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)。

文件名(<u>N</u>):	工作簿1				~
保存类型(工):	文本文件(制表符分隔)				~
▲ 隐藏文件夹			工具(1) ▼	保存(<u>S</u>)	取消

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。

BGI	Dr.Tom		c	TomAl 🕀 🔮 📀 🍨
[我的报告 示例报告 基础数据		~	
	Q 搜索项目名、项目编号等信息		Ŷ	+ 添加项目
	授任编号 项目典型 metagenomics 用品数 35 领理时期 2023-07-17 16:16 先效时期 2024-07-16 16:50 查看 ✔	投告编号 6 项目类型 metagenomics 杆品数 35 创建时间 2023-07-17 16:09 失效时间 2024-07-16 16:51	び 項目編号 物件 oryza_sativa 項目処理 RNAseq 程品数 6 生成対測 2023-04-23 14:12 重置 ◆	
	次目得う 物种 oryza_sativa 生成时间 2023-04-06 14:07		で 双目编号 物种 oryza_sativa 生成対阈 2023-04-06 14:05	

(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。

BGI	Dr.Tom						TomAl	•	0 9	2
	我的报告示例报告 基	基础数据								
		添加自定义项目				×	Y + 350	晒目		
	DEMO_Test2302172_(留意調整 64b4f8d8b1ab 期意思想 metagenomics ① 容易說 35 14個的別 2023-07-17 16 失致可測 2024-07-16 16 ②	 项目名称: 请 选择物种: 请 	当前可用次数	: 39/65 , 请注意超出后将无法新建 名	项目: (× 3	Ľ			
	Rice RiceRC, IGDBv1.		ACTURATION OF		AV.H	100E	2			
	项目编号 642e6184b1abbc9f 物种 oryza_sativa 生成时间 2023-04-06 14:07	f587a2c2		642e6167646c503b796a2ad6 oryza_sativa 2023-04-06 14:06	 原目編号 642e6144b1abbc9 物件 oryza_sativa 生成期間 2023-04-06 14:05 	ff587a1ab				

(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

230327 GCF_000001	405.39 🗠	-		C	1		C		
项目编号 6420f170646 物种 homo_sapier 生成时间 2023-03-27 (c501dbe3fcd74 is 19:29	项目编号 物种 生成时间	homo_sapiens 2023-01-04 09:52		项目编号 物种 生成时间	vitis_vinifera 2022-12-06 14:17			
目定义	直斎 🖉	自定义		查看。	自定义	ŧ	语 🖌		
(a month	1	-	A 1994	Ľ	-		C		
项目编号 物种 sus_scrofa 生成时间 2022-10-28 *	17:15	项目编号 物种 生成时间	rattus_norvegicus 2022-09-16 10:16		项目编号 物种 生成时间	oryza_sativa 2022-08-05 16:40			
自定义	查看 必	自定义		查看 🗸	自定义	1	语 🖉		

(4) 点击'上传'按钮,进入数据上传页面。

Dr.To	om					💽 TomAl	* & & = 0	@ *
主页							TEAS	纷析
			常规 请输入 基因	批量 ID成关键词 转录本 蛋白 全部	@: (
(+ 3 ⁺)	愛列 关联轴	頻					:: X	С
~	Gene ID 🖓	Gene Symbol 🖓	Type 🖓	KEGG Pathway Desc 🝸	KEGG Disease Desc \heartsuit	KEGG Module Desc 🖓	KEGG Reaction Desc \heartsuit	
	1	A1BG	mRNA	NA	NA	NA	NA	c
	10	NAT2	mRNA	00232///Caffeine metabolism	NA	NA	R07940///No_INFO 1.7-Dimethylx.	
	100	ADA	mRNA	00230///Purine metabolism 0	H00092///T-B-Severe combined	NA	R01560///Adenosine aminohydrol	
	1000	CDH2	mRNA	04514///Cell adhesion molecule	H00293///Arrhythmogenic right	NA	NA	ć
	10000	AKT3	mRNA	01521///EGFR tyrosine kinase in	H00027///Ovarian cancer H00	NA	NA	¢
Total 282	2,185 < 1	> 20祭/页 >					जीमा 2821	85

(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'。

Dr.Tom	y N) 🖻	0	e *
じ rice NCBI 已造廃物時: Oryza_sativa (39947, NCBI, GCF_001433935.1_IRGSP-1.0)	查看历史记录			
送 送 送 送 送 送 ご<				
上传文件 通貨探制表符分類的文本文件上传,最大成功店可在扩展列的「用户上传 音(1)也可以调用小工具、详情演现右侧说明文档、 NCBI Gene ID NCBI Gene ID Read counts ② 工作簿1.tx	查看文件说明			

(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。

Dr.Tom						0	@ *
	S rice NCBI				查着历史记录		
	已选择物种: Oryza_sativa (39947, I	NCBI, GCF_001433935.1_IRGSP	-1.0)				
	◎ 选择上传内容						
	X	Read Street	Co	×			
	基因 > 相关工具集合	转录本	蛋白	DMR			
	♀ 上传文件				查看文件说明		
			·	National Provide American Street Street			

Dr.T	om						X	2	1	• *
_							查看历史记录			
	历史上传记	录							×	
	序号	上传编号	类型	有效记录数	上传时间	录入状态	显示状态	查看详情		
	1	6444cbd1646c501d89a6683c	基因	15 行*6列	2023-04-23 14:10:25	🥝 alih	展示	直有		
			「「「「」」		¢o, ≣≐	M R				
		> 相关工具集合								
		♀ 上传文件					查看文件说明			
			11.27 - 201 (10.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11.27 - 11							

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量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。

Dr.Tom		×	EN	0	@ *
	> 相关工具集合				
	● 上传文件	<u> </u> 一直 一 一 一 一 一 一 一 一 一			
	请选择制表符分隔的文本文件上传,录入成功后可在扩展列的"用户上传"音看,也可以调用小工具。详情请见右侧说明文档。				
	NCBI Gene ID V 前选择文件类型 V				
	20. 填写方案	填写说明			
*	上传文件满足填写方案要求,请单击进入方案填写 仅支持 <u>填写一次方案,</u> 请上传全部样本的 read counts 后再填写方案 点击此处填写方案				

(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。

Dr.Tom						8	~	EN	0	@ *
	3.分组的样本数; 4.组名的字符数; 5.不同分组下包;	量需>=2; 不超过15个; 含的样本不可完全相同;								
	序号	分组名	分组包含样品		操作					
	1	A			删除					
	●		sample1 Read Count	~						
	1.比较方案不能 2.组名相同,比	相同; 较顺序不同将视为相同的比较方	sample4 Read Count	-						
		对照组	read_count_Control_2 Read Count	里组						
	Ð		read_count_Control_3 Read Count read_count_Control_4 Read Count							
	分析软件:	DEseq2								
	显音差异阈值 log2FC >=	1								
	hoger of y =	1.2 1								

(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, |log2FC|≥1, Qvalue≤0.05 填写完毕后,点击'确定'提交分析方案。

Dr.Tom		×	~	EN.	0	?
	Φ					
	差异方案					
	1.比较方案不能相同; 2.组名相同,比较顺序不同将视为相同的比较方案					
	对照组 处理组					
l	● A · · · B · · · ·	•				
		•				
ļ	分析统件: DEseq2 v					
L	显著差异阈值					
	log2FC >= 1					
	Qvalue <= 0.05					
	确定					

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。

Dr.Tom							8	e (?) 🧟"
主页 基因表达 基因性释 基础信息									我的分析
			常规 请输入 基因	批量 1D成关键词 转录本 蛋白 全部	\$	Q. (A) L49			
+ 57版列 关联转换 # 【 C									
×	Gene ID	Gene Symbol 🖓	Туре 🖓	KEGG Pathway Desc 🖓	KEGG Disease Desc \heartsuit	KEGG Module Desc 🖓	KEGG React	ion Desc 🖓	
	107275246	LOC107275246	mRNA	NA	NA	NA	N	A	
	107275247	LOC107275247	mRNA	NA	NA	NA	N	A	
	107275248	LOC107275248	mRNA	NA	NA	NA	N	A	¢
	107275249	LOC107275249	mRNA	NA	NA	NA	N	A	¢
	107275250	LOC107275250	mRNA	00190///Oxidative phosphorylat	H02086///Mitochondrial.compl	M00152///Cytochrome bc1 co	N	A	t
Total 33.	897 < 1	> 20祭/页 >						5746	33897