

Demo报告讨论

20250415

数据上传部分

1. 上传文件类型

- ☐ 上传的是搜库后的peaktable，如果按目前的搜库方案，在搜库阶段就会有group表，所以这里其实可以省去down下来重新传（客户自己提供peaktable除外），目前直接传搜库的group表还是很丝滑。
- ☐ 比较的组别，是否可以选多个比较方案？比如项目有ABC三个组，A是control，BC是相同突变体不同处理，需要对比AB，AC，BC，是否可以在这里同时选择？因为要重新对比其他组合，还要重新再来一遍传数据流程。
- ☒ 报错位置：在传完POS的peaktable和group后，一再传NEG，会容易出现POS表混乱的情况，只能刷新网页重新上传。

2. 数据预处理

这块内容，建议根据我们目前的peaktable和计算过程来填充默认值，客制化选择保留

- ☐ 缺失值类型：0
- ☐ 填充方式：均值
- ☐ 归一化方法：SUM
- ☐ RSD取log：None

结果部分

1. 数据总览（网页）

- ☐ 这里缺少一个信息，即peaktable的feature数和鉴定统计，只放了直接对比的，比如想知道这个项目正离子和负离子一共打到多少个feature，多少个能搜到名字的。
- ☐ 这里的raw_feature_num和remove_missing_num没有变化，和后面的preprocess_feature_num有啥区别？

2. 数据预处理

- ☐ （网页）目前点击这里的链接是提示404

2 数据预处理

- >打开缺失值过滤数据表_POS
- >打开缺失值过滤数据表_NEG
- >打开缺失值填充数据表_POS
- >打开缺失值填充数据表_NEG
- >打开归一化数据表_POS
- >打开归一化数据表_NEG
- >打开RSD过滤数据表_POS
- >打开RSD过滤数据表_NEG

- ☐ 在下载的结果表格里，每个表的第一列缺少表头，这里用的是啥ID？

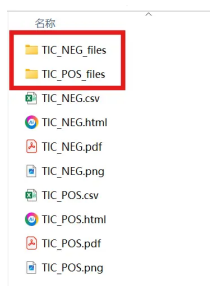
	A	B	C	D	E	F	G
1		RP.POS.1	RP.POS.10	RP.POS.100	RP.POS.101	RP.POS.102	RP.POS.10
2	7	18152	20112	25150	16772	16225	16273
3	76	3230	2739	2777	374	1898	18
4	176	23962	11921	17840	15383	22962	17752
5		9971	4368	6611	5612	8142	5467
6	185	963	557	401	751	1225	1305
7	203	202	155	169	104	114	8

□ 数据预处理提供4个文件夹，这3个里面表格缺失

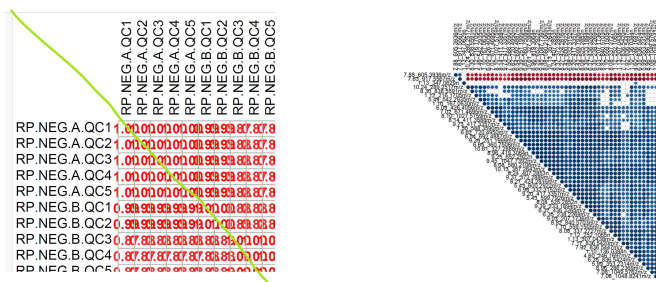


3. QC质控

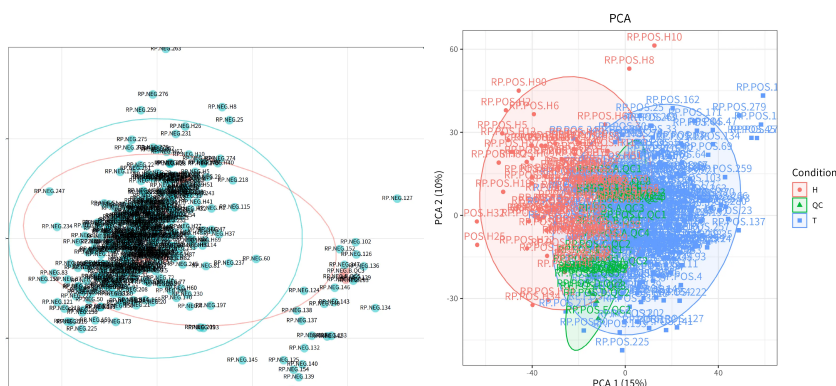
- 首先针对所有results（down下来的），建议按分析顺序编号，比如1.TIC，2.corr，3. RSD，4. PCA，5. PLD-DA。其他文件夹也是如此。
- results文件，TIC里面这两个文件夹是tmp文件？可以不要放在结果里。



□ corr，方框里建议不要显示数字，换成圆圈，另外整张图可以换成对角线切割，不用显示正方形，这样看起来不会很复杂



□ PCA有个问题，如果把所有样本归成一个组，再来算跟QC的PCA，跟按实际组来分后一起算的结果会有不同，建议直接分组计算，不然QC差异会被拉得很大，比如后面多元统计的这个PAC图，同样的数据，差异大小完全不同。反而多元统计的时候，不用放QC，只需要放样本分组的数据。



- PCA的图，如果这里只是要凸显QC的离散程度，样本不用打名字，不然根本看不清：
- QC的PLS-DA，我这边不太明白为啥要做，因为给不了什么参考信息

☐ QC和整体数据质量，建议增加一张图，可作为数据质量控制依据：

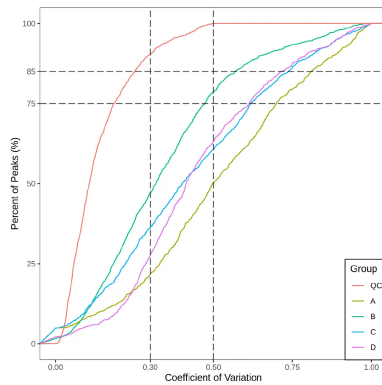
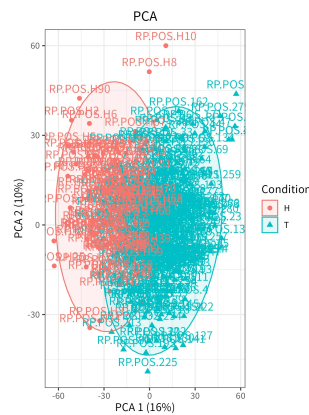


图 10: 各组样本 CV 分布图

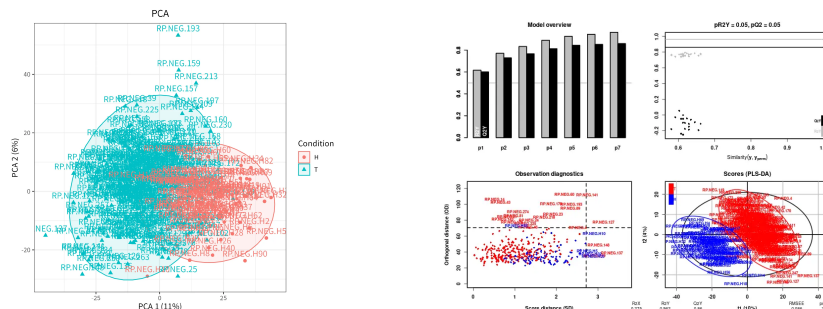
注：横坐标代表 CV 值，纵坐标表示小于对应 CV 值的物质数目占总物质数的比例，不同颜色代表不同的分组样本，QC 为质控样本，其中与 X 轴垂直的两条参考线对应的 CV 值为 0.3 和 0.5，与 X 轴平行的两条参考线对应物质数目占总物质数的 75% 和 85%。

4. 多元统计分析

☐ 组间对比PCA，首先建议在图里标注好POS or NEG，因为所有legend都一样，区分不了，其次这个图的比例可否调整为正方形？视觉上纵坐标会差异很大，其实纵坐标权重反而小；另外，点的名字，如果在置信区间里面是否可以不写？只标注离群点的名字即可：

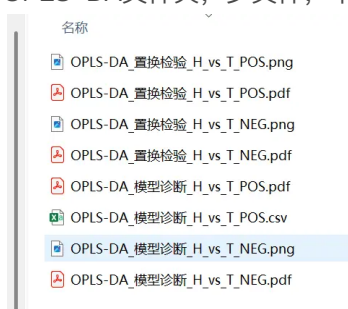


☐ PLS-DA，1. 首先同样建议图里面标好POS or NEG；2. 去除QC组，QC只放在QC部分，其他后面所有分析不要带；3. 结果里面的PLS-DA_H_vs_T_POS，依然是PCA结果，并不是PLS-DA结果，并且名字直接标的PCA；4. 需要的是，PLS-DA建模后，右下角的图，数据提取出来，重新画。



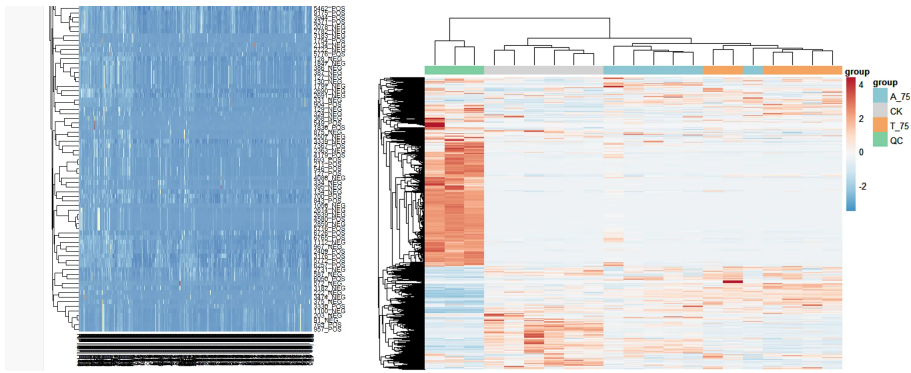
☐ PLS-DA诊断模型，少文件，没有POS的诊断模型png；

☐ OPLS-DA文件夹，少文件，不全；跟前面一样，不要放QC：



5. 差异筛选

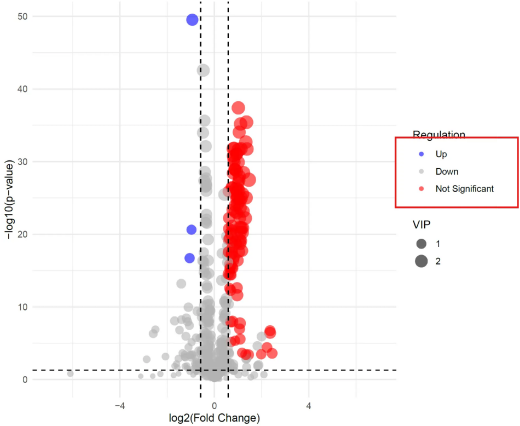
- ☐ HCA：这里作图的数据是什么数据？有做z-score运算吗？
- ☐ HCA：不建议显示很纵坐标名，另外做的时候需要分组：



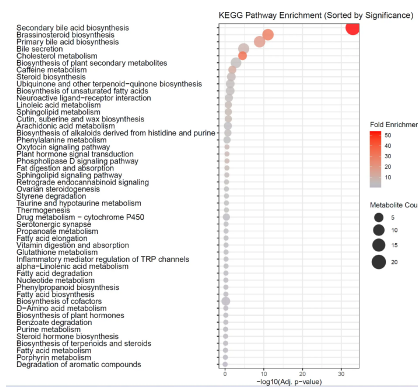
- ☐ HCA: HCA_heatmap_metabolite是什么？如果是用有name的，首先统一命名方式，其次，热图不用写名字，会给表的，而且name这里的合并逻辑是？ peaktable应该是一个feature对应一个name的可能存在不同feature对应1个name，但没有一个feature对应多个name：

	A	B	C	D
1		RP.ion.H1	RP.ion.H10	RP.ion.H11
2	2096_NEG_Glycocholic acid;Glyco-gamma-muricholic acid;Glyco-beta-muricholic acid	-0.1981716	-0.1977052	-0.2566108
3	2709_NEG-Tauro-beta-muricholic acid;Tauro-gamma-muricholic acid;Taurocholic acid	-0.1855516	0.02182371	-0.1821638
4	1936_NEG_Glycodeoxycholic acid;Glychenodeoxycholic acid	-0.3818028	-0.5865649	-0.1570776
5	4367_POS_Glychenodeoxycholic acid;Glycodeoxycholic acid	-0.464802	-0.6749387	-0.1478228
6	2540_NEG-Taurochenodeoxycholic acid;Taurodeoxycholic acid;Taurohyodeoxycholic acid (THDCA)	-0.3961696	-0.1606353	-0.2865981
7	5241_POS-Taurohyodeoxycholic acid (THDCA);Taurodeoxycholic acid;Taurochenodeoxycholic acid	-0.3418401	-0.2578456	-0.3528045
8	2098_NEG_Glycocholic acid;Glyco-gamma-muricholic acid;Glyco-beta-muricholic acid	-0.6259822	0.06819966	-0.0342954

- ☐ corr： 1. feature上次说了不用全放，只放系数最大的前50或者前20即可，其他的数据有提供表格； 2. 带name的（文件名要统一）也是，大于20的放20，小于20的放全部；
- ☐ corr： sample的相关性是啥意思？这个是怎么算的？
- ☐ VIP： 文件夹里面的表格，只给了VIP>1且有name的结果，这里不对，应该是所有feature的VIP值表；同样，distribution数量统计，需要统计所有feature的VIP分布
- ☐ volcano： 1. 少VIP筛选标准的火山图，这个是跟提交的筛选标准有关还是就不用？ 2. 火山图的legend错误。



- ☐ 差异代谢物表达量，这里面的两个图都没看明白
- ☐ 差异代谢物网络图：需要提供所有feature的，和有name的化合物的，目前只画了有name的
- ☐ KEGG富集： 1. 气泡图，建议调整画图用参数，把很坐标改成fold enrichment，颜色用-log10P，大小为化合物数量； 2. 通路名右对齐。



富集，增加MESA分析

20250427

网页结果展示部分

1. 数据总览

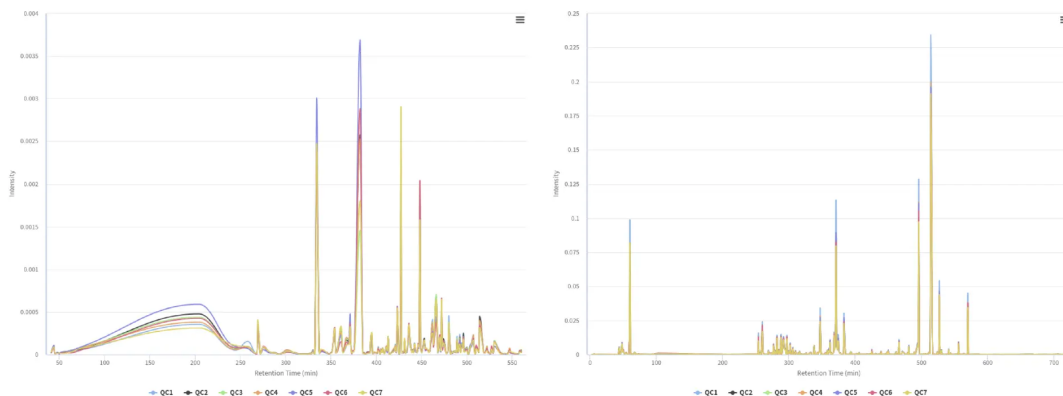
第二张表，显示组间对比上下调的统计表，可以只放POS&NEG，不用把两种离子模式单独列出，因为上下调筛选后的富集分析，用的是正负合并的结果，并不会针对单离子模式做分析。

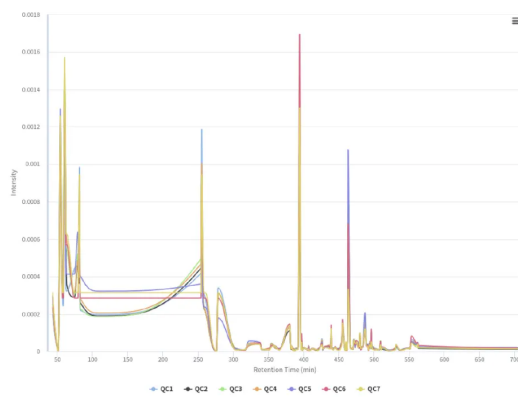
2. 数据预处理

点进去有表了，但是网页显示巨卡，是否可以换成下载链接，点击直接跳转下载？

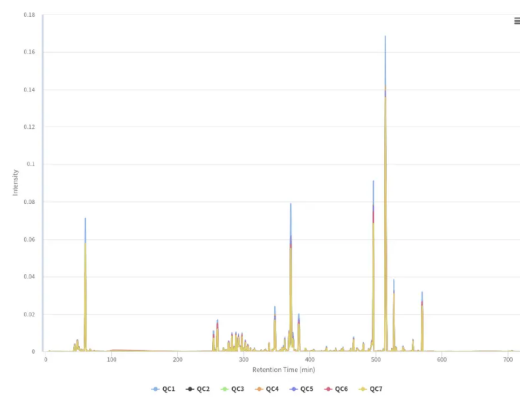
3. QC质控

a. 两次的的数据应该是一样的，为啥两次的QC TIC图不一样了？





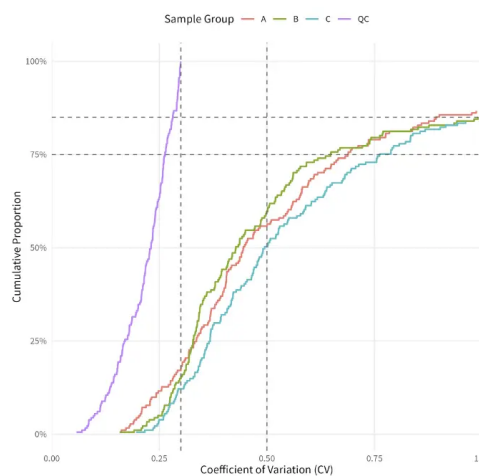
NEG模式总离子流图



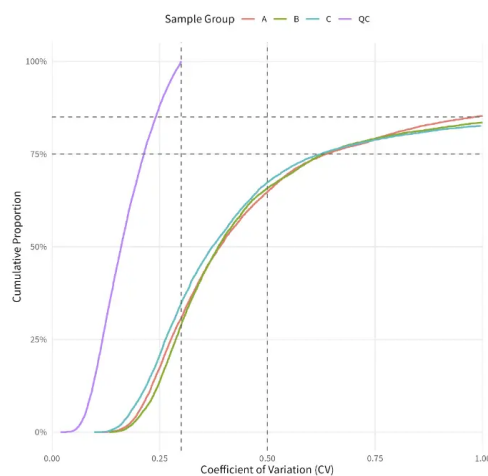
POS模式总离子流图

- 新加的这张CV图，纵坐标和横坐标加上去的虚线要在对应位置标出数值；
- CV的负离子数据是否有问题？为什么线条是锯齿状？
- CV图可以在QC的纵坐标值达到100%线条的时候，继续按横坐标填充100%的线条值，使整个图比较像抛物线。

3.5 QC和所有样本的CV分布



NEG模式CV图



POS模式CV图

4. 多元统计分析

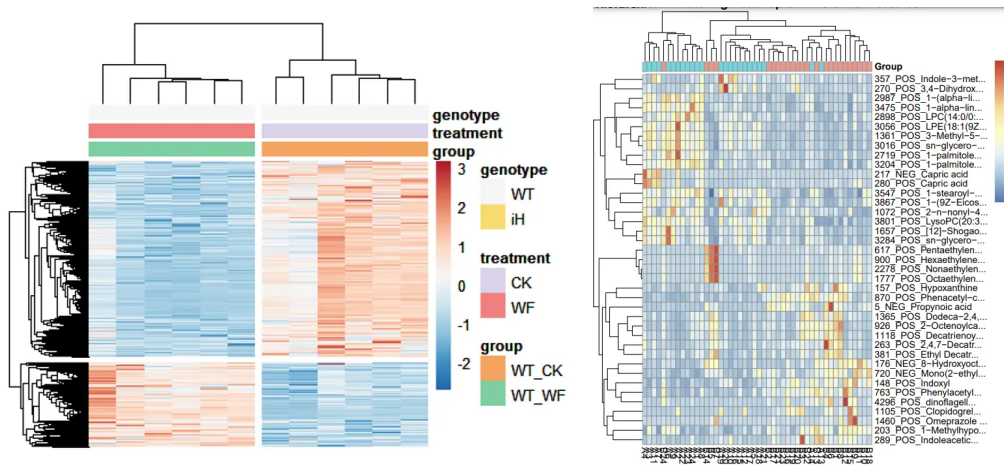
- 左侧目录栏有个小bug，第4部分显示了5.1的内容



- 这部分少了组间对比的OPLS-DA分析的图

5. 差异筛选

- all这一块，可以放到分组差异结果的后面
- 同样，链接打开的表格都是巨卡（分组差异结果表格），可以更换为下载链接
- 差异筛选后的热图，这里的筛选标准和数据处理方式是什么？上次也问过这个问题，因为展示出来的图不像是经过差异筛选的，比如同样是用差异筛选的feature作图，理论上两个组应该分的很开，像左边的这样，而现在感觉筛了没筛一样，没有用FC来筛？



d. 分组差异结果这部分，每个结果没有用标题隔开，现在是都放在一起的，建议隔开

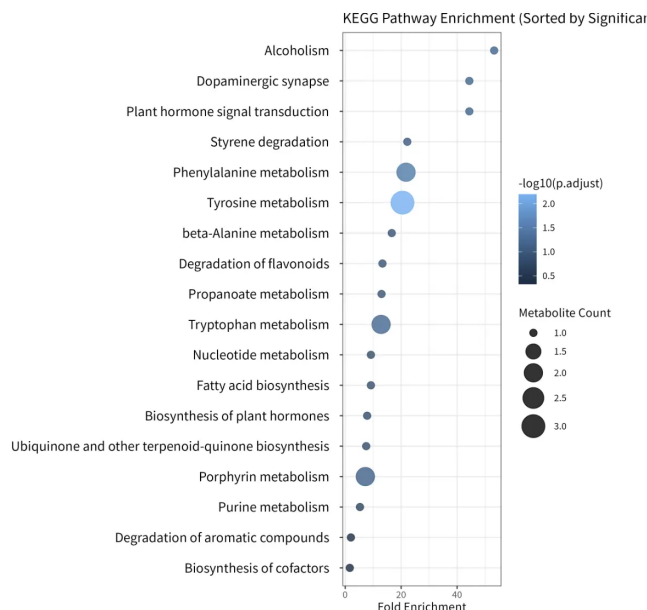
e. VIP下面的表，应该是做KEGG富集用的，表头需要更改一下，目前写的是gene

Show entries

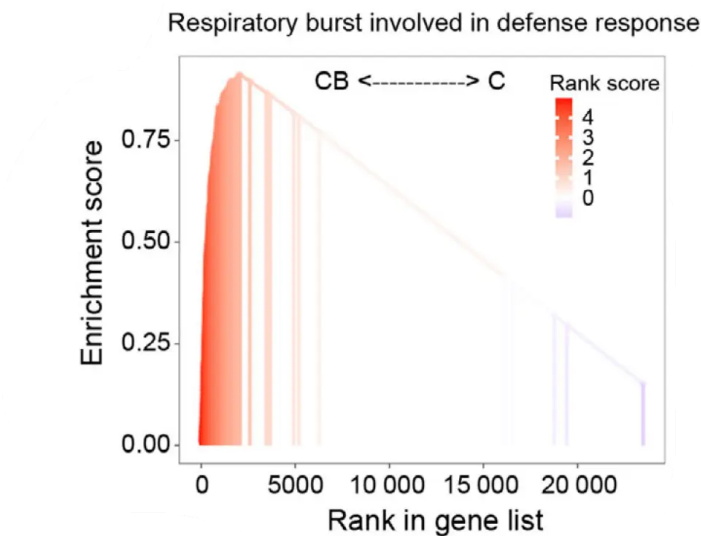
Search:

ID	Description	GeneRatio	BgRatio	RichFactor	FoldEnrichment	zScore
map01240	Biosynthesis of cofactors	1/12	328/6390	0.00304878048780488	1.6234756097561	0.502820799334497
map01220	Degradation of aromatic co...	1/12	263/6390	0.00380228136882129	2.02471482889734	0.736074486645674
map00230	Purine metabolism	1/12	101/6390	0.0099009900990099	5.27227722772277	1.87712902685598
map00860	Porphyrin metabolism	2/12	148/6390	0.0135135135135135	7.19594594594595	3.3078164713226
map00130	Ubiquinone and other terpe...	1/12	71/6390	0.0140845070422535	7.5	2.3888185506398

f. KEGG气泡图，count数量和圈圈大小最好成比例，目前数量只差一两个但是圈圈大小看起来差很多倍



g. MSEA分析的逻辑是什么？类似于GSEA的“排序→富集分数计算→显著性检验”？如果按GSEA的逻辑，即可以得到每个通路里面打到的化合物在该通路里面的“排序”（rank score），是否可以做类似GSEA的图：



下载的数据部分

1. 数据预处理

i. 缺失值过滤 + 缺失值填充

这里是填充的计算方式是？为什么不同feature填充的数值是一样的？

P	Q	R	S	T	U	V	W	X	Y	
A22	A23	A24	A3	A4	A5	A6	A7	A8	A9	B1
18860.9818	18641.1009	17849.4801	23104.7053	12159.045	20322.2943	16762.0689	16448.87	17408.2535	17578.4234	1
844.476868	1018.55923	1668.37933	2331.77795	2429.19769	1207.84755	1832.45564	1051.52639	1790.26968	3079.51622	26
6966.29343	4743.53787	2923.16537	5131.81796	5867.3189	3760.03773	6101.57137	3309.05342	6549.63424	6617.26346	31
10873.0228	16052.1532	11036.1773	8582.33984	11606.725	16403.4254	9736.59937	19383.1265	18899.7278	22089.9457	14
20272.1373	8602.91935	12763.3017	12491.0633	26245.8064	17555.0763	17543.4271	10859.6707	14736.4256	20385.1766	14
60820.4709	8186.70688	3839.0563	7935.96668	54386.5299	7201.61827	5530.62377	29795.0836	37952.0967	65746.128	67
4661.15056	12096.347	7285.17278	13521.5851	4002.21734	15578.6133	8496.5448	4283.163	6085.06535	19401.6714	65
9545.94746	10660.2387	1041.83597	1753.09383	9430.13519	1878.08374	3516.94288	3622.1067	6435.19525	2193.15933	12
2332.72085	5731.21366	950.61848	1676.99677	2046.12942	5056.60244	835.85923	1734.51621	12197.815	3691.221	16
5723.28817	6718.80521	8315.48275	3673.21434	4007.91372	3457.35101	3824.75984	3628.95725	8495.46704	3541.24755	17
92.2449099	384.628013	185.834623	599.137386	39.6927683	347.85184	592.155513	36.2777909	132.643616	707.74456	72
311.556631	1222.40466	3037.68282	824.77964	283.232594	2575.11777	477.423074	315.481837	267.604847	308.958819	33
100.468737	482.691018	469.716611	315.499251	64.2576787	754.736677	285.499822	223.630615	1070.22494	86.6899831	10
2951.08073	1936.45233	2175.92022	1573.88459	2371.13257	2295.78424	1522.08468	1441.4495	1363.55051	2328.22912	13
58131.4444	50765.5704	38330.2255	31935.1061	56423.9278	38640.5027	28601.868	54840.2098	78889.6467	64730.3583	45
21634.4169	7915.777	9548.22532	23581.7173	21634.4169	15857.8168	9597.26123	10230.7088	16287.5204	21634.4169	82
28911.1469	42079.9936	37068.8476	52117.0973	38100.99	37927.3126	34560.0151	35753.9665	86739.7574	57620.6514	17
357.600543	371.445148	440.233508	532.407484	510.585384	513.198687	464.129008	479.824688	269.930679	376.235607	81
582.678163	481.374167	444.803224	3028.41203	4190.74533	774.981811	750.171264	3883.42016	2412.79566	870.257495	74
319.692608	228.752645	367.4749	271.219615	409.544645	328.270289	301.984029	267.368492	164.825037	293.826041	72
6728.21382	5053.92584	2940.94956	16665.2116	13799.8193	24529.2659	15961.8419	16554.4223	12284.5178	45257.9944	89
582.859551	691.809016	2687.47307	766.950914	231.237128	432.864963	307.486602	621.126604	321.91505	722.889747	84
491.226149	628.189681	87.6753404	6214.73713	263.660213	13730.0853	40118.3744	1251.02174	484.773273	3463.06853	20
11315.2803	14145.7059	8553.53326	27263.0256	11029.2256	11706.3175	7661.72457	14406.1992	21705.7288	21634.4169	14
249.909804	2700.28287	446.471458	2089.81867	145.742294	2217.44747	2058.43284	4391.42839	1707.74871	16005.6882	30
530.498304	509.12569	185.678893	314.771551	346.083504	318.725	504.487615	212.029911	166.354126	348.661613	41
19131.7439	15320.6944	5833.74785	614.751832	8323.59149	11419.4717	17214.9478	20002.8602	27059.6799	22453.4969	83
20200.177	18139.9519	16882.8478	894.531896	5212.23378	8294.54786	17987.5	9764.57676	12123.0371	14668.0607	14
268533.144	249722.599	124314.207	61765.8888	104763.023	116642.118	171064.875	169703.017	113380.687	123953.143	21
4000.3403	4030.05005	5064.07434	3544.35403	4324.43120	2446.47400	2844.83547	7404.13200	8067.75700	4210.03257	44