

## 重要提示

您现在手上拿到的是欧易生物提供的 lncRNA 芯片的初步报告,项目结束以后,我们还将为您提供一次深入数据分析的服务,请将您还需要的分析要求汇总后,直接联系与您洽谈的销售代表,数据分析人员得到反馈后会尽快安排进一步分析工作。有任何疑问,请拨打公司总机 **021-38760130** 或服务热线 **4006-4008-26**,前台会为您转接至相关人员。

以下是欧易生物可以提供的深入分析的内容以及相应的分析要求,供您选择 (**请务必认真阅读并了解实现每项分析的前提条件**)。

## 目 录

1. 差异基因或 lncRNA 筛选.....	3
2. 差异整合分析(venn 分析).....	3
3. 关联分析.....	4
4. 聚类分析.....	5
1) 非监督层次聚类分析.....	5
2) 自组织映射聚类分析(SOM).....	6
5. 主成分分析(PCA) .....	7
6. GO 富集度分析.....	7
7. KEGG 富集度分析 .....	8
8. 蛋白网络分析.....	9

## 深入分析内容

### 1. 差异基因或 lncRNA 筛选

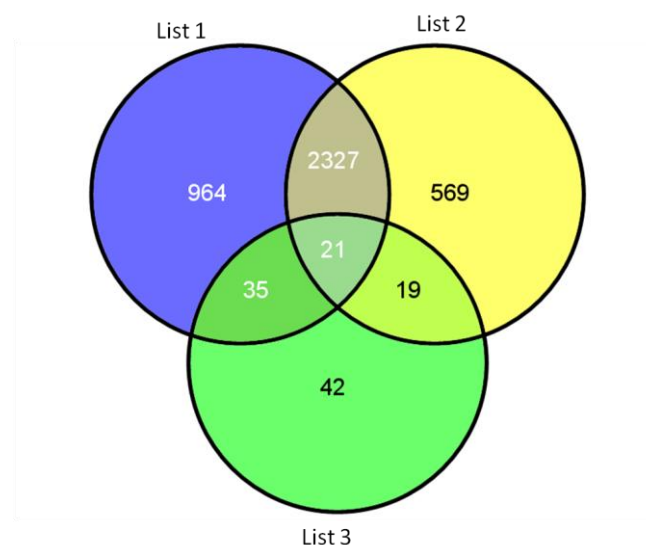
- 1) 在原有差异筛选的基础上，重新指定其他的实验组和对照组，进行差异筛选。
- 2) 改变芯片探针过滤条件和（或）差异倍数筛选条件对差异基因或 lncRNA 进行重新筛选。原有的芯片探针过滤标准和倍数筛选标准在初步报告“分析数据文件夹”下“**Agilent lncRNA 芯片分析文件查看说明**”中有详细说明。由于生物样本本身原因（个体差异比较大或者不明显）或者实验设计原因（比如没有设置生物学重复等），可能导致您项目中筛选得到的差异基因过多或者过少，这个时候您可以根据自己的需求对相应的阈值进行调整。

### 2. 差异整合分析(venn 分析)

对不同的比较进行交并集分析，找出几组比较之间共同或者各自特有的差异表达基因或 lncRNA。

**实现该分析的前提条件：需要您指定用于交并集分析的比较，每一次交并集分析仅针对 2 组以上(包括 2 组)、4 组以内(包括 4 组)比较的整合分析。**

结果形式：



结果说明:

差异整合分析	研究对象	基因个数
1	比较1的差异基因或lncRNA, 扣掉比较2、比较3中的差异基因或lncRNA	964
2	比较2的差异基因或lncRNA, 扣掉比较1, 比较3中的差异基因或lncRNA	569
3	比较3的差异基因或lncRNA, 扣掉比较1, 比较2中的差异基因或lncRNA	42
4	比较1与比较2中共有的差异基因或lncRNA, 扣掉比较3中的差异基因或lncRNA	2327
5	比较1与比较3中共有的差异基因或lncRNA, 扣掉比较2中的差异基因或lncRNA	35
6	比较2与比较3中共有的差异基因或lncRNA, 扣掉比较1中出现的差异基因或lncRNA	19
7	比较1, 比较2, 比较3中共有的差异基因或lncRNA	21

### 3. 关联分析

对指定的 lncRNA 和 mRNA 进行关联分析, 计算相关系数, 根据设定的阈值筛选 lncRNA 和 mRNA 关系对, 构建 lncRNA-mRNA 共表达网络。

**实现该分析的前提条件: 该项分析仅针对每组设置三个及以上生物学重复的实验, 并需要您指定用于关联的样本以及 lncRNA 和 mRNA。**

结果形式:

lncRNA-mRNA对					
lncRNA	pvalue	相关性系数	GenomicCoordinates	mRNA	PrimaryAcc
OE_P00800622	0.0070416	0.85335956	chr5:67704695-67704636	A_23_P100127	NM_170589
OE_P00807193	0.0052559	0.867473911	chr22:35836113-35836172	A_23_P100127	NM_170589
OE_P00316075	0.0022051	0.901668883	chr4:53579273-53579332	A_23_P100602	NM_005993
OE_P00317547	0.0012819	0.918284327	chr6:138144988-138144929	A_23_P100602	NM_005993
OE_P00319027	0.0023064	0.900146334	chr2:102067115-102067056	A_23_P100602	NM_005993
OE_P00321868	0.0038051	0.881435071	chr1:208039527-208039468	A_23_P100602	NM_005993
OE_P00322687	0.0020047	0.904819266	chr4:53617789-53617730	A_23_P100602	NM_005993
OE_P00326570	0.0053053	0.867045149	chr3:136792395-136792454	A_23_P100602	NM_005993
OE_P00327728	0.000657	0.934882391	chr6:111866669-111866728	A_23_P100602	NM_005993
OE_P00331767	0.0058382	0.862573555	chr3:158454371-158454430	A_23_P100602	NM_005993

结果说明:

将 lncRNA, mRNA 进行相关性分析, 筛选显著性 p 值小于 0.01 且相关系数大于等于 0.8 的 lncRNA-mRNA 对, 该筛选条件也可以根据您的需要进行更改。

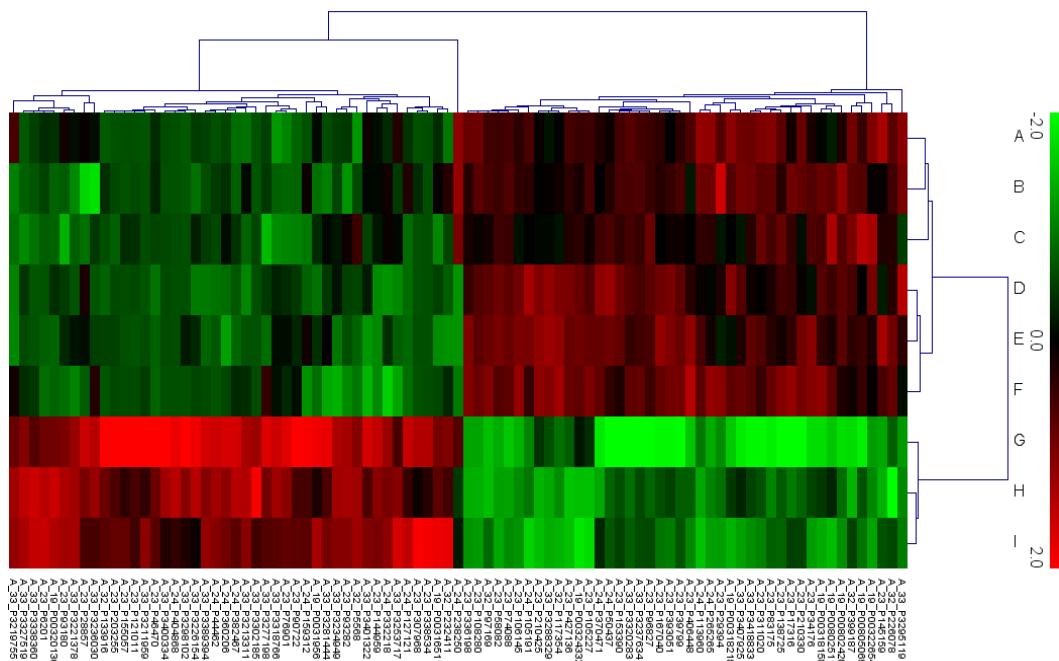
## 4. 聚类分析

### 1) 非监督层次聚类分析

计算多个样品两两之间的距离，构成距离矩阵，合并距离最近的两类为一新类，计算新类与当前各类的距离，再合并、计算，直至只有一类为止，用挑选的差异基因或 lncRNA 的表达情况来计算样品直接的相关性，一般来说，同一类样品能通过聚类出现在同一个簇中，聚在同一个簇中的基因或 lncRNA 可能具有相似的生物学功能。

**实现该分析的前提条件：需要您指定用于聚类分析的样本和基因或 lncRNA，每一次聚类包含至少 3 个以上(包含 3 个)样本，基因数量尽量不要超过 1000。我们建议您也可以自己利用聚类分析软件 cluster3.0 进行聚类分析，相应的安装程序以及软件具体的介绍和操作流程见“常用软件”文件夹内“cluster3.0 and treeview”文件。**

结果形式：



结果说明：

色阶表示基因表达量从相对低（绿）到相对高（红）变化。行名代表样本名称，列名代表探针名称。

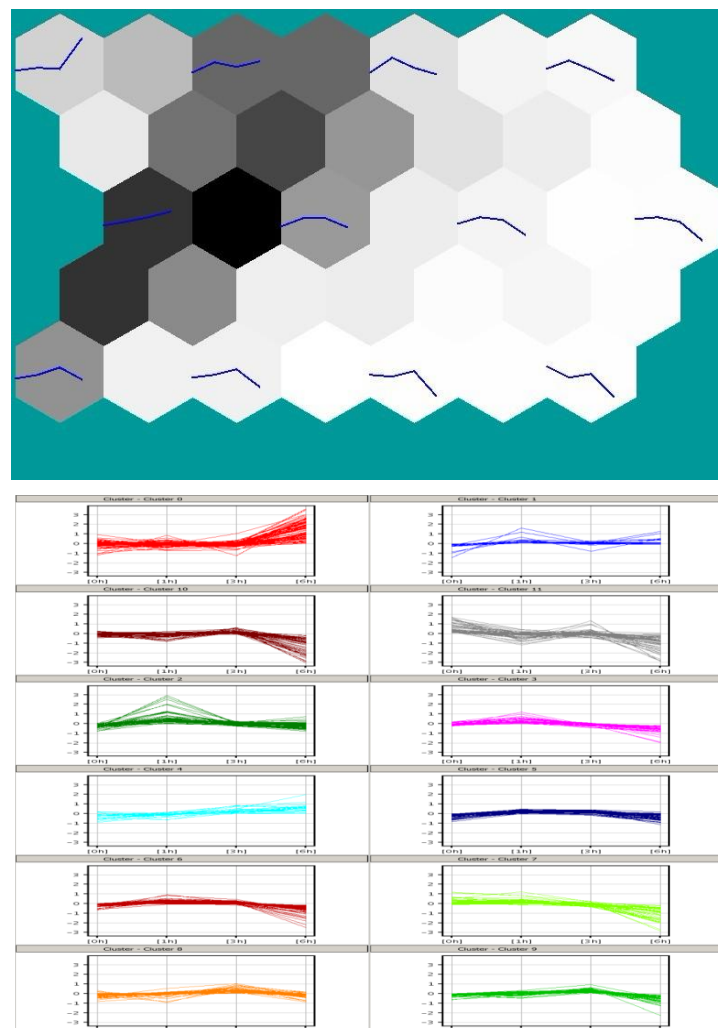


## 2) 自组织映射聚类分析(SOM)

自组织映射聚类是以神经网络算法为基础的一种分割性聚类方法。SOM 按照每一基因表达向量与各类中参照向量的相似性，将基因或者 lncRNA 分类，因此它与均值聚类的不同之处就是需要定义参照向量。在确定类的几何形状时，研究者就在有效地确定分类的数目。SOM 将高维的表达数据呈现在低维的空间中，结果便于解释。相邻节点的参照向量接近，结果图中保存了邻簇关系，更便于对结果进行调整。

**实现该分析的前提条件：该分析仅适用于设置了浓度梯度或者时间梯度的实验，需要您指定用于聚类分析的样本名称和顺序，每一次聚类包含至少 3 个以上(包含 3 个)样本。**

结果形式：



## 结果说明:

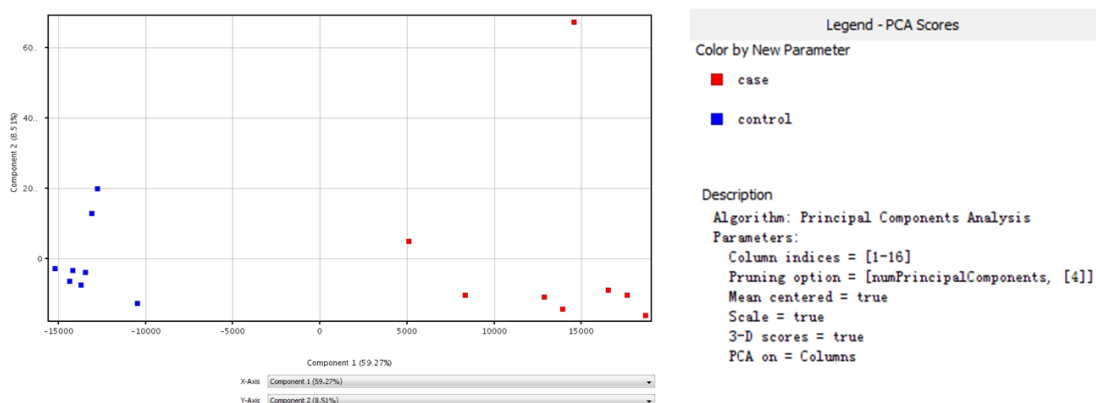
每个趋势用不同颜色的折线表示。每个趋势可以通过一个六边形同另外一个趋势连接。六边形的灰度表示相关性系数(correlation)的大小, 白色: correlation=1, 黑色: correlation=0, 灰度表示不同的相关性系数。通过连接相邻两个趋势的六边形的灰度可以了解它们之间的相关性系数。

## 5. 主成分分析(PCA)

通过 PCA 分析, 各组样本分布在二维空间的不同区域, 同组的样品在空间分布比较集中, 说明这些基因或者 lncRNA 选取具有代表性。

**实现该分析的前提条件: 需要您指定用于 PCA 分析的样本和基因或者 lncRNA。**

## 结果形式:



## 6. GO 富集度分析

通过 GO 分类号和 GO 数据库相关分析工具将分类与具体基因联系起来, 从而对这个基因的功能进行描述。在芯片的数据分析中, 研究者可以找出哪些变化基因属于一个共同的 GO 功能分支, 并用统计学方法检定结果是否具有统计学意义, 从而得出变化基因主要参与了哪些生物功能。

**实现该分析的前提条件: 在初步报告中已经对每一组差异筛选数据进行了 GO 富集度分析, 如果您需要对一些特定的基因进行 GO 富集分析, 请指定用于分析的基因。**

结果形式:

Terms

Rank	ID	Term	List Hits	List Total	Population Hits	Population Total	P-value	Genes ID's	Genes Names
1	GO:0055085	transmembrane transport	10	86	670	14200	7.13e-03	118980 140803	sideroflexin 2
2	GO:0007049	cell cycle	6	86	403	14200	3.54e-02	1104 23176	regulator of chromosome condensation 1
3	GO:0007608	sensory perception of smell	6	86	434	14200	4.79e-02	1813 219981	dopamine receptor D2
4	GO:0051301	cell division	5	86	262	14200	2.14e-02	1104 311	regulator of chromosome condensation 1
5	GO:0034329	cell junction assembly	3	86	89	14200	1.68e-02	10326 253559	signal-regulatory protein beta 1

## 7. KEGG 富集度分析

根据挑选出的差异基因，计算这些差异基因同 Pathway 的超几何分布关系，Pathway 分析会对每个有差异基因存在的 pathway 返回一个 p-value，小的 p 值表示差异基因在该 pathway 中出现了富集。Pathway 分析对实验结果有提示的作用，通过差异基因的 Pathway 分析，可以找到富集差异基因的 Pathway 条目，寻找不同样品的差异基因可能和哪些细胞通路的改变有关。

**实现该分析的前提条件：在初步报告中已经对每一组差异筛选数据进行了 KEGG 富集度分析，如果您需要对一些特定的基因进行 KEGG 富集分析，请指定用于分析的基因。**

结果形式:

Terms

Rank	ID	Term	List Hits	List Total	Population Hits	Population Total	P-value	Genes ID's	Genes Names
1	04110	Cell cycle	15	282	128	6790	2.49e-04	104394 11920	E2F transcription factor 4
2	04141	Protein processing in endoplasmic reticulum	15	282	171	6790	4.82e-03	12043 15467	ataxia telangiectasia mutated homolog (human)
3	04660	T cell receptor signaling pathway	15	282	112	6790	5.38e-05	12402 12525	cyclin A2
4	04062	Chemokine signaling pathway	14	282	187	6790	2.33e-02	109905 12540	cyclin D1
5	04510	Focal adhesion	14	282	200	6790	3.83e-02	109905 12043	proliferating cell nuclear antigen
6	04722	Neurotrophin signaling pathway	14	282	132	6790	1.09e-03	109905 12043	polo-like kinase 1 (Drosophila)
7	04380	Osteoclast differentiation	12	282	118	6790	3.48e-03	11433 14281	retinoblastoma-like 1 (p107)
8	04120	Ubiquitin mediated proteolysis	11	282	142	6790	3.36e-02	12402 15204	transforming growth factor, beta 1
9	04640	Hematopoietic cell lineage	11	282	83	6790	5.86e-04	12482 12494	CDC14 cell division cycle 14 homolog B (S. cerevisiae)
10	04210	Apoptosis	10	282	86	6790	2.82e-03	11920 12043	WEE1 homolog 1 (S. pombe)
11	05323	Rheumatoid arthritis	8	282	85	6790	2.42e-02	11433 11966	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide
12	04666	Fc gamma R-mediated phagocytosis	8	282	92	6790	3.67e-02	12540 13429	cell division cycle 6 homolog (S. cerevisiae)
13	04662	B cell receptor signaling pathway	7	282	76	6790	3.76e-02	14281 14784	origin recognition complex, subunit 5
14	05140	Leishmaniasis	7	282	65	6790	1.76e-02	14281 14969	anaphase promoting complex subunit 2



[www.oebitech.com](http://www.oebitech.com)

