

网络出版时间: 2017-1-20 11:13 网络出版地址: <http://www.cnki.net/kcms/detail/34.1065.R.20170120.1113.043.html>

前列腺癌基因表达谱芯片的生物信息学分析

王书鹏^{1,2} 吴松³ 蔡志明^{1,2}

摘要 目的 利用生物信息学方法分析前列腺癌基因表达谱芯片,研究前列腺癌差异表达基因的功能及调控网络。方法 采用 GEO 数据库中获取的前列腺癌基因表达谱芯片数据,利用 R 软件及 affy、limma、pheatmap、ggplot2 等 R 软件包进行数据挖掘及生物信息学分析。结合生物信息学工具 DAVID、GeneMANIA 对差异表达基因及其调控网络进行分析。结果 共筛选出前列腺癌与癌旁组织差异表达基因 56 个,表达上调 15 个,表达下调 41 个,前列腺癌与癌旁组织的差异表达基因被富集到不同的子集。其中 *cav1*、*slc16a2*、*cav2*、*slc16a5*、*magi2*、*ptrf*、*pdlim5*、*lmod1*、*abcc6* 等 9 个基因被富集到“细胞功能”分类下的“细胞膜组分”子集中,其中 *cav1*、*cav2*、*ptrf* 影响细胞膜内陷囊状结构的功能,可能在前列腺癌发生发展中发挥重要作用。结论 前列腺癌与癌旁组织的差异表达基因之间存在复杂的调控网络,生物信息学可以从中提取有效信息,为前列腺癌分子机制研究提供思路及数据基础。

关键词 前列腺癌;基因表达谱;生物信息学

中图分类号 R 737.25

文献标志码 A 文章编号 1000-1492(2017)02-0199-04
doi:10.19405/j.cnki.issn1000-1492.2017.02.010

前列腺癌是世界上男性第二大常见恶性肿瘤^[1]。在美国,每年约有 22 万例前列腺癌新发病例^[2]。在我国,前列腺癌的发病率逐步上升,前列腺癌已成为泌尿系统发病率最高的恶性肿瘤^[3]。前列腺癌容易发展成远处转移且预后差,因此早期诊断尤为重要^[4]。筛查前列腺癌生物标志物对前列腺癌的早期显示具有重要意义。

生物信息学作为结合分子生物学与信息技术的交叉学科得以飞速发展^[5]。基因表达谱芯片

作为一种高效、大规模获取基因数据的技术,广泛应用于基因芯片表达谱数据的采集。该研究利用生物信息学工具对基因芯片公共数据库 (gene expression omnibus, GEO) 的前列腺癌基因表达谱芯片进行数据分析,筛选出差异表达基因,并使用 DAVID 数据库、GeneMANIA 等工具进行相关的生物信息学分析及注释^[6],为进一步深入研究相关基因的生物功能及其参与前列腺癌发生发展的分子机制奠定基础。

1 材料与方法

1.1 材料 基因表达谱芯片数据来源于美国国家生物信息中心的 GEO 数据库中的前列腺数据集。该数据集是由 GPL570 芯片平台采集的 10 对前列腺癌及癌旁组织的基因芯片表达谱数据,前列腺癌的 Gleason 评分显示为中高度恶性,病理类型均为腺癌。

1.2 方法

1.2.1 数据处理及差异表达基因的筛选 从 GEO 数据库中获取前列腺癌芯片表达谱数据,利用 R 软件及 affy、limma、pheatmap、ggplot2 等 R 软件包进行数据挖掘及生物信息学分析。对基因芯片表达谱数据进行过滤和标准化,获取前列腺癌及癌旁组织的差异表达基因。选择统计学检验后差异最为显著的 56 个基因作为研究对象 (差异倍数 > 5.0 , $P < 0.05$)。

1.2.2 差异表达基因的功能富集分析 本研究中应用 DAVID 数据库对筛选的显著差异基因进行分析。DAVID 数据库集合多种基因注释数据库信息,可对基因数据集进行分析和注释。本研究主要应用 DAVID 数据库的基因本体 (gene ontology, GO) 模块对差异表达基因进行系统分析。

1.2.3 差异表达基因的调控网络分析 差异表达基因经 DAVID 数据库注释后获得 9 个与细胞膜结构相关的基因。考虑到细胞膜与肿瘤细胞的转移与侵袭力高度相关,本研究应用 GeneMANIA 数据库对这 9 个基因进行网络结构分析,获取基因之间的共定位、共表达、共蛋白结构域等方面的关系,研究

2016-10-17 接收

基金项目:国家重点基础研究发展计划(973 计划)项目(编号:2014CB745200);国家自然科学基金(编号:81301740)

作者单位:¹安徽医科大学深圳临床学院(深圳市第二人民医院)深圳 518035

²深圳市第二人民医院泌尿外科 深圳 518035

³深圳大学附属罗湖医院泌尿外科 深圳 518001

作者简介:王书鹏,男,硕士研究生;

蔡志明,男,主任医师,博士生导师,责任作者,E-mail:zhimingcai2000@126.com

这些基因的分子调控网络之间的关系。

2 结果

2.1 前列腺癌基因表达谱芯片的差异表达基因 对 GEO 数据库中的 10 对前列腺癌及癌旁组织的基因芯片表达谱数据进行分析,通过非监督式聚类分析绘制前列腺癌与癌旁组织差异表达基因热图,前列腺癌与癌旁组织相比,前列腺癌组织中基因表达量发生较为明显的改变,以基因表达量下调为主。同时获得了前列腺癌及癌旁组织差异有统计学意义的表达基因,共 56 个(差异倍数 > 5.0, $P < 0.05$),其中上调的达 15 个,下调的达 41 个(表 1)。热图中的颜色反映了基因表达量的水平,红色代表升高,绿色代表降低,颜色的亮度表示基因表达量升高或降低的程度(图 1)。

表 1 前列腺癌与癌旁组织的差异表达基因

基因	P 值	差异倍数	基因	P 值	差异倍数
<i>glyat1</i>	0.012	9.15	<i>rab17</i>	0.023	6.35
<i>dpysl3</i>	0.022	-8.81	<i>lrch2</i>	0.033	-6.31
<i>dlx1</i>	0.017	8.11	<i>itih5</i>	0.034	-6.31
<i>dlx2</i>	0.017	7.88	<i>hoxc4</i>	0.023	6.28
<i>mocs1</i>	0.019	-7.61	<i>calm1</i>	0.022	-6.27
<i>tbc1d1</i>	0.016	-7.51	<i>ferm2</i>	0.025	-6.24
<i>spock3</i>	0.047	-7.49	<i>vwa5a</i>	0.003	-6.21
<i>ptrf</i>	0.032	-7.43	<i>abcc6</i>	0.003	-6.21
<i>glyat1</i>	0.049	7.42	<i>cav1</i>	0.004	-6.15
<i>sri</i>	0.039	-7.38	<i>myocd</i>	0.001	-6.15
<i>hpn</i>	0.032	7.19	<i>epr1</i>	0.003	6.15
<i>stard4</i>	0.014	-7.07	<i>amacr</i>	0.011	6.13
<i>calm1</i>	0.026	-7.05	<i>tacc1</i>	0.004	-6.13
<i>gnl1</i>	0.015	-6.98	<i>ncs1</i>	0.012	-6.11
<i>rbfox2</i>	0.019	-6.94	<i>amacr</i>	0.003	6.09
<i>kctd9</i>	0.019	-6.91	<i>hoxd11</i>	0.021	-6.08
<i>hoxc6</i>	0.016	6.86	<i>cd40</i>	0.035	-6.05
<i>cdc42ep4</i>	0.021	-6.72	<i>lsm11</i>	0.034	-6.01
<i>spg20</i>	0.021	-6.71	<i>crispld2</i>	0.028	-6.01
<i>sowaha</i>	0.025	-6.65	<i>slc16a5</i>	0.005	-5.96
<i>slc16a2</i>	0.038	-6.58	<i>txn1</i>	0.001	-5.96
<i>acss3</i>	0.025	-6.53	<i>pdlim5</i>	0.006	5.94
<i>rhoq</i>	0.042	-6.52	<i>gal</i>	0.012	5.94
<i>magi2</i>	0.026	-6.51	<i>lmod1</i>	0.023	-5.94
<i>peg10</i>	0.028	-6.45	<i>sgcb</i>	0.005	-5.91
<i>fncl4</i>	0.027	-6.41	<i>pcat7</i>	0.036	5.90
<i>cav2</i>	0.034	-6.35	<i>slc43a1</i>	0.023	5.88
<i>tmem246</i>	0.035	-6.35	<i>nid1</i>	0.015	-5.72

2.2 差异表达基因的富集分析 从 GEO 数据库获得的差异表达基因需要进一步的注释,以揭示其内在关系。本研究中,采用 DAVID 数据库对差异表达基因进行生物信息学分析,结果显示,这些基因富集

在不同的功能注释模块中,根据 GO 分类,主要有分子功能、生物过程和细胞组成 3 个模块。在“细胞组成”模块中,*cav1*、*slc16a2*、*cav2*、*slc16a5*、*magi2*、*ptrf*、*pdlim5*、*lmod1* 及 *abcc6* 这 9 个基因被富集到“细胞膜组分”子集中,这些基因参与细胞膜的构建及相关功能的调节的,与细胞信号转导、细胞侵袭转移能力具有密切关联(表 2)。

2.3 富集于“细胞膜组分”子集的基因分子调控网络 上述 9 个基因被 GO 分析富集到“细胞膜组分”子集中,为进一步了解基因间的内在联系,本研究通过 GeneMANIA 数据库建立了这些基因的分子调控网络(图 2)。基因调控网络显示,*cav1* 和 *cav2* 存在共定位,*slc16a5* 和 *slc16a2* 存在相似的功能蛋白结构域,*cav2*、*slc16a2*、*cav1*、*magi2*、*ptrf*、*pdlim* 及 *lmod1* 存在密切的共表达关系,其中以 *cav1*、*cav2* 及 *ptrf* 三者关系最为密切,并且存在一定强度的物理交联。

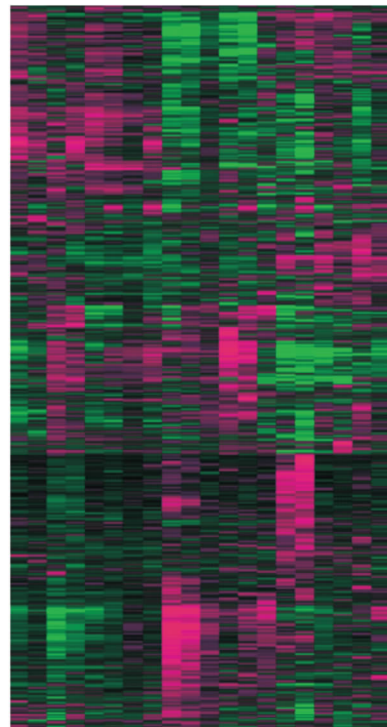


图 1 前列腺癌及癌旁组织的差异表达基因热图

绿色:基因低表达;红色:基因高表达;颜色深浅与表达量呈正相关性

3 讨论

根据美国国家癌症研究所发起的肿瘤基因组图谱计划的一系列高水平基因组学研究^[7]显示,肿瘤的发生发展往往伴随着一定数量的基因的高频突变,并且存在复杂的基因调控网络。传统的小规模

表2 差异表达基因被富集到“细胞组成”模块

项目	基因	P 值	富集倍数
GO:0005624	<i>cav2</i> , <i>slc16a2</i> , <i>cav1</i> , <i>slc16a5</i> , <i>magi2</i> , <i>ptrf</i> , <i>pdlim5</i> , <i>lmod1</i> , <i>abcc6</i>	0.013	4.06
GO:0005626	<i>cav2</i> , <i>slc16a2</i> , <i>cav1</i> , <i>slc16a5</i> , <i>magi2</i> , <i>ptrf</i> , <i>pdlim5</i> , <i>lmod1</i> , <i>abcc6</i>	0.018	3.91
GO:0044459	<i>cav2</i> , <i>cav1</i> , <i>hpn</i> , <i>magi2</i> , <i>fermt2</i> , <i>cd40</i> , <i>slc16a2</i> , <i>ketd9</i> , <i>slc16a5</i> , <i>ptrf</i> , <i>rab17</i> , <i>slc43a1</i> , <i>sgcb</i> , <i>abcc6</i>	0.029	2.32
GO:0045121	<i>cav2</i> , <i>cav1</i> , <i>ptrf</i> , <i>rhoq</i>	0.063	10.21
GO:0000267	<i>cav2</i> , <i>slc16a2</i> , <i>cav1</i> , <i>slc16a5</i> , <i>magi2</i> , <i>ptrf</i> , <i>pdlim5</i> , <i>lmod1</i> , <i>abcc6</i>	0.064	3.03
GO:0005856	<i>pdlim5</i> , <i>fermt2</i> , <i>ncs1</i> , <i>rhoq</i> , <i>dpysl3</i> , <i>lmod1</i> , <i>cde42ep4</i> , <i>tace1</i> , <i>calm1</i> , <i>sgcb</i>	0.083	2.64
GO:0005886	<i>sri</i> , <i>cav2</i> , <i>cav1</i> , <i>hpn</i> , <i>magi2</i> , <i>fermt2</i> , <i>rhoq</i> , <i>dpysl3</i> , <i>cd40</i> , <i>ketd9</i> , <i>slc16a2</i> , <i>slc16a5</i> , <i>ptrf</i> , <i>rab17</i> , <i>slc43a1</i> , <i>calm1</i> , <i>sgcb</i> , <i>abcc6</i>	0.095	1.74
GO:0005887	<i>cav2</i> , <i>ketd9</i> , <i>slc16a2</i> , <i>cav1</i> , <i>hpn</i> , <i>slc16a5</i> , <i>cd40</i> , <i>slc43a1</i> , <i>sgcb</i>	0.011	2.76
GO:0031226	<i>cav2</i> , <i>ketd9</i> , <i>slc16a2</i> , <i>cav1</i> , <i>hpn</i> , <i>slc16a5</i> , <i>cd40</i> , <i>slc43a1</i> , <i>sgcb</i>	0.012	2.71

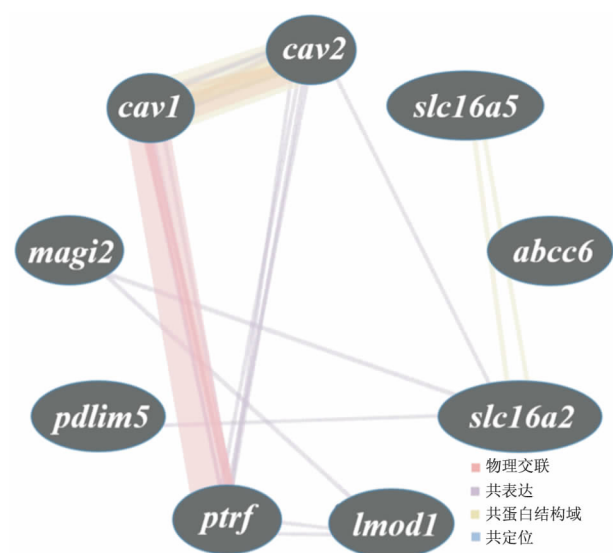


图2 差异基因中涉及细胞膜成分的基因调控网络结构图

基因之间存在4种关系:物理交联(橙色);共表达(紫色);共蛋白结构域(黄色);共定位(蓝色)

分子生物学实验通常仅能研究少数的几个基因的功能。如何研究多基因调控事件在肿瘤发生发展中的作用,是生物信息学研究的重点。

本研究从基因芯片公共数据库获取前列腺癌及癌旁组织的基因表达谱芯片数据,通过R语言等生物信息学工具分析绘制前列腺癌基因表达热图,直观展示基因差异表达的总体情况,并筛选出前列腺癌及癌旁组织差异有统计学意义表达的基因56个。同时利用DAVID数据库将差异表达基因按生物学功能划分为多个子集,其中有9个基因富集到“细胞膜组分”这一子集。其中的*cav2*、*slc16a2*、*cav1*、*slc16a5*、*magi2*、*ptrf*、*pdlim5*、*lmod1*、*abcc6*等差异基因影响细胞膜的组分,从而改变细胞膜的结构与功能,进一步影响细胞膜的侵袭能力,使前列腺癌细胞具有较强的侵袭力,并参与前列腺癌发生和发展。研究^[8]显示,其中的*magi2*基因编码的蛋白质与细胞

膜表面相关的鸟苷酸激酶同源物具有结构相似性,是前列腺癌潜在的分子标志物。

为深入研究这些基因间的相互作用关系,本研究通过GeneMANIA构建这9个基因的分子网络图,显示出基因之间的关系网络。在网络图中,*cav1*、*ptrf*、*cav2*这3个基因显示出密切的相关性。*cav1*和*cav2*基因位于7号染色体,彼此位置相邻,同属于Caveolin家族,该家族参与编码细胞膜表面的内陷囊状结构(Caveolae)的骨架蛋白^[9]。*cav1*编码的Caveolae的骨架蛋白是细胞膜微囊的主要成分,参与调控细胞的分裂与凋亡,*cav1*不仅在胞吞、细胞转运、细胞黏附、细胞信号转导中起着重要作用,而且参与有丝分裂级联反应的反馈调控。研究^[7]显示,*cav1*是一个抑癌基因,*cav1*突变与胃癌、甲状腺癌等肿瘤的发生发展相关。*cav2*同样参与构建Caveolae的骨架蛋白,直接与G蛋白 α 亚单位交联,激活MAPK1及STAT3信号通路并调控有丝分裂及细胞周期进程,*cav2*突变与乳腺癌等肿瘤发病相关^[10-11]。*ptrf*基因促进转录复合物的转录与解离,*ptrf*基因编码的蛋白也定位于细胞膜内陷囊状结构,与*cav1*一起参与细胞膜内陷囊状结构的形成,影响细胞的转运与黏附,与前列腺的转移存在一定的关联^[12]。可见,*cav1*、*cav2*、*ptrf*这3个基因与细胞膜表面的内陷囊状结构的结构和功能密切相关,可能影响细胞的黏附、侵袭能力,与肿瘤的发生发展存在密切联系。

本研究通过生物信息学方法分析前列腺癌及癌旁组织的基因表达谱芯片数据,通过基因表达谱热图直观展现了整体的基因表达差异,筛选出显著性差异表达基因,并将这些基因按生物学功能进行分类。通过GeneMANIA对“细胞功能”分类下的“细胞膜组分”子集中的基因进行网络调控结构的展示,结果表明,*cav1*、*cav2*和*ptrf*存在紧密的关联,并

且发现这 3 个基因参与细胞膜表面的内陷囊状结构,调控细胞的黏附、侵袭等功能,可能在前列腺癌的发生发展过程中起到一定的作用,为前列腺癌分子机制研究提供思路及数据基础。

参考文献

- [1] Attard G, Parker C, Eeles R A, et al. Prostate cancer[J]. *Lancet*, 2015, 387(10013):70-82.
- [2] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2015 [J]. *CA Cancer J Clin*, 2015, 65(1):5-29.
- [3] 中华医学会泌尿外科学分会前列腺癌联盟. 中国前列腺癌早期诊断专家共识 [J]. *中华泌尿外科杂志*, 2015, 36(8):561-4.
- [4] 邵胜, 徐凌凡, 徐明, 等. Propachlor 对前列腺癌细胞 PC3 及 LNCaP 的作用及其可能机制 [J]. *安徽医科大学学报*, 2015, 50(8):1045-8.
- [5] Petryszak R, Burdett T, Fiorelli B, et al. Expression Atlas update - a database of gene and transcript expression from microarray and sequencing-based functional genomics experiments [J]. *Nucleic Acids Res*, 2014, 42(Database issue):D926-32.
- [6] Huang da W, Sherman B T, Lempicki R A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources [J]. *Nat Protoc*, 2009, 4(1):44-57.
- [7] Cancer Genome Atlas Research Network, Weinstein J N, Collisson E A, et al. The cancer genome atlas pan-cancer analysis project [J]. *Nat Genet*, 2013, 45(10):1113-20.
- [8] Mahdian R, Nodouzi V, Asgari M, et al. Expression profile of MAGI2 gene as a novel biomarker in combination with major down-regulated genes in prostate cancer [J]. *Mol Biol Rep*, 2014, 41(9):6125-31.
- [9] Sugie S, Mukai S, Yamasaki K, et al. Significant association of caveolin-1 and caveolin-2 with prostate cancer progression [J]. *Cancer Genomics Proteomics*, 2015, 12(6):391-6.
- [10] Ma X, Liu L, Nie W, et al. Prognostic role of caveolin in breast cancer: a meta-analysis [J]. *Breast*, 2013, 22(4):462-9.
- [11] Yamasaki T, Seki N, Yoshino H, et al. MicroRNA-218 inhibits cell migration and invasion in renal cell carcinoma through targeting caveolin-2 involved in focal adhesion pathway [J]. *J Urol*, 2013, 190(3):1059-68.
- [12] Moon H, Lee C S, Inder K L, et al. PTRF/cavin-1 neutralizes non-caveolar caveolin-1 microdomains in prostate cancer [J]. *Oncogene*, 2014, 33(27):3561-70.

An integrative bioinformatics study of gene expression profile in prostate cancer

Wang Shupeng^{1,2}, Wu Song³, Cai Zhiming^{1,2}

(¹The Shenzhen Clinical Institute of Anhui Medical University, The Second People's Hospital of Shenzhen, Shenzhen 518035; ²Dept of Urology, The Second People's Hospital of Shenzhen, Shenzhen 518035; ³Dept of Urology, The Affiliated Luohu Hospital of Shenzhen University, Shenzhen 518001)

Abstract Objective To analyze gene expression profile for exploring the function and regulatory network of differentially expressed genes in prostate cancer by bioinformatics. **Methods** The data of gene expression profile in prostate cancer were obtained from GEO database. R software and *affy*, *limma*, *pheatmap*, *ggplot2* and other R packages were applied for data mining and bioinformatics analysis. Combined with DAVID and GeneMANIA, differentially expressed genes and their regulatory networks were annotated. **Results** These differentially expressed genes with statistical significance were 56 genes, 15 upregulated genes, 41 downregulated genes; these genes were enriched into different subgroups. *cav1*, *slc16a2*, *cav2*, *slc16a5*, *magi2*, *ptrf*, *pdlim5*, *lmod1* and *abcc6* were enriched into the “cell membrane component” subgroup of “cell component” category. *cav1*, *cav2* and *ptrf* regulated the function of caveolae, they may play an important role in the occurrence and development of prostate cancer. **Conclusion** Differentially expressed genes between prostate cancer and adjacent tissues assemble a complex regulatory network. Bioinformatics is a tool for data mining of the regulatory network, which provides ideas and data for the molecular mechanisms in prostate cancer.

Key words prostate cancer; gene expression profile; bioinformatics