

研究报告

Research Report

多种类型癌症基因表达数据 Meta 分析以预测生物标志物和药物靶标

Shashank K.S.¹ Mamatha H.R.¹ Prashantha C.N. ✉

¹ 信息科学系, PES 技术研究所, 班加罗尔, 印度

² 生物科学系, 生物科学中心, 班加罗尔, 印度

✉ 通讯作者: prashantha.cn@gmail.com

计算分子生物学, 2015 年, 第 4 卷, 第 11 篇 doi: 10.5376/cmb.cn.2015.04.0011

本文首次发表在 Computational Molecular Biology 上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

建议最佳引用格式:

Shashank K.S., Mamatha H.R., and Prashantha C.N., 2015, Meta Analysis of Gene Expression Data of Multiple Cancer Types To Predict Biomarkers and

Drug Targets Interactions in Ovarian Cancer, Computational Molecular Biology, 5(5): 1-9

摘要 多种癌症(例如: 乳腺癌, 结肠癌和卵巢癌)的基因表达数据的 Meta 分析, 可用于鉴定预后和分子诊断标志物等功能基因的特征。但是, 寻找一种可靠的能够鉴别不同癌症类型的基因特征的方法仍然是一种挑战。本研究的目的是开发微阵列统计数据分析方法和 SVM 分类器, 以确定在不同的癌症类型中的差异表达基因。使用我们的方法来测试 16 组数据, 如 6 个乳腺癌, 4 结肠癌和 6 卵巢癌的数据。我们利用 4 种方法对结果进行分析: (a) 对数据进行预处理, 通过删除空白值和非有效值以确定表达数据的质量($p < 0.05$); (b) 利用统计学分析基因差异表达情况以预测上调和下调的基因表达情况; (c) 根据癌症的类型进行数据的分组; (d) 基因网络预测, 确定基因-基因的相互作用以了解生物标记。我们已经预测了乳腺癌中的 8 个标志物, 结肠癌中的 10 个标志物和卵巢癌中的 16 个标志物, 为癌症的诊断和治疗发展提供了新的方向。

关键词 乳腺癌; 结肠癌; 卵巢癌; 基因芯片; 统计; Limma; Biocoductor; geNETClassifier

Meta Analysis of Gene Expression Data of Multiple Cancer Types To Predict Biomarkers and Drug Targets

Shashank K.S.¹ Mamatha H.R.¹ Prashantha C.N. ✉

¹ Department of Information Science, PES Institute of Technology, Bangalore, India

² Department of Biological sciences, Scientific Bio-Minds, Bangalore, India

✉ Corresponding author, prashantha.cn@gmail.com

Abstract Meta-analysis of gene expression data of multiple cancer types such as breast, colon and ovary used to identify gene signatures that functionally used as a marker to prognosis and molecular diagnostics. There is a reliable identification of gene signatures is associated with different cancer types remains a challenge. The aim of this study is to develop microarray statistical data analysis methods and SVM classifiers to identify differentially expressed genes in different cancer types. Using our method to perform 16 datasets such as 6 breast cancers, 4 colon cancer and 6 ovarian cancers of different datasets. Our results is analyzed in 4 different methods (a) preprocess the data to identify quality expression of datasets by removing null values and non-significant values ($p < 0.05$) (b). Differential gene expression analysis using statistical analysis to predict upregulation and downregulated gene signatures (c) subgrouping of datasets that has been classified based on cancer types (d) gene network prediction to identify gene-gene interaction to understand biological markers. We have predicted 8 markers in breast cancer, 10 markers in colon cancer and 16 markers in ovarian cancer is providing new direction for diagnostics and therapeutic development.

Keywords Breast cancer; Colon cancer; Ovarian cancer; Microarray; Statistics; Limma; Biocoductor; geNETClassifier

收稿日期: 2014 年 8 月 17 日

接受日期: 2014 年 9 月 25 日

发表日期: 2014 年 10 月 16 日

癌症是疾病中很大的一个家族, 能够危害人类的健康和生活。据调查显示, 世界范围内, 2014-15 年度有 22% 的疾病死亡是由于癌症引起的(癌症的病例 297 例, WHO, 2014)。在印度, 癌症是仅次于

心脏病的第二大致死性疾病。关于癌症的统计调查显示, 82%的女性受到乳腺癌的影响, 62%的男性和女性受到结肠癌的影响, 90%的女性受到卵巢癌的影响(Matsushita K et al., 2010)。在早期, 癌症诊断技术已呈现出下降的趋势, 在 21 世纪前十年, 分子生物标志物的出现有助于在早期识别癌症疾病。在目前的研究中, 使用乳腺癌、结肠癌和卵巢癌这三种癌症模型是有助于利用微阵列技术确定分子标记。

在本年度内, 全球范围内有 910 万名女性受到乳腺癌的影响, 此外, 每年有 232 670 名女性在被诊断患有乳腺癌。其中有 30%的女性人口是由于遗传异常引起的, 如 BRCA1 和 BRCA2 基因突变(Dumitrescu et al., 2005)。此外, 还有一些其他的致癌基因如 p53、PTEN、K-RAS、NBS1 等也能够引起乳腺癌(Honrado E et al., 2006)。结肠癌也经常影响到其他组织, 例如肺部和乳腺组织。K-ras、p53、APC、GSK-3 β 和 β -catenin 基因的改变, 主要影响 Wnt- β -acatenin 信号通路, 进而影响乳腺癌和卵巢癌(Vogelstein et al., 1988; Fearon et al., 1990)。对于女性而言, 卵巢上皮癌也是一种非常危险的癌症类型, p53、MDM2、MCI-2、BCL-XL、EGFR 和 NOXA 等基因的突变主要参与卵巢癌(Baekelandt et al., 1999; Kupryjanczyk et al., 2003; Nielsen et al., 2004)。一些遗传标记已经被提出用于识别癌症, 如乳腺癌的 BRCA1、BRCA2 基因, 结肠癌的 APC、GSK-3 β 基因和卵巢癌的 CA125 基因。此外, 还有很多的血清标志物, 有助于临床上对于乳腺癌、结肠癌和卵巢癌的诊断。然而, 其检测更多的遗传标记的有效性却受到了怀疑, 广大的研究者建议应该使用更先进的技术, 例如 DNA 微阵列技术对成千上万个与癌症显著性相关的基因进行遗传识别(Golub., 2001; Elvidge., 2006)。

在此, 我们在微阵列数据的基础上提出了计算方法来预测三种类型癌症的遗传标记。采用不同的算法对这些数据进行预处理, 生物显著预测强度值显示筛选的数据集的质量; 利用统计技术来预测上调和下调的差异基因表达。基于高层次的计算能力, 我们比较了 Meta 分析来预测多种癌症类型的标记。使用基因-基因网络研究显示了不同的显著基因, 特别是对多种癌症类型的具体检测。我们相信这些新的基因, 显示基因图谱的表达将提供高价值的标志物, 在癌症的诊断和治疗上提供新的方法。

1 材料与方法

乳腺癌, 结肠癌和卵巢癌这三种不同的癌症疾病的 Meta 分析的数据从 GEO 数据库中检索。乳腺癌的数据(GEO ID: GSE30543) 6 例, 关于 sum149 与 siRNA 的对照和 siRNA 对 TIG1 复制子的靶向作用(Wang et al., 2013)。结肠癌的数据(GSE34299)包括 4 株 HT29 亲代细胞株和 HT29RC PLX4720 耐药细胞株, 其是通过提高药物的浓度以提高获得菌株的性耐药(Mao et al., 2012)。卵巢癌数据(GSE35972)包括 6 个未经处理的 TOV112D 细胞与不同生物学重复处理的 NSC319726 细胞样品(Yu et al., 2012)。所有不同的样本数据进行了 GPL570 (HG-U133_PLUS_2) Affymetrix 人类全基因组芯片平台分析。在不同的转录突变体中, HG-U95Av2 探针是相同的复制, 所有的 RNA 探针均来自 RefSeq、dbEST 和 GenBank。序列簇从 UniGene 数据库获取构建, 基因的名称来自于公开可用的数据库。运用统计分析软件如 R 和 Bioconductor 对乳腺癌、卵巢癌和结肠癌的基因进行分析预处理和差异表达基因分析, 以作为潜在的药物靶点。

1.1 原始微阵列数据的预处理

使用 Bioconductor 包的 Affy 和 Affycoretools 用于预处理数据(Team., 2012; Robinson et al., 2010), 利用不同的算法, 如使用 RMA 和 MAS5 算法对数据进行预处理, 有助于消除所有探针数据的前景和背景。使用不同的统计技术方法如常量、变量等处理标准化的探针, 以对 PM 和 MM 值进行预测修正。然而, MM 探针的信号强度通常大于 PM 探针, MM 探针可以检测到真正的信号和背景信号。校正后的信号强度水平用于分析每种疾病的差异基因表达。

1.2 差异基因表达分析

数据进行预处理后, 生成的 CEL 文件用于分析细胞差异基因的表达情况。利用 Limma 软件包对来自于 microRNA 样品的差异基因表达数据进行分析。对于两个样品的对照组和癌症组出现的差异水平的改变进行分析, 发现基因表达发生了上调或者下调。对样品中表达模式相同的一类基因进行聚类分析, 或者对具有相似表达模式基因的样品进行聚类, 每个样本进行多组重复。一个探针在组中的表达水平被认定为该组所有重复处理的平均表达水平。因此, 差异表达的问题就是方法的比较。当存在两个样本组的时候, 这仅仅是一个测试(Prashantha et al., 2013)。聚类试验能够对上调和下调的基因进行聚类

分析, 相关系数 B 值能够对空白-癌症组数据进行差异表达基因的层次聚类分析。用树状图表示各个目标之间的关系(Parker et al., 2009), 使用 GO.db 包工具对差异表达基因进行注释。利用 GO 注释的 HGU 133+2 包能够有助于了解发生差异表达基因所参与的生物过程、分子功能或者细胞组分等系统分类功能。

1.3 差异表达基因的比较分析

利用遗传学分类算法, 通过基因网络分析, 对不同的疾病进行差异表达基因分类。全基因组关联研究对表达的基因群, 或者等级基因的表达矩阵文件, 以及不同变量的探针进行优化整理。利用基于分类器的多种 SVM 对所需要的基因进行分类选择; 对于基因间共有的信息(相互作用)和共表达的信息(有关系的)利用算法进行统计和分析。这些分析能够评估可变量因素间的关联度, 据此可以为每个等级内的基因建立一个网络。这些网络是可以绘制的, 根据每种疾病的特征提供具有综合该属性的基因群。

1.4 功能注释和富集分析

为了从细胞水平获得这些差异表达基因的功能, 我们利用 GO 数据库对这些基因的功能和所在细胞的位置进行了分类。我们通过使用 Cluster Profiler 执行 GO 聚类分析, 来分析这些差异表达基因在细胞内在分子功能和生物学过程中对细胞产生的影响。注释的数据库、可视化以及数据集成工具被用来确定代表生物功能和途径的差异表达基因。

2 结果与讨论

本研究的重点是对于三种最常见的癌症类型的数据分析, 例如从公共数据库 GEO 中获取的 6 组乳腺癌、4 组结肠癌和 6 组卵巢癌的微阵列数据。该数据集包含有与对照组相对应的癌症基因组数

据, 有助于预测个体癌症类型或癌症基因与基因之间的相互作用的药物靶点。

2.1 单个癌症类型的药物靶点预测

我们搜集了每种癌症类型的个体数据, 根据癌症类型和其控制的组织对基因表达模式进行了分类。具体而言, 所有的数据被分类, 根据特别区分的癌症类型和其控制的组织进行药物靶点的预测。此外, 我们对根据基因与基因之间的相互作用而分类的显著表达的上调或下调的 K -genes 进行了分类。

2.1.1 乳腺癌

乳腺癌的数据集中包括 6 个样本, 其中 3 组 SUM149 细胞转染 siRNA 作为对照组, 另 3 组 SUM149 细胞转入 siRNA 靶点 tarzarotene 诱导的 TIG1 基因。这 6 个样本利用 hgu133+2 对 54675 个基因进行了注释, 利用传统的方法, 通过 P 值的显著性来筛选基因, 我们筛选出 54675 个基因, 其中 12788 个基因与表达显著相关。在乳腺癌中发生表达差异的基因中, 有 1220 个基因表达上调, 11568 个基因表达下调。利用 SVM 分类法, 我们已经鉴定了 1275 个最常见的与乳腺癌显著相关的基因。在 1275 个基因中有 751 个编码蛋白质的基因, 这些蛋白质编码基因, 有助于预测疾病的靶基因, 进而帮助药物靶标控制疾病(图 1)。利用功能注释和富集分析 130 个显著上调的基因, 例如 SHISA2、FBXO23、mmp7、fn1、Cfi、Egr1、DCLK1、DCN、SERPINB3、SERPINB4、MAP3K4、ITGBL1、OLFML3、NPY1R 和 PHLDA1 基因, 这些基因主要与乳腺癌的转录调控相关(表 1)。使用基因与基因间的相互作用研究各类基因发现 30 个与基因调控显著相关的基因。有 8 个基因如 fbxo23、MMP7、FN1、CFI、DCN、SERPINB3、SERPINB4 和 MAP3K4 在血清内乳腺细胞表达, 可以作为潜在的血清标志物(表 1 和表 2, 图 2)。

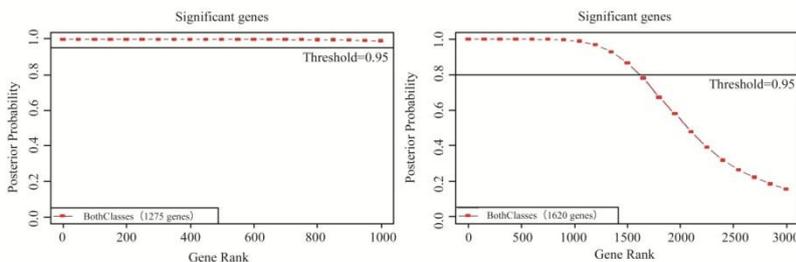


图 1 乳腺癌细胞中疾病组和对照组发生差异表达的重要基因

Figure 1 Significant genes of breast cancer cells that differentially expressed in disease and control tissues

表 1 差异表达基因与乳腺癌的基因突变显著相关

Table 1 differentially expressed genes are significantly associated with genetic mutation of breast cancer

Genes	pValue	FDR B&H	FDR B&Y	Bonferroni
SHISA2, FBXO32, MMP7, FN1, CF1, EGR1, DCLK1	4.81E-10	9.85E-07	8.08E-06	9.85E-07
FBXO32, FN1, DCN, CF1, SERPINB3, SERPINB4, MAP3K4	1.30E-09	1.34E-06	1.10E-05	2.67E-06
ITGBL1, EBXO32, MMP7, FN1, DCN, CF1, OLFML3, MAP3K4	6.36E-09	4.35E-06	3.57E-05	1.30E-05
NPY1R, FN1, DCN, CF1, PHLDA1, EGR1, DCLK1	2.99E-08	1.53E-05	1.26E-04	6.13E-05
FBXO32, FN1, DCN, CF1, SERPINB3, SERPINB4	7.49E-08	3.07E-05	2.52E-04	1.54E-04

表 2 乳腺癌中参与基因与基因相互作用的重要基因

Table 2 Significant genes in breast cancer that is involved in gene-gene interaction

Gene Ranking	Gene ID	Mean Diff	Express UP/Down	Redundant
1	241763_s_at	1.92	UP	FALSE
2	1558048_x_at	1.4645	UP	TRUE
3	241762_at	1.4404	UP	TRUE
4	225803_at	1.352	UP	TRUE
5	211906_s_at	1.3315	UP	TRUE
6	202270_at	1.3159	UP	TRUE
7	231577_s_at	1.2727	UP	TRUE
8	205440_s_at	1.2299	UP	TRUE
9	214702_at	1.3266	UP	TRUE
10	202269_x_at	1.1202	UP	TRUE
11	204259_at	1.2022	UP	TRUE
12	227404_s_at	1.1577	UP	TRUE
13	209720_s_at	1.1845	UP	TRUE
14	229800_at	1.0121	UP	TRUE
15	225328_at	1.0491	UP	TRUE
16	218162_at	1.0409	UP	TRUE
17	238623_at	1.0733	UP	TRUE
18	230493_at	-0.8728	DOWN	FALSE
19	209719_x_at	0.966	UP	TRUE
20	210413_x_at	0.9609	UP	TRUE
21	222940_at	0.9466	UP	TRUE
22	201694_s_at	0.9359	UP	TRUE
23	211896_s_at	0.8786	UP	TRUE
24	217997_at	-0.9454	DOWN	TRUE
25	205422_s_at	1.0944	UP	TRUE
26	219553_at	-0.8652	DOWN	TRUE
27	203854_at	0.8962	UP	TRUE
28	224549_x_at	0.8446	UP	TRUE
29	209335_at	0.8693	UP	TRUE
30	232573_at	0.8754	UP	TRUE

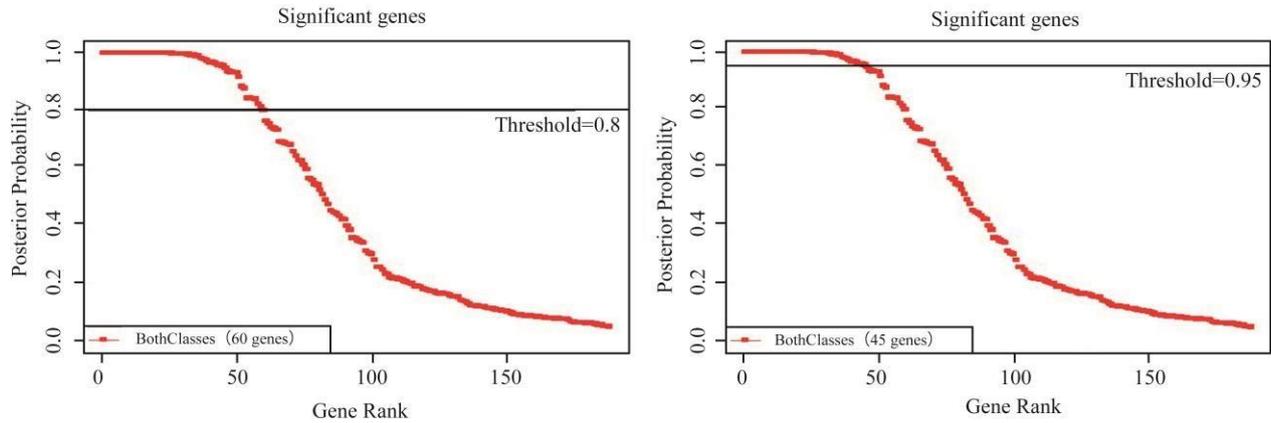


图 4 结肠癌细胞中疾病组和对照组发生差异表达的重要基因

Figure 4 Significant genes of colon cancer cells that differentially expressed in disease and control tissues

2.1.2 结肠癌

我们的分析采用 4 个数据集，其中 2 个样品为 HT29 亲代细胞系，另外 2 个样本为 HT29RC PLX4720 耐药细胞株。我们已经发现，在 HT29RC 细胞系中，有 268 个基因的表达发生了上调，有 1268 个基因的表达下调与细胞系和所控制的组织有关。利用 SVM 分类法从结肠癌数据中鉴定出具有显著差异表达的基因 60 个，其中有 45 个基因具有蛋白表达，这些基因结合在一起在结肠癌组织中

发生上调和下调表达(图 4)。我们发现 10 个基因在结肠组织的血清中普遍表达，并有可能作为一个最好的预测功能的生物标志物。对这些差异表达基因的功能注释和富集分析结果显示，KRAS、DPT、PROM1、MMP1、MMP7、FBN2、MAOB、SPRR3、PHLDB2、EMP1、DCLK1 和 AKAP12 是结肠癌的调节基因(表 3 和表 4)。MMP1 和 MMP7 基因主要参与结肠组织中的免疫调节，K-ras 基因是一个致癌基因，其通过抑制 p53 基因表达与 MDM2 蛋白受体而进行转录调控以引起结肠癌(图 3)。

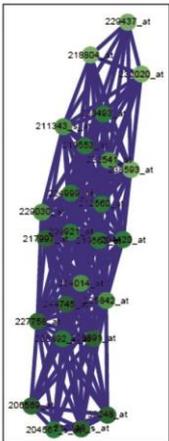


图 2 利用 geNETClassifier 预测乳腺癌数据中基因-基因相互作用网络

Figure 2 Gene-gene interaction network predicted using geNETClassifier on breast cancer data

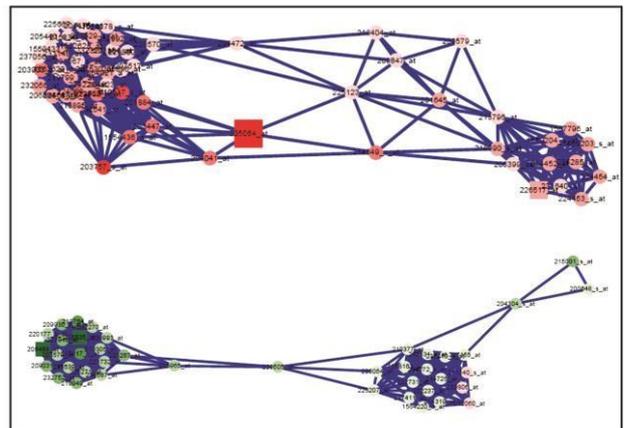


图 3 利用 geNETClassifier 预测结肠癌数据中基因-基因相互作用网络

Figure 3 Gene-gene interaction network predicted using geNETClassifier on colon cancer

表 3 差异表达基因与结肠癌的遗传突变显著相关

Table 3 Differentially expressed genes is significantly associated with genetic mutation of colon cancer

Genes	pValue	FDR B&H	FDR B&Y	Bonferroni
MMP1, MAOB, SPRR1A	1.08E-04	2.82E-01	1.09E-02	9.16E-02
SPRR1A, SPRR1B	1.38E-04	3.61E-01	1.18E-02	9.92E-02
EMP1, MAOB, DCLK1, AKAP12	2.70E-04	7.04E-01	1.46E-02	1.23E-01
PROM 1	1.18E-03	6.49E+00	3.26E-03	2.99E-02
KRAS	1.18E-03	6.49E+00	3.26E-03	2.99E-02

表 4 结肠癌中参与基因-基因相互作用的重要基因

Table 4 Significant genes in colon cancer that is involved in gene-gene interaction

Gene Ranking	Gene ID	Mean Diff	Express UP/Down	Redundant
1	205064_at	2.385	UP	FALSE
2	205488_at	-2.2332	DOWN	FALSE
3	206505_at	-2.1166	DOWN	TRUE
4	226517_at	0.9886	UP	FALSE
5	227529_s_at	1.0332	UP	FALSE
6	224009_x_at	1.5783	UP	TRUE
7	1557796_at	1.1084	UP	TRUE
8	210517_s_at	1.7642	UP	TRUE
9	214452_at	0.8011	UP	TRUE
10	1559203_s_at	0.9882	UP	TRUE

2.1.3 卵巢癌

卵巢癌的研究也是设置 6 个样本，其中 3 个是对 TOV112D 细胞利用 p53 靶向治疗，另 3 个样本是对 NSC319726 细胞利用 p53 靶向治疗。在所有 54675 个基因中，有 1566 个基因的差异性表达与卵巢癌有关。显著性检验，其中在卵巢组织中有 810 个基因发生表达上调，756 个基因发生表达下调(图 5)。使用 SVM 分类的方法我们对所有的数据的显著性基因进行了分类，结果显示只有 309 个基因的

表达与卵巢癌有关。其中 175 个基因是与 TOV112D 细胞和 NSC319726 细胞有关的 p53 突变基因，这些基因能够靶向治疗卵巢癌，可以作为最好的生物标志物(表 5 和表 6)。研究结果显示的部分与卵巢癌具有相关性的基因有 CDKN1A、PTEN、MDM2、DDB2、GADD45A、FANCC、HRAS、MLH1、DNMT1、VDR、PMS2、APC、TP53I3、MSH2、IGFBP3、EGFR、APC、MSH2、MET、CHMP4C、BIRC5、EGFR、TP53TP63 等(图 6 和图 7)。

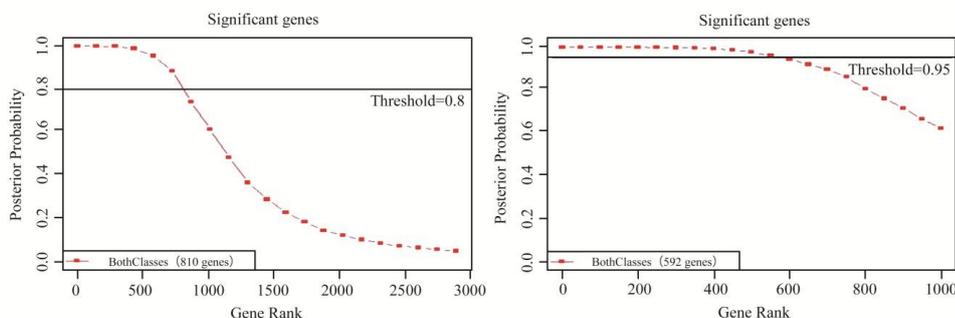


图 5 卵巢癌细胞中疾病组和对照组发生差异表达的重要基因

Figure 5 Significant genes of ovarian cancer cells that differentially expressed in disease and control tissues

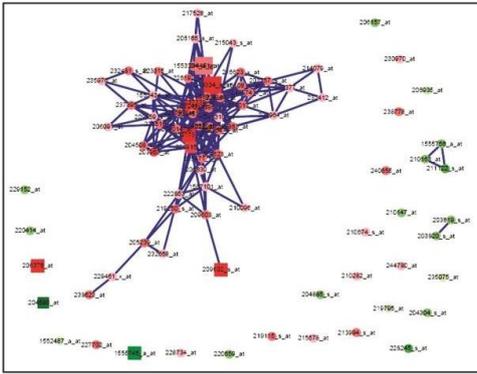


图 6 利用 geNETClassifier 预测卵巢癌数据中基因-基因相互作用网络

Figure 6 Gene-gene interaction network predicted using geNETClassifier on ovarian cancer

表 5 卵巢癌中参与基因-基因相互作用的重要基因

Table 5 Significant genes in ovarian cancer that is involved in gene-gene interaction

Gene Ranking	Gene ID	Mean Diff	Express UP/Down	Redundant
1	209173_at	1.2714	UP	FALSE
2	204580_at	-1.2938	DOWN	FALSE
3	206378_at	1.1423	UP	FALSE
4	1555745_a_at	-1.1253	DOWN	FALSE
5	240304_s_at	1.1246	UP	FALSE
6	209602_s_at	1.1043	UP	FALSE
7	215729_s_at	-0.8966	DOWN	FALSE
8	228241_at	1.0233	UP	TRUE
9	205048_s_at	0.997	UP	FALSE
10	227662_at	1.027	UP	FALSE
11	228915_at	1.0024	UP	TRUE
12	214451_at	0.8899	UP	TRUE
13	237339_at	0.9032	UP	TRUE
14	219580_s_at	0.8417	UP	TRUE
15	205239_at	0.8214	UP	FALSE
16	237086_at	0.8232	UP	TRUE
17	204667_at	0.7949	UP	TRUE
18	223864_at	0.7965	UP	TRUE
19	214774_x_at	0.8386	UP	FALSE
20	216623_x_at	0.8309	UP	TRUE
21	226597_at	0.767	UP	TRUE
22	206157_at	-0.7531	DOWN	FALSE
23	238778_at	0.7077	UP	FALSE
24	238017_at	0.7999	UP	TRUE
25	203820_s_at	-0.8356	DOWN	FALSE
26	1553394_a_at	0.7044	UP	TRUE
27	237395_at	0.7001	UP	FALSE
28	232481_s_at	0.7613	UP	FALSE
29	205471_s_at	0.7132	UP	TRUE
30	215108_x_at	0.7184	UP	TRUE

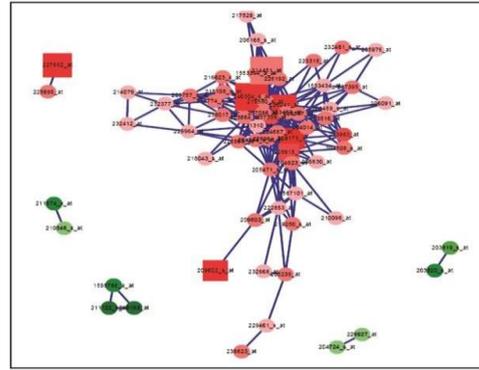


图 7 卵巢癌中前 52 位基因特征相互作用网络

Figure 7 Top 52 gene signature interaction network on ovarian cancer types

表 6 差异表达基因与卵巢癌的遗传突变显著相关

Table 6 Differentially expressed genes is significantly associated with genetic mutation of Ovarian cancer

Genes	FDR B&H	FDR B&Y	Bonferroni
NOS3, EDN2	1.13E+01	1.33E-02	1.16E-01
PTEN, MLH1, PMS2, APC, MSH2	8.71E-03	1.53E-04	1.23E-03
BAX, FANCC, MLH1, VDR, APC	1.94E+01	2.65E-02	2.32E-01
FAS	4.51E+01	4.70E-02	4.12E-01
BAX, PTEN, FANCC, MLH1, VDR	2.38E+01	3.07E-02	2.69E-01
CDKN1A, PTEN, MDM2, DDB2, GADD45A, FANCC, HRAS, MLH1, DNMT1, VDR, PMS2, APC, TP53I3, MSH2, IGFBP3, EGFR, TP53	1.38E-14	1.97E-15	2.27E-14
PTEN, MDM2, TP53	3.36E-03	1.10E-05	1.27E-04

3 结论

本研究利用统计学的方法来对与乳腺癌、结肠癌和卵巢癌显著相关的差异表达基因进行预测分析。这种不同的计算方法，能够在癌症组织中像单个基因标志物那样预测出生物标志物。在所有癌症类型的分析中使用功能富集分析已经确定了不同的功能特性的基因，特别是有助于生物标志物。使用这种方法，有助于在感染疾病的早期以及疾病发展中进一步鉴定生物标志物。该方法基于个体基因来提供分析信息，这对于阐明癌症途径，以及加快特定癌症的潜在药物靶点的搜索有助于提供有用的信息。

致谢

我们感谢创始人兼首席执行官、总经理、科学家 Prashantha C.N.,先生，感谢班加罗尔实验室配套的设计项目、实验方法和对本稿件所做出的贡献。我们也感谢计算机科学中心的 Mamatha 给我们提供实际的方法，对我们目标的一步一步实现提供了具有价值的指导。我也感谢我的父母对我的支持，使我的项目得以成功。我们更感谢 Prashantha 对我们项目的成功指导。

利益冲突

所有作者都接受出版这篇文章。

参考文献

Baekelandt M, Kristensen GB, Nesland JM, Tropé CG, Holm R. 1999. "Clinical significance of apoptosis-related factors p53, Mdm2, and Bcl-2 in advanced ovarian cancer." *J Clin*

Oncol, 17: 2061

Cancer Fact sheet N°297". Feb 2014. "World Health Organization", Retrieved 10 June 2014.

Dumitrescu RG, Cotarla I., 2005, "Understanding breast cancer risk--where do we stand in 2005?" *J Cell Mol Med*, 9:208–21

Elvidge G. 2006. "Microarray expression technology: from start to finish." *Pharmacogenomics* 7: 123–134.

Fearon ER and Vogelstein B. 1990, "A genetic model for colorectal tumorigenesis." *Cell*, 61, 759–767.

Golub TR. 2001. "Genome-wide viewsofcancer." *N Engl J Med* 344: 601–602

Honrado E, Osorio A, Palacios J, Benitez J. 2006. "Pathology andgene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations." *Oncogene*, 25:5837–45

Kupryjańczyk J, Szymańska T, Madry R, Timorek A, Stelmachów J, Karpińska G, Rembiszewska A, Ziółkowska I, Kraszewska E, Debniak J, Emerich J, Ułańska M, Płuzańska A, Jedryka M, Goluda M, Chudecka-Głaz A, Rzepka-Górska I, Klimek M, Urbański K, Breborowicz J, Zieliński J, Markowska J. 2003. "Evaluation of clinical significance of TP53, BCL-2, BAX and MEK1 expression in 229 ovarian carcinomas treated with platinum-based regimen." *Br J Cancer*, 88: 848–854

Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, Coresh J, Gansevoort RT. Chronic Kidney Disease Prognosis Consortium. 2010. "Association of estimated glomerular filtration rate and

- albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis.” *Lancet*, 375:2073–2081.
- Mao M, Tian F, Mariadason JM, Tsao CC, Lemos RJr, Dayyani F, Gopal YN, Jiang ZQ, Wistuba II, Tang XM, Bornman WG, Bollag G, Mills GB, Powis G, Desai J, Gallick GE, Davies MA, Kopetz S. 2012. “Resistance to BRAF inhibition in BRAF-mutant colon cancer can be overcome with PI3K inhibition or demethylating agents.” *Clin Cancer Res*, 19(3):657-67.
- Nielsen JS, Jakobsen E, Hølund B, Bertelsen K, Jakobsen A. 2004, “Prognostic significance of p53, Her-2, and EGFR overexpression in borderline and epithelial ovarian cancer.” *Int J Gynecol Cancer* 14: 1086–1096
- Prashantha Nagaraja, Kavya Parashivamurthy, Nandini Sidnal, Siddappa Mali, Dakshyani Nagaraja, and Sivarami Reddy. 2013, “Analysis of gene expression on ngn3 gene signaling pathway in endocrine pancreatic cancer.” *Bioinformatics*, 9(14): 739–747.
- Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D. 2009. “Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes.” *Journal of Clinical Oncology* 27: 1160–1167
- R Core Team, 2012. “R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.”
- Robinson MD, McCarthy DJ, Smyth GK. 2010. “EdgeR: a bioconductor package for differential expression analysis of digital gene expression data.” *Bioinformatics*, 26:139–140
- Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM and Bos JL. 1988. “Genetic alterations during colorectal-tumor development.” *N. Engl. J. Med.*, 319, 525–532.
- Wang X1, Saso H, Iwamoto T, Xia W, Gong Y, Pusztai L, Woodward WA, Reuben JM, Warner SL, Bearss DJ, Hortobagyi GN, Hung MC, Ueno NT. 2013. “TIG1 promotes the development and progression of inflammatory breast cancer through activation of Axlkinase.” *Cancer Res*, 73(21):6516-25.
- Yu X, Vazquez A, Levine AJ, Carpizo DR. 2012. “Allele-specific p53 mutant reactivation.” *Cancer Cell*, 21(5):614-25.