

## 数据上传步骤说明

Dr. Tom 系统支持用户上传基因、转录本、蛋白以及差异甲基化区域(DMR)的相关属性，方便客户在 Dr. Tom 系统上分析自有数据。

本文档以上传基因表达量为例，介绍数据上传的详细步骤。具体的文件要求请查看 Dr. Tom 系统帮助页面-RNA、蛋白、单细胞、WGBS-功能介绍-数据上传文件说明。

基因表达量通常是在做完基因集的比对和定量之后得到的。根据是否进行数据的标准化及标准化方法的不同，会得到不同的表达量数据，如 Read count、FPKM、TPM、RPKM、CPM 等。Dr. Tom 支持用户上传 Read count、FPKM、TPM 进行数据的可视化。

### 1. 准备需要上传的表达量文件

文件的第一列为基因 ID，第一行为样本名称，该表格展示的是基因在样本中的表达量，如下图所示：

ID	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6
8693	6619.16	5833.86	6487.22	8363.36	5078.79	1895.82
100533467	4435.23	9811.50	8970.88	7172.33	9682.41	6750.40
109504726	141.38	8195.87	3663.18	3900.02	7942.50	8246.08
79008	6233.01	4321.49	2312.22	6103.42	8092.53	2593.47
101059918	9838.78	7165.28	5714.44	5536.26	7148.92	6679.49
150094	8389.71	5024.98	9918.34	8565.71	1773.94	7877.94
100526772	1811.06	6780.09	5695.26	8266.29	193.98	9857.68
151742	4341.98	3407.90	1639.30	8308.01	8459.63	7671.89
145781	5113.20	2465.99	439.02	6533.93	7830.16	9024.45
106865373	157.16	7089.87	4614.55	3227.41	6169.06	6638.04
116804918	5470.72	7782.15	263.75	6746.95	6989.37	5605.16
100526832	57.84	6092.46	9775.87	3709.64	5754.11	423.26
55096	8837.17	5591.61	467.75	2308.44	5335.85	6769.45

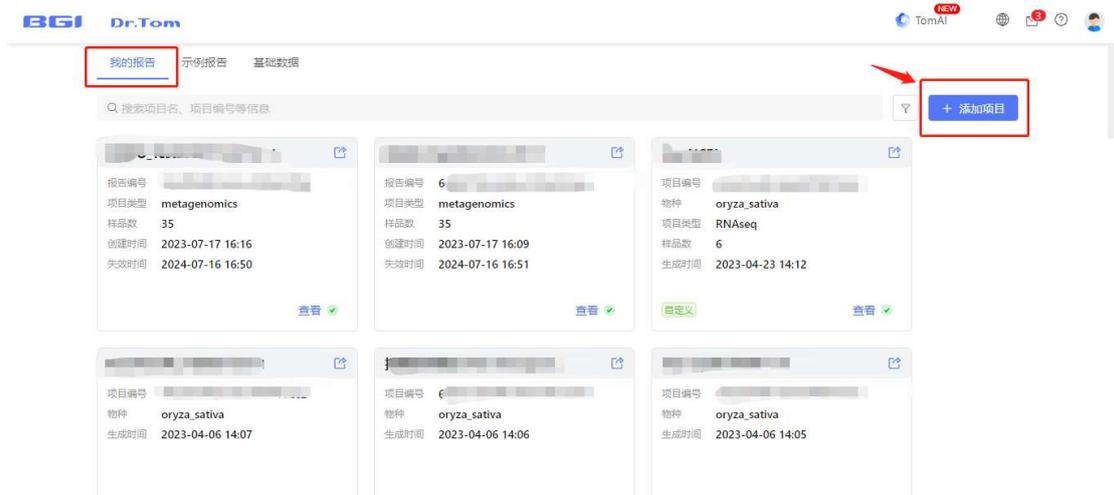
第一列的表头为“ID”，后续每一列的表头为样本名称。

表格需保存成文本文件（制表符分隔）。

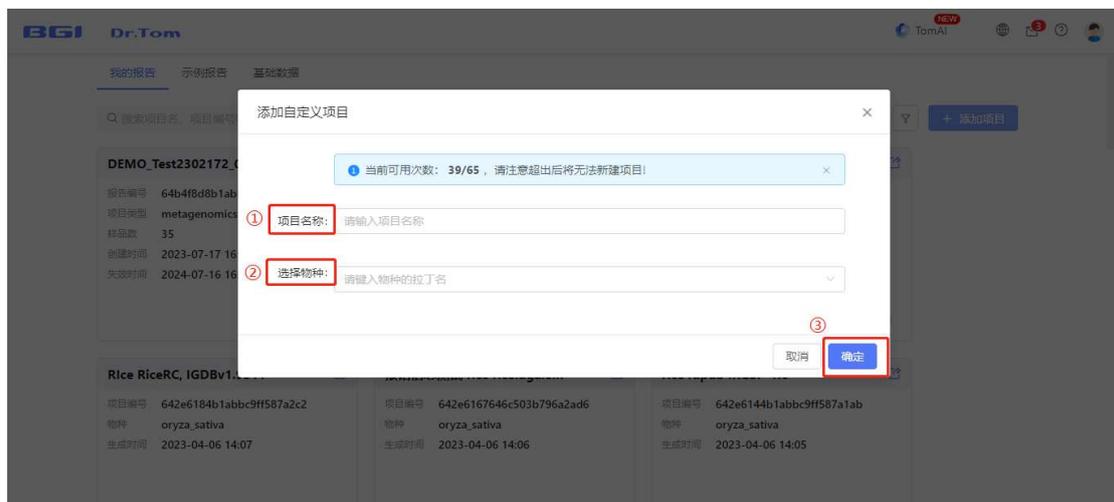


### 2. 上传表达量文件

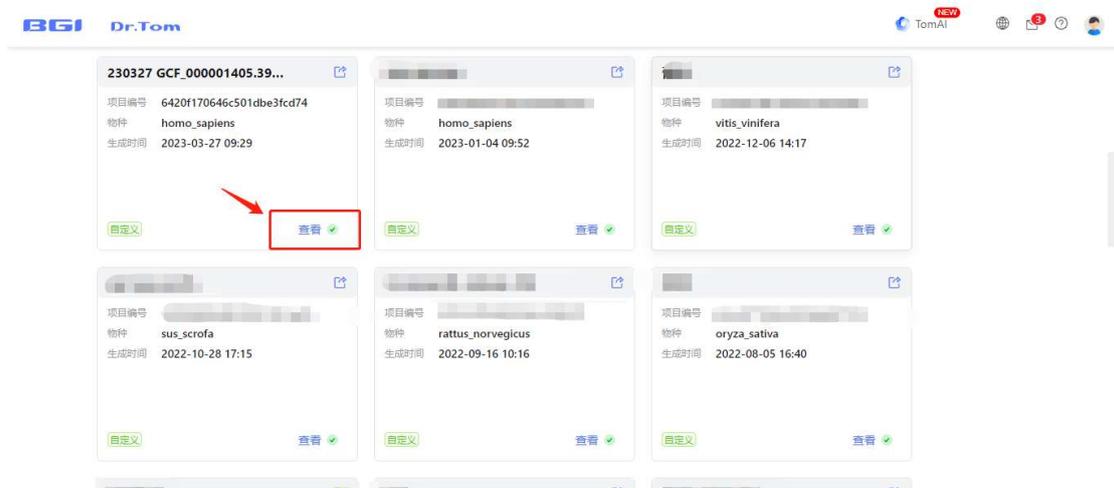
(1) 登录 Dr. Tom 系统。首先，在‘我的报告’选项卡下，点击‘添加项目’按钮以创建一个新的项目。



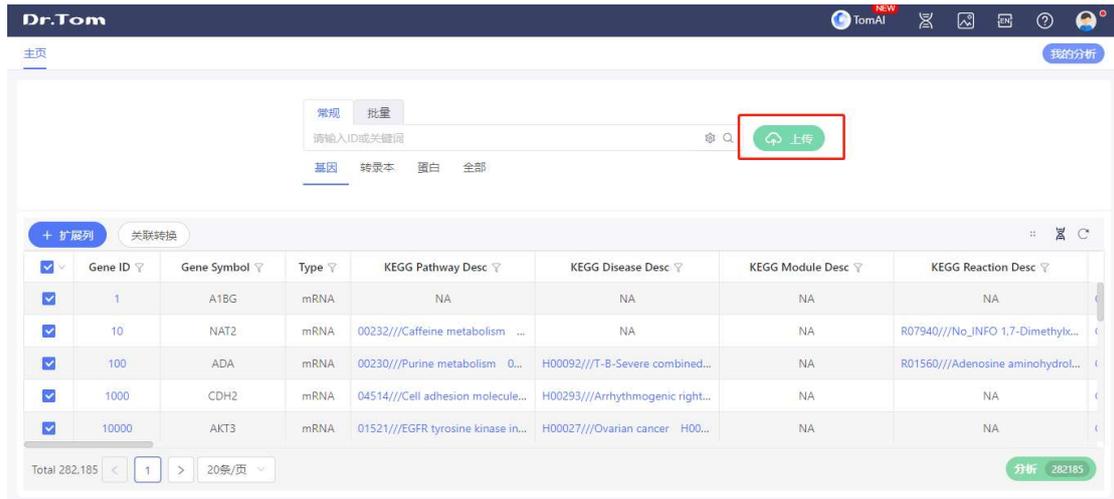
(2) 填写项目名称以及该项目所使用的参考基因组，点击‘确定’，创建项目。参考基因组必须与比对及定量时所使用的参考基因组一致，否则系统无法识别文件中的数据。需要注意的是，项目一旦创建，参考基因组就无法修改了，需谨慎选择。



(3) 在‘我的报告’选项卡下可看到创建的项目。点击‘查看’，进入项目



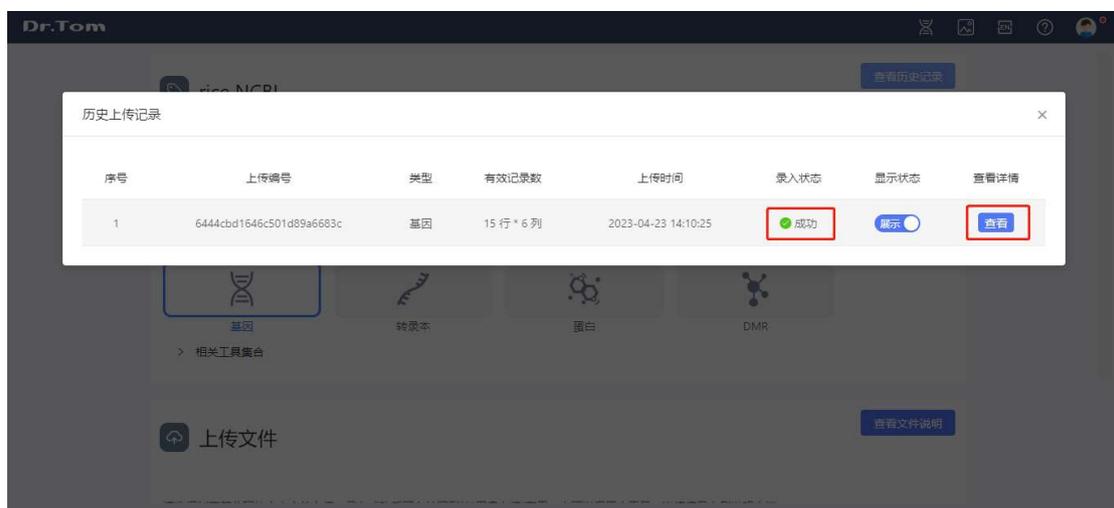
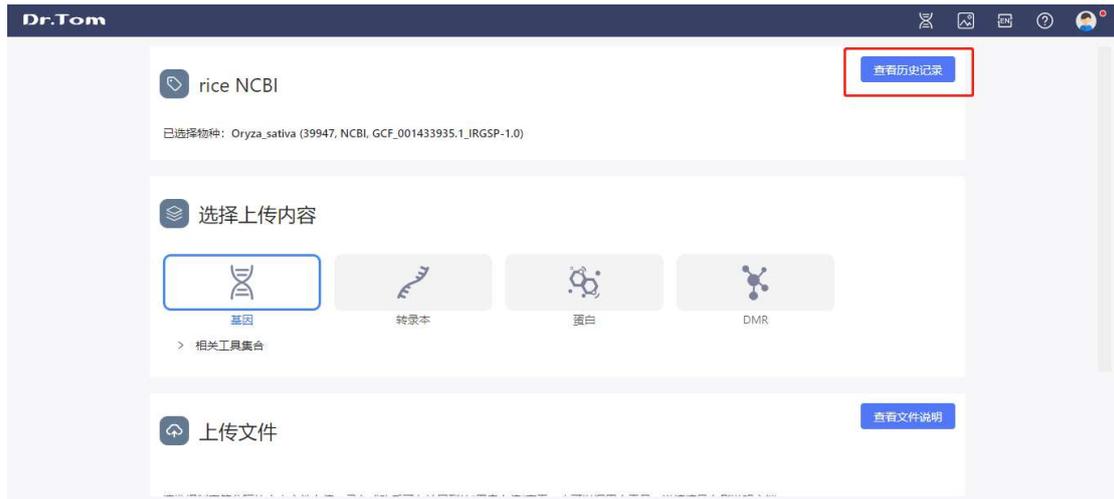
(4) 点击‘上传’按钮，进入数据上传页面。



(5) 在'选择上传内容'部分，选择要上传的内容（此例中选择基因）；在'上传文件'部分，选择 ID 形式（NCBI Gene ID, Gene Symbol 或 Ensembl Gene ID，以实际页面显示及文件内容为准）、文件类型；并在上传框中需要上传的文件。点击'上传'。



(6) 文件上传后，可点击'查看历史记录'按钮，确认文件是否成功上传。如上传失败，可点击'查看'按钮查看失败原因。



### 3. 填写分析方案

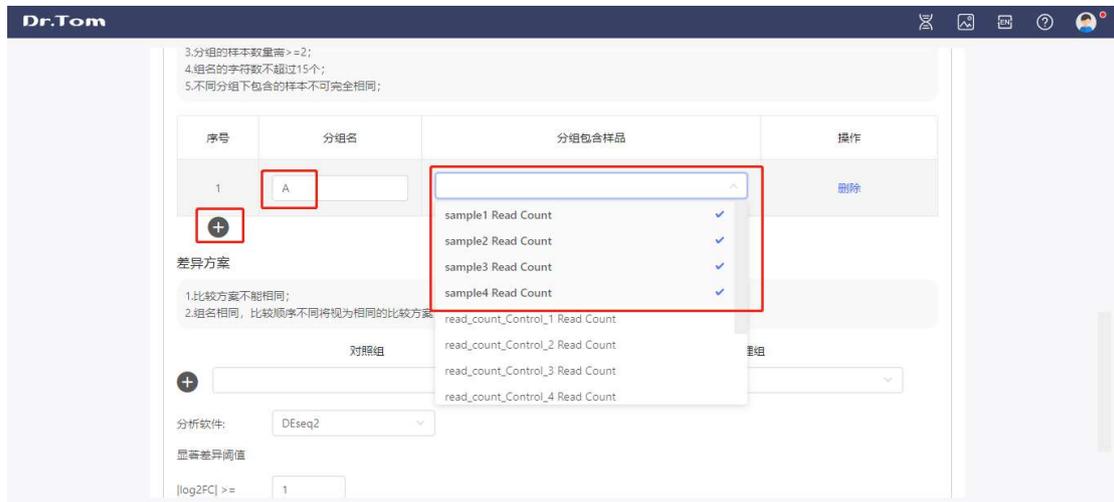
(1) 当上传的文件是 Read count 时，会出现‘填写方案’模块。单击‘点击此处填写方案’，则会出现方案填写界面。因差异分析输入的表达式数据需为未经过标准化的 read count，如果上传的是其他表达量（FPKM 或 TPM），则不会出现该按钮。



(2) 根据实际项目情况填写分析方案。

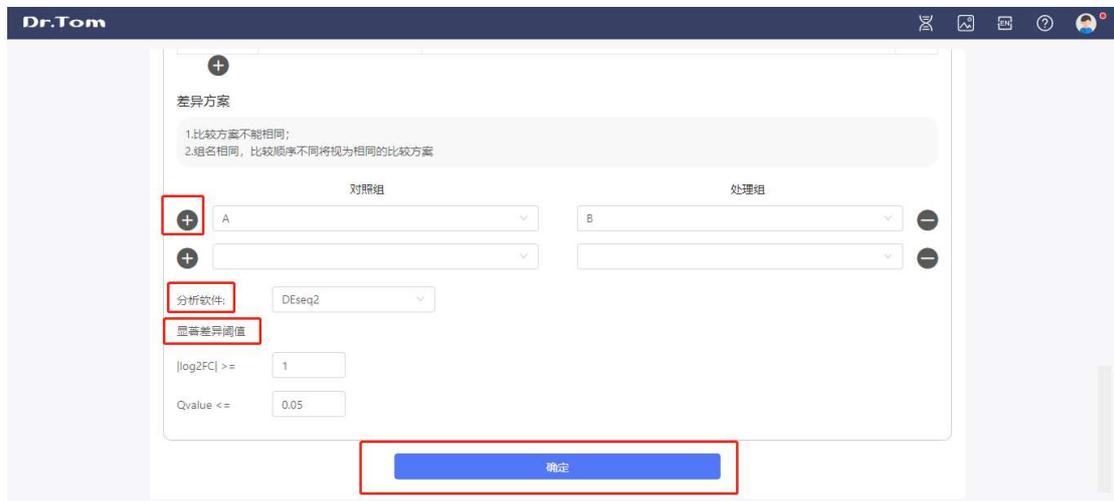
如果实验涉及中有生物学重复，则勾选‘有生物学重复’，反之，则勾选‘无生物学重复’。(本例中选择有生物学重复)

在‘分组方案’部分，设置组别。点击‘+’可添加组别。进行组别的命名，并将生物学重复归属为同一组。



(3) 设置好组别后，在‘差异方案’部分设置差异比较方案。对应设置好对照组及处理组。若有多组差异比较方案，则点击‘+’添加。

选择差异分析软件，并设置好差异阈值。通常默认选择 DEseq2， $|\log_2FC| \geq 1$ ， $Qvalue \leq 0.05$  填写完毕后，点击‘确定’提交分析方案。



#### 4. 查看分析结果并使用 Dr. Tom 进行数据可视化

提交分析方案后，系统需要一定的时间分析您的数据。分析完毕后，您会收到来自 Dr. Tom 系统的邮件提醒。登录 Dr. Tom 系统，进入该项目，即可在导航栏模块查看分析结果。

常规 批量  
请输入ID或关键词

[基因](#) [转录本](#) [蛋白](#) [全部](#)

[+ 扩展列](#) [关联转换](#) ::  

<input checked="" type="checkbox"/>	Gene ID	Gene Symbol	Type	KEGG Pathway Desc	KEGG Disease Desc	KEGG Module Desc	KEGG Reaction Desc
<input checked="" type="checkbox"/>	107275246	LOC107275246	mRNA	NA	NA	NA	NA
<input checked="" type="checkbox"/>	107275247	LOC107275247	mRNA	NA	NA	NA	NA
<input checked="" type="checkbox"/>	107275248	LOC107275248	mRNA	NA	NA	NA	NA
<input checked="" type="checkbox"/>	107275249	LOC107275249	mRNA	NA	NA	NA	NA
<input checked="" type="checkbox"/>	107275250	LOC107275250	mRNA	00190//Oxidative phosphorylat...	H02086//Mitochondrial compl...	M00152//Cytochrome bc1 co...	NA

Total 33,897 < 1 > 20条/页 分析 33897