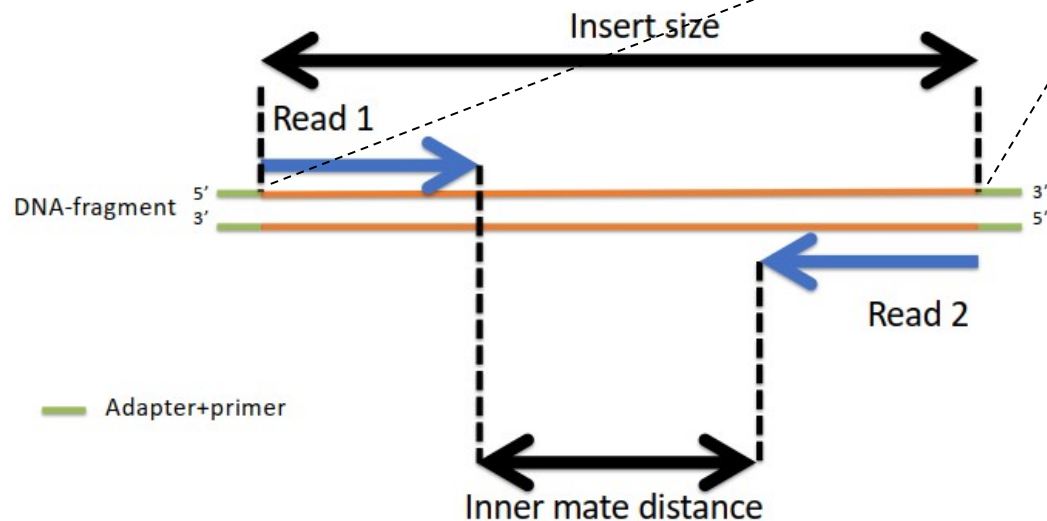
RNA



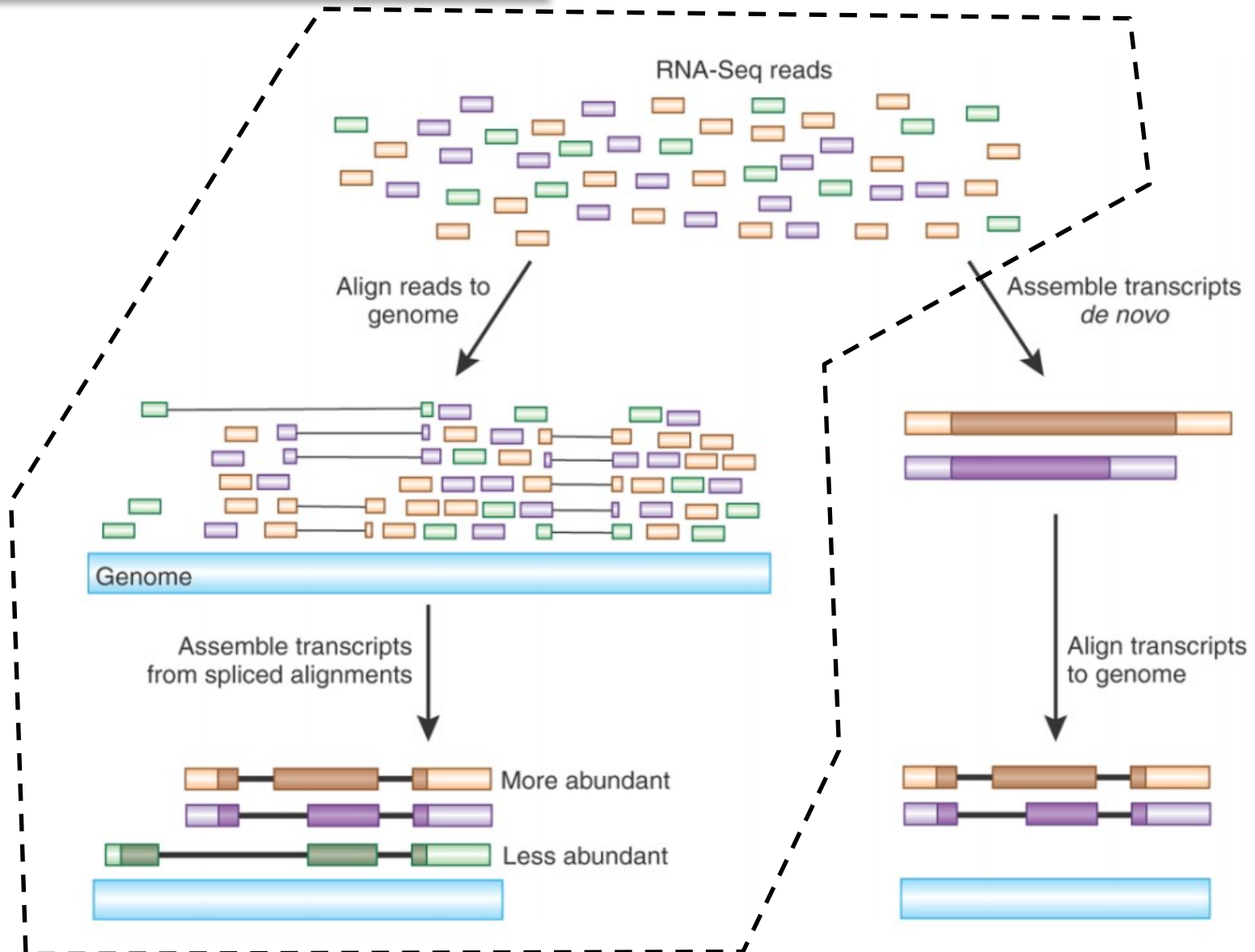
RNA 片段



DNA 片段



# 转录组测序



# 测序数据分析



www.shipin.com

185-25100513999421997179

FASTQ

FASTA

GTF / GFF

SAM / BAM

BED

WIG / BIGWIG / BEDGRAPH

MATRIX (EXCEL)

TXT

BWA / BOWTIE2 / HISAT2 / TOPHAT2

SRATOOLS

FASTQC

ComBat

MACS2

BEDTOOLS

FeatureCounts

SAMTOOLS

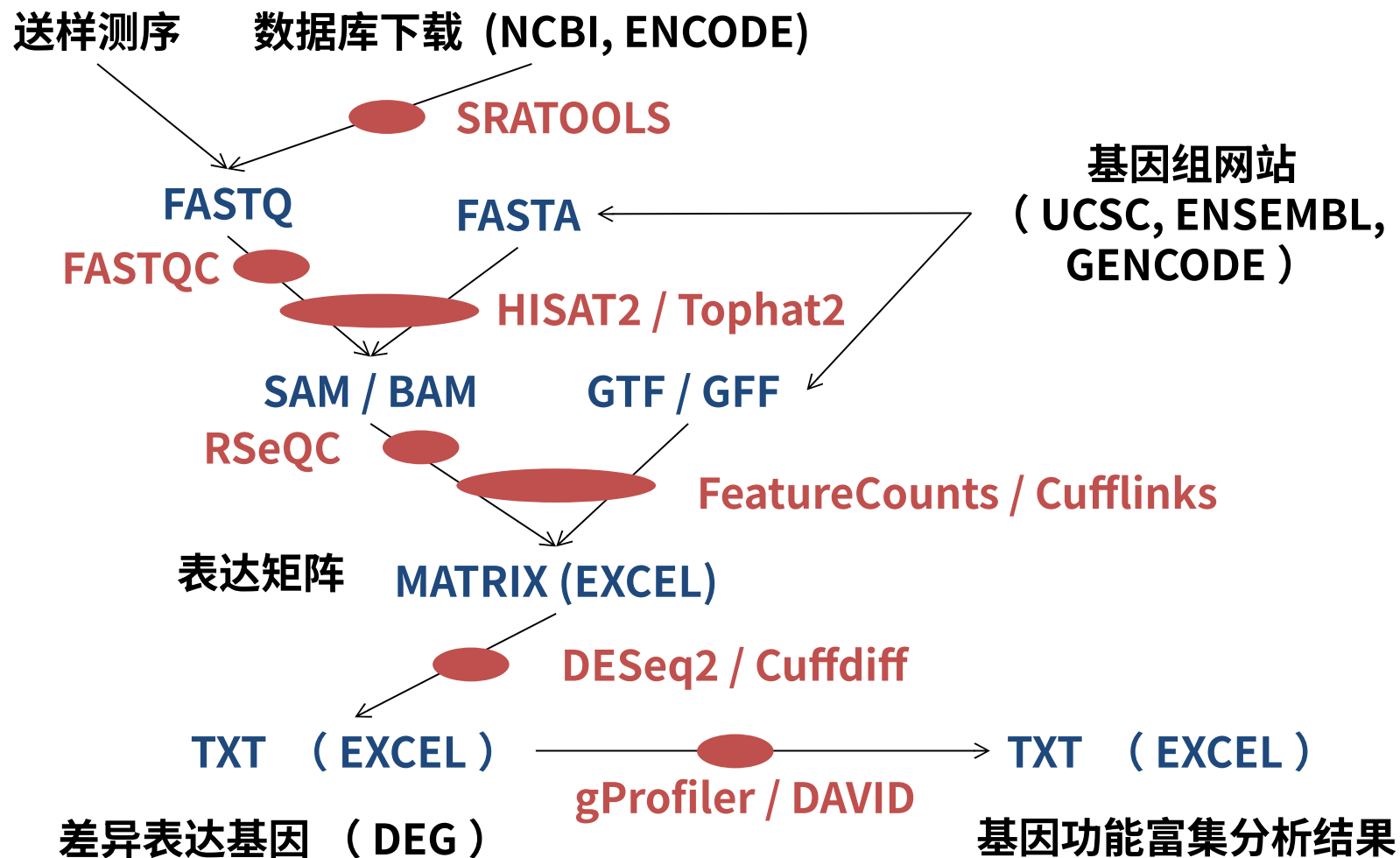
SALMON

CUFFDIFF / DESeq2

GSEA / gProfiler / DAVID / EnrichR

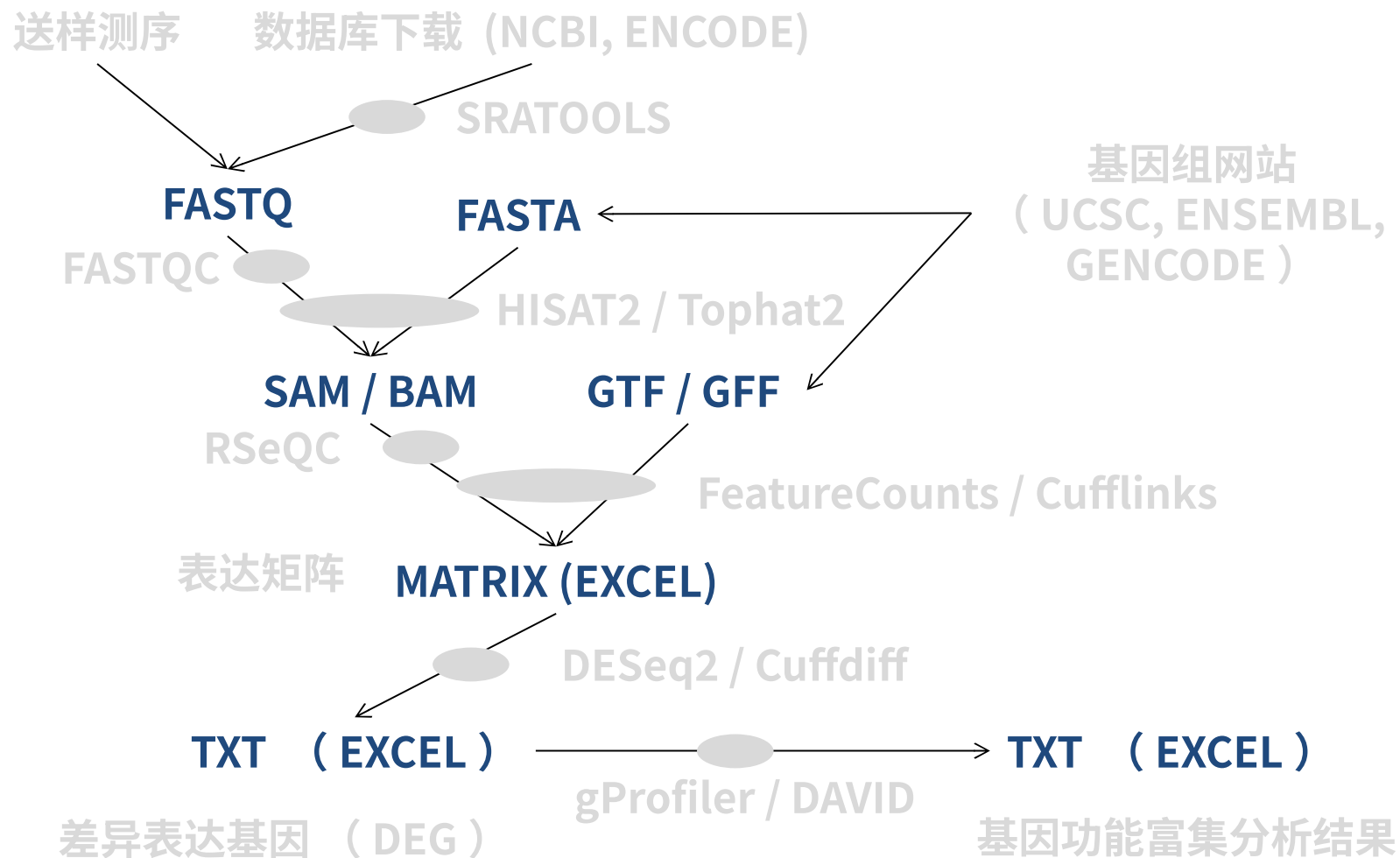
# 文件与工具

流程多变，不唯一!!!

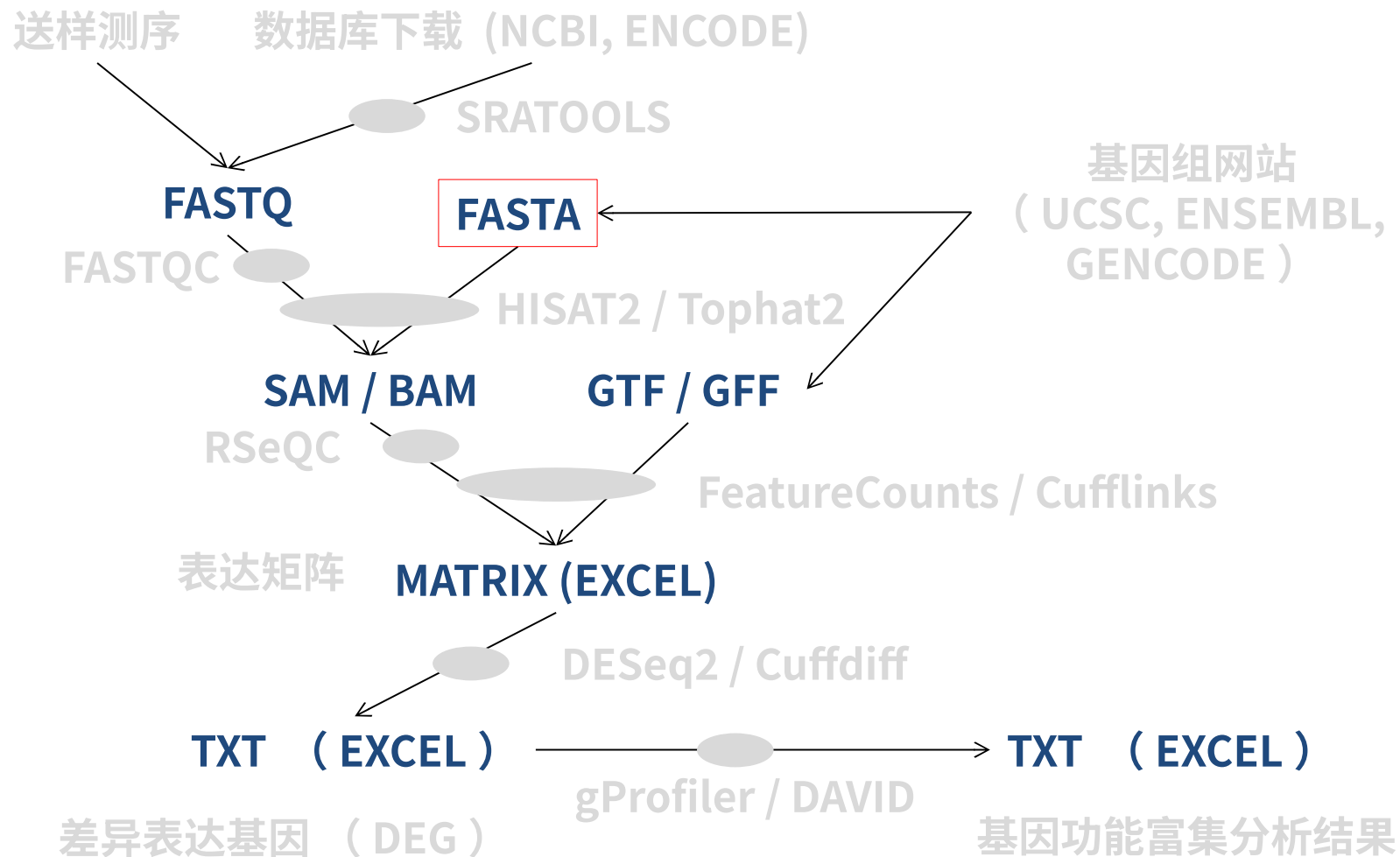




# 文件



# FASTA

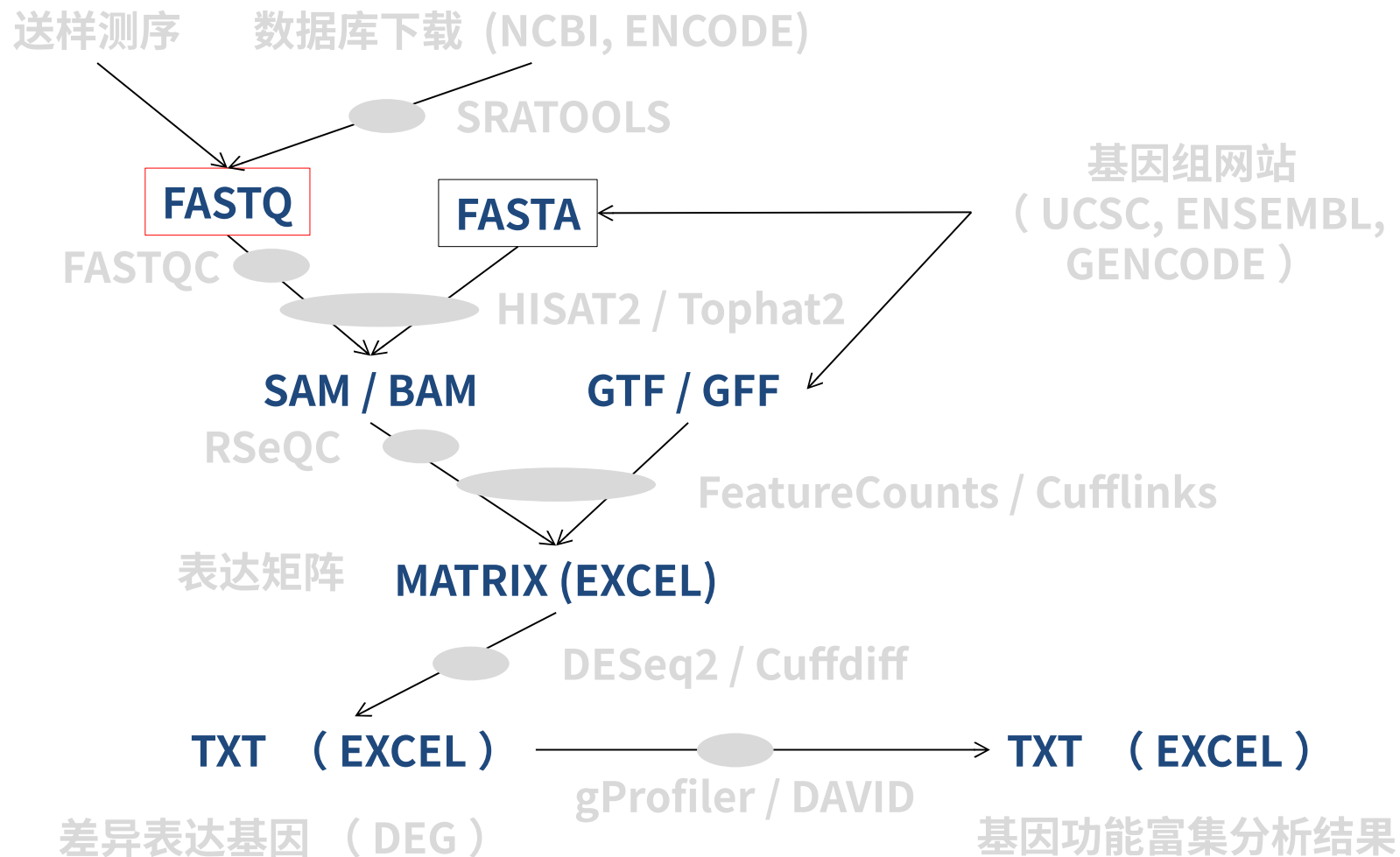


# FASTA

注意染色体（或转录本）名字， 注意碱基类型大小写含义， 人类基因组碱基数量 ~3GB  
一定要留意基因组版本号（比如人类的 hg19 ）

```
>chrM
GATCACAGGTCTATCACCTATTAACCACTCACGGGAGCTCTCCATGCAT
TTGGTATTTTCGTCTGGGGGGTGTGCACGCGATAGCATTGCGAGACGCTG
...
CTATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACCTACTA
AAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATAACAATTGAAT
GTCTGCACAGCCGCTTTCCACACAGACATCATAACAAAAAATTTCCACCA
>chr1
AACCCCCCTCCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGC
CAAACCCCAAAAACAAAGAACCCTAACACCAGCCTAACAGATTTCAAAT
TTTATCTTTAGGCGGTATGCACTTTTAACAGTCACCCCCCACTAACACA
TTATTTTCCCCTCCCCTCCACTCCCACTACTACTAATCTCATCAATACAACCCC
GCCCATCCTACCCAGCACACACACACCGCTGCTAACCCCATACCCCGAAC
...
```

# FASTQ



# FASTQ

```
@SRR8179797.116 116 length=151
GGCTAGTG ... GCCGTCT
+SRR8179797.116 116 length=151
AAAA.FFFF ... F.A<7)..FF7
```

每四行记录了一个读段

文件较大，一般会压缩后存储（.gz, .sra）

第三行可以只保留“+”

碱基测序质量打分（ASCII 转码）：

$$Q = S - N \text{ (N 一般是 33)}$$

$$P = 10^{(-Q/10)}$$

ASCII可显示字符

二进制	十进制	十六进制	图形	二进制	十进制	十六进制	图形	二进制	十进制	十六进制	图形
0010 0000	32	20	(空格) ( )	0100 0000	64	40	@	0110 0000	96	60	`
0010 0001	33	21	!	0100 0001	65	41	A	0110 0001	97	61	a
0010 0010	34	22	"	0100 0010	66	42	B	0110 0010	98	62	b
0010 0011	35	23	#	0100 0011	67	43	C	0110 0011	99	63	c
0010 0100	36	24	\$	0100 0100	68	44	D	0110 0100	100	64	d
0010 0101	37	25	%	0100 0101	69	45	E	0110 0101	101	65	e
0010 0110	38	26	&	0100 0110	70	46	F	0110 0110	102	66	f
0010 0111	39	27	'	0100 0111	71	47	G	0110 0111	103	67	g
0010 1000	40	28	(	0100 1000	72	48	H	0110 1000	104	68	h
0010 1001	41	29	)	0100 1001	73	49	I	0110 1001	105	69	i
0010 1010	42	2A	*	0100 1010	74	4A	J	0110 1010	106	6A	j
0010 1011	43	2B	+	0100 1011	75	4B	K	0110 1011	107	6B	k
0010 1100	44	2C	,	0100 1100	76	4C	L	0110 1100	108	6C	l
0010 1101	45	2D	-	0100 1101	77	4D	M	0110 1101	109	6D	m
0010 1110	46	2E	.	0100 1110	78	4E	N	0110 1110	110	6E	n
0010 1111	47	2F	/	0100 1111	79	4F	O	0110 1111	111	6F	o
0011 0000	48	30	0	0101 0000	80	50	P	0111 0000	112	70	p
0011 0001	49	31	1	0101 0001	81	51	Q	0111 0001	113	71	q
0011 0010	50	32	2	0101 0010	82	52	R	0111 0010	114	72	r

## 质量控制 (FASTQC)

### FastQC Report

#### Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content
- Kmer Content

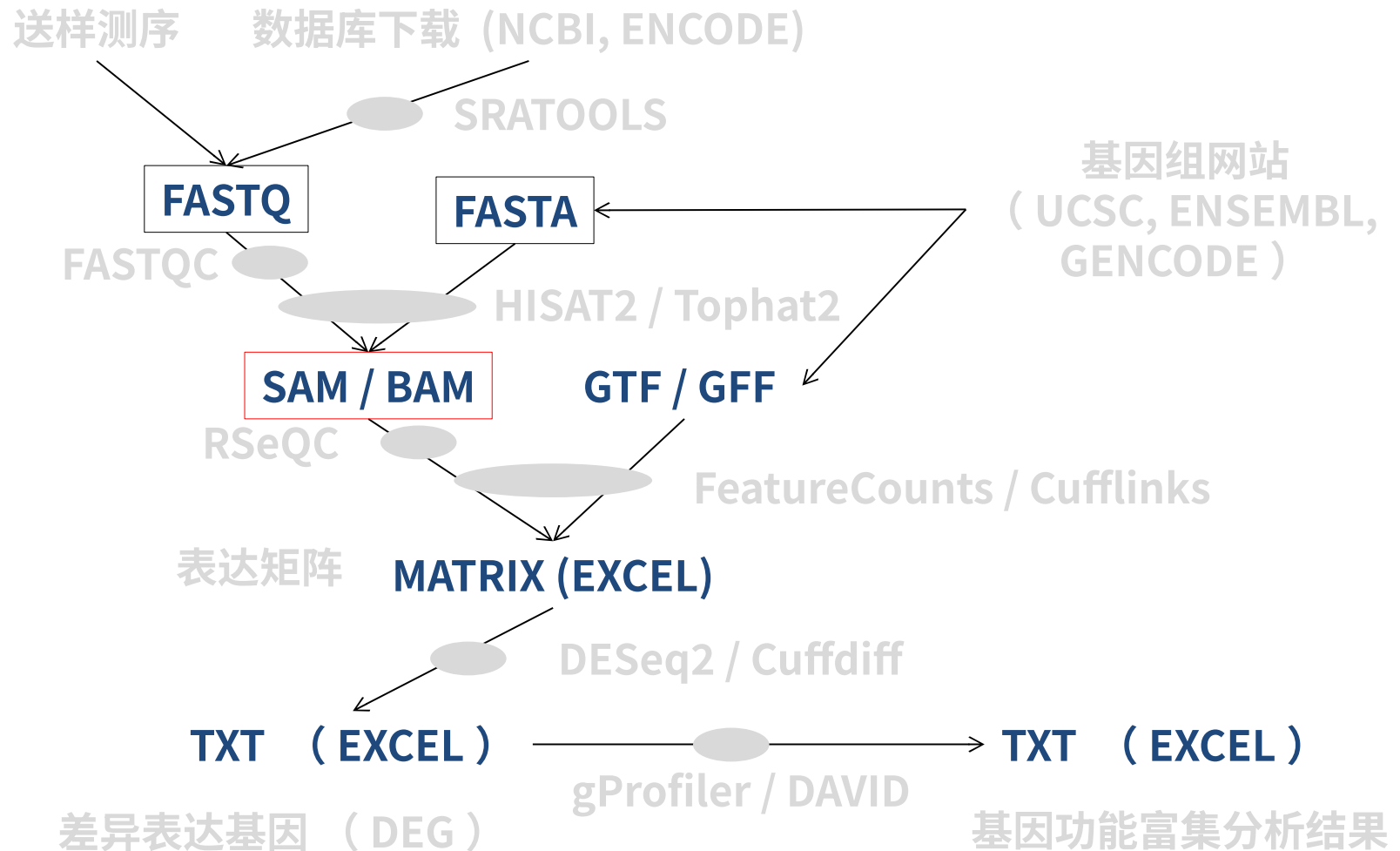
#### Basic Statistics

Measure	Value
Filename	L1.clean.fq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	5487063
Sequences flagged as poor quality	0
Sequence length	150
%GC	65

#### Per base sequence quality



# SAM / BAM



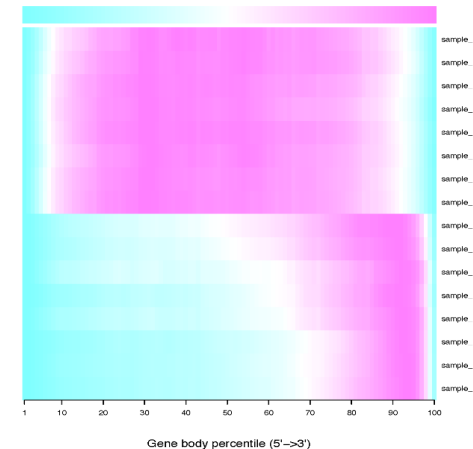
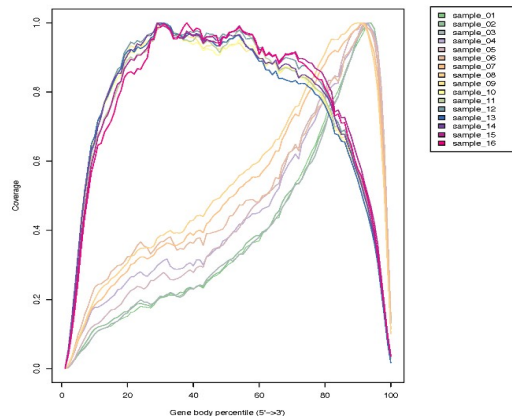
# SAM / BAM

SAM 文件记录了每一个读段 “如何” 比对到 “何处”，也保存了碱基类型和质量信息

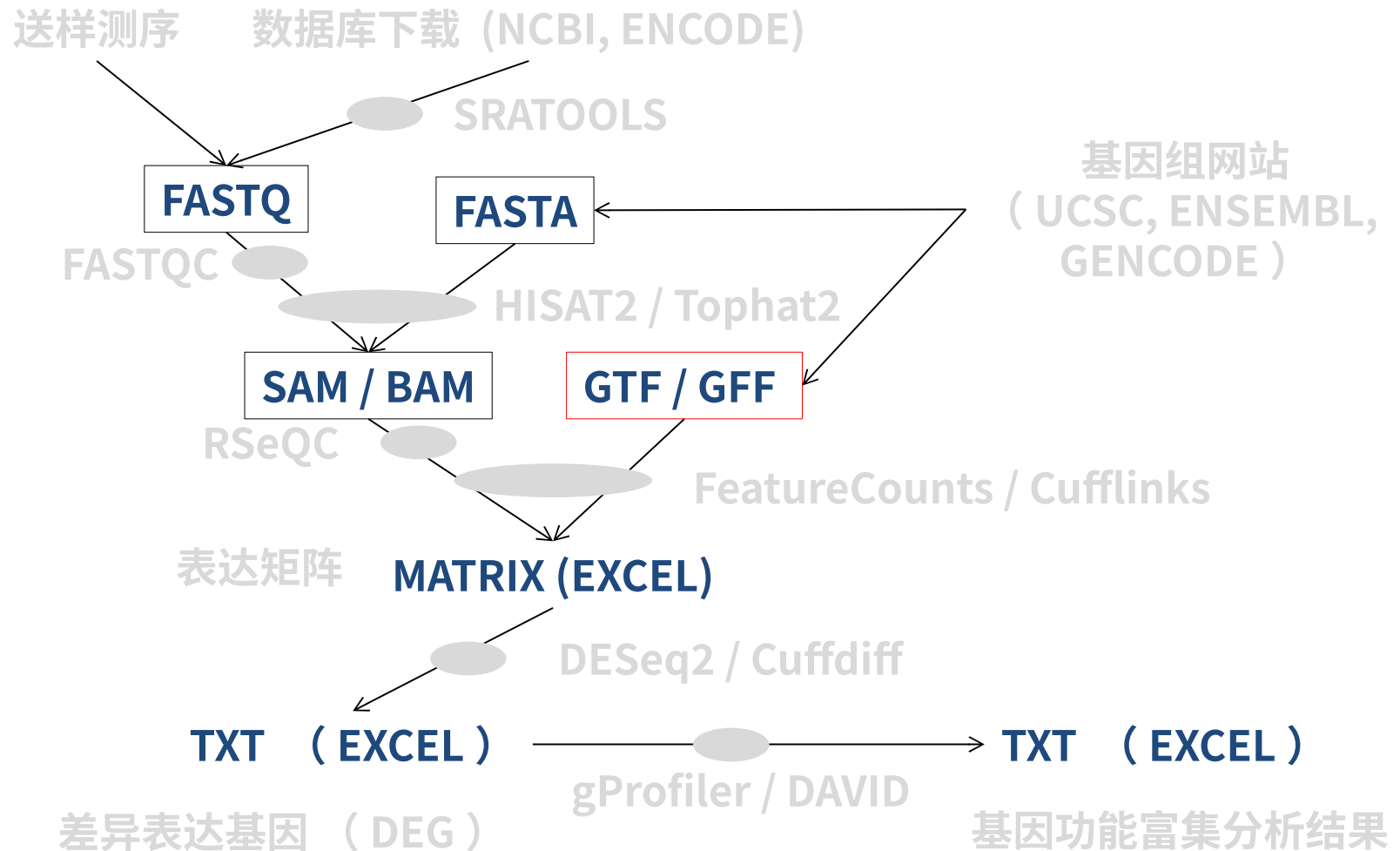
```
SRR8179797.52782089 99 chrM 44 60 143M8S = 44 -159
CATGCATTTGGTATTTTCGTCTGGGGGGTGTGCACGCGATAGCATTGCGAGACGCTGGAGCCGG
AGCACCTTATGTCGCAGTATCTGTCTTTGATTCTGCCTCATCCTATTATTATAGCACCTACGTTCA
ATATTACAGGCAGATCGGA
AAAAAAFF.FFFFFFFF<FFFFFFFF7FFAFFFFFFFFFAFFFFFFFAFFFFFFFAAFFFFFFFFF<FFFFAAFFFFFFF<)FFFFFF7A
<FFF<FF.FFF7FFFFFFFFF7<FFFFFFFF.FFF.AFF<F<FAF)FFFFFF<F<<AFFAFF7A7FF7<F.FFFFFFF<7
AS:i:-20 XN:i:0 XM:i:3 XO:i:0 XG:i:0 NM:i:3 MD:Z:70C35T11C24 YS:i:-18
YT:Z:CP NH:i:1
```

SAM 文件较大，一般使用 SAMTOOLS 将 SAM 格式转化为二进制的 BAM 格式

质量控制 (QC)



# GTF / GFF



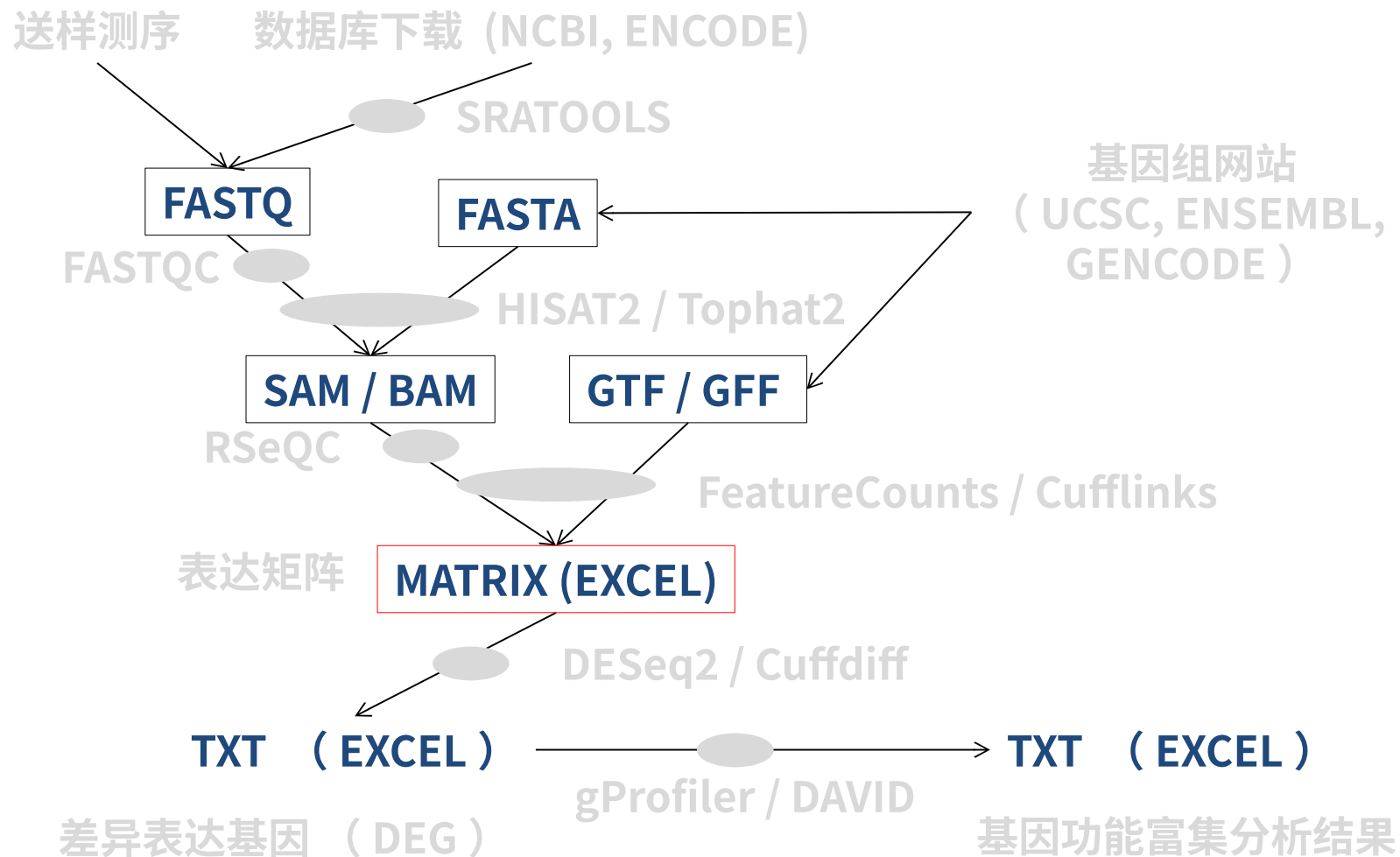


# GTF / GFF

GTF 文件记录了每一个基因组元件的 “性质” 和 “位置”

```
chr1 protein_coding CDS 69091 70005 . +  
0 gene_id "ENSG00000186092"; transcript_id  
"ENST00000335137"; exon_number "1"; gene_name  
"OR4F5"; gene_source "ensembl_havana";  
gene_biotype "protein_coding"; transcript_name  
"OR4F5-001"; transcript_source "ensembl_havana";  
tag "CCDS"; ccds_id "CCDS30547"; protein_id  
"ENSP00000334393";
```

# MATRIX



# MATRIX 表达

读段个数矩阵 ( read count matrix )

	Sample1	Sample2	Sample3	...
Gene1	5	9	1	
Gene2	353	247	271	
Gene3	3777	2364	25	
...				

↓  
归一化

FPKM / RPKM / RPM / TPM

不同样本同基因可比

↓  
对数变换

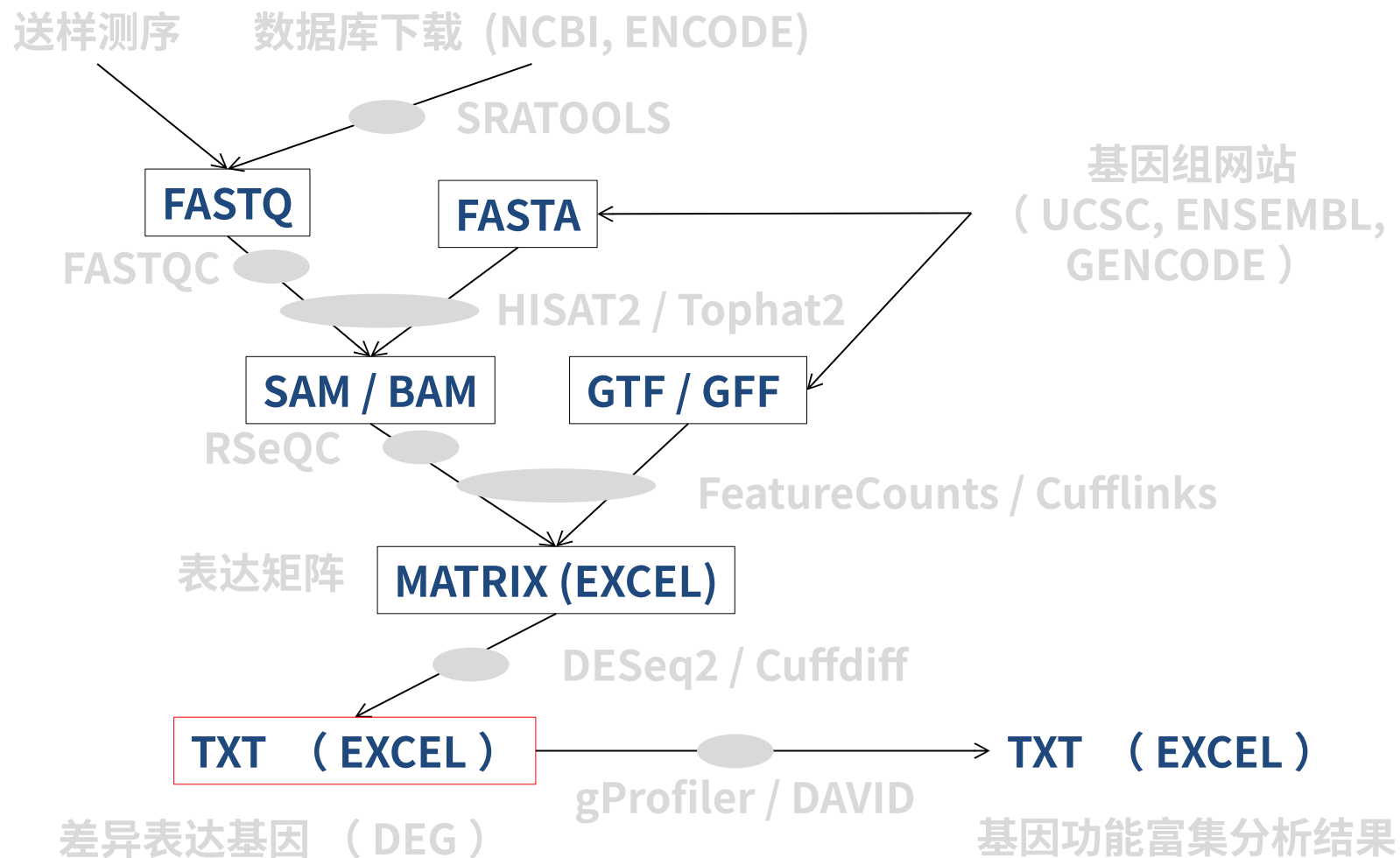
类正态，易做检验

↓  
标准化

每一行减去 均值 除以 标准差

PCA / Heatmap

# TXT 差异



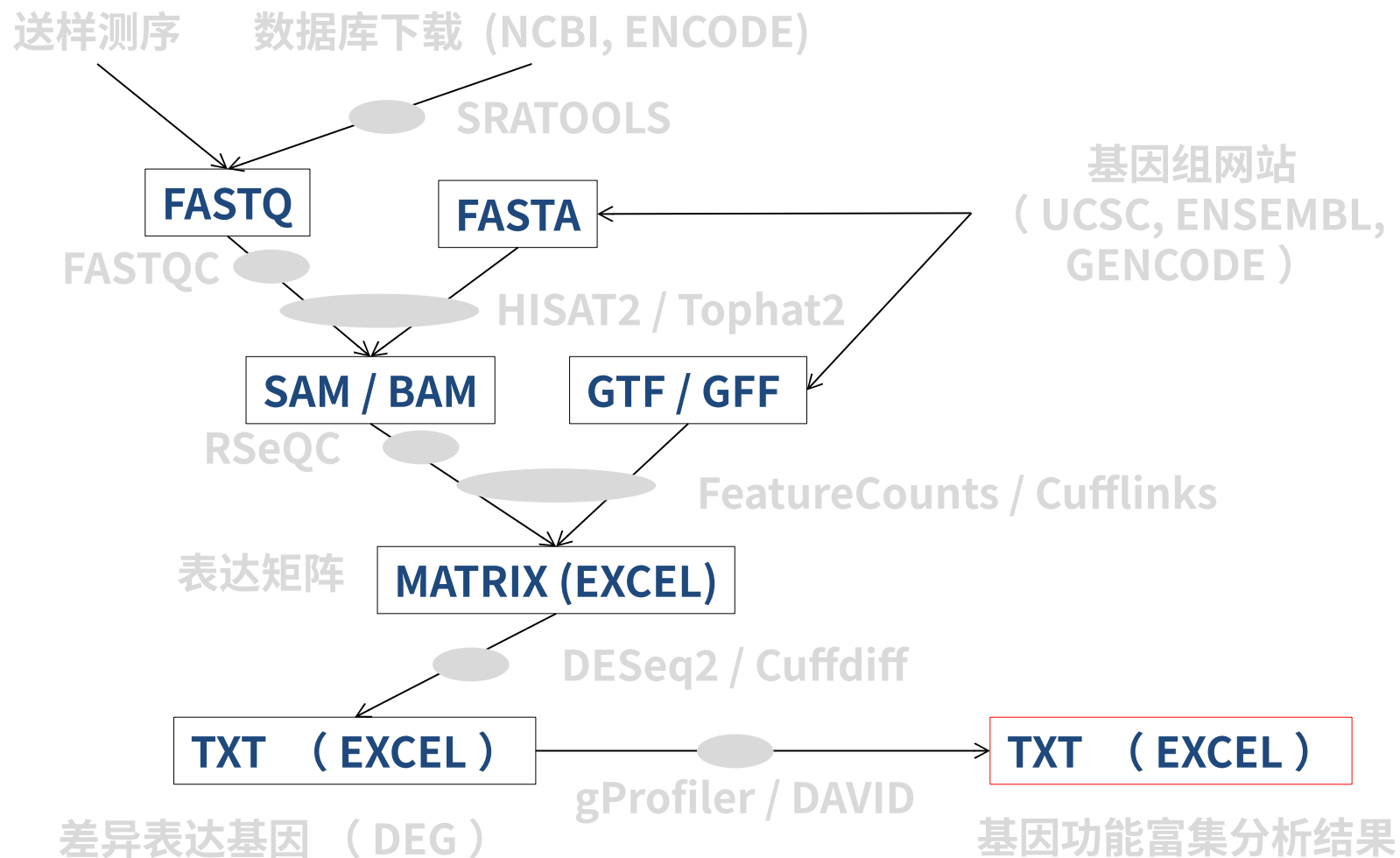
# TXT 差异

一般情况下， Fold Change 和 Pvalue (Padj) 都要考虑

## DESeq2 结果

	Case 1	Case 2	Ctrl 1	Ctrl 2	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Cx3cr1	13	7	363	691	309.4573221	-6.253625195	0.51093498	-12.2396	1.91E-34	2.43E-30
5830411N06Rik	1082	821	8	20	391.3886748	5.589308195	0.472684251	11.82461	2.91E-32	1.86E-28
Igkc	50	20	3819	1850	2027.636391	-7.19943295	0.611847719	-11.7667	5.79E-32	2.46E-28
Cd163l1	1147	912	17	20	427.7605469	5.216160719	0.45457985	11.47469	1.77E-30	5.63E-27
Jchain	8	8	1442	651	753.2178054	-7.844916069	0.693993672	-11.304	1.25E-29	3.19E-26
Zeb2	18	7	255	750	278.0539922	-5.789824356	0.539203416	-10.7377	6.77E-27	1.44E-23
Gzma	33	30	915	715	555.5517287	-5.402705083	0.516218347	-10.4659	1.24E-25	2.25E-22
Gzmk	10	12	204	383	175.0640247	-5.243874941	0.529835716	-9.89717	4.28E-23	6.82E-20
S1pr5	25	15	327	367	226.5920639	-4.769446278	0.509896752	-9.35375	8.46E-21	1.20E-17

# TXT 富集



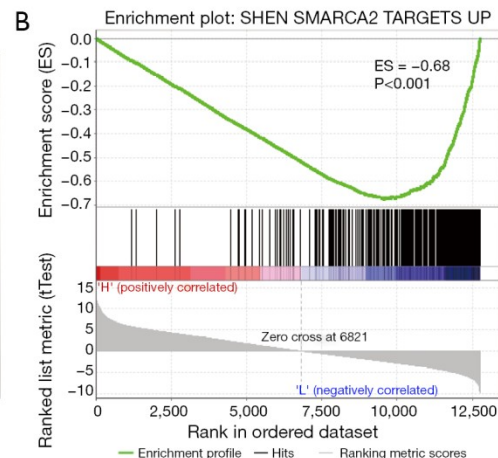
# TXT 富集

## gProfiler 结果

一般情况下，只考虑 **Padj**，也可以综合考虑**基因个数**

[illegible]

## 第二课内容



## 富集柱状图

# GSEA



# 临床应用

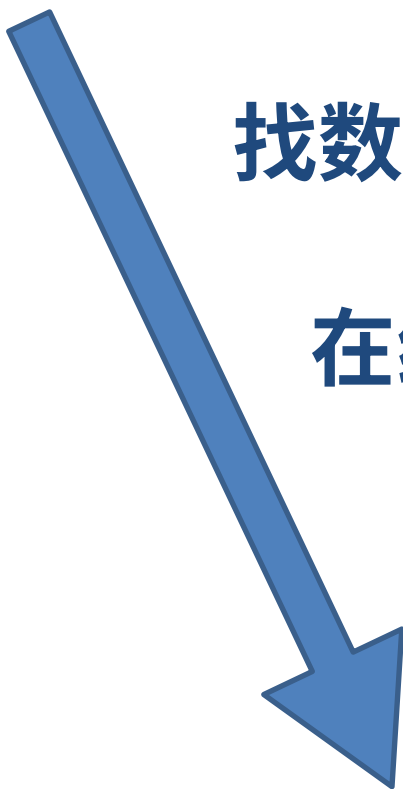
想法 ( idea )

找数据?

在线分析?

可视化?

产出 ( ppt , paper , funding )



# 临床应用 1

## 找数据

癌症 + 表达矩阵 + 临床信息

<http://www.cbioportal.org/datasets>



[Data Sets](#)

[Web API](#)

[R/MATLAB](#)

[Tutorials/Webinars](#)

[FAQ](#)

[News](#)

[Visualize Your Data](#)

[About](#)

[cBioPortal Installations](#)

Login

### Datasets

The table below lists the number of available samples per cancer study and data type. It also provides links to download the data for each study. For alternative ways of downloading, see the [Download Documentation](#).

<input type="text"/>		<input type="button" value="Columns"/>			
Name ^	Reference	All	Mutations	CNA	RNA-Seq
Acinar Cell Carcinoma of the Pancreas (JHU, J Pathol 2014)	<a href="#">Jial et al. J Pathol 2014</a>	23	23	0	0
Acral Melanoma (TGEN, Genome Res 2017)	<a href="#">Liang et al. Genome Res 2017</a>	38	38	38	36
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)	<a href="#">Andersson et al. Nat Genet 2015</a>	93	93	0	0
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)	<a href="#">Zhang et al. Nat Genet 2016</a>	73	73	0	0
Acute Myeloid Leukemia (OHSU, Nature 2018)	<a href="#">Tyner et al. Nature 2018</a>	672	622	0	451
Acute Myeloid Leukemia (TCGA, Firehose Legacy)		200	197	191	173
Acute Myeloid Leukemia (TCGA, NEJM 2013)	<a href="#">TCGA, NEJM 2013</a>	200	200	191	173
Acute Myeloid Leukemia (TCGA, PanCancer Atlas)	<a href="#">TCGA, Cell 2018</a>	200	200	191	173
Acute myeloid leukemia or myelodysplastic syndromes (WashU, 2016)	<a href="#">Welch et al. N Engl J Med. 2016</a>	136	136	0	0
Adenoid Cystic Carcinoma (FMI, Am J Surg Pathl. 2014)	<a href="#">Ross et al. Am J Surg Pathl 2014</a>	28	28	28	0
Adenoid Cystic Carcinoma (JHU, Cancer Prev Res 2016)	<a href="#">Rettig et al, Cancer Prev Res 2016</a>	25	25	0	0
Adenoid Cystic Carcinoma (MDA, Clin Cancer Res 2015)	<a href="#">Mitani et al. Clin Cancer Res 2015</a>	102	65	0	0
Adenoid Cystic Carcinoma (MGH, Nat Gen 2016)	<a href="#">Drier et al. Nature Genetics 2016</a>	10	10	0	0
Adenoid Cystic Carcinoma (MSKCC, Nat Genet 2013)	<a href="#">Ho et al. Nat Genet 2013</a>	60	60	60	0
Adenoid Cystic Carcinoma (Sanger/MDA, JCI 2013)	<a href="#">Stephens et al. JCI 2013</a>	24	24	0	0

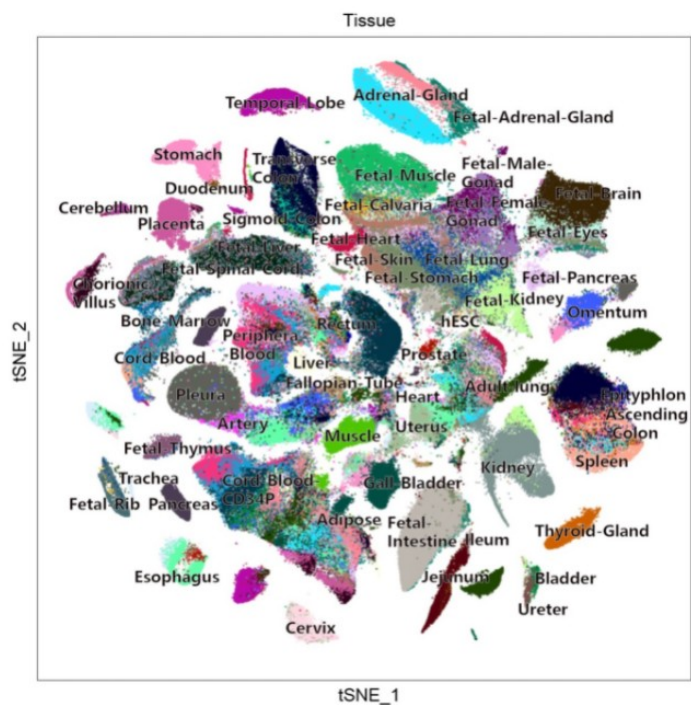
# 临床应用 2

## 找数据

### 单细胞转录组数据（HCL）

<https://db.cngb.org/HCL/index.html>

### Human Cell Landscape



### scHCL panel

Upload your own DGE

+ Add files...

### File Info

#### notifications

File size: <=200mb


File extension: txt/csv

# 临床应用 3

## 在线分析

关系网络绘制

<https://string-db.org/>

 Search Download Help My Data

Protein by name >

Protein by sequence >

Multiple proteins >

Multiple sequences >

Proteins with Values/Ranks >

Organisms >

Protein families ("COGs") >

Examples >

Random entry >

### SEARCH

Multiple Proteins by Names / Identifiers

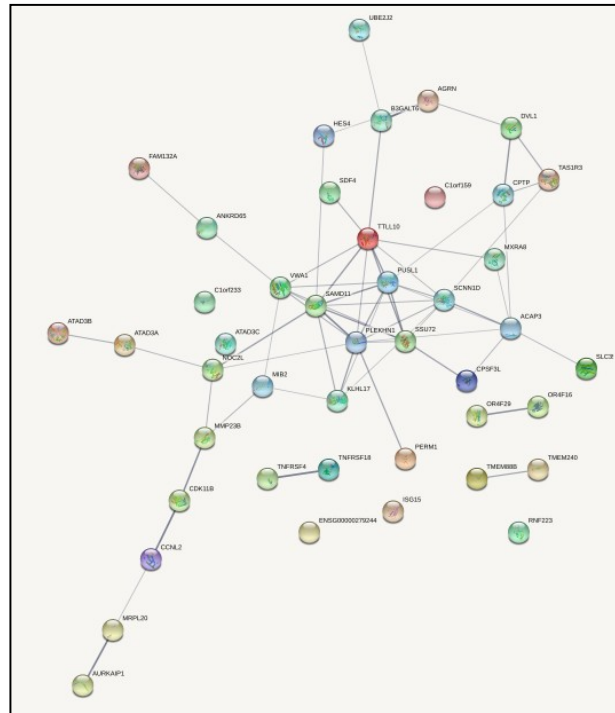
List Of Names: (one per line, examples: #1 #2 #3)

... or, upload a file:

Organism: auto-detect

Advanced Settings

SEARCH




CytoScape

# 临床应用 4

## 在线分析

<http://gepia.cancer-pku.cn/>

### TCGA 数据在线分析



# GEPIA

Gene Expression Profiling Interactive Analysis

Single Gene Analysis

Cancer Type Analysis

Multiple Gene Analysis

Enter gene name:

The indicators in search box are "symbol" or "alias (newest symbol)".

e.g. ERBB2/ENSG00000141736/2064

GoPIA!

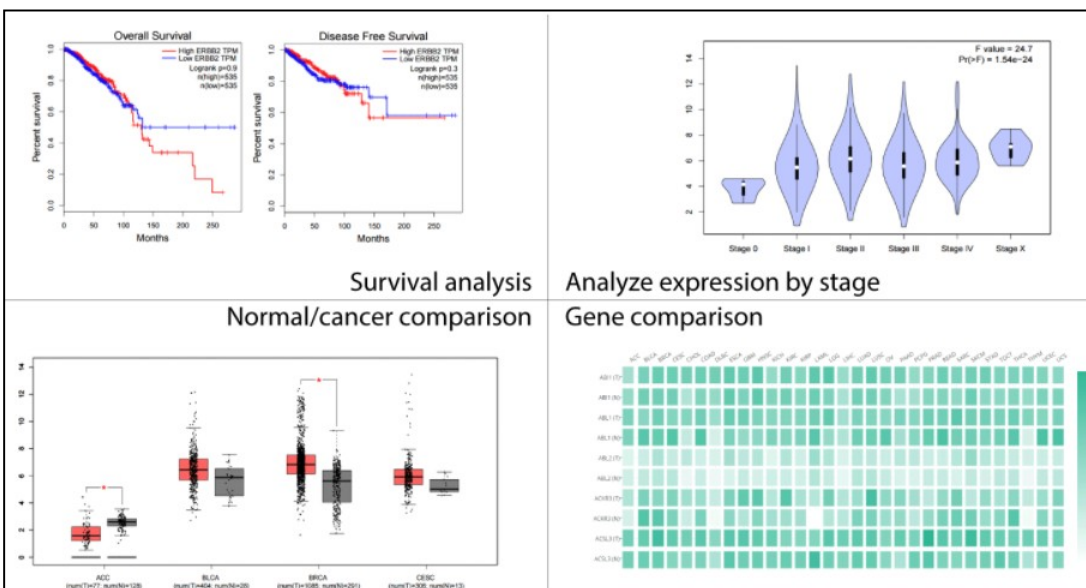
Profile

Boxplots

Stage Plots

Survival Analysis

Similar

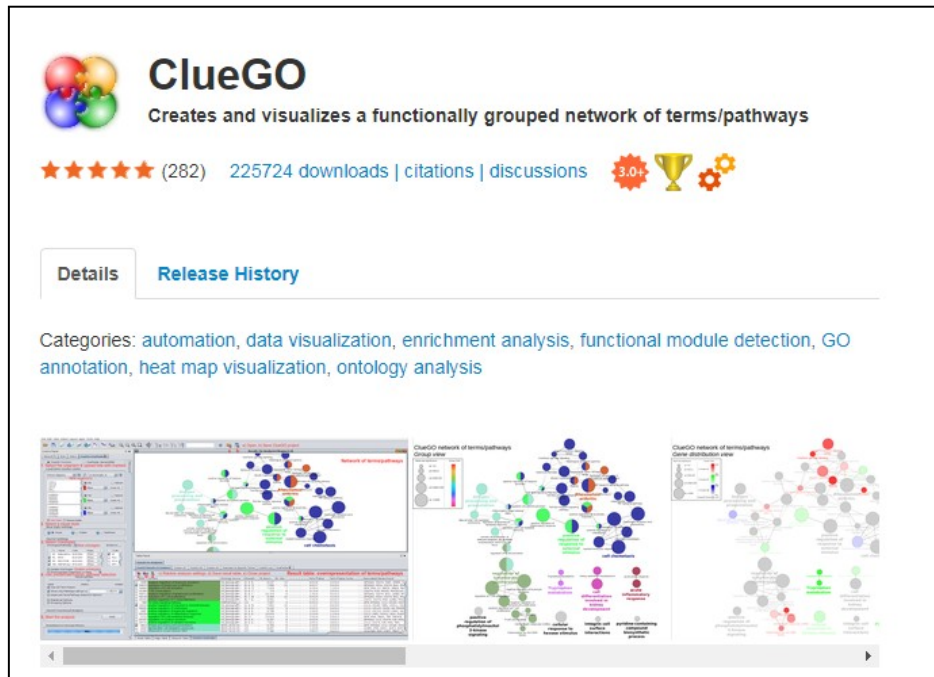


# 临床应用 5

## 可视化

富集 + 网络展示 ClueGO

<https://apps.cytoscape.org/apps/cluego>



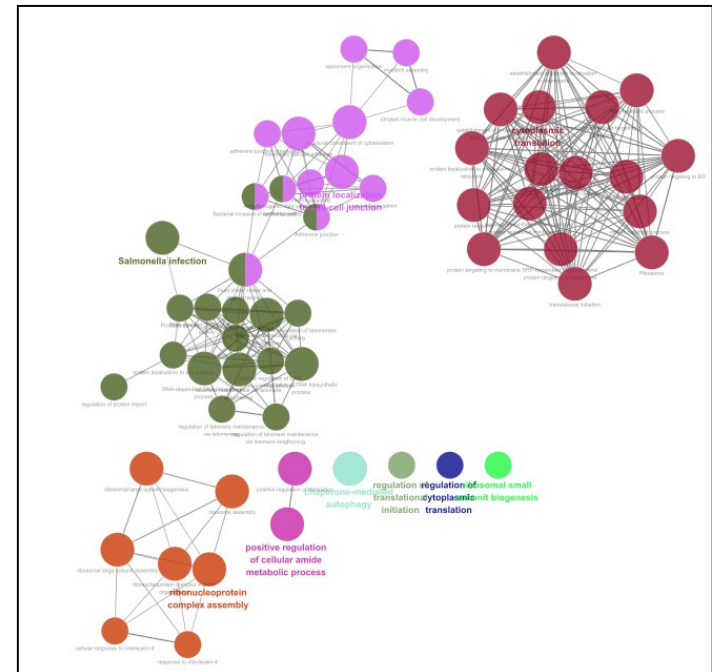
**ClueGO**  
Creates and visualizes a functionally grouped network of terms/pathways

★★★★★ (282) 225724 downloads | citations | discussions 3.0+

Details Release History

Categories: automation, data visualization, enrichment analysis, functional module detection, GO annotation, heat map visualization, ontology analysis

The screenshot displays the ClueGO web application interface. It features a header with the ClueGO logo and a brief description. Below this, there are statistics including a star rating, download count, and version number. A navigation bar includes 'Details' and 'Release History' tabs. The main content area lists various categories of the application's functionality. At the bottom, there are three small thumbnail images showing different network visualizations generated by the software.





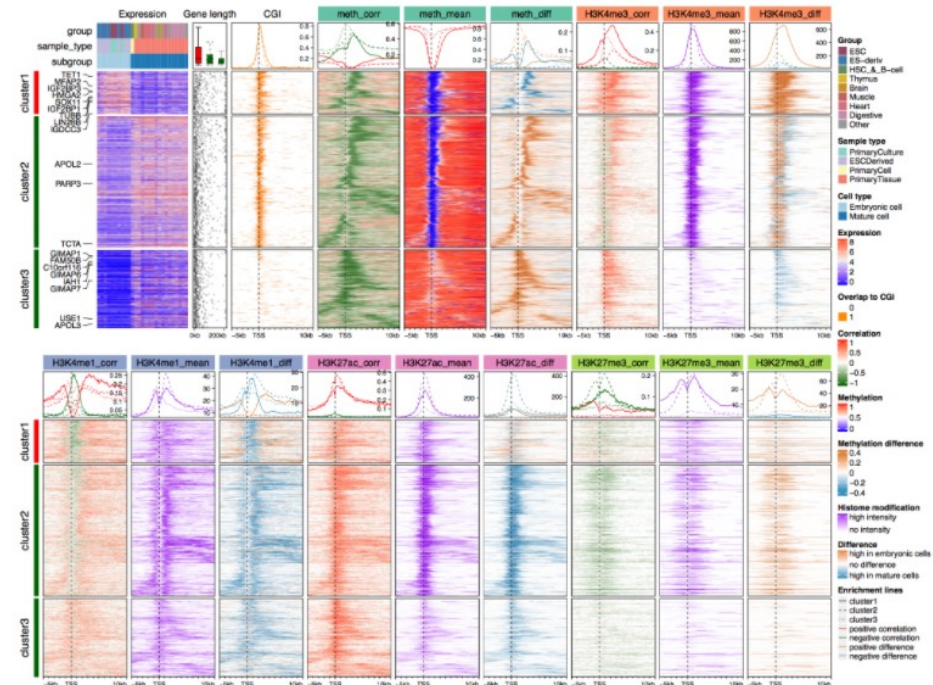
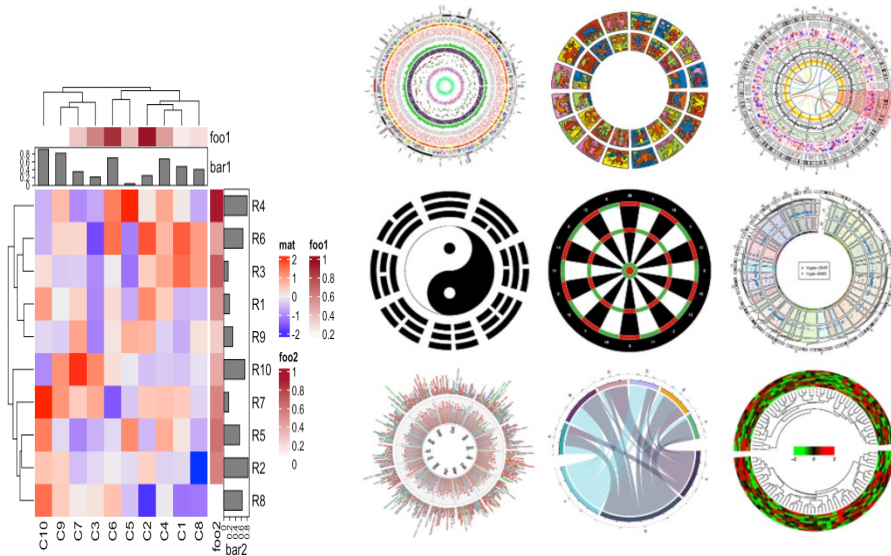
# 临床应用 6

## 可视化

ComplexHeatmap + Circlize +  
EnrichedHeatmap + ...

Zuguang Gu

<https://github.com/jokergoo>



# 临床应用 7

## 可视化

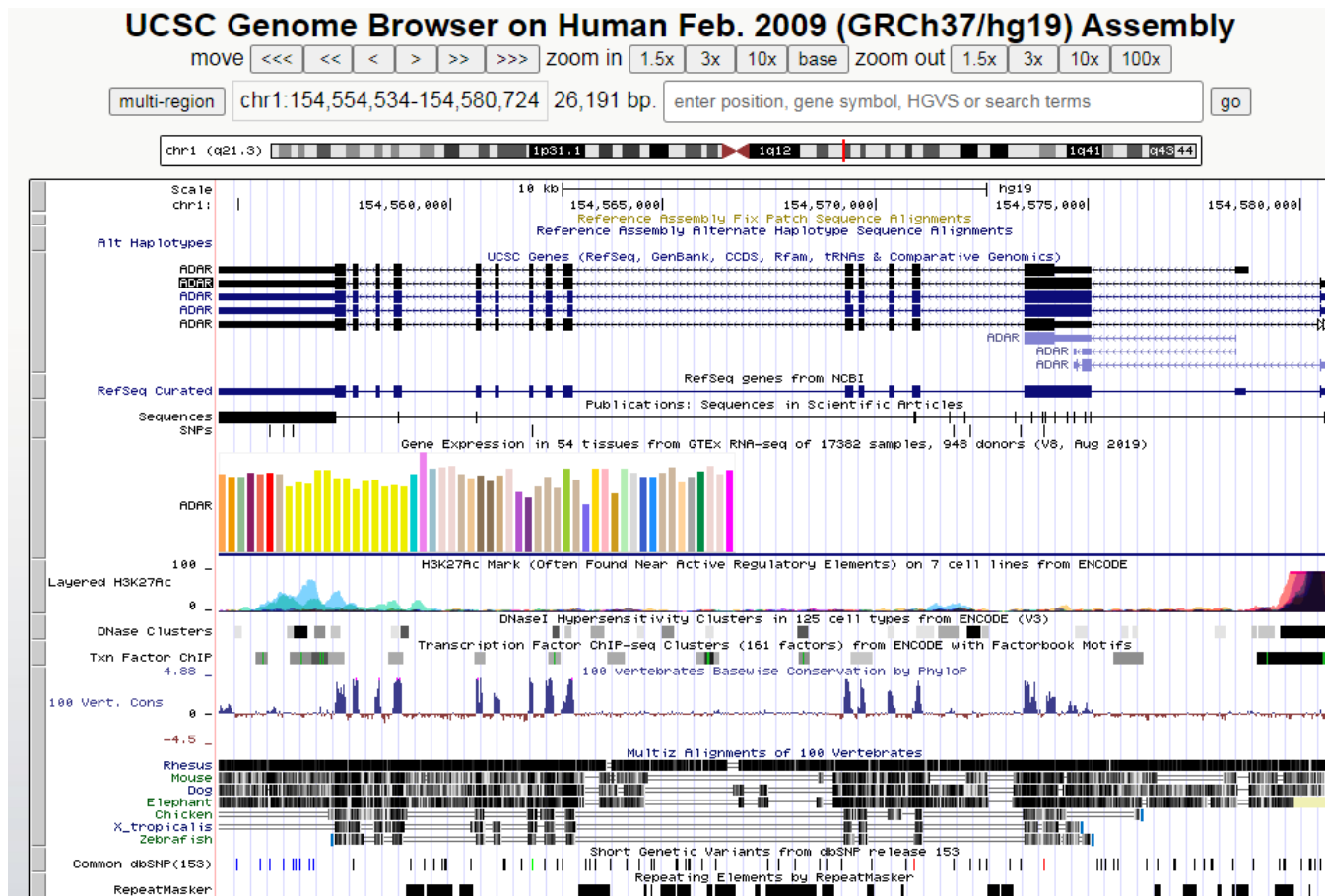
UCSC + WashU + IGV

<http://genome.ucsc.edu/>

<http://epigenomegateway.wustl.edu/>

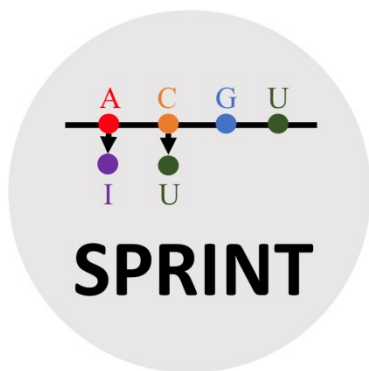
<https://software.broadinstitute.org/software/igv/>

Genome Viewer

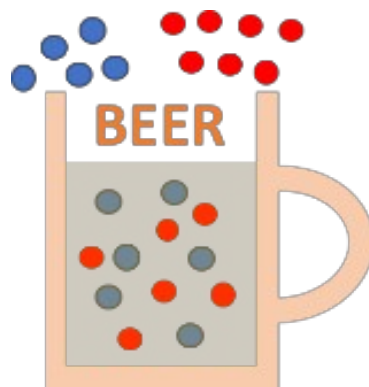




# 介绍研究



RNA 编辑位点鉴定



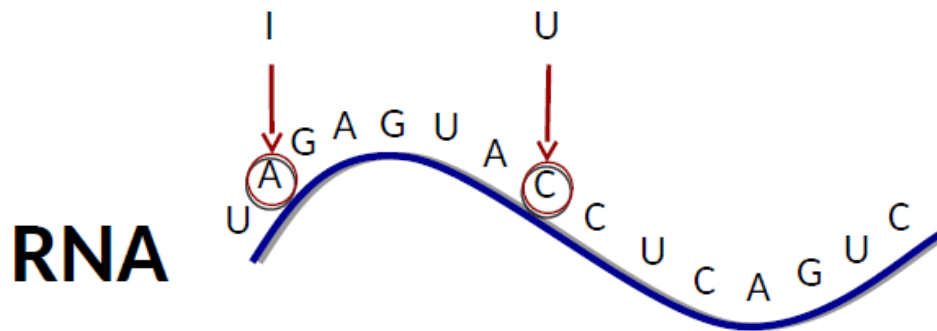
单细胞批次效应



# SPRINT

**RNA Editing** : RNA editing generates post-transcriptional sequence alterations, primarily the modification of RNA nucleotides<sup>1</sup> .

**Primary types** : adenosine-to-inosine ( **A-to-I**, detected as A-to-G )<sup>2</sup> 、 cytosine-to-uracil ( **C-to-U**, detected as C-to-T )<sup>3</sup>



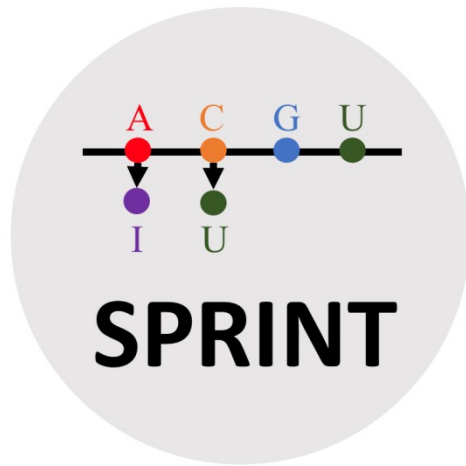
1. Farajollahi, S. & Maas, S. Molecular diversity through RNA editing: a balancing act. *Trends Genet.*

2. Zipeto, M. A., Jiang, Q., Melese, E. & Jamieson, C. H. RNA rewriting, recoding, and rewiring in human disease. *Trends in molecular medicine*

3. Blanc, V. et al. Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. *Genome biology*

# SPRINT

<https://github.com/jumphone/SPRINT>



输入 : RNA-seq 测序数据 (.fastq)

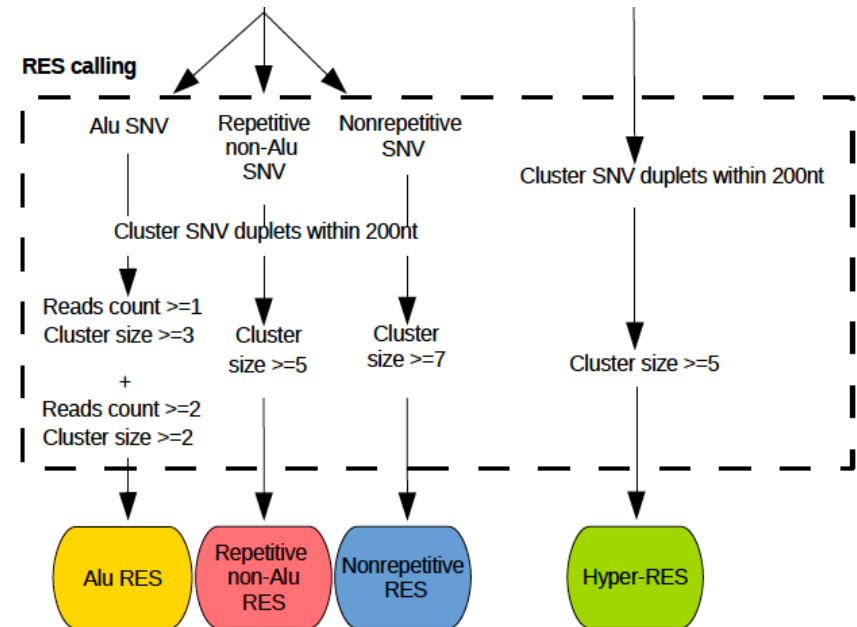
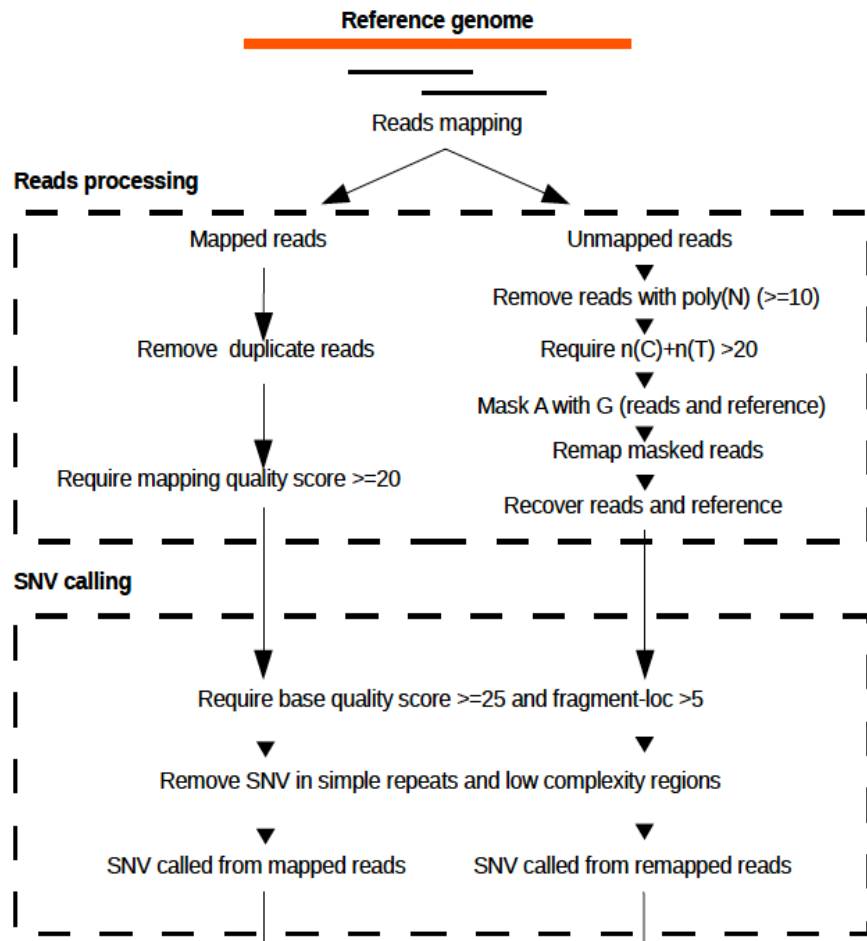
输出 : RNA 编辑位点 (BED)

SNP-free RNA editing Identification Toolkit

Ref. SPRINT: an SNP-free toolkit for identifying RNA editing sites, *Bioinformatics*, 2017

# SPRINT

<https://github.com/jumphone/SPRINT>



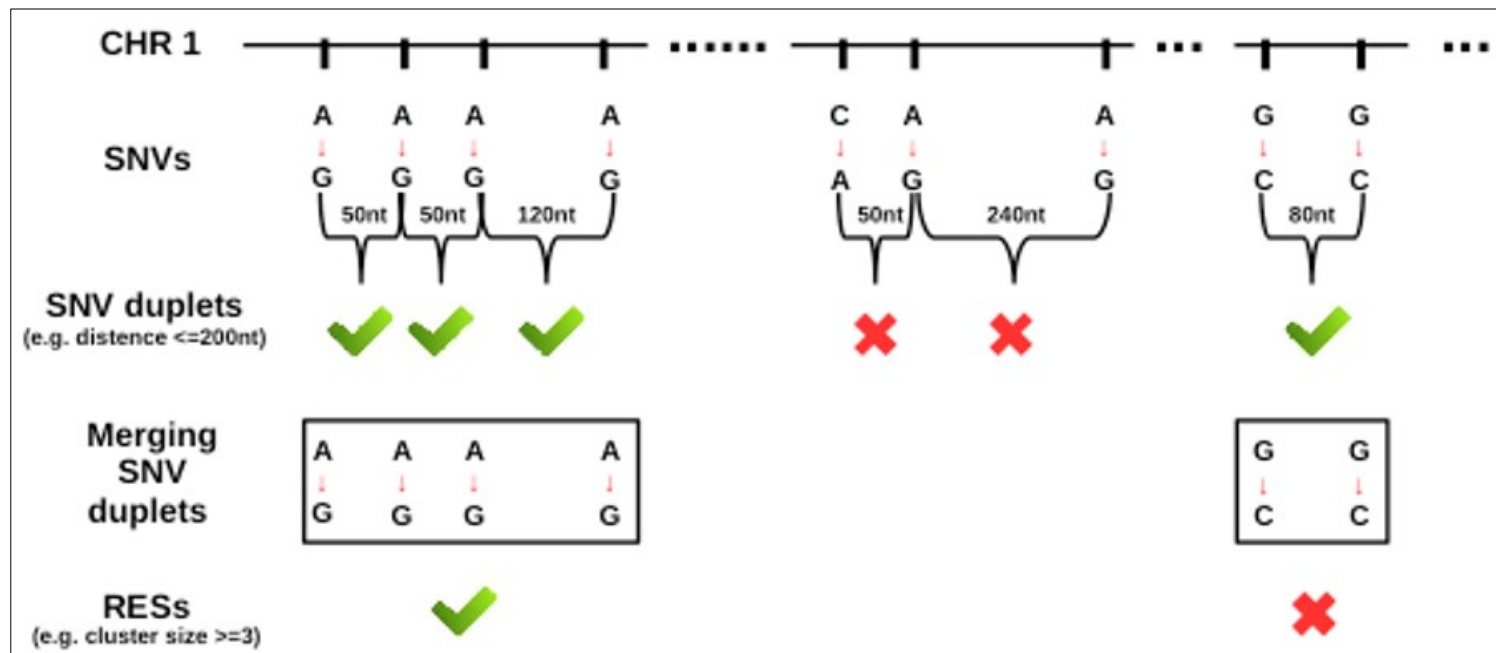
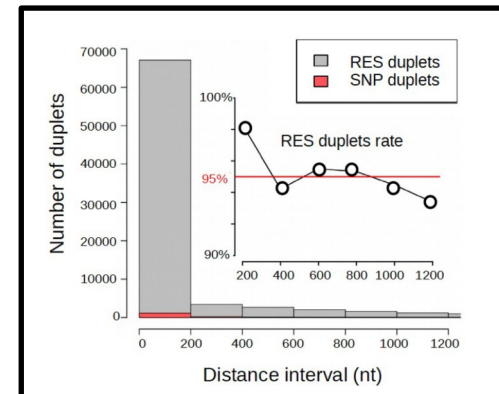
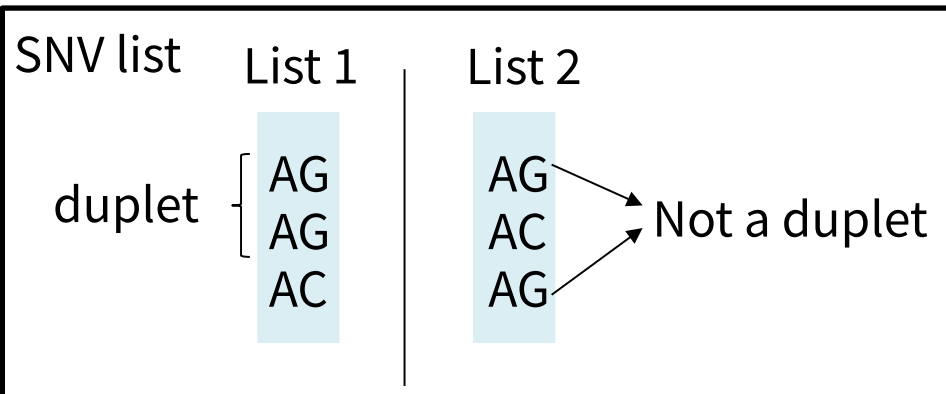
**SPRINT consists of three major steps:**

1. Reads processing
2. SNV calling
3. RES calling

# SPRINT

<https://github.com/jumphone/SPRINT>

SNV calling ⇒

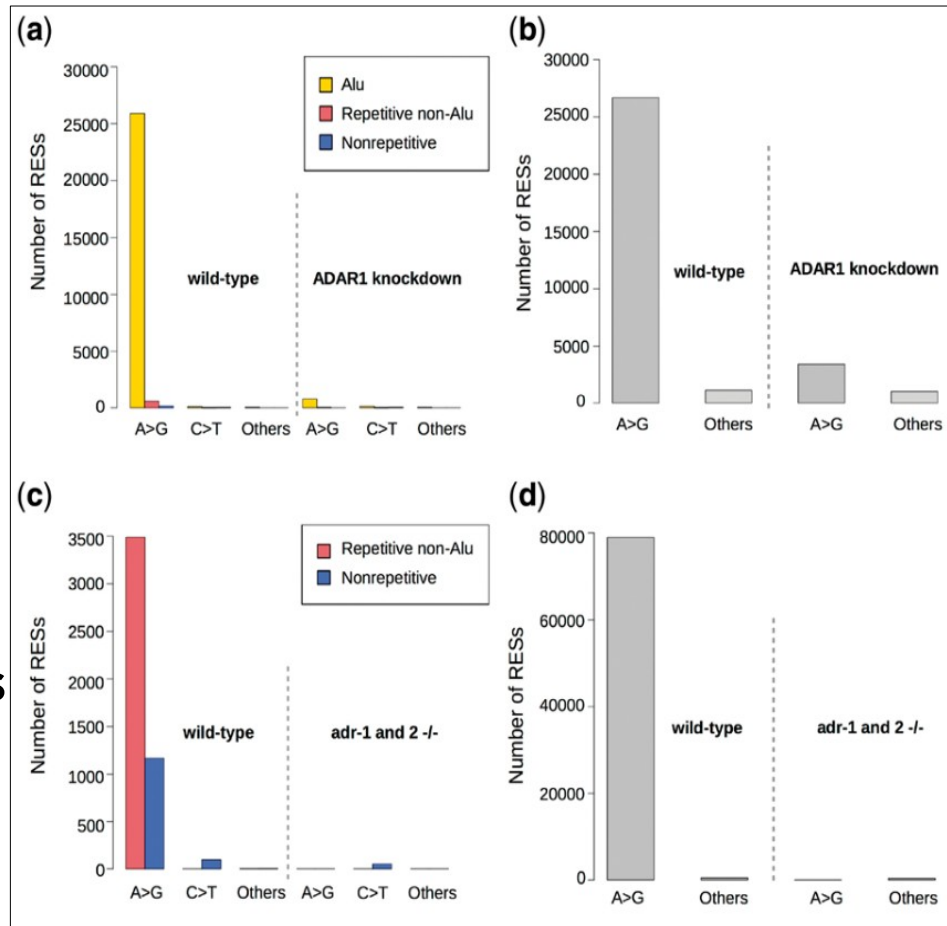


# SPRINT

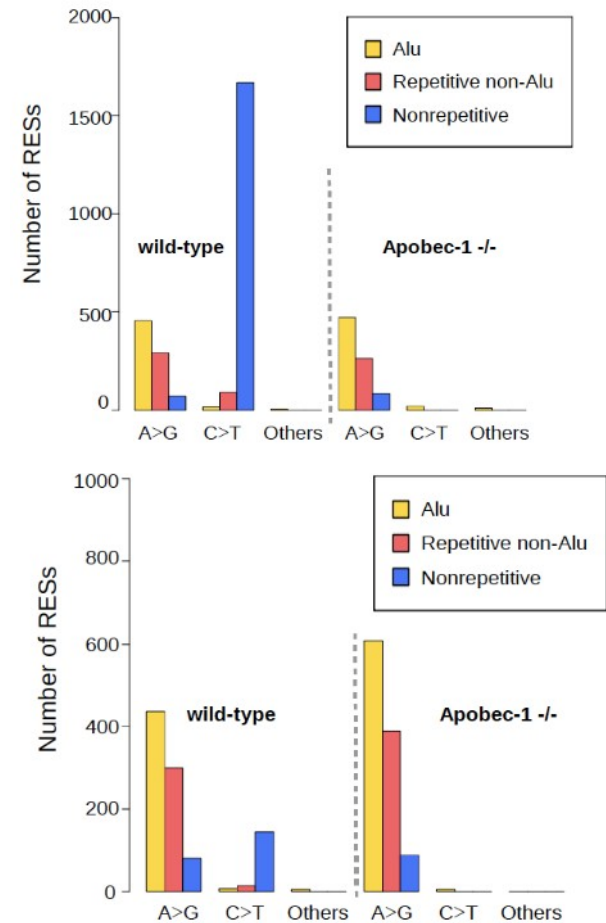
<https://github.com/jumphone/SPRINT>

## A-to-I

Human



## C-to-U Mouse



# SPRINT

<https://github.com/jumphone/SPRINT>

Cell lines	Tools	Alu sites				Repetitive non-Alu sites				Nonrepetitive sites			
		Known SNP (%)	Total	A- to-G (%)	Precision (%)	Total	A- to-G (%)	Precision (%)	FDR (%)	Total	A-to- G (%)	Precision (%)	FDR (%)
GM12878, cell	SPRINT	0	336 304	97.7	96.5	14 019	87.8	97.2	-	5407	49.8 <sup>a</sup>	96.8	-
	Ramaswami <i>et al.</i> *	100	147 029	95.8	-	2385	97.4	-	-	1451	86.6	-	-
GM12878, cytosolic	SPRINT	0	359 725	98.9	96.9	5469	96	96.8	-	2081	75.9	95.5	-
	GIREMI*	70	36 131	99	99.4	267	83.7	84.3	-	1193	82.8	73.8	-
	GIREMI*	100	39 757	99.7	-	260	88.6	-	-	1010	73.5	-	-
U87MG	SPRINT	0	48 085	99.6	96.2	988	99.5	97.1	4.5	296	87.8	91.2	0
	GIREMI	100	2152	99.8	-	114	96.5	-	9	509	88.6	-	53
	RNAEditor	100	62 979	-	-	6142	-	-	42.3	155	-	-	55.5
	REDIttools (de novo)	100	628	96.5	-	238	46.2	-	80	14 949	39.7	-	100
	JACUSA (RRD)	100	2154	94.7	-	331	39	-	-	4527	20.8	-	-

# SPRINT

<https://github.com/jumphone/SPRINT>

	Input data	Re-align & hyper-RES calling	Reads alignment assisting	SNV calling	RES calling
SPRINT	RNA-seq	√	√	√	√
GIREMI	dbSNP + RNA-seq				√
RNAEditor	dbSNP + RNA-seq		√	√	√
REDIttools	dbSNP + RNA-seq or DNA-seq + RNA-seq			√	√
RED	dbSNP + RNA-seq or DNA-seq + RNA-seq				√
RES-Scanner	DNA-seq + RNA-seq		√	√	√
JACUSA	DNA-seq + RNA-seq			√	√



# SPRINT

<https://github.com/jumphone/SPRINT>

## SPRINT: an SNP-free toolkit for identifying RNA editing sites.

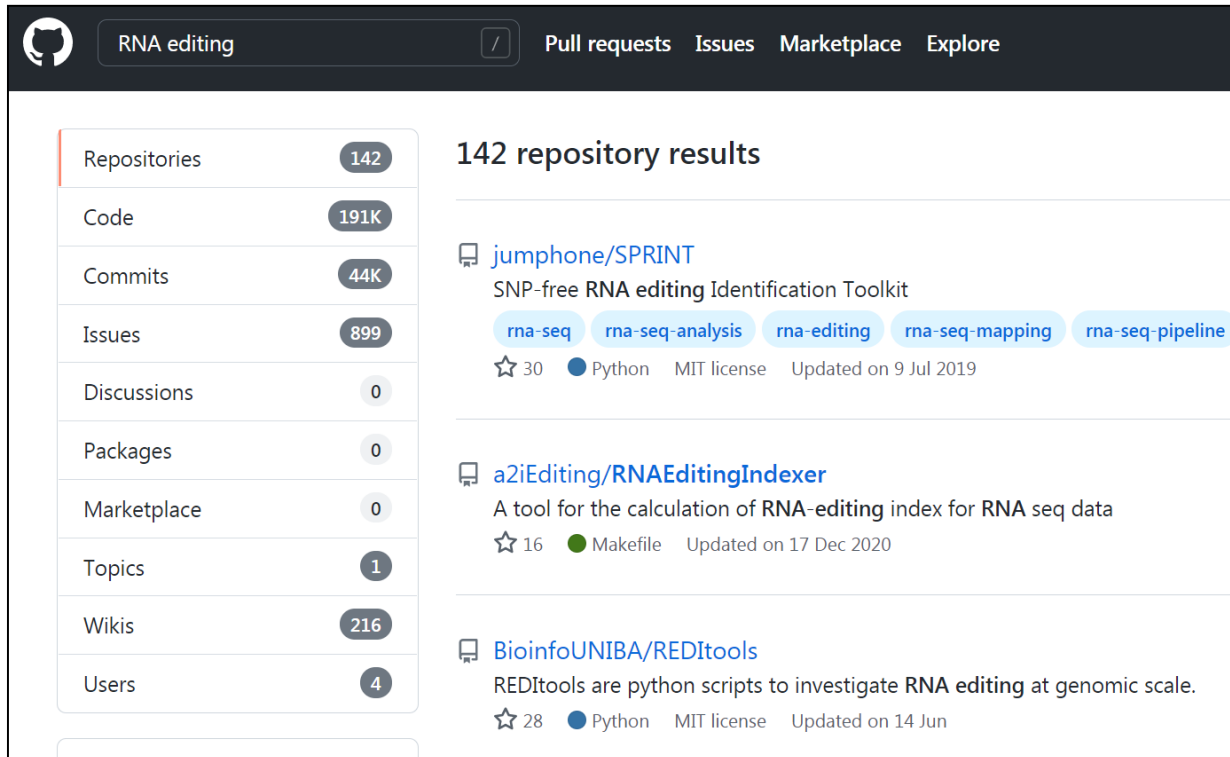
2017 *Bioinformatics* | Volume: 33, Issue: 22, pp 3538-3548 | DOI: 10.1093/BIOINFORMATICS/BTX473

Feng Zhang <sup>1</sup>, Yulan Lu <sup>2</sup>, Sijia Yan <sup>3</sup>, Qinghe Xing <sup>3</sup>, Weidong Tian <sup>3</sup>

<sup>1</sup> State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development,

<sup>2</sup> The Molecular Genetic Diagnosis Center, Shanghai Key Lab of Birth Defect, Translational Medicine Research Center of Children Development and Diseases, Pediatrics Research Institute., <sup>3</sup> Fudan University

 56 References  29 Citations \*



The screenshot shows the GitHub search interface for the query "RNA editing". The search results are displayed on the right, showing 142 repository results. The left sidebar shows the search filters: Repositories (142), Code (191K), Commits (44K), Issues (899), Discussions (0), Packages (0), Marketplace (0), Topics (1), Wikis (216), and Users (4). The search results list the following repositories:

- jumphone/SPRINT**: SNP-free RNA editing Identification Toolkit. Tags: rna-seq, rna-seq-analysis, rna-editing, rna-seq-mapping, rna-seq-pipeline. 30 stars, Python, MIT license, Updated on 9 Jul 2019.
- a2iEditing/RNAEditingIndexer**: A tool for the calculation of RNA-editing index for RNA seq data. 16 stars, Makefile, Updated on 17 Dec 2020.
- BioinfoUNIBA/REDIttools**: REDIttools are python scripts to investigate RNA editing at genomic scale. 28 stars, Python, MIT license, Updated on 14 Jun.

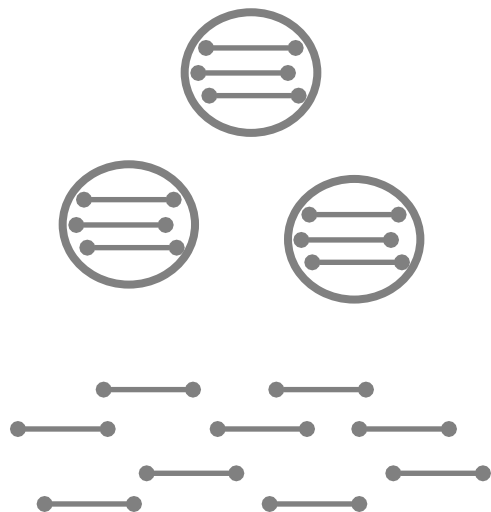
# BEER

<https://github.com/jumphone/BEER>

单细胞测序技术是**对单个细胞的序列信息进行高通量探测的技术**。与大细胞量（bulk）测序不同，每一条测序读段都有独特的标记，用以区分来自不同细胞的测序读段。

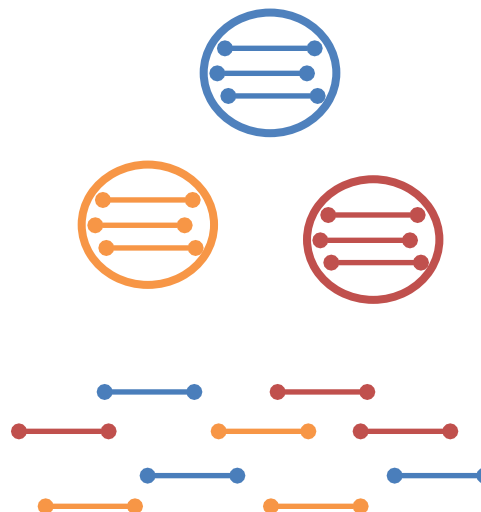
2005 年至今

大细胞量测序技术



2012 年至今

单细胞测序技术



# BEER

<https://github.com/jumphone/BEER>

单细胞分析工具**非常多**

<https://www.scrna-tools.org/tools>

The screenshot displays the 'Tools' page of the scRNA-tools website. The header includes the logo and navigation links: Table, Tools (highlighted), Analysis, Updates, Submit, and FAQs. The main heading is 'Tools' with a subtitle 'View entries for individual tools'. On the left, there is a 'Sorting & Filters' section with a 'Sort by' dropdown set to 'Name' and a 'Filter by category' section with a list of categories including Alignment, Allele Specific, Alternative Splicing, Assembly, Cell Cycle, Classification, Clustering, Differential Expression, Dimensionality Reduction, Gene Filtering, Gene Networks, Gene Sets, Haplotypes, and Immune. A 'FILTER' button is at the bottom of the filter section. The main content area shows a list of tools under the letter 'A', including 'acorde', 'ACTINN', 'ACTION', 'ACTIONet', 'ACTIVA', and 'ADImpute'. A vertical sidebar on the left of the tool list contains letters A through Z. A cookie consent banner is visible at the bottom right.

scRNA-tools

Table Tools Analysis Updates Submit FAQs

## Tools

View entries for individual tools

### Sorting & Filters

Sort by

Name

Filter by category

Select multiple categories and click FILTER below

- Alignment
- Allele Specific
- Alternative Splicing
- Assembly
- Cell Cycle
- Classification
- Clustering
- Differential Expression
- Dimensionality Reduction
- Gene Filtering
- Gene Networks
- Gene Sets
- Haplotypes
- Immune

FILTER RESET

A

- acorde
- ACTINN
- ACTION
- ACTIONet
- ACTIVA
- ADImpute

We use cookies

We use cookies and other tracking technologies to enhance your navigation, analyze site usage, and assist in our marketing efforts. (Privacy Policy)

OK Change my preferences

# BEER

降噪、降维、展示

Seurat: <https://satijalab.org/seurat/>

有 3 个阶段的数据是消除批次效应的突破口

数据预处理

标准化, 基因数量、读段数量, 线粒体基因表达, 细胞周期, doublet

线性降维

	Cell1	Cell2	Cell 3	...
Gene1				
Gene2				
Gene3				
...				

线性

	Cell1	Cell2	Cell3	...
PC1				
PC2				
PC3				
...				

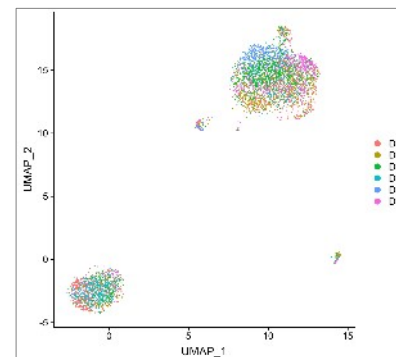
$$C1 * \text{Gene1} + C2 * \text{Gene2} + C3 * \text{Gene3} + \dots = \text{PC1}$$

非线性降维  
(展示)

选择解释方差较大  
的一定数量的 PCs  
(top20, 50, or 100)

非线性

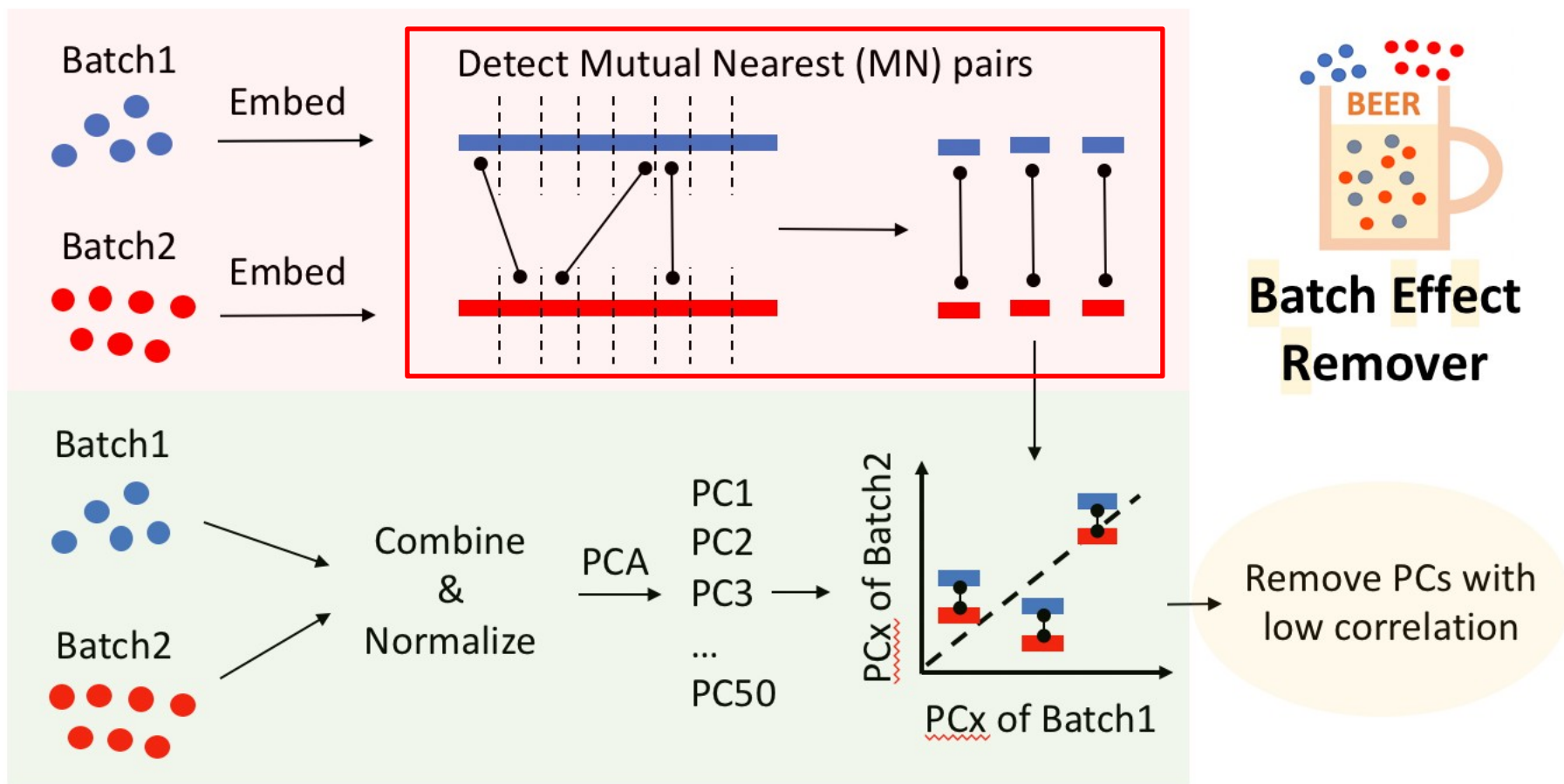
tSNE  
UMAP  
...



# BEER

<https://github.com/jumphone/BEER>

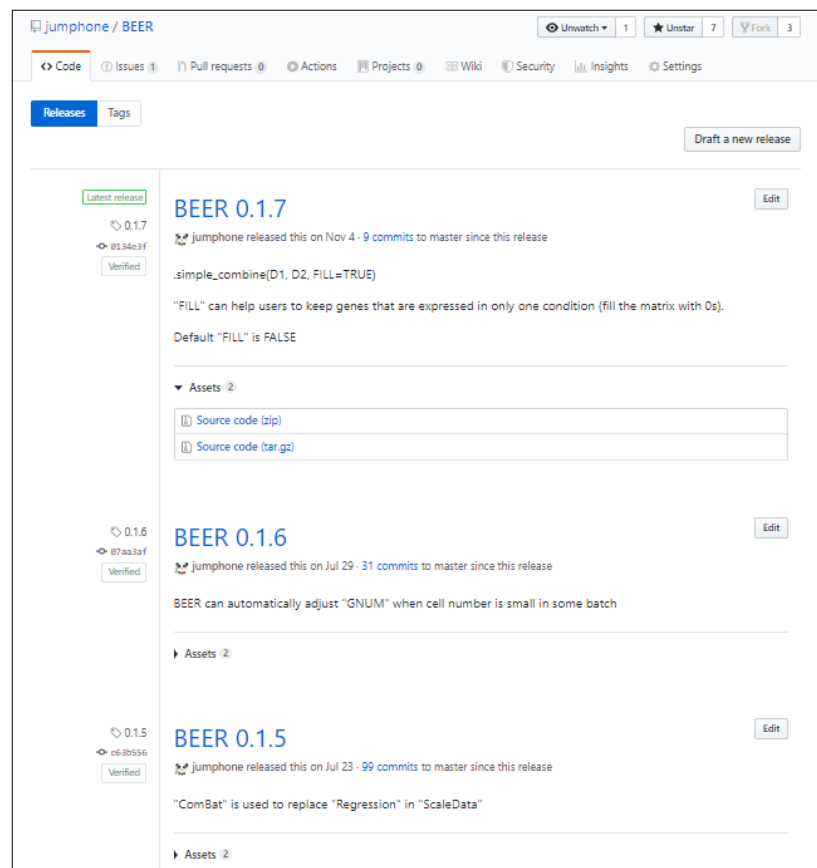
寻找互最近邻对（不同批次中相似的细胞类型）



# BEER

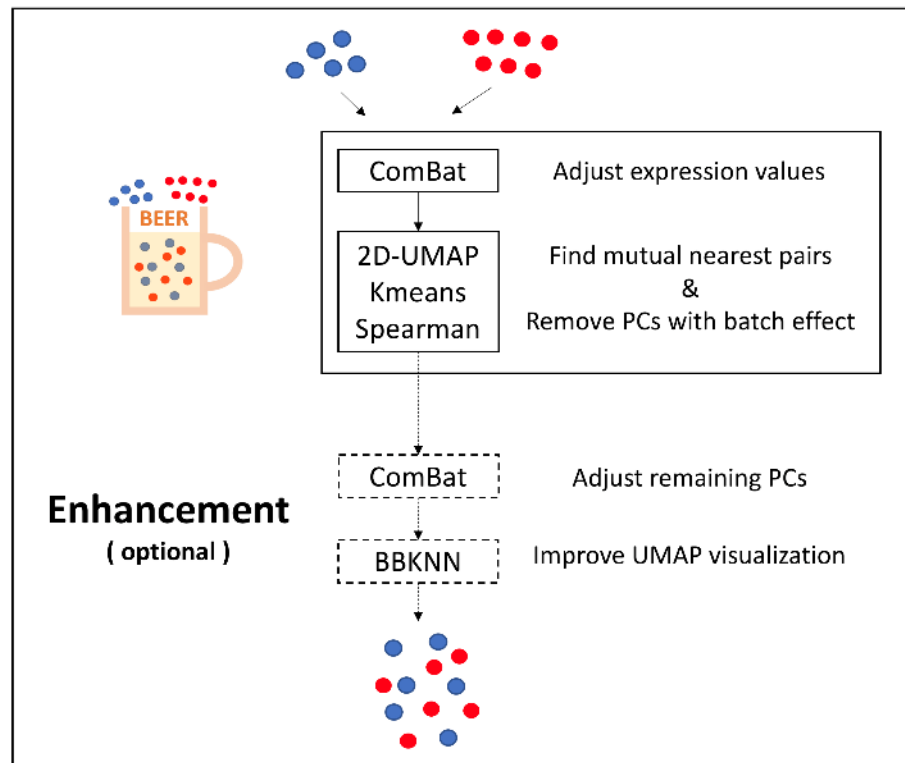
<https://github.com/jumphone/BEER>

## Recent update



The screenshot shows the GitHub repository for jumphone/BEER. The 'Releases' tab is selected, displaying a list of versions. The latest release is BEER 0.1.7, released on Nov 4, with 9 commits since the previous release. The release description includes the command `simple_combine(D1, D2, FILL=TRUE)` and explains that 'FILL' helps users keep genes expressed in only one condition. Assets for source code (zip and tar.gz) are provided. Below it, BEER 0.1.6 is shown, released on Jul 29, with 31 commits. Its description mentions automatic adjustment of 'GNU' when cell number is small. Assets are also provided. At the bottom, BEER 0.1.5 is shown, released on Jul 23, with 99 commits. Its description mentions using 'ComBat' to replace 'Regression' in 'ScaleData'. Assets are also provided.

## 提出了加强版分析流程 (应对非线性批次效应)



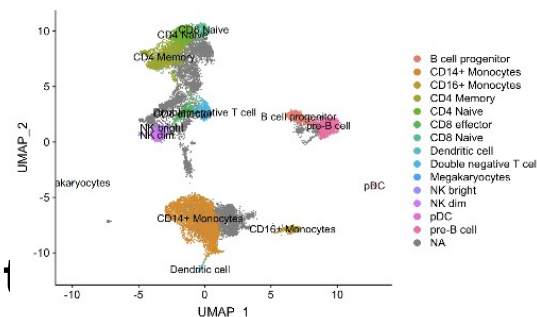
# BEER

<https://github.com/jumphone/BEER>

## 合并 scATAC 与 scRNA

1. Find mutual nearest (MN) pairs
2. Remove PCs with “batch effect”
3. Use Combat & BBKNN to further adjust the output

原版 BEER



No imputation ( 没有使用插值【预测的值】 )

No cell removing ( 没有删除细胞 )

加强版 BEER

