

• 基础研究 •

人心力衰竭心肌细胞基因表达谱分析方法初探

吴晓霞 应晓敏 李伍举 万涛 吴加金 刘惠亮

【摘要】 目的 通过对多中心人类心力衰竭心肌细胞基因表达谱数据整理、分析,探索识别不同表达谱中相同结构蛋白基因的方法。方法 选取14个已发表的原发病为扩张型心肌病和(或)缺血性心肌病的心力衰竭基因表达谱,筛选出各表达谱中差异表达 ≥ 1.4 倍的心肌结构蛋白基因及功能蛋白基因。编写程序自动下载相关基因片段的序列,并建立数据库。根据表达谱中提供的识别号及相关注释信息,用BatchGenAna网站在染色体基因组上进行序列定位,通过定位图显示同一基因不同片段的定位信息。结果 通过GenBank序号及芯片中克隆序号得到不同表达谱中相同序号的基因仅有23组,检出率低,漏检了相同基因的不同片段。通过染色体基因组比对、定位,检出不同表达谱中定位在同一染色体区域的相同基因的不同片段共51组,且集中定位在1、2、9、11、12、16号染色体上。通过定位明确显示此基因的编码区及调控区序列。结论 基因组定位方法可识别不同表达谱中相同基因的不同片段,明确定位基因编码区及调控区,为后续的功能基因组研究提供可靠信息。

【关键词】 心力衰竭;肌细胞,心脏;基因表达谱

Preliminary exploration for a method of analyzing gene expression profiles in myocardium cells of heart failure patients

WU Xiaoxia*, YING Xiaomin, LI Wujun, et al

* Department of Cardiology, General Hospital of Chinese People's Armed Police Forces, Beijing 100039, China

【Abstract】 Objective To explore a method of identifying the same structural gene in different gene expression profiles of human heart failure myocardium cells. Methods Fourteen available gene expression profiles in myocardium of heart failure patients were selected. These patients suffered dilated cardiomyopathy and/or ischemic cardiomyopathy. Over 1.4 times differentially expressed structural and functional genes were screened out. A local database of these genes with known nucleotide sequences was established. According to the Accession Number and other information of genes provided in the profiles, the genomic location of these genes on the chromosome was performed by BatchGenAna. The location information of different fragments of the same gene was shown in location map. Results Only 23 groups of genes were obtained by the Accession Number in distinct profiles, and the different fragments of the same gene were left out. By sequence comparison and genomic location, 51 groups of genes were identified, including different fragments of the same gene. The genes mostly distributed on chromosome 1, 2, 9, 11, 12, 16. The coding region and regulation region were also distinguished. Conclusion Genomic location is effective in identifying the different fragments of the same gene, and in locating definitely the coding region and transcriptional regulation region. The accurate sequences of genes in these regions will provide reliable information for further investigation of the functional genomics.

【Key words】 heart failure; myocyte, cardiac; gene; expression monitoring

心力衰竭(heart failure, HF)是一种复杂的临床综合征,目前认为多种分子途径共同参与、介导了

收稿日期:2008-05-06

基金项目:国家自然科学基金资助项目(30300130)

作者单位:100039 北京市,武警总医院内科(吴晓霞、刘惠亮);100850 北京市,军事医学科学院基础医学研究所(应晓敏、李伍举、吴加金);100853 北京市,解放军总医院肝胆外科(万涛)。吴晓霞,Tel:010-88276537

HF过程中基因的表达变化,最终通过引起功能蛋白数量的改变影响心脏功能,因此对HF发生机制的深入了解需从心脏分子病理角度出发。基因芯片(microarray)技术为研究疾病过程中基因转录、表达变化提供了工具。目前已有多个中心采用此技术对人类HF心肌标本进行检测,显示HF发展过程及其终末期基因表达模式的差异^[1~3]。

基因表达谱的数据繁杂、海量,不同芯片公司提供的基因片段不同,即使是相同基因在不同表达谱中其基因名称、数据库编号(accession No.)亦不相同,因此增加了对表达谱数据分析的难度。本研究通过对不同HF心肌表达谱数据整理、分析,探索识别不同表达谱中相同结构蛋白基因的方法。

1 资料与方法

1.1 研究对象 选取2002-2009年发表的14个^[2~15]原发病为扩张型心肌病和(或)缺血性心肌病的HF基因表达谱的数据为研究对象。入选标准:(1)各表达谱中与对照组比较HF时差异表达的基因;(2)涉及心肌结构、细胞骨架、细胞外基质、氧化应激/细胞凋亡途径的基因。排除标准:各表达谱中的EST序列;受体及抗原基因;各种生物酶基因;通用的转录因子、翻译延长因子及非人类种属的基因。

1.2 方法 (1)通过编写自动检索、下载程序,按照每个基因的基因名称或功能描述(description)从美国国立生物医学网站的核酸数据库(NCBI; www.ncbi.nlm.nih.gov)中自动下载所有相关基因片段的全序列及其基因调控区序列,建立数据库,并对数据库中相同基因名称的不同片段进行相似性比较,删除冗余数据,建立数据库。(2)筛选各表达谱中HF时差异表达变化 ≥ 1.4 倍的基因纳入本研究,以下称为HF相关基因。(3)利用入选基因在表达谱中的数据库编号或识别号(identifier),通过编写程序在NCBI网站核酸数据库下载基因序列及相关注释信息,整合为新的数据库,并与前述数据库进行比较。根据GenBank数据库中基因碱基序列及相关注释信息,通过BatchGenAna网站(<http://biosrv1.bmi.ac.cn/BatchGenAna>)在人类染色体基因组上进行序列定位。根据入选基因在基因组中的定位信息,与位于相同转录区域的参考基因(RefSeq基因)比对,通过图形显示同一条染色体上不同表达谱中相同基因不同片段的定位信息。人类基因组数

据采用NCBI build 36.2(下载自ftp://ftp.ncbi.nih.gov/genomes/H_sapiens),EST、mRNA和RefSeq基因的注释信息采用NCBI Map Viewer数据(下载自ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/mapview)。BatchGenAna网站为军事医学科学院基础医学研究所计算生物学中心(应晓敏等)创建,相关论文将另文发表。

2 结果

通过比较各表达谱中的基因编号或识别号,筛选得到至少在两个表达谱中出现的基因仅为23组,检出率低,包括:AB004885、AF00987、AF001548、AF051850、AL021155、D00596、D17408、D88687、K03460、L05096、L25286、L38941、M22430、M23115、M30262、M33308、P35293、U03688、U03688、U03688、U76702、X83703、Y09862,漏检了相同基因的不同片段。

通过染色体基因组定位,另检出不同表达谱中定位在同一染色体区域相同基因的不同片段共51组(不包括上述23组),明确显示此基因的编码区及调控区序列,同时检出此基因在GenBank数据库中的其他所有转录本信息(图1)。通过比对可以区分同一蛋白的不同亚型的定位信息(图2)。

通过染色体基因组定位得到的各表达谱中HF时差异表达的心肌结构及功能蛋白的基因集中定位在1号、2号、9号、11号、12号、16号染色体上。

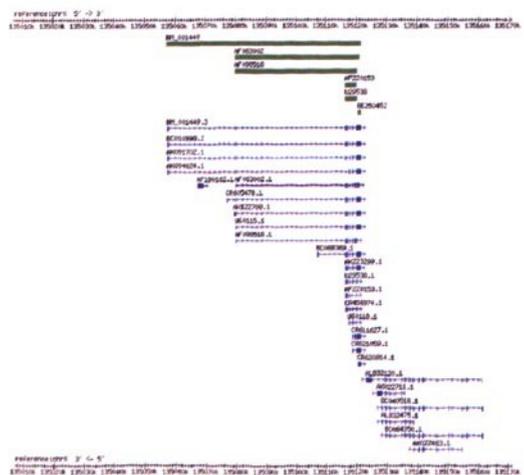
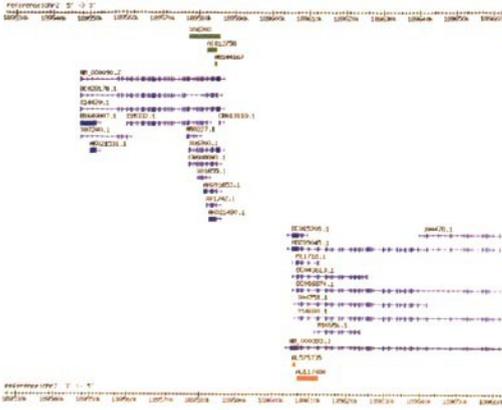


图1 相同基因的不同转录本定位在同一染色体的同一基因组上(绿色条带),蓝色条带为RefSeq基因及既往提交过的其他所有转录本信息



绿色为Ⅲ型α1,红色为V型α2

图2 胶原蛋白的不同亚型

3 讨论

慢性 HF 在人群中的发病率为 1.5%~5.6%，65 岁以上人群达 7.4%，5 年死亡率高达 82%，严重威胁人类健康。内科治疗虽能使心脏功能得到短期改善，但患者 5 年的死亡率仍接近 50%^[16]，心脏移植因供体来源有限，不能满足广大患者的需要。因此对 HF 发病机制及其发展过程中病理生理改变的深入认识，将为其临床治疗探索新的思路与方法。近年来 HF 的基因治疗越来越受到关注。

自 2000 年以来，已有多家研究中心采用基因芯片技术对人类 HF 的心肌标本进行检测，并发表了完整的表达谱数据。通过对 HF 基因表达谱差异基因的分析，显示出 HF 发展过程中心脏功能所需的转录需求^[1]。

人类全基因组芯片点样的探针数通常在 2 万以上，其结果是海量的表达谱数据的不断涌现，如何分析这些数据为基因治疗提供可靠的信息成为亟待解决的问题。生物信息学通过对生物学实验数据的获取、加工、存储、检索与分析，达到提取数据中所蕴含的生物学意义的目的^[17]。本研究从已发表的多个 HF 心肌表达谱数据的生物信息学分析入手，利用 NCBI 网站的核酸数据库中的信息，通过自动下载程序，建立了包含基因名称、基因识别号、基因序列及相关注释信息的较为完整的 HF 相关基因数据库。

在对表达谱数据的分析过程中，笔者发现国外不同中心发表的表达谱数据之间存在较大差异，考

虑与以下因素有关：(1)由于各中心选用的基因芯片不同，不同芯片所提供的基因片段亦不相同，其原始序列数据编号(即 accession No.)及信息缺乏统一标准，存在相同基因不同命名，部分基因命名区域序列不完整的问题。因此仅从各中心发表的表达谱数据中较难获得一致的信息。例如：图 1 中显示的 NM_001449、AF063002、AF098518、AF220153、U29538、BE25045 几条基因分别取自不同的表达谱，其数据库编号差异很大，分别来自 GenBank、EMBL、UniGene 等数据库。又如：心钠肽(ANP)在不同的表达谱中分别提供了 M31776、AL021155、M25296 等识别号。对于肌动蛋白(actin)、肌球蛋白重链(myosin heavy chain)等有多重亚型的蛋白，其差异更为繁复。(2)基因芯片的结果分析采用比较扫描器所记录的荧光强度的方法，许多因素都将影响观测到的强度，如两标本间 mRNA 总体数量的不同、浓度、亮度、染色标记的有效性、暴露时间、杂交及洗脱的严格性、扫描相机的灵敏度等。因此各表达谱中差异表达的基因不尽相同，表达变化的倍数差别亦较大。(3)各中心的心肌标本之间存在种族及个体差异。因此仅根据已发表的表达谱数据进行分析，很难从中发掘一致性的信息，同时也丢失了大量有用的信息。

2003 年人类基因组序列图绘制成功，识别了人类基因组 DNA 碱基对序列，阐明人类基因组在染色体上的位置。本研究通过 BatchGenAna 网站将相同基因的不同片段定位在染色体上，避免了漏检。同时通过定位图形可以清晰显示不同基因片段的大小及其相互关系，并通过与 GenBank 数据库中的参考基因比对，明确此基因的编码区及调控区序列，为功能基因组学的研究提供了可靠的数据信息。

人类基因并不是平均分布在染色体中，而是在有些区域成簇出现，有些区域分布很少。本研究显示 HF 时差异表达的心肌结构及功能蛋白的基因集中定位在 1、2、9、11、12、16 号染色体上，而在 3、8、13、15、18、20、22 号染色体上分布较少，由此提示 HF 发展过程中，人体根据心脏自身功能需要通过调控基因的转录来适应体内病理生理环境的需要。

本研究通过生物信息学分析将已发表的不同表达谱数据进行整理、分析，将 HF 时差异表达的基因进行染色体定位，可以减少各中心因标本的种族、个

体差异及基因芯片技术本身缺陷造成的数据偏差,使表达谱数据分析的可信度提高。另一方面,药物可能通过对基因表达的调控、抑制、对基因产物的修饰等途径发挥作用。通过对各表达谱数据的准确、全面分析,将可能为从基因水平上治疗 HF 的药物设计找到新的靶点,为 HF 的基因治疗提供理论支持。

本研究的局限性:由于部分表达谱数据发表较早,根据其提供的识别号下载的序列与染色体基因组中序列差别较大,未能将其定位在染色体上,可能因此出现了部分漏检。此问题随着核酸数据库更新及 BatchGenAna 网站功能的日益完善有望解决。

参考文献

- [1] 吴晓霞,智光.心力衰竭时心肌基因表达谱变化的特点及其意义.中华老年心脑血管病杂志,2006,8:497-499.
- [2] Steenman M, Chen YW, Le-Cunff M, et al. Transcriptional analysis of failing and nonfailing human hearts. *Physiol Genomics*, 2003, 12: 97-112.
- [3] Kaab S, Barth AS, Margerie D, et al. Global gene expression in human myocardium oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure. *J Mol Med*, 2004, 82: 308-316.
- [4] Yang J, Moravec CS, Sussman MA, et al. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation*, 2000, 102: 3046-3052.
- [5] Hwang JJ, Allen PD, Tseng GC, et al. Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics*, 2002, 10: 31-44.
- [6] Tan FL, Moravec CS, Li JB, et al. The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci USA*, 2002, 99: 11387-11392.
- [7] Barrans JD, Allen PD, Stamatou D, et al. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular based cDNA microarray. *Am J Pathol*, 2002, 160: 2035-2043.
- [8] Grzeskowiak R, Witt H, Drungowski M, et al. Expression profiling of human idiopathic dilated cardiomyopathy. *Cardiovasc Res*, 2003, 59: 400-411.
- [9] Boheler KR, Volkova M, Morrell C, et al. Sex- and age-dependent human transcriptome variability: implications for chronic heart failure. *Proc Natl Acad Sci USA*, 2003, 100: 2754-2759.
- [10] Blaxall BC, Tschannen-Moran BM, Milano CA, et al. Differential gene expression and genomic patient stratification following left ventricular assist device support. *J Am Coll Cardiol*, 2003, 41: 1096-1106.
- [11] Chen YJ, Park S, Li YF, et al. Alterations of gene expression in failing myocardium following left ventricular assist device support. *Physiol Genomics*, 2003, 14: 251-260.
- [12] Kittleson MM, Ye SQ, Irizarry RA, et al. Identification of gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation*, 2004, 110: 3444-3451.
- [13] Margulies KB, Matiwala S, Cornejo C, et al. Mixed messages: transcription patterns in failing and recovering human myocardium. *Circ Res*, 2005, 96: 592-599.
- [14] Kittleson MM, Minhas KM, Irizarry RA, et al. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. *Physiol Genomics*, 2005, 21: 299-307.
- [15] 吴晓霞,万涛,吴红金,等.全基因组芯片结合通路分析筛选心力衰竭患者心肌细胞差异基因.中华心血管病杂志,2009,37:120-125.
- [16] Mckinsey TA, Olson EN. Toward transcriptional therapies for the failing heart: chemical screens to modulate genes. *J Clin Invest*, 2005, 115: 538-546.
- [17] Goodman N. Biological data becomes computer literate: new advances in bioinformatics. *Curr Opin Biotechnol*, 2002, 13: 68-71.