

乳腺癌转移相关基因表达调控网络的 线性微分方程模型的初步研究*

倪春生^{1)**} 孙保存^{1)***} 冯玉梅¹⁾ 张丹芳¹⁾ 李晓青¹⁾ 董骝焕^{2)**} 张亮³⁾

¹⁾天津医科大学附属肿瘤医院肿瘤研究所, 天津 300060;

²⁾南开大学数学科学院, 天津 300071;

³⁾生物芯片北京国家工程研究中心, 北京 100084)

摘要 初步构建乳腺癌转移相关基因表达调控网络的线性微分方程模型, 并分析模型的可靠性和生物学意义. 采用基因芯片技术, 分别对 30 例伴有淋巴结转移的乳腺癌组织及其相应淋巴结转移癌组织进行基因表达谱的比较, 选择差异基因通过线性微分方程方法构建表达调控网络模型. 差异表达基因共 27 个, 其中 $\text{Ratio} > 3$ 的明显上调基因 14 个, 而 $\text{Ratio} < 0.33$ 的明显下调基因 13 个. 比较伴有淋巴结转移的乳腺癌组织和其相应淋巴结转移癌组织, 分析筛选了 27 个表达差异基因, 应用数学线性微分方程方法初步构建乳腺癌转移相关基因表达调控网络的线性微分方程模型, 通过分析模型中重要节点、通路的生物学意义, 判定网络的数学特性, 初步表明, 调控网络的可靠性和乳腺癌转移的形成是与多基因、多通路异常引起的细胞恶性转化相关.

关键词 乳腺癌, 基因芯片, 转移, 调控网络, 线性微分方程模型

学科分类号 R737.9

乳腺癌是常见的妇科恶性肿瘤, 严重危害妇女生命健康, 近年患病率上升很快, 在妇女恶性肿瘤中占第一位. 近年来生物信息学和肿瘤分子生物学的快速发展, 带来了乳腺癌诊断、预后和治疗的进步. cDNA 表达微阵列 (基因芯片) 是大规模、平行、快速检测和分析肿瘤样本多肿瘤相关基因及分子标志的高效方法. 为研究乳腺癌转移的机制, 我们采用高通量的 cDNA 表达微阵列技术, 对乳腺癌的肿瘤相关基因作初步探索, 采用数学的线性微分方程模型对乳腺癌的基因表达谱数据进行初步研究.

1 材料与方法

1.1 组织材料

收集 2003 年 1 月至 12 月间, 天津医科大学附属肿瘤医院乳腺科收治的 30 例淋巴结转移阳性乳腺癌患者行乳腺根治术切除的标本. 每个病例标本分别取原发癌和相应腋窝淋巴结转移癌组织. 30 例标本均经过病理组织学证实乳腺原发癌的病理类型和转移的腋窝淋巴结数目, 且淋巴结转移癌中癌细胞达 75% 以上. 结合临床病例资料进一步确定 TNM 分期和国际标准临床分期 (表 1). 手术切取的标本组织立即投入液氮中速冻后, -80°C 保存备用.

1.2 总 RNA 的提取

采用 Trizol 一步抽提法 (Trizol, Invitrogen 公司) 提取组织细胞总 RNA, RNeasy mini kit (Qiagen 公司) 纯化 RNA, GeneQuant pro RNA/DNA Calculator (Pharmacia Biotech) 定量 RNA 浓度, 琼脂糖凝胶电泳检测 RNA 完整性.

1.3 探针标记与杂交

逆转录合成 cDNA 后随机引物法标记, 每例病例的乳腺原发癌组织用 Cy3-dCTP 标记, 对应配对的淋巴结转移癌组织用 Cy5-dCTP 标记. 标记产物用 QIAquick PCR purification Kit (Qiagen 公司) 纯化后用于杂交.

1.4 基因芯片

芯片由北京博奥生物芯片有限公司提供. 芯片包含 23 232 个 oligo 序列.

1.5 杂交和洗涤

将表达谱芯片和杂交探针分别置于 95°C 水浴中变性 5 min, 冰浴后立即将探针加在芯片上, 盖

*国家高技术研究发展计划(863)资助项目(2002AA2Z2011).

共同第一作者. * 通讯联系人.

Tel: 022-27587000, E-mail: nknini@vip.sina.com.cn

收稿日期: 2005-06-02, 接受日期: 2005-06-30

上盖玻片，置于杂交盒内. 于 42℃ 恒湿杂交箱内杂
交过夜. 依次用 2×SSC + 0.2% SDS 溶液、0.1%

SSC + 0.2% SDS 和 0.1% SSC 分别洗涤 10 min，室
温晾干.

Table 1 Pathological diagnosis and clinical staging of 30 breast cancer specimens

No.	Pathological type	Grading	Numbers of metastatic lymphnode	TNM staging	Clinical staging
1	Carcinoma simplex	II	2	T2N0M0	2A
2	TMC	II	2	T2N2M0	3A
3	IDC	II	1	T2N0M0	2A
4	Carcinoma simplex	II	3	T2N0M0	2A
5	Carcinoma simplex	II	2	T2N2M0	3A
6	Carcinoma simplex	II	4	T2N2M0	3A
7	Carcinoma simplex	II	20	T3N1M0	3A
8	Carcinoma simplex	II	2	T3N2M0	3A
9	TMC	III	1	T2N0M0	2A
10	IDC	III	11	T3N1M0	3A
11	IDC	II	14	T2N2M0	3A
12	IDC	II	4	T1N0M0	1
13	Carcinoma simplex	II	1	T2N0M0	2A
14	Carcinoma simplex	II	13	T2N0M0	2A
15	Apocrine carcinoma	III	28	T3N2M0	3A
16	TMC	II	1	T3N1M0	3A
17	Carcinoma simplex	III	2	T3N1M0	3A
18	Carcinoma simplex	III	17	T3N2M1	4
19	Carcinoma simplex	III	9	T2N1M1	4
20	IDC	II	14	T3N2M0	3A
21	TMC	III	27	T2N1bM0	2B
22	Carcinoma simplex	I	16	T2N2M0	3A
23	TMC	III	36	T3N2M0	3A
24	Carcinoma simplex	II	2	T2N0M0	2A
25	Carcinoma simplex	II	10	T2N1M0	2B
26	Carcinoma simplex	II	20	T2N1bM0	2B
27	Carcinoma simplex	II	2	T2N0M0	2A
28	Carcinoma simplex	II	3	T1N0M0	1
29	IDC	II	4	T2N1M0	2B
30	Carcinoma simplex	III	10	T4bN1bM0	3B

TMC: Typical medullary carcinoma; IDC: Infiltrating ductal carcinoma.

1.6 芯片扫描与分析验证

经严格洗片后的杂交芯片用 ScanArray 4000 扫描仪 (General Scanning 公司产品) 进行扫描，获得荧光信号图像，用 Gene Pix PRO4.0 软件对图像进行处理，分析 Cy3 和 Cy5 两种荧光信号的强度，线性归一化后，Cy5/Cy3 两者的比值用于异常表达判断. 基因差异表达的标准：a. Cy3 和 Cy5 荧光强度值其中之一必须 $\geq 1\ 000$ ，若两者均小于 1 000 认为在该病例中原发癌和转移癌均不表达此基因；b. Cy5/Cy3 比值 ≤ 0.33 或 ≥ 3.0 ；c. Cy5/Cy3 比值 ≤ 0.33 或 ≥ 3.0 样本数必须超过总样本的 1/2，即在

30 例病例中有 15 例以上共同出现 Cy5/Cy3 比值 ≤ 0.33 或 ≥ 3.0 .

芯片实验数据已经经过美国 M.D. Anderson 癌症中心的生物信息学处理，将淋巴结表达相关差异基因滤除. 本研究还对芯片结果中的差异基因 IGFBP5、FN、MMP1、MMP3、MMP9 等进行了 Real-time PCR、免疫组化和蛋白质印迹的验证工作，结果证实应用癌细胞达 75% 以上的淋巴结转移癌与原发癌比较筛选差异表达基因的方法是可靠的.

1.7 生物信息学分析

我们用矩阵 $X(n \times m)$ 来表示 n 个基因在 m 个时间点上的表达数据. 其中 x_{ij} 分别表示第 i 个基因在第 j 个时间点的表达值. 同时, 基因的表达经常受到其他基因及自身表达的促进或者抑制, 这是一种相互影响、相互调控的复杂关系. 这种关系构成了复杂的基因表达调控网络模型, 基因表达调控网络模型通过基因表达矩阵的信息, 提取基因调控矩阵 (gene regulation matrix), 进而描述基因间的相互调控作用和环境改变对某个基因的影响. 我们使用加权矩阵 $W(n \times n)$ 来表示基因调控矩阵. w_{ij} 表示第 j 个基因对第 i 个基因的影响, 正数表示促进, 负数表示抑制, 零表示没有影响. w_{ij} 的绝对值越大, 表示影响越强, 选取线性微分方程模型如下:

$$\frac{\Delta x_i(t)}{\Delta t} = \sum_{j=1}^n w_{ij} \times x_j(t) + \xi(t), \forall i; \forall t,$$

其中, t 表示第 t 个时间点, $x_i(t)$ 和 $x_j(t)$ 分别表示第 i 个基因和第 j 个基因在第 t 个时间点的表达值, $\xi(t)$ 表示误差.

此次实验 t_1 至 t_6 时间点的划分是依据国际抗癌联盟 (UICC) 制订的 TNM 分期法^[1] 所得到的国际标准临床分期. $t_1 \sim t_6$ 分别对应: t_1 对应 I 期; t_2 对应 II A 期; t_3 对应 II B; t_4 对应 III A 期; t_5 对应 III B 期; t_6 对应 IV 期.

我们使用线性微分方程模型, 由于临床分期只有 6 个时间点, 但有 27 个基因. 因此需要对其表达

数据进行三次样条插值, 然后运用最小二乘法, 解出 w_{ij} , 并选择权值在 2σ 外的显著点, 构建出初步的乳腺癌转移相关基因网络模型加权矩阵 W . 根据 W , 获得基因网络模型拓扑图.

2 结 果

本研究中所设的 40 个管家基因均为阳性, 所设阴性对照和空白对照为阴性, 说明实验无污染, 检测体系可靠.

依照本研究设定的基因差异表达的判断标准: a. Cy3 和 Cy5 荧光强度值其中之一必须 $\geq 1\,000$, 若两者均小于 1 000 认为在该病例中原发癌和转移癌均不表达此基因; b. Cy5/Cy3 比值 ≤ 0.33 或 ≥ 3.0 ; c. Cy5/Cy3 比值 ≤ 0.33 或 ≥ 3.0 样本数必须超过总样本的 1/2, 即在 30 例病例中有 15 例以上共同出现 Cy5/Cy3 比值 ≤ 0.33 或 ≥ 3.0 , 在本次研究 30 例病例的 30 张基因芯片中, 乳腺癌组织与淋巴结转移癌组织表达有差异的基因共有 27 个, 差异表达的基因占 0.001 1% (27/23 232), 其中淋巴结转移癌组织中表达降低 (下调趋势) 的基因有 13 个 (表 2), 表达增高 (上调趋势) 的基因有 14 个 (表 3), 各占基因点数的 0.000 5% (13/23 232) 和 0.000 6% (14/23 232).

我们应用 30 例病例中筛选出的 27 个基因 (表 4) 构建初步的基因网络模型 (表 5), 所得基因网络模型拓扑图如图 1.

Table 2 Down-regulation genes in metastatic breast-cancer tissue of lymph node

GB.accession	HUGOname	Down-regulation cases
AF311912	Secreted frizzled-related protein 2	23
NM002091	Gastrin-releasing peptide	22
AK026320	Homo sapiens cDNA : FLJ22667 fis, clone HSI08385	22
AK022342	Homo sapiens cDNA : FLJ12280 fis, clone MAMMA1001744	22
NM002427	Matrix metalloproteinase 13 (collagenase 3)	21
NM002421	Matrix metalloproteinase 1 (interstitial collagenase)	20
AK022198	Homo sapiens cDNA : FLJ12136 fis, clone MAMMA1000312	20
AF156100	Fibulin 6	20
NM001723	Bullous pemphigoid antigen 1 (230/240 ku)	19
AK026784	Homo sapiens cDNA : FLJ23131 fis, clone LNG08502	18
NM000422	Keratin 17	16
NM002422	Matrix metalloproteinase 3 (stromelysin 1,progelatinase)	16
NM001854	Collagen, type XI, alpha 1	24

Table 3 Up-regulation genes in metastatic breast-cancer tissue of lymph node

GB.accession	HUGOname	Up-regulation cases
NM000587	Complement component 7	27
NM006205	Phosphodiesterase 6h, cGMP-specific, cone, gamma	25
NM021950	Membrane-spanning 4-domains, subfamily A, member 1 (MS4A1), transcript variant 3	25
NM002561	Purinergic receptor P2X, ligand-gated ion channel, 5 (P2RX5), transcript variant 1	23
NM021966	T-cell leukemia/lymphoma 1A	22
NM002371	Mal, T-cell differentiation protein	22
NM001877	Complement component (3d/Epstein Barr virus) receptor 2 (CR2), transcript variant 2	22
AL049265	Homo sapiens mRNA: cDNA DKFZp564F053 (from clone DKFZp564F053)	22
NM001781	CD69 antigen (p60, early T-cell activation antigen)	21
AF213459	EphA3	21
NM002922	Regulator of G-protein signalling 1	21
NM001770	CD19 antigen	21
NM006183	Neurotensin	18
NM014031	VLCS-H1 protein	17

Table 4 Time-course gene expression data of gene related to breast cancer metastasis

NO.	Name	GB.accession	t1	t2	t3	t4	t5	t6
1	C7	NM_000587	3.341 21	2.529 46	1.149 62	2.601 46	3.267 23	2.745 18
2	MS4A1	NM_021950	2.201 79	1.927 89	0.852 07	1.798 55	1.718 94	4.237 66
3	TCL1A	NM_021966	2.451 35	2.909 27	1.506 98	1.210 93	0.813 90	0.742 40
4	PDE6H	NM_006205	2.106 73	2.402 64	1.174 21	1.050 09	0.969 03	0.344 85
5	C8	AL049265	2.463 96	2.008 01	1.556 23	1.431 18	1.396 12	0.185 83
6	CR2	NM_001877	2.594 25	2.335 56	1.677 39	1.219 59	1.006 53	-0.298 9
7	EPHA3	AF213459	2.308 12	0.776 83	1.068 21	1.424 11	1.460 94	1.102 92
8	P2RX5	NM_002561	2.192 51	1.212 43	0.591 90	1.289 85	0.000 01	1.067 93
9	MAL	NM_002371	1.584 32	1.383 82	0.657 15	1.247 37	0.335 75	1.309 13
10	RGS1	NM_002922	1.816 27	1.599 15	0.308 23	1.511 97	2.010 35	-0.152 0
11	CD69	NM_001781	2.537 68	1.392 57	0.547 90	1.146 42	0.957 27	0.834 20
12	CD19	NM_001770	1.448 99	1.527 44	0.812 32	0.982 93	1.077 32	0.020 47
13	NTS	NM_006183	2.160 28	2.105 08	1.416 63	1.062 12	1.676 74	4.489 89
14	VLCS-H1	NM_014031	1.580 25	1.655 33	-0.127 5	1.229 35	0.207 98	0.522 90
15	COL11A1	NM_001854	-3.316 1	-2.080 4	-3.067 8	-1.777 6	-0.564 5	-0.164 0
16	GRP	NM_002091	-1.311 6	-2.092 5	-1.629 0	-2.193 4	-3.130 4	-0.396 4
17	MMP13	NM_002427	-1.192 2	-1.690 3	-2.310 4	-1.425 0	-1.372 4	-1.914 2
18	SFRP2	AF311912	-1.735 7	-1.751 7	-1.701 8	-1.694 3	-0.441 9	-1.448 1
19	C9	AK026320	-1.707 8	-1.984 1	-0.826 9	-1.315 8	0.554 05	-1.537 3
20	FIBL-6	AF156100	-1.213 4	-1.770 2	-2.124 9	-0.850 0	-0.273 7	-1.515 2
21	BPAG1	NM_001723	-1.988 3	-2.164 9	-1.643 8	-1.793 0	0.000 01	-0.233 7
22	MMP1	NM_002421	-2.090 7	-1.441 2	-2.688 5	-1.577 5	-0.366 2	1.523 46
23	C10	AK022342	-3.079 5	-1.828 2	-1.967 6	-1.278 1	-0.044 3	-0.854 0
24	C11	AK022198	-2.310 2	-2.190 9	-0.553 1	-1.106 4	-0.737 3	-0.937 4
25	C12	AK026784	-0.853 8	-1.526 8	-0.718 9	-1.425 5	-0.597 4	-0.214 0
26	KRT17	NM_000422	-2.579 7	-1.255 8	-0.681 3	-1.438 6	-0.123 0	-0.166 0
27	MMP3	NM_002422	-1.491 4	-1.516 6	-2.218 6	-0.782 0	-1.991 4	-1.952 5

Table 5 The weight matrix of genetic network of gene related to breast cancer metastasis

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	-1	-1	0	-1	0	0	-1	-1	-1	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	-1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	-1	-1	0	-1	0	0	0	0	-1	1
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

*The genes of NO.1~27 point are corresponding for NO.1~27 genes in the Table 4.

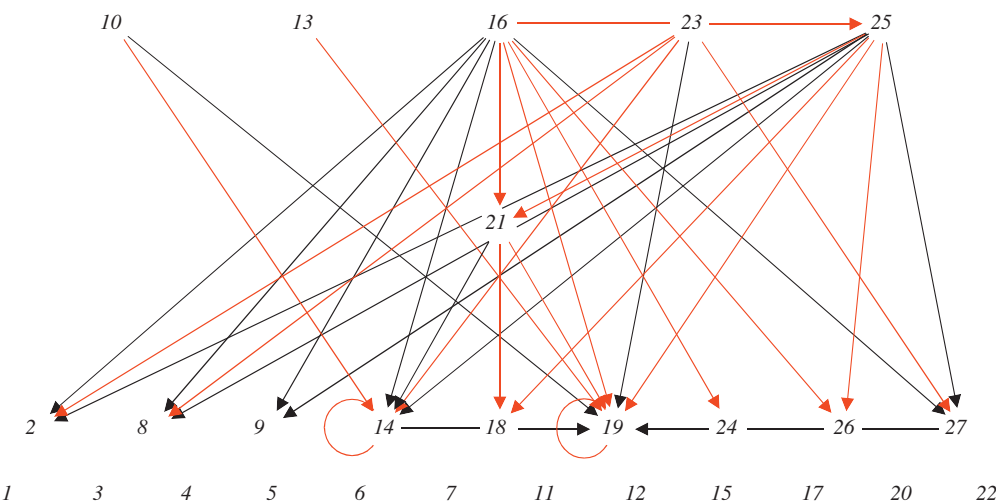


Fig. 1 Genetic network of gene related to breast cancer metastasis

The genes of NO.1~27 point are corresponding for NO.1~27 genes in the Table 4. The red line means upper gene-induced upregulation of lower genes. The black line means upper gene-induced downregulation of lower genes.

3 讨 论

乳腺癌的发生、发展和治疗具有很强的个体特异性, 现在利用 cDNA 芯片等技术可获得细胞基因表达谱(gene expression profiling), 鉴定各样本在 mRNA 水平上的表达差异, 从而能更精确地对患者实施个性化诊疗. 一般来说, 基因的表达经常受到其他基因及自身表达的促进或者抑制, 这是一种相互影响、相互调控的复杂关系. 这种关系构成了复杂的基因表达调控网络, 基因表达调控网络模型是表现基因表达调控网络中生物分子相互作用及作用路径等生物信息的新兴的生物信息学技术. 本研究采用表达谱基因芯片对该肿瘤基因表达谱进行分析, 收集表达差异基因, 应用数学线性微分方程方法初步构建乳腺癌转移相关基因表达调控网络的线性微分方程模型, 判定模型的可靠性并分析相关表达差异基因之间关系, 发现了一些与乳腺癌转移密切相关的基因, 为乳腺癌转移机制的研究提供一些新的实验支持.

对于构建初步基因网络、提取网络中最显著的信息的问题, 线性微分方程模型是一个方便快捷的模型. 即使有噪声, 线性微分方程模型也能很准确地预测网络模型的关系^[2]. D'haeseleer 等^[3]在 1999 年就应用线性微分方程模型成功地构建了一个包含 65 个 mRNA 的初步网络模型.

根据基因网络拓扑图, 1, 3, 4, 5, 6, 7, 11, 12, 15, 17, 20, 22 号这 12 个基因被排除在结构框架之外, 为非节点基因, 其余参与模型框架构建的 15 个基因为节点基因.

分析基因网络模型拓扑图, GRP 基因 (geneID: NM_002091) 位于 16 号节点, 对 2、8、9、14、21、19、24、26、27 这 9 个基因处于上游调控位置, 在调控网络的数学模型中属于重要的上游节点基因. 该基因明显下调在生物学上也同样具有重要的作用. GRP 参与神经肽信号通路, 在细胞周期中起到检查点 (check point) 的作用, 为 DNA 依赖的转录调控因子, 具有生长因子活性、信号传导分子活性、蛋白激酶活性, 该基因及其受体经常在乳腺癌等癌症中有表达, 作为有丝分裂原对肿瘤细胞的增殖和分化起到重要作用. 有学者认为 GRP 在乳腺癌组织中的表达明显低于在相关正常组织的表达, 作为一种起到形态原 (胚胎组织中可影响形态的物质) 功能的癌胚抗原, 来影响肿瘤细胞的增殖和调控肿瘤细胞的分化^[4].

BPAG1 基因 (geneID: NM_001723) 位于 21 号节点, 处于 16、25 这两个节点基因的下游位置, 对 14、18、19 这三个基因处于上游调控位置, 在调控网络的数学模型中属于位置较特殊的节点基因. BPAG1 参与基底膜构成和细胞基质连接 - 半桥粒构成, 为钙离子结合蛋白, 具有结构蛋白活性, 参与细胞骨架 - 中间丝构成, 参与构成肌动蛋白复合物, 同时具有细胞粘附分子活性, 参与细胞间黏附, 在细胞周期中起到调控作用. BPAG1 是半桥粒的主要组分, 半桥粒是一种跨膜复合物, 介导上皮细胞和基底膜之间的粘附. 半桥粒的异常表达已经被报道与许多类型的癌症相关, 在癌症的浸润与转移中起到重要的作用. Lo 等^[5]报道使用基因差异显示技术鉴定了在鼻咽癌细胞中 BPAG1 的表达下调. Vanaja 等^[6]使用寡核苷酸基因芯片技术, 发现 BPAG1 的表达量与前列腺癌的发展进程呈强烈的负相关.

SFRP2 基因 (geneID: AF311912) 位于 18 号节点, 处于 21 这个节点基因的下游位置. SFRP 家族由 5 种分泌型糖蛋白构成, SFRP2 参与细胞膜的组成和细胞外间隙的构成, 为跨膜受体, 具有信号传导分子活性, SFRP2 起到拮抗 Wnt 信号通路的作用. Wnt 信号通路的异常激活和上调是许多癌症的重要特征. 因此, SFRP2 作为 Wnt 信号通路的负性调控因子对预示肿瘤生成有重要作用. 在乳腺癌、间质瘤、肺癌等多种肿瘤组织中, SFRP2 的下调与肿瘤生成相关^[7]. Lee 等^[8]报道 SFRP2 在原发性乳腺导管癌等乳腺癌细胞系中有明显的表达下调, 可以作为乳腺肿瘤细胞中的肿瘤分子标志物.

综合分析基因网络模型拓扑图中上述 GRP 基因、BPAG1 基因、SFRP2 基因的位置和生物学功能发现, GRP 具有蛋白激酶活性, 在肿瘤细胞转移发生中可能具有信号传导和调控其他基因表达的作用. BPAG1 为钙离子结合蛋白, 可能具有钙调蛋白的功能, 参与基底膜构成和细胞基质连接 - 半桥粒构成, 具有细胞粘附分子活性参与细胞间黏附, 具有结构蛋白活性参与细胞骨架 - 中间丝构成, 参与构成肌动蛋白复合物, 该基因可能与肿瘤浸润转移相关. SFRP2 参与细胞膜的组成和细胞外间隙的构成, 该基因可能与肿瘤浸润转移相关. 因此, 初步提出 GRP 基因、BPAG1 基因、SFRP2 基因间的生物学调控通路存在的假设: GRP 通过其蛋白激酶活性调控 BPAG1 的表达, BPAG1 可以与钙离子结合, 其效应可能导致细胞黏附能力和细胞骨架改

变, 还可调节 SFRP2 的表达, 使肿瘤细胞浸润转移能力改变.

金属蛋白酶基因 (MMP) 家族中的 MMP3 (geneID: NM_002422), 位于 27 号节点, 处于 16、23、25 这 3 个上游重要的节点基因的下游位置. MMP3 属于金属蛋白酶家族中的基质溶解蛋白亚型, 具有基质降解酶 1 活性, 与锌离子结合参与胶原及其他细胞外基质大分子的降解. MMP3 的表达与乳腺癌的发生、侵袭和转移明显相关^[9]. 而本次研究中 MMP3 表达明显下调, 这可能是由于 MMP3 的表达随着乳腺肿瘤新生、发展等不同的生长过程而出现不同的变化^[10].

Mal 基因 (geneID: NM002371), 位于 9 号节点, 处于 16、25 这 2 个上游重要的节点基因的下游位置. MAL 参与细胞表面受体相关信号传导和细胞膜的构成, 为跨膜受体, 具有结构分子活性, 裂解酶和氧化还原酶活性, MAL 基因编码的 Mal 蛋白包含有许多潜在的酪氨酸激酶 II 和蛋白激酶 C 磷酸化位点, Mal 蛋白在 RNA 加工的磷酸化过程中起到重要作用. 肿瘤中 MAL 的过表达和过多 Mal 蛋白的产生将打破免疫系统的平衡和削弱免疫杀伤能力. 由于 MAL 的过表达可能是肿瘤逃避免疫和产生免疫耐受的关键步骤, 因此在乳腺肿瘤病人体内检测到 Mal 蛋白或 Mal 蛋白的抗体就表明一种存在潜在恶性的可能性^[11].

总结 MMP3, Mal, KRT17^[12], P2RX5^[13], MS4A1^[14]等下游基因的生物学特性: MS4A1、P2RX5、Mal、KRT17、SFRP2 为细胞膜组分, MAL、KRT17 具有结构分子活性, 参与细胞骨架构成, 可能与肿瘤浸润转移中的变形运动相关; MAL、P2RX5、SFRP2 为跨膜受体, 在肿瘤细胞转移发生中可能具有信号传导的作用; MMP3 具有基质降解酶 1 活性参与胶原及其他细胞外基质的降解, 与肿瘤浸润转移中的基质降解相关. 故 MS4A1, P2RX5, Mal, KRT17, MMP3 等下游节点基因作为效应器表现出的以上这些特点均与乳腺癌转移的发生密切相关, 因此初步反映以上基因的生物学特性与所处的基因网络拓扑图中的位置具有一定的相关性和对应关系.

综上所述, GRP 基因、MAL 基因、MMP3 基因和 SFRP2 基因等均已经被证实为乳腺癌发生、转移相关基因. 网络图中初步表明乳腺癌转移的形成是与多基因异常引起多条传导通路异常致使细胞恶性转化相关, 这与生物学、医学上公认的看法是

相一致的. 图 1 中 23 号、25 号、19 号节点基因处于调控和被调控的关键位置, 但此三位之上的基因具体名称和功能尚未明确, 所以这三个基因将是我们以后验证和分析的目标. 因此, 利用基因芯片等技术筛选表达差异基因, 由此借助数学方法建立调控网络的线性微分方程模型具有一定的可靠性, 在认识乳腺癌转移机制, 乳腺癌诊断和治疗方面具有一定的理论和应用价值.

参考文献

- 1 Silva O E, Zurrida S. Breast Cancer: A Practical Guide. 2nd. Amsterdam: Elsevier Science Ltd, 2000. 66~70
- 2 Wessels L F, Van Someren E P, Reinders M J, *et al.* A comparison of genetic network models. Pacific Symposium on Biocomputing, 2001, **6**: 508~519
- 3 D'haeseleer P, Wen X, Fuhrman S, *et al.* Linear modeling of mRNA expression levels during CNS development and injury. Pacific Symposium on Biocomputing, 1999, **4**: 41~52
- 4 Jensen J A, Carroll R E, Benya R V, *et al.* The case for gastrin-releasing peptide acting as a morphogen when it and its receptor are aberrantly expressed in cancer. Peptides, 2001, **22** (4): 689~699
- 5 Lo A K, Yuen P W, Liu Y, *et al.* Downregulation of hemidesmosomal proteins in nasopharyngeal carcinoma cells. Cancer Lett, 2001, **163**(1): 117~123
- 6 Vanaja D K, Chevillet J C, Iturria S J, *et al.* Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res, 2003, **63** (14): 3877~3882
- 7 Lee A Y, He B, You L, *et al.* Expression of the secreted frizzled-related protein gene family is downregulated in human mesothelioma. Oncogene, 2004, **23** (39): 6672~6676
- 8 Lee J L, Chang C J, Chueh L L, *et al.* Expression of secreted frizzled-related protein 2 in a primary canine mammary tumor cell line: a candidate tumor marker for mammary tumor cells. In Vitro Cell Dev Biol Anim, 2003, **39** (5~6): 221~227
- 9 Duffy M J, Maguire T M, Hill A, *et al.* Metalloproteinases: role in breast carcinogenesis, invasion and metastasis. Breast Cancer Res, 2000, **2**: 252~257
- 10 Rudolph-Owen L A, Matrisian L M. Matrix metalloproteinases in remodeling of the normal and neoplastic mammary gland. J Mammary Gland Biol Neoplasia, 1998, **3** (2): 177~189
- 11 Dalmau J, Gultekin S H, Voltz R, *et al.* Mal, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders. Brain, 1999, **122** (1): 27~39
- 12 Ceratto N, Dobkin C. Human type I cytokeratin genes are a compact cluster. Cytogenet Cell Genet, 1997, **77** (3~4): 169~174
- 13 Luo A, Kong J. Discovery of Ca²⁺-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. Oncogene, 2004, **23** (6): 1291~1299

14 Liang Y, Tedder TF. Identification of a CD20-, FcepsilonRIbeta-, and HTm4-related gene family: sixteen new MS4A family members

expressed in human and mouse. *Genomics*, 2001, **72**(2): 119~127

Primary Research of Linear Differential Model for The Genetic Regulatory Networks of Gene Related to Breast Cancer Metastasis*

NI Chun-Sheng^{1)**}, SUN Bao-Cun^{1)***}, FENG Yu-Mei¹⁾, ZHANG Dan-Fang¹⁾,

LI Xiao-Qing¹⁾, DONG Liu-Huan^{2)**}, ZHANG Liang³⁾

¹⁾*Cancer Institute, Tianjin Medical University Cancer Hospital, Tianjin 300060, China;*

²⁾*School of Mathematical Sciences, Nankai University, Tianjin 300071, China;*

³⁾*Biochip Beijing National Engineering Center, Beijing 100084, China)*

Abstract Bioinformatics and gene chip are two new technologies in the field of life science. Bioinformatics is very important to the data analysis of gene chip. The expression and regulation of gene have a profound influence on progression and metastasis of malignant tumors. In the present report, the oligonucleotide microarray technique was used to identify gene expression profiling and screen differential expression genes in breast cancer with a special emphasis on metastasis factors. Then a linear differential model is employed to construct rough genetic regulatory network of gene related to breast cancer metastasis. An available gene chip system was used to analyze surgical samples, including both breast cancer primary tissue and metastasis tissue, collected from 30 patients in different clinical staging. 27 differential expression genes were identified. 14 of the total are up-regulation genes whose Ratio is large for 3, and the rest are down-regulation gene whose Ratio is small for 0.33. In search of putative metastasis and tumor progression factors with Bioinformatics method, linear differential model, a linear model estimated -by Lest Squares Method, was used to construct rough regulatory network of gene related to breast cancer metastasis. With the analysis of the rough regulatory network constructed by the linear differential model, the reliability of regulatory network and abnormality of signal transmission pathways caused by multiple gene abnormal expression, which maybe correlate to breast cancer metastasis, were manifested.

Key words breast cancer, gene chip, metastasis, regulatory networks, linear differential model

*This work was supported by a grant from Hi-Tech Research and Development Program of China (2002AA2Z2011).

**Both authors contribute to this work equally.

***Corresponding author . Tel: 86-22-27587000, E-mail: nknini@vip.sina.com.cn

Received: June 2, 2005 Accepted: June 30, 2005