

基于通路表达谱数据的 乳腺癌分型研究及其药物筛选

汇报人:华红丽

研究背景

已有研究表明，乳腺癌是由多种因素引发的复杂疾病，且在基因层面存在着广泛的异质性。但现有的研究中，对乳腺癌亚型的研究与分析大多是单纯地基于差异表达基因或者解剖学结构，忽略了非差异基因的影响以及生物功能层面的一致性。

在临床上，乳腺癌患者在直接使用或术后使用化疗药物时，大多数是医生在解剖学特征的基础上开处方，由于个体差异性的存在，这种方式并非所有患者都是适用的。

因此，我们将生物学通路和基因表达数据结合起来，提出一种乳腺癌亚型分析的新思路。根据分类结果对患者所使用的化疗药物进行筛选，得到针对不同类群患者的最优化疗药物组合。

研究材料

数据来源:

TCGA	KEGG	HPRD, IBIS, BioGRID, IntAct, MINT
590个样本 (529个疾病样本, 61个正常样本)	289条生物学通路 (现291条)	16,257个蛋白质
17,814个基因	6,696个基因	238,429个互作关系

使用工具:

R包: iGraph, 用于计算网络中心性, 评估基因重要性

R包: Clusterv, 用于聚类分析及其结果评估

研究方法

基因表达谱通路化：（都转化为gene symbol）

P_1	g_{11}	g_{12}	g_{13}	g_{14}	...	$g_{1 \cdot n_1}$
P_2	g_{21}	g_{22}	g_{23}	g_{24}	...	$g_{2 \cdot n_2}$
P_3	g_{31}	g_{32}	g_{33}	g_{34}	...	$g_{3 \cdot n_3}$
P_4	g_{41}	g_{42}	g_{43}	g_{44}	...	$g_{4 \cdot n_4}$
\vdots	\vdots	\vdots	\vdots	\vdots	...	\vdots
P_{289}	$g_{289 \cdot 1}$	$g_{289 \cdot 2}$	$g_{289 \cdot 3}$	$g_{289 \cdot 4}$...	$g_{289 \cdot n_{289}}$

共591个通路基因表达谱

研究方法

计算网络中心性，评估基因重要性：

度： $Degree_i = \frac{n_i}{N}$

介数： $B_v = \sum_{i \neq j \neq v \in V} \frac{\sigma_{ivj}}{\sigma_{ij}}$

紧密度： $C_v = \frac{1}{n-1} \sum_{j \neq v \in V} d_{vj}$

N步覆盖度： $\omega_{in} = \frac{d_{in}}{N} \quad (n = 2, 3, 4, 5)$

各网络中心性的pearson相关系数分析:

Correlation Value	Degree	Betweenness	Closeness	Coverage_2	Coverage_3	Coverage_4	Coverage_5
Degree	1	0.8113	0.0361	0.2254	0.1303	0.0317	0.0184
Betweenness	0.8113	1	0.0051	0.0339	0.0131	0.0029	0.0017
Closeness	0.0361	0.0051	1	0.1795	0.4535	0.9135	0.9923
Coverage_2	0.2254	0.0339	0.1795	1	0.6925	0.1779	0.1054
Coverage_3	0.1303	0.0131	0.4535	0.6925	1	0.5757	0.4032
Coverage_4	0.0317	0.0029	0.9135	0.1779	0.5757	1	0.9299
Coverage_5	0.0184	0.0017	0.9923	0.1054	0.4032	0.9299	1

研究方法

判断通路上/下调：

本课题设定非线性最小二乘拟合函数值为0，根据正常样本的通路基因表达谱数据，得到函数的拟合参数 α_i, β ，即可得到目标函数：

$$y = \sum_{i=1}^n \alpha_i X_i + \beta$$

研究方法

判断通路上/下调:

再将其他样本的通路基因表达值带入目标函数中, 得到 y 值, 若 $y > 0$, 判断通路为上调; 若 $y < 0$, 则判断通路为下调。

$$p_{ij} = \begin{cases} 1 & \text{if } \sum_{k=1}^n \alpha_k g_{ik} + \beta \geq 0 \\ -1 & \text{if } \sum_{k=1}^n \alpha_k g_{ik} + \beta < 0 \end{cases}$$

其中, p_{ij} 表示样本 j 在通路 i 上的方向;

α_k, β 表示通过正常样本拟合得出的拟合参数;

g_{ik} 表示样本 j 的通路 i 的第 k 个基因的表达值;

n 表示通路 i 的基因数目。

研究方法

计算通路失常得分 (PDS-pathway deregulation score)

$$d'_{ij} = \sqrt{\sum_{k=1}^n \omega_k (g_{ijk} - g_{imk})^2}$$

质与量结合:

$$d_{ij} = d'_{ij} \times p_{ij} = \pm \sqrt{\sum_{k=1}^n \omega_k (g_{ijk} - g_{imk})^2}$$

其中, d'_{ij} 表示样本j的第i条通路的失调得分;

ω_k 表示基因k的权重 (度、介数、紧密度、2-5步覆盖度);

g_{ijk} 表示样本j的第i条通路的第k个基因的表达值;

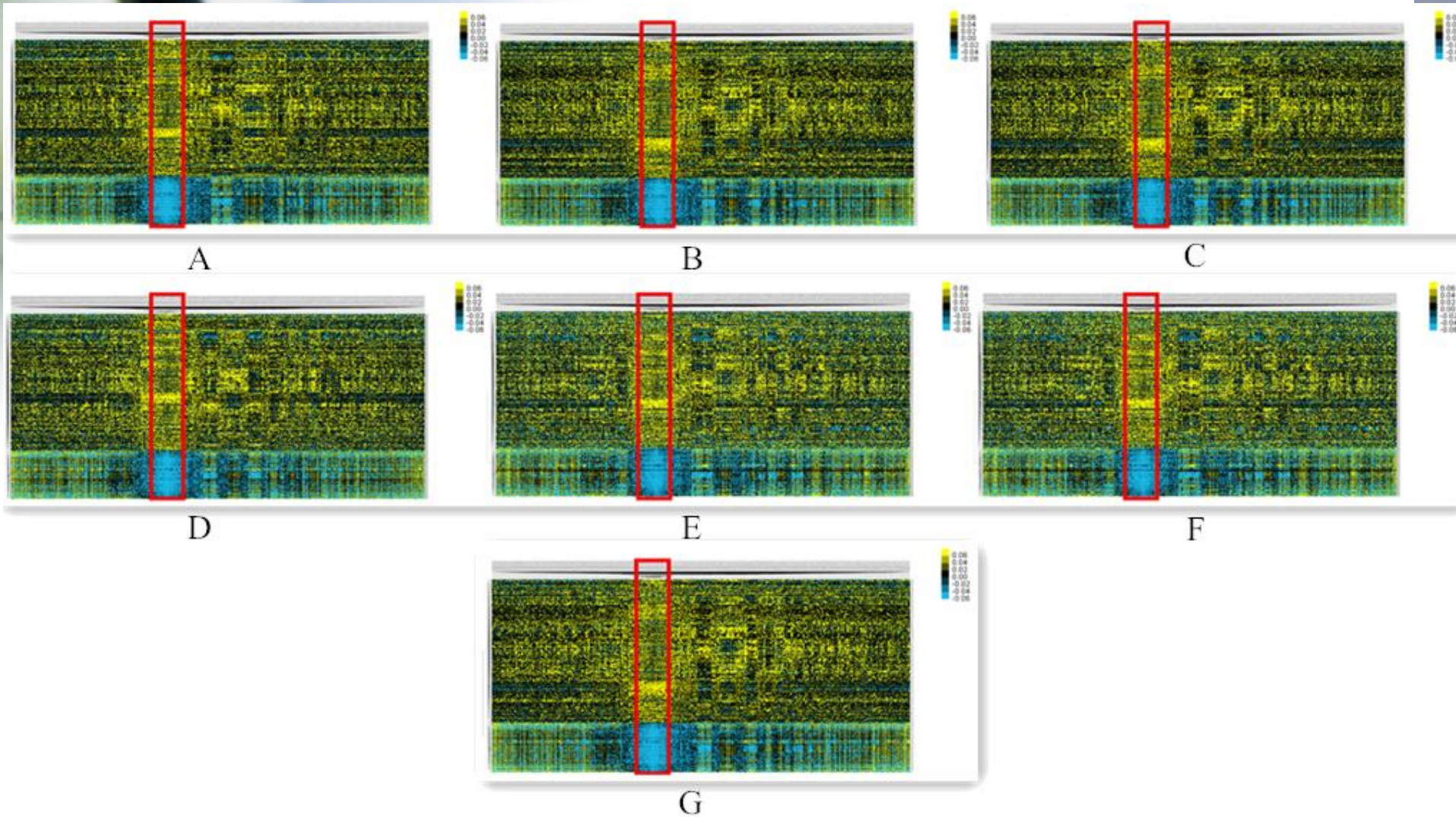
g_{imk} 表示所有正常表达值均值处理后的第i条通路的第k个基因的表达值;

研究方法

度量通路失调情况：

P	S_1	S_2	S_3	S_4	...	S_{590}
P_1	d_{11}	d_{12}	d_{13}	d_{14}	...	$d_{1.590}$
P_2	d_{21}	d_{22}	d_{23}	d_{24}	...	$d_{2.590}$
P_3	d_{31}	d_{32}	d_{33}	d_{34}	...	$d_{3.590}$
P_4	d_{41}	d_{42}	d_{43}	d_{44}	...	$d_{4.590}$
\vdots	\vdots	\vdots	\vdots	\vdots	...	\vdots
P_{289}	$d_{289.1}$	$d_{289.2}$	$d_{289.3}$	$d_{289.4}$...	$d_{289.590}$

基于7种网络中心性的通路失调表达谱



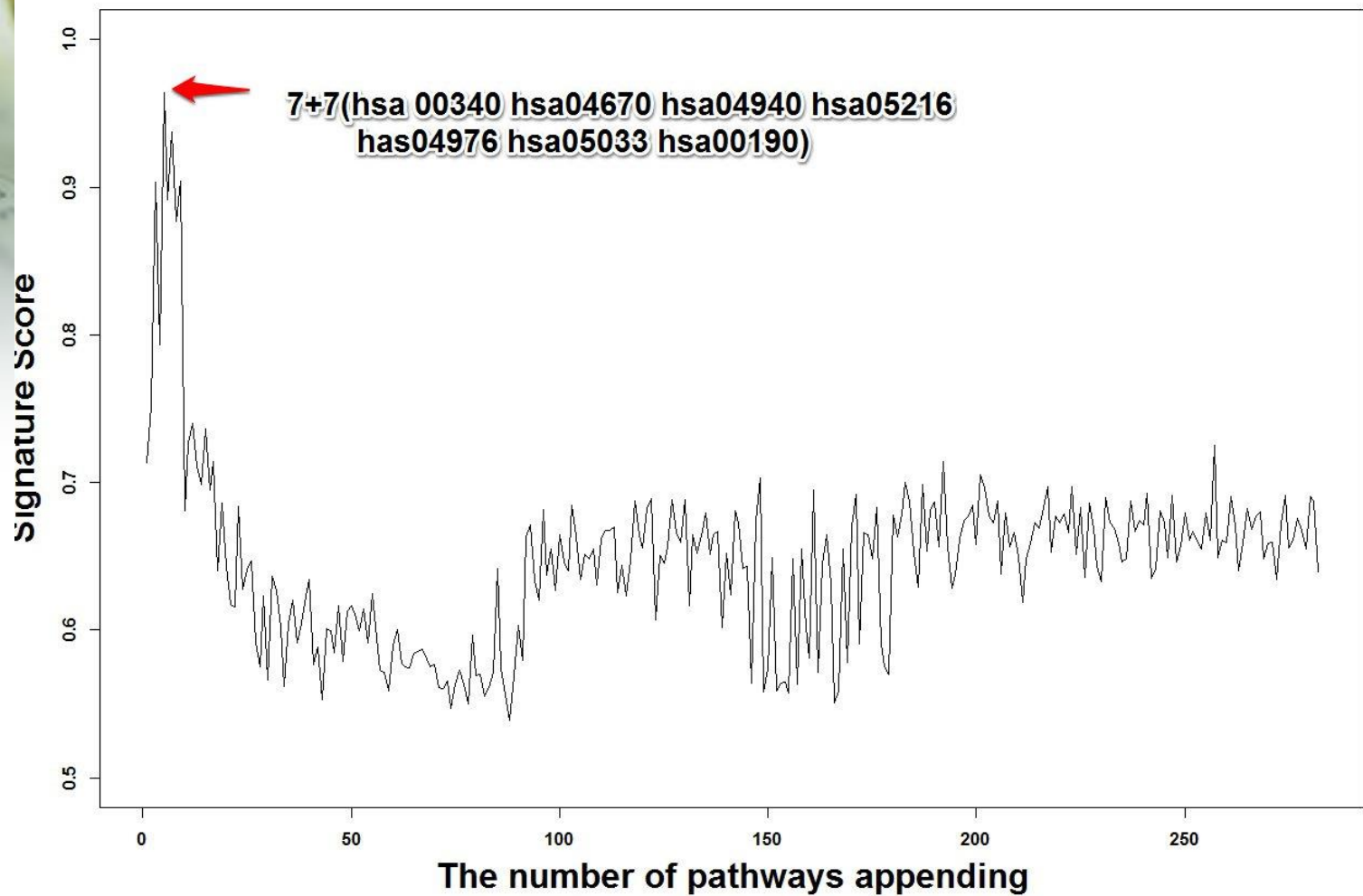
研究方法

通路优化:

- 1) 人工筛选出7条与癌症相关的通路作为种子通路集;
- 2) 用clusterv计算并衡量其他282条通路对分类的稳定性并排序;
- 3) 根据排序一步步扩充种子通路, 寻找到最优分类的通路集合。

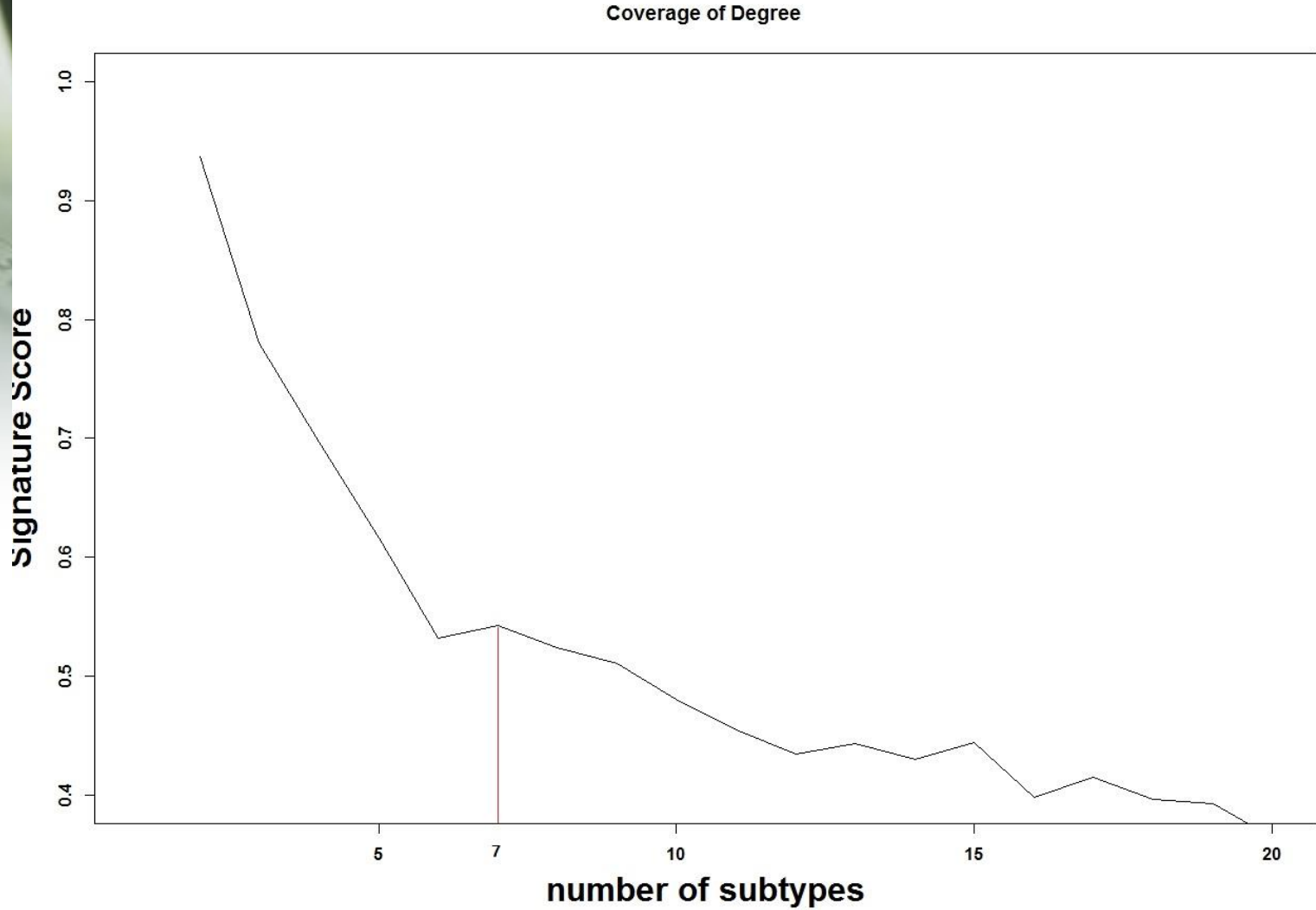
KEGG ID of pathway	description
hsa05200	Pathways in cancer
hsa05202	Transcriptional misregulation in cancer
hsa05203	Viral carcinogenesis
hsa05204	Chemical carcinogenesis
hsa05205	Proteoglycans in cancer
hsa05206	MicroRNAs in cancer
hsa05217	Basal cell carcinoma

Pathways Optional in Coverage of Degree




追加的通路及其通路名

Pathway ID	Pathway Name	Involved in PDS
hsa04964	Proximal tubule bicarbonate reclamation	C;C5;C4;C3;C2
hsa00360	Phenylalanine metabolism	C2
hsa00910	Nitrogen metabolism	C3
hsa00232	Caffeine metabolism	C;C5;C4
hsa00340	Histidine metabolism	D
hsa04670	Leukocyte transendothelial migration	D
hsa04940	Type I diabetes mellitus	D
hsa05216	Thyroid cancer	D
hsa04976	Bile secretion	D
hsa05033	Nicotine addiction	D
hsa00190	Oxidative phosphorylation	D
hsa03010	Ribosome	B
hsa05219	Bladder cancer	B
hsa05210	Colorectal cancer	B
hsa05213	Endometrial cancer	B
hsa01210	2-Oxocarboxylic acid metabolism	B



$$\underline{Signature\ Score(N - 1) < Signature\ Score(N) > Signature\ Score(N + 1)}$$

分类结果



Data of classification	Number of classifications
Degree	7
Betweenness	8
Coverage of 2 steps	8
Coverage of 3 steps	7
Coverage of 4 steps	3

A close-up photograph of a black fountain pen with gold accents, resting on a document. The document has some faint, mirrored text visible. Overlaid on the right side of the image is the Chinese phrase '谢谢大家!' in a large, stylized font with a blue-to-yellow gradient and a white outline.

谢谢大家!