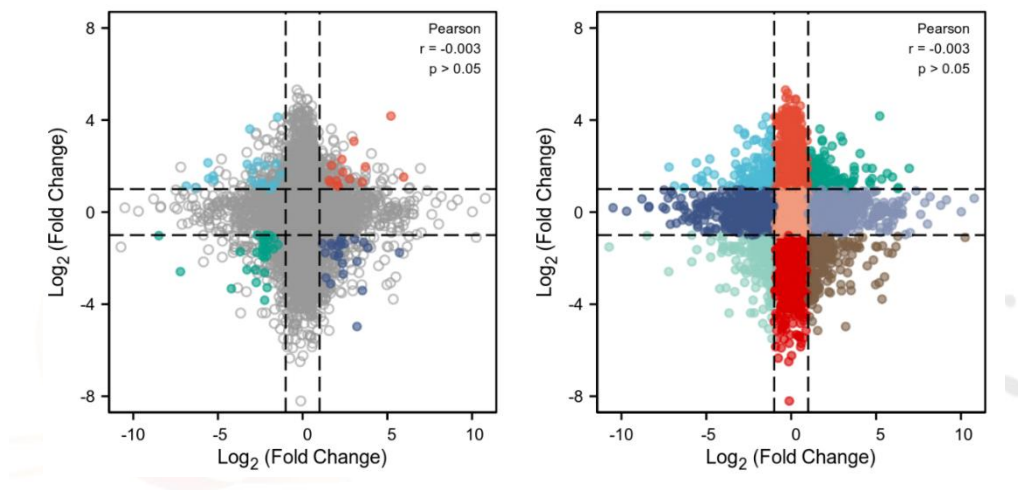


表达差异 -九象限图



网址: <https://www.xiantao.love>



更新时间: 2023.09.28

目录

基本概念	3
应用场景	4
结果解读	5
数据格式	6
参数说明	7
统计方法	7
阈值	8
点	9
分子 ID 标注	10
相关性标注	11
标题	12
坐标轴	13
风格	14
图片	15
结果说明	16
主要结果	16
补充结果	17
方法学	18
如何引用	19
常见问题	20

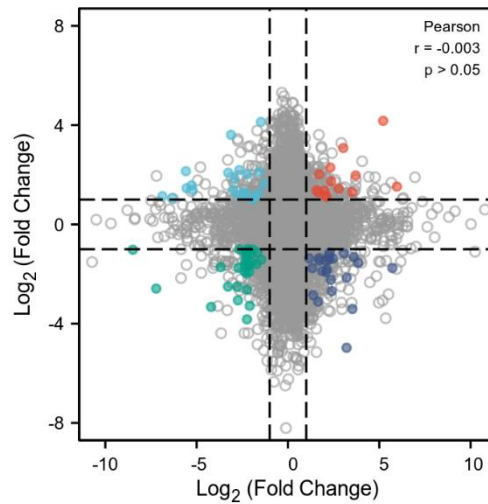
基本概念

九象限图：散点图的一种，在九象限图中，横轴和纵轴分别表示两个组学的表达水平变化（logFC）情况。用黑色虚线，从左至右、从上至下，依次分为 1-9 个象限。

一些问题：

- **Fold change**：差异倍数，简单来说就是分子在一组样品中的表达值的均值除以其在另一组样品中的表达值的均值。
- 为什么要做 **Log 2** 转换？两个数相除获得的结果（fold change）要么大于 1，要么小于 1，要么等于 1。对于基因差异，简单说，大于 1 表示上调（可以描述为上调多少倍），小于 1 表示下调（可以描述为下调为原来的多少分之多少）。大于 1 可以到多大呢？多大都有可能。小于 1 可以到多小呢？最小到 0。用原始的 fold change 描述上调方便，描述下调不方便。绘制到图中时，上调占的空间多，下调占的空间少，展示起来不方便。所以一般会做 Log 2 转换。默认我们都会用两倍差异（fold change == 2 | 0.5）做为一个筛选标准。Log 2 转换的优势就体现出来了，上调的基因转换后 Log 2 (fold change) 都大于等于 1，下调的基因转换后 Log 2 (fold change) 都小于等于 -1。无论是展示还是描述是不是都更方便。

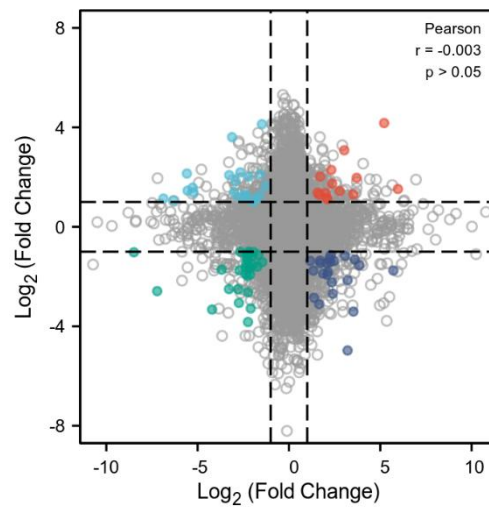
➤ 图形构成



应用场景

九象限图可直观地展示两种组学数据的相关性，帮助研究者快速定位到与研究相关的关键分子。比如转录组与蛋白组，通过九象限图可直观得到样本中基因的表达情况和基因表达后翻译的情况，帮助研究者快速定位到与研究相关的关键基因。

主要结果



- 图中横坐标代表横轴组学对应的 **logFC**，纵坐标代表纵轴组学对应的 **logFC** 值。图中每个点表示两个组学共同分子 ID。不同象限的颜色代表不同表达趋势，灰色表示非共同差异分子（即在两个组学都不差异和只有单个组学有差异的情况）。
- 整体来看，共同差异分子可能有相同的表达趋势，也可能有不同的表达趋势，图整体是以 $X=0$ 的垂线左右对称， $Y=0$ 的横线上下对称。点越偏离中心，表示在两个组学中的差异倍数都越大。
- 图的带颜色的两个组学共有差异点是我们关注的重点。

数据格式

	A	B	C		A	B	C
1	gene	logFC	padj	1	protein	logFC	pvalue
2	A1BG	-2.69328	0.009398	2	AC005041	0.30771	0.727498
3	A1CF	-0.93186	0.336871	3	AC007663	-0.36905	0.378184
4	A2ML1	-4.71165	0.055393	4	AC008132	4.458616	0.00025
5	A4GALT	1.792873	0.089303	5	AC008764	-0.73399	0.062908
6	A4GNT	-4.94415	0.060014	6	AC009533	0.47112	0.313581
7	AAAS	0.530048	0.634438	7	AC010323	-4.0775	0.302567
8	AACS	2.016416	9.04E-05	8	AC010531	0.416454	0.918709
9	AADAC	-0.21983	0.875698	9	AC011511	0.539225	0.121866
10	AADAT	-4.23047	3.26E-05	10	AC062028	0.973721	0.013241
11	AAGAB	0.898257	0.26768	11	AC069335	-1.99827	0.000281
12	AAK1	0.962854	0.162753	12	AC073896	1.400237	0.002188
13	AAMDC	-0.31294	0.858966	13	AC087235	1.3886	0.730501
14	AAMP	-0.06671	0.95575	14	AC090152	0.818298	0.134406
15	AANAT	2.054735	0.423973	15	AC091390	-1.7988	0.156625
16	AAR2	0.333749	0.720683	16	AC093323	0.602813	0.330959
Sheet1				Sheet2			

表格：两个组学的差异分析结果

- 两个 sheet 表。
- 每个表3列数据: ID 列 | logFC 列 | p 值或者校正后 p 值(p.adj)列 (注意列名).
数据会过滤掉 logFC 或者是 p 值列有缺失的数据。
- 每个表至少 3 行，至多 70000 行。
- 两个 sheet 表的 ID 列至少有 3 个共同分子 ID。
- 若验证数据时返回报错，需要对上传数据进行相应的调整，然后再上传数据。

参数说明

(说明：标注了颜色的为常用参数。)

统计方法

统计方法	▼
相关性统计方法	Pearson ▼

- 统计方法：可选择在图中标注两个组学匹配上的分子 ID 对应 logFC 相关性方法。



阈值

阈值 

横轴logFC阈值 1

纵轴logFC阈值 1

横轴p值阈值 0.05

纵轴p值阈值 0.05

- 横轴 logFC 阈值：可以控制图中划分是否有显著差异表达对应的 logFC 阈值
- 纵轴 logFC 阈值：可以控制图中划分是否有显著差异表达对应的 logFC 阈值
- 横轴 p 值阈值：可以控制图中划分是否有显著差异表达对应的 p 值阈值
- 纵轴 p 值阈值：可以控制图中划分是否有显著差异表达对应的 p 值阈值

点

点

填充色

▼

▼

▼

▼

▼

▼

▼

▼

描边色

▼

▼

▼

▼

▼

▼

▼

▼

样式

圆形

▼

大小

1

不透明度

0.6

差异点

☒

- 填充色：点的填充色
- 描边色：点的描边色
- 样式：可以选择圆形、正方形、菱形等形状
- 大小：点的大小
- 不透明度：0 为完全透明，1 为完全不透明。
- **差异点**：是否突出展示两个组学共同差异分子（差异筛选条件即阈值参数设定的范围）。默认展示。

分子 ID 标注

分子ID标注 ▼

标注id(数据第1列) ?
 可以输入想要标注的id,
 1行1个 ✎

标注大小 5pt ▼

标记数量 🔴

- 标注 id（第一列）：可以标注想要标注的[共同分子 ID](#)
- 标注大小：默认是 5pt
- 标记数量：可以在图中展示每个象限的分子 ID 数量。[参数 点-差异点](#)开启时，标记的是[共同差异分子 ID](#) 个数及其所在象限。[参数 点-差异点](#)关闭时，标记的是两个组学能够匹配上的所有[共同分子 ID](#) 个数及其象限

相关性标注

相关性标注 

是否展示 

标注位置 右上 

字的大小 5pt 

- 是否展示: 可以选择是否展示匹配上的分子 ID 对应两个组学 logFC 的相关性, 默认展示
- 标注位置: 相关性的标注位置可以选择右上, 右下, 左上, 左下。默认右上
- 字的大小: 默认是 5pt



标题

标题	
大标题	大标题内容
x轴标题	x轴标题内容
y轴标题	y轴标题内容

- 大标题：大标题文本
- X 轴标题：X 轴标题
- Y 轴标题：Y 轴标题

补充：在要换行的中间插入\n。如果需要上标，可以用两个英文输入法下的大括号括住，比如 {{2}}；如果需要下标，可以用两个英文输入法下的中括号括住，比如 [[2]]

坐标轴

坐标轴	
x轴范围	逗号隔开
y轴范围	逗号隔开

- X 轴范围：可以控制 x 轴范围和刻度，提供 2 个值来控制范围。形如 $-1, 1$
(注意不要调整的过大或过小)
- Y 轴范围：可以控制 y 轴范围和刻度，提供 2 个值来控制范围。形如 $-1, 1$
(注意不要调整的过大或过小)。



风格



- 边框：是否显示主图边框，默认显示
- 网格：是否添加网格
- xy 颠倒：x 轴和 y 轴是否转换
- 文字大小：图中的文字部分的大小（包括标签文字和刻度数），默认是 7pt

图片

图片	▼
宽度 (cm)	6
高度 (cm)	6
字体	Arial ▼

- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体



结果说明

主要结果

主要结果 补充结果 方法学

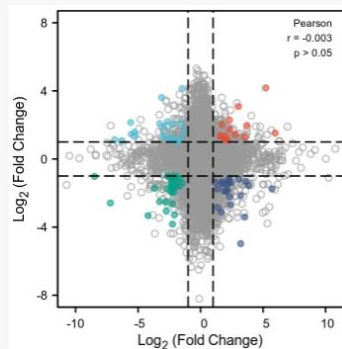
保存结果

下载整份报告

九象限图

九象限图: 可视化有ID关联的多组学差异分析结果。以转录蛋白联合分析为例, 每个蛋白或基因在图中都表示为一个点, 可以使用不同的颜色或形状来区分不同的变化趋势组。通过对图表的观察, 可以快速确定转录组和蛋白组之间的关系和趋势, 以及特定基因或蛋白质的表达模式。

横轴阈值: $\log_2(1)$ 纵轴阈值: $\log_2(1)$



九象限图.pdf

九象限图.tiff

共同差异.xlsx

主要结果格式为图片格式, 提供 PDF、TIFF 格式下载, 结果报告可以下载包括 pdf 以及说明文本的内容。

	A	B	C	D	E	F	G
1	ID	table1_logFC	table1_padj	table2_logFC	table2_pval	part	diff
2	ACAT1	-1.61264097	0.04199792	1.25949988	0.000058	part1	opp
3	ACVR1C	-3.01060672	0.01285956	1.39316913	0.00148103	part1	opp
4	ADGRA3	-2.26189095	0.00642082	-1.08820281	0.00729434	part7	homo
5	ALDH4A1	-1.9400242	0.00385163	-1.01504403	0.03704756	part7	homo
6	ARHGEF10L	-1.46232602	0.0473516	-1.4221217	0.04132219	part7	homo
7	ARHGEF39	2.13689013	0.00297122	-1.82006336	0.02944666	part9	opp
8	ARMC7	1.7048989	0.03132187	-1.35958187	0.00377592	part9	opp
9	ASF1B	1.65251703	0.02661198	-3.10819965	0.00723867	part9	opp
10	ATG2B	-1.82658224	0.03282881	1.11245996	0.02046705	part1	opp
11	ATP7B	-2.26892597	0.01045711	-1.9807155	0.00355556	part7	homo
12	AURKA	2.21255098	0.00143075	-1.28118585	0.00591677	part9	opp
13	BANF2	5.71742098	0.0149591	-1.75997115	0.00000544	part9	opp
14	BICDL1	3.06037223	0.0087768	-1.16918366	0.00811063	part9	opp
15	BYSL	1.70772714	0.01596681	2.02576386	0.00712227	part3	homo
16	C1R	-2.21032123	0.00076567	2.02338318	0.02349285	part1	opp
17	C1RL	-2.36945826	1.3198E-05	1.29715507	0.01507323	part1	opp
18	CCN1	-2.34710675	7.329E-06	-1.63767393	0.00035734	part7	homo
19	CCT6A	1.16820464	0.03494994	-1.36869606	0.00314282	part9	opp
20	CD38	-2.28259367	0.0205192	-1.30859826	0.00893968	part7	homo
21	CD69	-2.34663618	0.01803048	-1.498928	0.01865543	part7	homo

另外, 提供共同差异表下载, 是我们研究的重点。

补充结果

相关性分析

同时提供Pearson和Spearman统计方法，可以根据需要选择标注在图中

方法	统计量	自由度	P值	相关性系数	置信区间
Pearson	-0.30037	8334	0.7639	-0.0032903	-0.024756 - 0.018178
Spearman	9.67e+10		0.8812	-0.0016374	

相关系数为正，说明两个组学的logfc之间存在正相关关系；相关系数为负，说明两个组学的logfc之间存在负相关关系；
相关系数绝对值代表相关程度，0-0.3代表弱或者不相关；0.3-0.5代表若相关；0.5-0.8代表中等程度相关；0.8-1代表强相关
相关是否有统计学意义还需要结合p值来查看

此表格为共同分子 ID 在两个组学 logFC 的相关性统计结果，可选择是否标注在图中。

差异趋势统计

根据logFC和P值（或校正后的P）统计两个组学中的差异趋势

差异方向相同	差异方向相反
43	44

差异方向相同表示分子不仅在两个组学中同时符合logFC和P值（或校正后的P）的阈值筛选且都是上调或者都是下调（即图中第3，7象限部分点）；
差异方向相反表示分子不仅在两个组学中同时符合logFC和P值（或校正后的P）的阈值筛选但一个组学上调另一个组学下调（即图中第1，9象限部分点）。

此表格对应共同差异表中差异方向是否相同的统计。

方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包: `ggplot2`[3.3.6] 进行可视化。



如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 (www.xiantao love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



常见问题

1. 为什么有时画出的图不是 9 个象限?

答：如果输入的 logFC 阈值不正确或为零，那么无论表达水平如何，都不能正确划分为上调或下调，从而导致不足 9 个象限的情况发生。需要注意的是，具体设置 logFC 阈值的值应根据具体的实验设计、研究领域和数据集来确定，可结合相关文献进行选择。

2. 如果研究组学的差异筛选条件不仅是 logFC 和 P 值（或矫正后的 P）怎么办?

答：比如代谢组学中除了 logFC 和矫正后的 P 值，常以 VIP 值一起来筛选差异代谢物。如果是类似情况，可以自己先过滤掉不符合的 VIP 值所在行，用处理后的数据画九象限图。

3. 如何把不同组学的分子 ID 统一?

答：在将不同组学的分子 ID 进行统一时，可以考虑以下几种方法：

- 基于基因 ID：对于转录组学和蛋白组学等组学研究，蛋白 ID 通常存在对应的基因 ID。可以通过将两者的基因 ID 进行匹配，从而实现分子 ID 的统一。这样可以将蛋白组学数据映射到相应的基因上，使得不同组学层面之间的关联更加清晰。
- 使用公共数据库：利用公共数据库（如 UniProt、NCBI、Ensembl 等）提供的信息，根据分子的特征或描述进行关联和匹配。这些数据库提供了丰富的生物信息资源，可以帮助将不同组学层面的分子 ID 进行统一。
- 基于相关性进行匹配：如果在你的研究中存在已知的关联信息，比如转录组学数据中的基因表达与蛋白组学数据中的蛋白表达存在相关性，可以利用这些关联信息来对不同组学的分子 ID 进行匹配。例如，可以基于相关性分析不同组学的分子 ID 进行对应。

需要注意的是，在进行分子 ID 统一时，要确保选择合适的方法和策略，并对匹配结果进行验证和确认，以确保统一后的数据是准确、可靠的。

