1）问题模型：疾病类型、数据有多少、以及该论文中从事的那个数据挖掘问题研究

2）基本方法：给出该论文使用的算法是什么、及初步解释

3）结果描述：该论文得到的“好结果”是什么性能、如何与其他方法进行比较的

4）改进思路：你觉得如何进一步优化改进该论文的结果

Pedraza, V., J. A. Gomez-Capilla, et al. (2010). "Gene Expression Signatures in Breast Cancer Distinguish Phenotype Characteristics, Histologic Subtypes, and Tumor Invasiveness." Cancer **116**(2): 486-496.

乳腺癌病人:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 病人类型编号 | T分期 | N分期 | ER 类型 | 病人例数 |
| #Ph1 | T1,T2,T3 | N0 | ER阳性 | 8例 |
| #Ph2 | T1,T2,T3 | N0 | ER 阴性 | 10例 |
| #Ph3 | T1,T2,T3 | N+(N1,N2,N3) | ER 阴性 或ER阳性 | 13例 |
| #Ph4 |  |  |  |  |

153 个选择的基因被评估。

研究目标：

1，分析phenotypes与gen expression 之间的关联。

2，分析不同期别的肿瘤与基因序列表达分化的关联。

3，研究代谢的基因通路。

4，寻找与疾病进展相关的有生物学意义的基因序列。

所分析的变量：

1，TumorSize： T1,T2,T3

2，Axillary lymph node：N0,N1，N2，N3

3，ER Status： ER+, ER-

4，Tumor Phenotype：

5，Tumor differentiated：Grade1, Grade2, Grade3

6, Tumor histology: ductal carcinoma, lobular carcinoma

分析算法:

Tumor phenotypes VS Normal samples

Fold-change ratio(FC) = 2/power(mean(IG1 - IG2)) 单值表示

Student t statistic 去检测 是否有明显的差异.

Different Tumor samples : TumorX Vs TumorY

FC\_x = 1/mean(IGX) FC\_y = 1/mean(IGY)

Student t test for significant.

multiple test :the False Discovery Rate(FDR)

Tumor vs Normal : FDR < 0.01 认为有显著差异

Tumor1 vs Tumor2 ：FDR < 0.05 认为有显著差异

Unsupervised hierachial clustering analysis for tumor vs normal

Pearson correlation used for similarity measure

Complete linkage for the agglomerative method

|  |  |
| --- | --- |
| [GSM272923](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272923) | Control Con\_rep1 |
| [GSM272924](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272924) | Control Con\_rep2 |
| [GSM272925](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272925) | Control Con\_rep3 |

|  |  |
| --- | --- |
| [GSM272926](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272926) | Control Con\_rep4 |
| [GSM272927](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272927) | Control Con\_rep5 |
| [GSM272928](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272928) | Control Con\_rep6 |
| [GSM272929](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272929) | Control Con\_rep7 |
| [GSM272930](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272930) | Control Con\_rep8 |
| [GSM272931](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272931) | Control Con\_rep9 |
| [GSM272932](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272932) | Control Con\_rep10 |

|  |  |
| --- | --- |
| [GSM272923](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272923) | Control Con\_rep1 |
| [GSM272924](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272924) | Control Con\_rep2 |
| [GSM272925](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272925) | Control Con\_rep3 |

|  |  |
| --- | --- |
| [GSM272926](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272926) | Control Con\_rep4 |
| [GSM272927](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272927) | Control Con\_rep5 |
| [GSM272928](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272928) | Control Con\_rep6 |
| [GSM272929](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272929) | Control Con\_rep7 |
| [GSM272930](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272930) | Control Con\_rep8 |
| [GSM272931](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272931) | Control Con\_rep9 |
| [GSM272932](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272932) | Control Con\_rep10 |
| [GSM272933](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272933) | Control Con\_rep11 |
| [GSM272934](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272934) | Control Con\_rep12 |
| [GSM272935](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272935) | Control Con\_rep13 |
| [GSM272936](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272936) | Control Con\_rep14 |
| [GSM272937](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272937) | Control Con\_rep15 |
| [GSM272938](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272938) | Control Con\_rep16 |
| [GSM272939](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272939) | Control Con\_rep17 |
| [GSM272940](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272940) | Control Con\_rep18 |
| [GSM272941](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272941) | Control Con\_rep19 |
| [GSM272942](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272942) | Control Con\_rep20 |
| [GSM272943](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272943) | Control Con\_rep21 |
| [GSM272944](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272944) | Control Con\_rep22 |
| [GSM272945](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272945) | Control Con\_rep23 |
| [GSM272946](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272946) | Control Con\_rep24 |
| [GSM272947](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272947) | Control Con\_rep25 |
| [GSM272948](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272948) | Control Con\_rep26 |
| [GSM272949](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272949) | Control Con\_rep27 |
| [GSM272950](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272950) | Tumor Tum\_Ph1\_rep1 |
| [GSM272951](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272951) | Tumor Tum\_Ph1\_rep2 |
| [GSM272952](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272952) | Tumor Tum\_Ph1\_rep3 |
| [GSM272953](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272953) | Tumor Tum\_Ph1\_rep4 |
| [GSM272954](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272954) | Tumor Tum\_Ph1\_rep5 |
| [GSM272955](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272955) | Tumor Tum\_Ph1\_rep6 |
| [GSM272956](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272956) | Tumor Tum\_Ph1\_rep7 |
| [GSM272957](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272957) | Tumor Tum\_Ph1\_rep8 |
| [GSM272958](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272958) | Tumor Tum\_Ph1\_rep9 |
| [GSM272959](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272959) | Tumor Tum\_Ph1\_rep10 |
| [GSM272960](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272960) | Tumor Tum\_Ph2\_rep1 |
| [GSM272961](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272961) | Tumor Tum\_Ph2\_rep2 |
| [GSM272962](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272962) | Tumor Tum\_Ph2\_rep3 |
| [GSM272963](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272963) | Tumor Tum\_Ph2\_rep4 |
| [GSM272964](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272964) | Tumor Tum\_Ph2\_rep5 |
| [GSM272965](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272965) | Tumor Tum\_Ph2\_rep6 |
| [GSM272966](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272966) | Tumor Tum\_Ph2\_rep7 |
| [GSM272967](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272967) | Tumor Tum\_Ph2\_rep8 |
| [GSM272968](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272968) | Tumor Tum\_Ph2\_rep9 |
| [GSM272969](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272969) | Tumor Tum\_Ph2\_rep10 |
| [GSM272970](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272970) | Tumor Tum\_Ph2\_rep11 |
| [GSM272971](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272971) | Tumor Tum\_Ph2\_rep12 |
| [GSM272972](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272972) | Tumor Tum\_Ph2\_rep13 |
| [GSM272973](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272973) | Tumor Tum\_Ph2\_rep14 |
| [GSM272974](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272974) | Tumor Tum\_Ph2\_rep15 |
| [GSM272975](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272975) | Tumor Tum\_Ph3\_rep1 |
| [GSM272976](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272976) | Tumor Tum\_Ph3\_rep2 |
| [GSM272977](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272977) | Tumor Tum\_Ph3\_rep3 |
| [GSM272978](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272978) | Tumor Tum\_Ph3\_rep4 |
| [GSM272979](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272979) | Tumor Tum\_Ph4\_rep1 |
| [GSM272980](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM272980) | Tumor Tum\_Ph4\_rep2 |