

文章编号: 1007- 2985(2006) 02- 0088- 04

基因芯片癌与癌旁组织分析中药品混合对结果的影响

庄远红^{1,2}, 张亮³, 孟欣³, 任永红³

(1. 湖南省第一师范学校, 湖南长沙 410000; 2. 湖南师范大学生命科学学院, 湖南长沙 410081; 3. 生物芯心北京国家工程研究中心, 北京 100081)

摘要: 基因芯片技术已经广泛地应用到生物科学与医学研究的多个领域中. 由于实验费用的制约, 在使用芯片数目较少的情况下, 一些研究者倾向于把经过同一种处理的多个个体的 RNA 进行混合, 希望能消除个体差异. 但应用一个研究肿瘤的具体例子, 比较单个个体与多个个体的 RNA 混合后得到的基因芯片的结果表明, 多个样品的 RNA 混合后得到的结果和单个样品得到的结果, 无论从正确率、假阳性、假阴性来比较, 在统计学上都没有明显差异, 因此: 结合多个因素考虑, 建议进行基因芯片分析的 RNA 样品不用进行混合.

关键词: 基因芯片技术; 样品混合; 单个个体

中图分类号: Q523; R730. 23

文献标识码: A

基因芯片技术用很少的生物样本便能一次平行分析成千上万个基因的表达情况, 利用基因芯片技术可以从研究单个或少数几个基因而转向从基因组的水平来研究生命现象, 因而越来越多的生物和医学研究工作者正在把基因芯片技术应用到自己研究课题中来. 基因芯片技术比较成功的应用包括用它来寻找肿瘤分子标记物^[1]、研究肿瘤抗药性产生的机理^[2]、分析不同亚型细胞发育的过程^[3]以及研究不同中草药作用机理^[4]等等. 在微阵列基因芯片实验中, 一些研究者趋向于把多个相同处理的生物个体样品进行混合, 其原因主要是基于每次基因芯片分析的花费较大或者是由于单个 RNA 样品的量太少. 目前由于基因芯片技术中样品标记技术的改进, 一次典型的全基因组基因芯片分析只需要 5~ 10 μg 的总 RNA, 利用双重扩增技术甚至可以分析单细胞的 RNA 表达情况^[5-6]. 样品量少不再是基因芯片技术应用的瓶颈, 研究者最关心的是基因芯片技术花费的问题. 一些研究者相信把样品进行混合后, 可以消除个体之间的差异, 因此能更真实地反映所研究的实验因素. 目前国际上已有一些研究针对样品混合与单个个体进行芯片实验的优劣进行了比较^[7-8], 大多认为样品最好不予混合. 在此文中, 笔者用一个具体的基因芯片实验从正确率、假阳性率和假阴性率的指标上比较了单个样品与样品混合的结果.

1 材料与方法

1.1 材料

10 个胃癌患者的癌组织及其对应的癌旁组织(癌组织 5 cm 之外, 病理学正常组织)由北京肿瘤研究所提供. 基因芯片为北京博奥生物芯片有限责任公司制备, 芯片上含有人的 2 万 2 千个基因. 每个基因设计了其对应的一条 70 碱基寡聚核苷酸. 基因芯片扫描仪为北京博奥生物芯片公司生产的 LuxScan10 KTM 双通

* 收稿日期: 2005- 11- 25

基金项目: 国家“863”计划资助项目(2002AA2Z2011)

作者简介: 庄远红(1971-), 女, 湖南师范大学生命科学学院研究生, 湖南省第一师范学校讲师, 主要从事基因芯片技术研究.

道激光共聚焦扫描仪.

1.2 基因芯片实验过程

首先每个癌组织和其对应的癌旁组织 RNA 进行双荧光标记后和芯片进行杂交. 然后, 每个癌组织样品都取等量 RNA 进行混合, 每个癌旁组织也取等量 RNA 进行混合, 混合后的癌与癌旁组织 RNA 双荧光标记后和芯片杂交, 进行荧光交换, 使用 2 张芯片. 有关 RNA 的抽提、RNA 的标记、芯片的杂交、清洗、扫描具体过程可参照网址: <http://www.capitalbio.com/Product/protocols.htm>.

1.3 差异基因的选择

芯片扫描得到的 TIFF 图片用美国 Axon 公司的 GenePix Pro 软件(4.0 版)分析得到基因在 cy5 和 cy3 通道的杂交数据信号, 采用与基因信号强度及基因所在芯片位置相关的 LOWESS 归一化方法对基因在癌与癌旁组织中的表达信号比值进行归一化^[9]. 在用统计学方法(例如 t- test 方法)决定某个基因的表达是否发生变化时, 往往需要此基因在同样的一对样品中至少有 4 个重复的数据点, 相对于 22 K 的全基因组芯片, 芯片上每个基因只有一个点, 因此每对样品就需要至少做 4 张芯片, 这在实际的实验中难以做到; 目前在全基因组芯片分析差异表达基因时, 一对样品往往做 2 张基因芯片分析, 这 2 张芯片之间进行荧光交换以消除 cy5 和 cy3 荧光标记的偏向性^[10]. 笔者在测试每对癌组织及其对应的癌旁组织时也进行荧光交换, 使用 2 张芯片. 在芯片平台验证的实验中, 经过多次分析同一份 RNA 自己和自己做比较, 在检测到 40% ~ 48% 的基因数目时, 发现变化 2 倍以上的基因数目在 40~ 50 个之间(即假阳性的基因数目), 远远低于单通道技术 Affymetrix 公司基因芯片技术平台的假阳性率^[11]. 因此结合基因芯片平台检测的精确性, 在对每张芯片进行分析时, 把变化在 2 倍以上的基因认为是变化基因; 而对于一对样品荧光交换的 2 张芯片之间, 只有在 2 张芯片上变化趋势是一致的基因才认为是在此对样品中真正变化的基因.

2 结果与讨论

10 对癌组织和对应的癌旁组织比较后, 通过等阶聚类分析, 可以发现一类基因在癌组织中都是上调的, 也有一类基因在癌组织中全都下调了(如图 1).

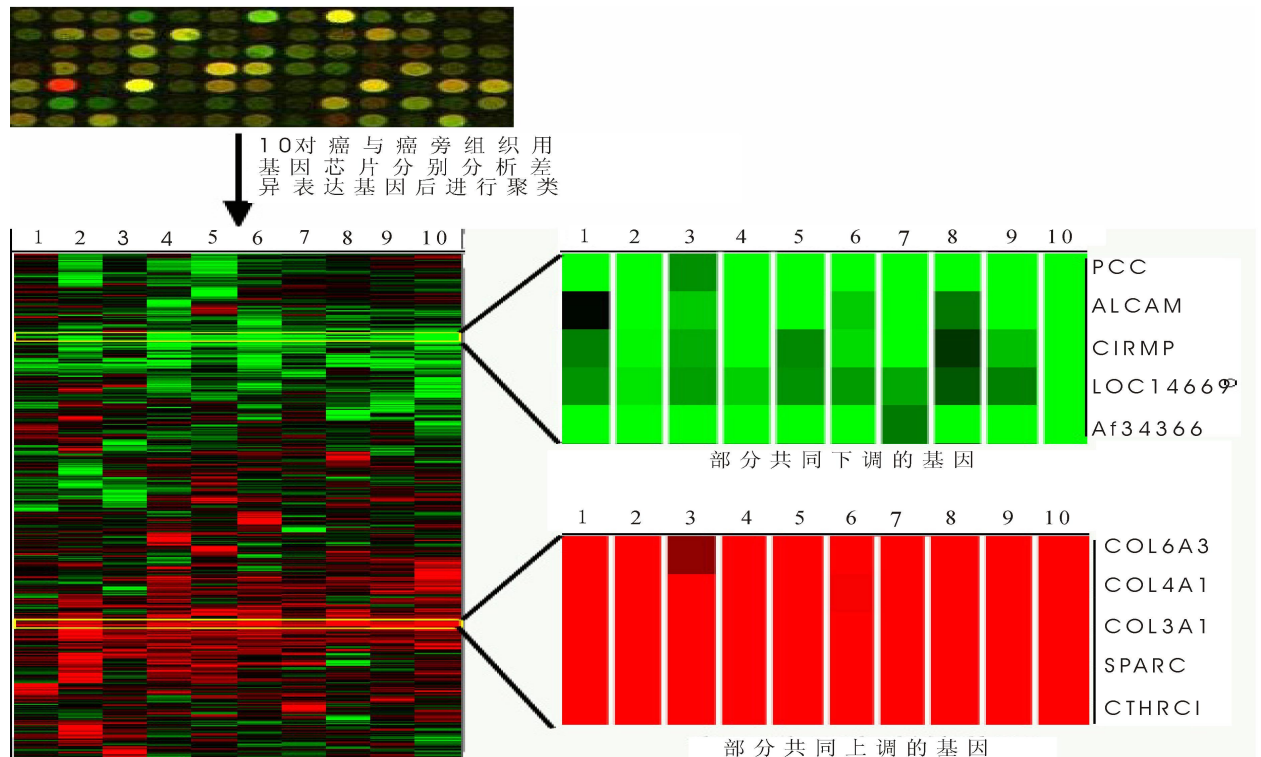


图 1 用等阶聚类法分析在样品中有相同变化趋势的基因

结果表明至少在 8 对样品中都发生同样变化的基因, 数目是 316 个, 在技术上可认为这 316 个基因是与个体差异无关而与所研究的肿瘤因素相关. 在本文的个体和样品混合的比较研究中, 笔者从芯片技术筛选的角度称此 316 个基因为与肿瘤相关的“标准基因”, 每个个体的癌组织与癌旁组织, 10 个个体的癌组织混合物与癌旁组织混合物得到的结果都与此类“标准基因”进行比较, 比较结果见表 1.

表 1 不同个体的癌组织与癌旁组织之间的差异基因数目

个体编号	实际检测的 差异基因数目	正确基因 数目	正确率/%	假阳性 基因数目	假阳性率/%	假阴性 基因数目	假阴性率/%
1	955	166	17.4	789	82.6	150	15.7
2	1 411	142	10.1	1 269	89.9	174	12.3
3	1 532	169	11.0	1 363	89.0	147	9.6
4	1 944	285	14.7	1 659	85.3	31	1.6
5	1 749	221	12.6	1 528	87.4	95	5.4
6	1 766	294	16.6	1 472	83.4	22	1.2
7	1 273	213	16.7	1 060	83.3	103	8.1
8	2 083	253	12.1	1 830	87.9	63	3.0
9	1 588	232	14.6	1 356	85.4	84	5.3
10	2 718	270	9.9	2 448	90.1	46	1.7
10 个个体 等量混合	1 856	243	13.1	1 613	86.9	73	3.9

表中正确基因为检测出有差异并包含在标准基因中的基因; 正确率为包含在标准基因中的数目/实际检测的差异基因数目; 假阳性基因为检测出有差异但并非包含在标准基因中的基因; 假阳性率为假阳性基因数目/实际检测的差异基因数目; 假阴性基因为没有检测到的标准基因; 假阴性率为假阴性基因数目/实际检测的差异基因数目.

首先可以用 SPSS10.0 for windows 统计学软件分析 10 个个体的癌组织和癌旁组织得到差异基因与标准基因比较, 在 95% 显著性水准上, 10 个个体之间的正确率、假阳性率和假阴性率都符合正态分布(经正态性检验, 结果显示 p 均大于 0.05), 因此可以用 t 检验判断 10 对样品单独杂交与样品混合杂交相比, 结果有无显著性差异. 在基因芯片中, 为了比较单个个体的癌组织与癌旁组织 RNA 得到的结果与 10 个个体的癌组织与癌旁组织 RNA 混合得到的结果, 采用统计学单样本 t 检验公式 $t = \frac{x - x_0}{sd/\sqrt{n}}$, 从正确率、假阳性率和假阴性率来

判断混合的样品与 10 个个体单独比较有无显著性差异. 以正确率为例, x 是 10 个个体癌组织与癌旁组织分别比较得到的差异基因数目的正确率的均数, x_0 是 10 个个体癌组织混合与对应的癌旁组织混合比较得到的差异基因数目的正确率, n 是样品量(这里为 10), sd 是 10 对样本正确率的标准差. 计算出 t 值后, 根据 t 值和自由度查统计用表, 可以得到相应的 p 值, 结果见表 2

表 2 10 对样品单独杂交与混合杂交比较

项目	正确率	假阳性率	假阴性率
t 值	0.553	-0.553	1.577
p 值	0.594	0.594	0.149

结果显示, 在 95% 显著性水准上, 10 对样品单独杂交与样品混合杂交比较, 在正确率、假阳性率和假阴性率上都无显著性差异(p 值均大于 0.05).

进一步, 若在 10 对样品中找出分别至少在 5 对、6 对、7 对、9 对、10 对样品中都有同样变化趋势的基因作为“标准基因”, 得到的与肿瘤相关的“标准基因”数目分别是 1 711 个、1 059 个、605 个、136 个、47 个, 也分别采用统计学单样本 t 检验, 从正确率、假阳性率和假阴性率来判断混合的样品与 10 个个体单独比较有无显著性差异, 各种情况下得到的 p 值见表 3.

表 3 不同“标准基因”下 10 对样品单独杂交与混合杂交比较的 p 值

“标准基因”样品数目	正确率	假阳性率	假阴性率
5 对	0.241	0.241	0.211
6 对	0.653	0.653	0.159
7 对	0.684	0.684	0.145
9 对	0.331	0.331	0.160
10 对	0.210	0.210	0.165

从表 3 可以看出, 即使用不同的标准来筛选与所研究的肿瘤因素相关的“标准基因”, 在 95% 显著性水准上, 10 对样品单独杂交与样品混合杂交在正确率、假阳性率及假阴性率上都无显著性差异(p 值均大于 0.05)。

从以上寻找与肿瘤发生相关基因一个实际的例子可以看出, 单个个体和多个个体样品混合后与基因芯片进行杂交反应, 实际上都只是一次实验, 每一次杂交得到的变化基因绝大多数和所研究的肿瘤因素没有关系(假阳性率超过 75%, 见表 1), 因而不能把从一对样本得到的结果来解释所研究的生物学因素, 而要选择多个个体中有共同变化趋势的基因来分析所研究的实验因素。在用基因芯片技术从分子水平研究疾病的机理、药物对疾病的治疗作用、细胞对外界的应激反应等过程中, 一些研究者趋向于把多个个体的样品进行混合, 希望先做少数几张芯片以做一个初步的探讨; 笔者的数据显示把样品进行混合后, 从正确率、假阳性率和假阴性率来看, 和单个个体实验并无显著差异。因此建议即使是初步的探讨性实验, 也可用个体动物的 RNA 材料来进行基因芯片分析, 一方面可以节省动物材料, 另一方面, 若进行后续研究, 则最初的动物个体得到的数据可以和后续的个体动物之间形成延续性, 方便后续的数据分析。笔者用以上双通道激光的基因芯片技术从肿瘤混合样品与单个样品比较得到的结果和 Affymetrix 芯片公司单通道激光的基因芯片技术得到的结果是类似的(http://www.affymetrix.com/support/technical/technote/pooling_technote.pdf)。

参考文献:

- [1] VEER L J, DAI H, VAN DE VIJVER M J, et al. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer [J]. *Nature*, 2002, 415(6871): 530–536.
- [2] CHEN C D, WELSBIE D S, TRAN C, et al. Molecular Determinants of Resistance to Antiandrogen Therapy [J]. *Nat. Med.*, 2004, 10(1): 33–39.
- [3] CHIANG M K, MELTON D A. Single-Cell Transcript Analysis of Pancreas Development [J]. *Dev. Cell*, 2003, 4(3): 383–393.
- [4] IIZUKA N, OKA M, YAMAMOTO K, et al. Identification of Common or Distinct genes Related to Antitumor Activities of a Medicinal Herb and Its Major Component by Oligonucleotide Microarray [J]. *Int. J. Cancer*, 2003, 107(4): 666–672.
- [5] KAMME F, SALUNGA R, YU J, et al. Single-Cell Microarray Analysis in Hippocampus CA1: Demonstration and Validation of Cellular Heterogeneity [J]. *J. Neurosci*, 2003, 23(9): 3607–3615.
- [6] DIXON A K, RICHARDSON P J, PINNOCK D, et al. Gene-Expression Analysis at the Single-Cell Level [J]. *Trends. Pharmacol. Sci.*, 2000, 21(2): 65–70.
- [7] SHIH J H, MICHALOWSKA A M, DOBBIN K, et al. Effects of Pooling mRNA in Microarray Class Comparisons [J]. *Bioinformatics*, 2004, 20(18): 3318–3325.
- [8] KENDZIORSKI C M, ZHANG Y, LAN H, et al. The Efficiency of Pooling mRNA in Microarray Experiments [J]. *Biostatistics*, 2003, 4(3): 465–477.
- [9] YANG Y H, DUDOFF S, LUU P, et al. Normalization for cDNA Microarray Data: a Robust Composite Method Addressing Single and Multiple Slide Systematic Variation [J]. *Nucleic Acids. Res.*, 2002, 30(4): 15.
- [10] WEIGELT B, GLAS A M, WESSELS L F, et al. Gene Expression Profiles of Primary Breast Tumors Maintained in Distant Metastases [J]. *Proc. Natl. Acad. Sci. USA*, 2003, 100(26): 15901–15905.

(下转第 100 页)

- [12] FORSYTH M, WONG, S, NAIRN K M., et al. NMR Studies of Modified Nasicon-Like, Lithium Conducting Solid Electrolytes? [J]. Solid State Ionics, 1999, 124: 213– 219.
- [13] BEST A S, NEWMAN P J, MACFARLANE D R, et al. Characterisation and Impedance Spectroscopy of Substituted $\text{Li}_{1.3}\text{Al}_{0.3}\text{Ti}_{1.7}(\text{PO}_4)_3-x(\text{ZO}_4)_x$ ($Z= \text{V}, \text{Nb}$) Ceramics [J]. Solid State Ionics, 1999, 126: 191– 196.
- [14] BIRKE P, SALAM F, DRING S, et al. A First Approach to a Monolithic All Solid State Inorganic Lithium Battery [J]. Solid State Ionics, 1999, 118: 149– 157.
- [15] PADHI A K, NANJUNDASWAMY K S, GOODENOUGH J B. Phospho-Olivines as Positive-Electrode Materials for Rechargeable Lithium Batteries [J]. J. Electrochem Soc., 1997, 144: 1 188– 1 194.
- [16] MASQUELIER C, PADHI A K, NANJUNDASWAMY K S, et al. New Cathode Material for Rechargeable Lithium Batteries: The 3-D Framework Structures $\text{Li}_3\text{Fe}_2(\text{XO}_4)_3$ ($X= \text{P}, \text{As}$) [J]. J. Solid State Chemistry, 1998, 135: 228– 234.

Preparation of $\text{Li}_{1.3}\text{Al}_{0.3}\text{Ti}_{1.7}(\text{PO}_4)_3$ Thin Films by Sol-Gel Technique and the Film Characteristics

XIAO Zhuo-bing, MA Ming-you

(College of Chemistry and Chemical Engineering, Jishou University, Jishou 416000, China)

Abstract: Solid electrolyte of $\text{Li}_{1.3}\text{Al}_{0.3}\text{Ti}_{1.7}(\text{PO}_4)_3$ thin films were prepared by sol-gel technique. The phase identification and surface morphology were investigated by X-ray diffraction and scanning electron microscopy. Electrochemical window and electronic insulating properties were examined by cyclic voltammetry and current testing instrument. The results show that the films with good crystallinity are homogeneous, dense and crack-free. The electrochemical window is more than 2.4V, and the electronic conductivity is $10^{-11} \text{ S} \cdot \text{cm}^{-1}$. The film possesses wide electrochemical window and good electronic insulating.

Key words: lithium ion; thin films; electrochemistry; cyclic voltammetry

(责任编辑 易必武)

(上接第 91 页)

- [11] PIPER M D, DARAN-LAPUJADE P, BRO C, et al. Reproducibility of Oligonucleotide Microarray Transcriptome Analyses [J]. An Interlaboratory Comparison Using Chemostat Cultures of *Saccharomyces Cerevisiae*. J. Biol. Chem., 2002, 277(40): 37 001– 37 008.

Effects of Pooling RNA on the Research in the DNA Microarray Analysis of Cancer and Its Corresponding Adjacent Tissues

ZHUANG Yun-hong^{1,2}, ZHANG Liang^{3*}, MENG Xin³, REN Yong-hong³

(1. Hunan First Normal College, Changsha 410000, China; 2. College of Life Science, Hunan Normal University, Changsha 410081, China; 3. National Engineering Research Center for Beijing Biotech Technology, Beijing 100081, China)

Abstract: Microarray technology has been applied in a variety of biological and medical research fields. Due to the relative higher costs, some researchers are apt to pooling RNA samples extracted from different subjects onto a single microarray to defray the cost of microarray experiment. The basic assumption of pooling is that the pooling RNA can lessen the diversity between individual subjects. Employing an example in cancer research, the authors analyzed the differential expressed genes between gastric cancer tissues and its corresponding adjacent normal tissues using ten individual RNA and their pooling RNA. The results indicate that there are no significant variances between pooling and individual RNA concerning of the numbers of correct gene, false positive gene and false negative gene based on a statistical approach. So the authors suggest using RNA from individual subject instead of pooling them even if in the primary studies.

Key words: DNA microarray; pooling RNA; individual subject

(责任编辑 易必武)