



转录组与蛋白质组关联分析help文档



@2019 BGI All Rights Reserved

运行过程如果遇到任何问题，请发送邮件至luoxing@genomics.cn

Contents

Contents	1
1 软件使用说明	2
1.1 运行步骤	2
1.2 correlation_config.json文件参数说明	2
2 流程文件与文件夹说明	4
2.1 以比较组为名字的文件夹	4
2.2 Data_Analysis文件夹	4
2.3 submit文件夹	5
3 流程代码解析	5

1 软件使用说明

1.1 运行步骤

- 1.根据时间和项目号，任务单号 在你的工作目录中创建一个为项目号的文件夹（以下统称该文件夹为root）。
- 2.拷贝/Idfssz1/SP_MSI/Pipeline/Pipeline/TP_Cor/bin文件下的run.sh和correlation_config.json到root。
- 3.根据项目内容，修改相应的correlation_config.json内的参数。
- 4.将RNA_Seq所有比较组文件夹拷贝到你的目录中命名为trans（以下统称该文件夹为transcription_root）。
- 5.将iTRAQ所有比较组文件夹拷贝到你的目录中命名为pro（以下统称该文件夹为protein_root）。
- 6.将protein_root中的对应的比较组根据json文件的Comparison_Group参数设置对应的文件名，注意每个比较组需要_C.xls，_F.xls，_P.xls，_path.xls，_Down.xls，_Up.xls,xls。
- 7.将transcription_root中的对应的比较组根据json文件的Comparison_Group参数设置对应的文件名，注意每个比较组需要_C.txt，_F.txt，_P.txt，_path.txt，.GeneDiffExp.xls，如果转录组id为gene，则需要相应的参考基因集文件refMrna.fa和id转换文件gene2tr，如果Whether the transcriptome ko file exist参数为yes，则需要对应的.ko文件。
- 8.sh run.sh即可。

注意：如果输出BGI smile,则运行成功，否则失败，找到原因后，设置json文件的step从对应步骤开始即可，如果2-6步运行时间超过一小时，则很可能运行失败，你需要检测配置错误，重新从该步运行。json中的Comparison_Group参数命名的id可以是任意id，但是必须要与转录组与蛋白质组所对应的比较组命名一致。转录因子流程需要你查阅附件1中看是否有该物种再决定是否进行转录因子分析。转录组需要提供相应的ko文件，如果找不到ko文件，可以将Whether the transcriptome ko file exist设置为no，流程会再跑一遍转录组的Pathway通路富集流程。如果转录组蛋白质组没有做相应的注释与富集，可以设置相应参数为no，该流程会自动进行注释与富集，并且会生成相应的ko文件。

1.2 correlation_config.json文件参数说明

表1-2 流程参数配置说明：配置文件不区分大小写，所有分割符请使用英文符号，*号参数代表维护人员的配置参数，使用者可以忽略。

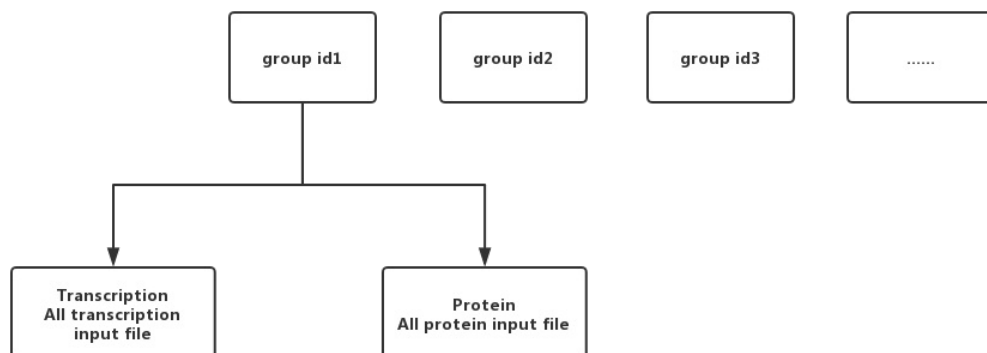
Basis_Information	基础参数配置说明
Language	根据报告需要生成中文或者英文项目报告
Comparison_Group	每个关联分析比较组的名字，使用逗号隔开
Reference_data_set	转录组参考基因集文件的名字
Date	流程的开始时间
SGEProject	提交任务时所需的CPU名字
SGEQueue	提交任务时所需的队列名字
step	流程运行步骤，最大为0-6

Transcriptome ID type	转录组的起始id类型
Transcriptome_path	所有转录组的文件夹路径
Proteomics_path	所有蛋白质组的文件夹路径
Whether to empty the process file	是否删除流程生成的中间文件
*R_HOME_PATH	R路径
*perl_HOME_PATH	perl路径
*Python_HOME_PATH	python3路径
*Correlation analysis process server path	关联分析流程所在路径
Filter Parameters	差异倍数与Blast配置说明
ProteinFoldChange	蛋白组差异倍数，注意与itraq参数一致
GeneFoldChange	转录组差异倍数，注意与RNASeq参数一致
Protein p value Threshold	蛋白组p或者q阈值，注意与itraq参数一致
Gene p value Threshold	转录组p阈值，注意与RNASeq参数一致
Gene Non-difference label	非差异转录组标记
BlastEvaluate	Blast E值
Blast_Memory	Blast任务所需内存
Transcription factor	转录因子配置说明
Whether to do transcription factor analysis	根据附件1是否进行转录因子分析
TF_Species_Name	比较组物种学名
TF_Species_type	比较组物种类型: animal, plant, microorganism
TF_Memory	转录因子提交任务所需内存
*TF_plant_database	转录因子植物数据库路径
*TF_animal_database	转录因子动物数据库路径
Annotation	功能注释富集配置说明
Whether to do transcriptome annotation and enrichment	转录组是否做注释与富集
Whether to do protein annotation and enrichment	蛋白质组是否做注释与富集

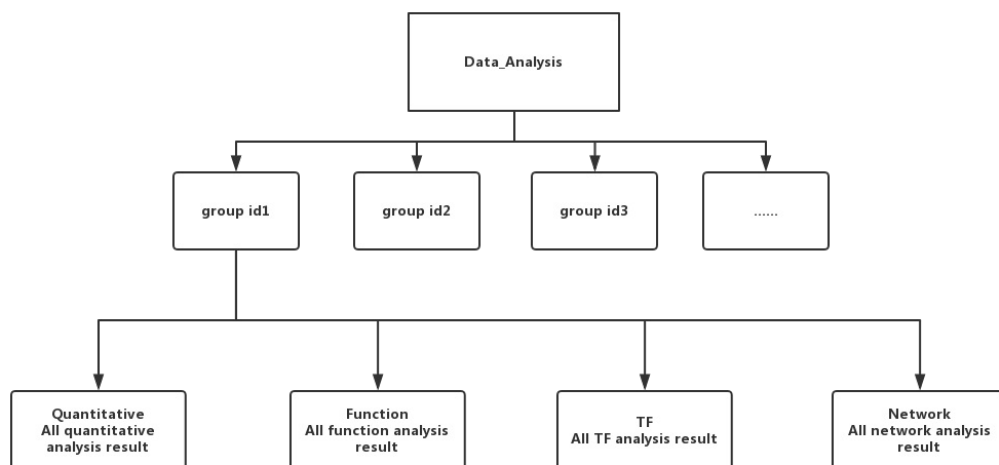
NR_Species_type	根据附件2填写NR数据库中物种的分类
AnnotationMemory	功能富集提交任务所需内存
TF_Memory	转录因子提交任务所需内存
Whether the transcriptome ko file exist	转录组是否存在ko文件
*root_GO	GO数据库路径
*root_KEGG	KEGG数据库路径
Network	关联网络配置说明
NetworkMemory	网络分析提交任务所需内存
String_Species_type	根据附件3填写网络分析物种名
*root_database	String数据库路径

2 流程文件与文件夹说明

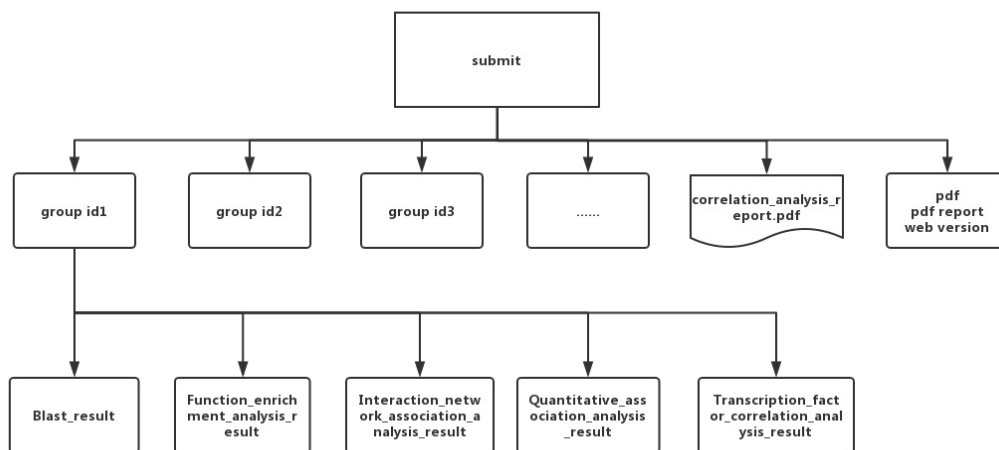
2.1 以比较组为名字的文件夹



2.2 Data_Analysis文件夹



2.3 submit文件夹



3 流程代码解析

表3-1 流程代码0-6步功能

step	功能
0	将每个比较组分别拷贝到对应文件夹，生成对应的Data_Analysis文件夹与submit文件夹，程序初始化
1	进行定量分析Blast,关联网分析Blast，【转录因子分析Blast】（可选）
2	进行定量分析，生成关联火山图，饼图，散点图
3	进行功能富集分析，生成功能火山图，功能气泡图，功能关联图，Pathway整合分析通路图
4	进行关联网分析，生成对应Pathway通路互作网络图，PPI关联互作网络图
5	转录因子分析，生成TF饼图，TF条形图，TF小提琴图，TF卡方检验统计表
6	生成pdf报告，【删除流程中间非必需文件】（可选）
