科爱传播
KE'AI COMMUNICATIONS
生命科学

·导读版·

DNA Microarrays Part B:
Databases and Statistics

DNA芯片(B辑):
数据和分析

Alan Kimmel, Brian Oliver

ELSEVIER
原版引进

科学出版社
www.sciencep.com

Methods in Enzymology, Volume 411
《酶学方法》第 411 卷

DNA Microarrays Part B: Databases and Statistics

DNA 芯片(B 辑):数据和分析

Edited by
Alan Kimmel & Brian Oliver
NIDDK
National Institutes of Health
Bethesda, Maryland

科 学 出 版 社
北 京

图字:01-2006-7331 号

This is an annotated version of
DNA Microarrays Part B: Array Databases and Statistics
(Methods in Enzymology Volume 411)
Edited by Alan Kimmel & Brian Oliver
Copyright © 2006, Elsevier Inc.
ISBN-13: 978-0-12-182816-5
ISBN-10: 0-12-182816-6

图书在版编目(CIP)数据

DNA 芯片(B 辑):英文/(美)基梅尔(Kimmel, A.)等编著. —影印本. —北京:科学出版社,2007

ISBN 978-7-03-018226-5

Ⅰ.D… Ⅱ.基… Ⅲ.脱氧核糖核酸—英文 Ⅳ.Q523

中国版本图书馆 CIP 数据核字(2006)第 148102 号

责任编辑:孙红梅/责任印制:钱玉芬/封面设计:耕者设计工作室

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com
中国科学院印刷厂 印刷
科学出版社发行 各地新华书店经销
*

2007 年 1 月第 一 版 开本:787×1092 1/16
2007 年 1 月第一次印刷 印张:30 1/4 插页:16
印数:1—2 000 字数:717 000

定价:90.00 元

(如有印装质量问题,我社负责调换〈科印〉)

对 DNA 芯片技术的应用具有实际指导意义的两卷丛书

——评《酶学方法:DNA 芯片 A,B 两辑》

张亮

(生物芯片北京国家工程研究中心,北京,102206,Email:lzhang@capitalbio.com)

荷兰 Elsevier 出版社出版的《酶学方法》一直是针对生命科学领域业已成熟的生物技术进行最系统介绍的技术方法类丛书。自 20 世纪 90 年代初期出现,迄今约 20 年的时间里,生物芯片技术的内涵一直在不断扩大,技术方法也一直在不断地完善。DNA 芯片是生物芯片的一个类别,泛指固体基片载体上固定的是 DNA 序列的固相芯片;除 DNA 芯片外,还有在固相基片表面固定多糖的多糖芯片,固定蛋白的蛋白芯片等。由于 DNA 芯片技术发展相对成熟,2006 年 9 月,《酶学方法》中的第 410 和 411 卷对 DNA 芯片技术进行了系统介绍。

在 DNA 芯片技术中,按照用途来分,又可以分为基因表达谱芯片、MicroRNA 检测芯片、甲基化位点检测芯片、转录因子 DNA 吸附位点检测芯片,等等;可以预计,在 DNA 芯片技术中,能高通量并行分析成千上万个基因表达的 DNA 基因表达谱芯片技术在未来 10 年内将会成为分子生物学实验室中常规的实验技术。其原因除了基因表达谱芯片技术的重复性、精确性、灵敏度等技术指标逐步改善,能为分子生物学家所接受外,还将得益于具有自主知识产权的国产关键生物芯片仪器的广泛普及。分子生物学家若想在自己实验室构建 DNA 芯片技术平台,在没有经验积累的情况下,从最初的实验设计,例如所研究的问题只是简单寻找处理和对照样品中差异表达的基因,还是在含有多个分析因素的实验中寻找特异表达的基因;在什么情况适合设计寡聚核苷酸的 DNA 芯片,什么情况下适合设计 cDNA 芯片;到中间的实验实施细节,包括如何选择生物芯片仪器;选择 DNA 点样液和点样用的基片时需要注意哪些因素;如何获得信噪比良好的 DNA 芯片杂交图片;到 DNA 芯片湿实验结束后,如何把芯片的数据和所研究的生物学现象关联起来等一系列问题,潜在的 DNA 芯片技术应用者都希望得到全面的指导。

Elsevier 出版社此次出版的《酶学方法》DNA 芯片技术的 A,B 两辑,是由出版社组织了 DNA 芯片技术领域各个技术环节的专家对 DNA 芯片技术进行的一次全方位介绍。读者在读完这两辑以后,将从中获得关于 DNA 芯片技术具体而实用的操作指导,使读者在 DNA 芯片实验中少走弯路。具体来说,DNA 芯片技术的 A,B 两辑内容的重要特点包括:

(1)系统性。由于 DNA 芯片技术是在融合了多门学科的基础上产生的交叉学科,涉及了光学、精密仪器、电子微加工、化学、生物学、生物信息学等学科的内容,没有人能对所有这些领域的内容都特别精通,因此本书汇集了各个技术环节的专家,各自编写自己擅长的部分。并且在章节的结构安排上,是先进行全面的介绍,然后深入到具体环节,逐点进行详细的阐述,即由面及点地介绍。例如先从整个面上介绍 DNA 芯片技术的现状,包括一些主要的商业化 DNA 芯片技术平台,然后引伸到研究者若要自己构建 DNA 芯片技术

平台，需要具体注意哪些事项。

(2)继承性。DNA 芯片技术实际上根植于传统的分子生物学，与很多经典的分子生物学技术，诸如用于基因表达分析的 DNA 芯片技术，和传统的 Northern blot 方法等，均有很多相似的地方，包括均需要 RNA 的完整性，杂交是在固液相之间发生的，所得到的数据也需要归一化，等等。本书在介绍 DNA 芯片技术的实验方法时，多借用分子生物学家业已熟悉的传统分子生物学方法，从这些传统方法的基本原理入手，然后过渡到在 DNA 芯片技术操作中，有哪些原理和方法和传统方法类似，而有哪些环节有所不同，在操作的时候是需要具体注意的，让即使还没有接触过 DNA 芯片技术的人也不会感到陌生。

(3)实用性。作为 DNA 芯片技术方法类的指导丛书，读者在翻阅以后，希望能够在各个技术环节上得到具体的建议，或者能够按照指导丛书的指导，亲自动手进行 DNA 芯片实验操作，亦或能够对 DNA 芯片实验产生的海量数据进行分析，得到规律性的结论。考虑到本书的读者主要是面向分子生物学科研人员，一方面，本书在湿实验方面介绍得非常详细，从样品的 RNA 抽提，到 RNA 标记方法的选择，再到杂交过程常见问题的解决对策等；另一方面，在介绍 DNA 芯片数据分析方法时，并没有从如何编写生物信息学的程序入手进行深奥的介绍，而是结合生物学实验设计和特点，主要讲述进行生物信息分析时需要注意的一些原则，并在此基础上，向读者介绍如何使用一些业已被广泛接受的分析软件和一些公共数据库来进行芯片的数据分析和注释。这样的结构安排就比较符合分子生物学家的需求：在湿实验过程中，分子生物学研究人员能够结合已有的分子生物学实验经验和技巧，很容易地领悟和掌握 DNA 芯片实验操作的技巧；而对于编写生物信息学需要的各种源程序则感到无从下手，若只是让他们掌握如何使用已被公认的生物信息分析软件，就相对简单得多。这样科研人员在进行 DNA 芯片实验设计、实验实施、到论文发表的整个科研过程会感到简便易行。

(4)前瞻性。DNA 芯片技术是伴随着功能基因组学的兴起而发展起来的。生物芯片技术的本质特点就是高通量、半定量。尽管到目前为止，分子生物学家多用 DNA 芯片技术来并行分析成千上万的基因表达，但随着生物学各个领域的应用需求，DNA 芯片技术已经不再局限于基因表达的分析，而是在细胞的 DNA 水平，乃至 RNA 水平，再到蛋白水平，都得到了广泛应用。例如，用以比较肿瘤组织和正常对照组织中基因组 DNA 的缺失或扩增的比较基因组杂交技术；高通量检测非编码小 RNA 表达情况的 MicroRNA 芯片检测技术；研究转录因子蛋白在基因组 DNA 上吸附位点的 ChIP-on-chip 技术，等等，本书都进行了介绍，以开拓读者的视野。引导读者在了解 DNA 芯片技术向上述应用领域延伸的过程中，体会到只要有高通量的需求，就有对生物芯片技术应用的需求。

随着组学的发展，人们才有机会从系统生物学的角度来诠释生命现象。生物芯片技术的高通量特点是和组学研究对技术的要求非常匹配的。掌握 DNA 芯片技术基本原理和具体操作过程，分子生物学实验室能适应组学的发展。笔者相信此次 Elsevier 出版社出版的《酶学方法》中 DNA 芯片技术的 A，B 两辑，将极大地促进这种高通量的生物学技术在诠释生命现象中的应用。

Contributors to Volume 411

Article numbers are in parentheses and following the name of contributors.
Affiliations listed are current.

JUSTEN ANDREWS (3), *Department of Biology, Indiana University, Bloomington, Indiana*

BERNARD F. ANDRUSS (1, 2), *Asuragen, Inc., Austin, Texas*

JULIEN F. AYROLES (11), *Department of Genetics, North Carolina State University, Raleigh, North Carolina*

TANYA BARRETT (19), *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland*

TIM BEISSBARTH (18), *The Walter and Eliza Hall Institute of Medical Research, Bioinformatics Group, Victoria, Australia*

GEORGE W. BELL (22), *Bioinformatics and Research Computing, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts*

MICHEL BELLIS (21), *CRBM–CNRS, Montpellier, France*

NIRMAL K. BHAGABATI (9), *The Institute for Genomic Research, Rockville, Maryland*

KEVIN BOGART (3), *Drosophila Genomics Resource Center, Indiana University, Bloomington, Indiana*

JOHN C. BRAISTED (9), *The Institute for Genomic Research, Rockville, Maryland*

ALVIS BRAZMA (20), *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

VINCENT J. CAREY (8), *Harvard University, Boston, Massachusetts*

AMY CASH (3), *Department of Biology, Indiana University, Bloomington, Indiana*

JAMES COSTELLO (3), *Drosophila Genomics Resource Center, Indiana University, Bloomington, Indiana*

GREGORY E. CRAWFORD (14), *Department of Pediatrics, Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina*

ADELE CUTLER (23), *Department of Mathematics and Statistics, Utah State University, Logan, Utah*

SEAN DAVIS (14), *Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland*

TIMOTHY S. DAVISON (2), *Asuragen Discovery Services, Asuragen, Inc., Austin, Texas*

TOM DOWNEY (13), *Partek Incorporated, Saint Louis, Missouri*

BRIAN EADS (3), *Department of Biology, Indiana University, Bloomington, Indiana*

RON EDGAR (19), *National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland*

OLOF EMANUELSSON (15), *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut*

MARK B. GERSTEIN (15), *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut*

GREG GIBSON (11), *Department of Genetics, North Carolina State University, Raleigh, North Carolina*

JEREMY GOLLUB* (10), *Department of Biochemistry, Stanford University Medical School, Stanford, California*

JÉRÔME HENNETIN (21), *CRBM–CNRS, Montpellier, France*

ELEANOR A. HOWE (9), *Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts*

CHARLES D. JOHNSON (2), *Asuragen Discovery Services, Asuragen, Inc., Austin, Texas*

MISHA KAPUSHESKY (20), *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

GARY J. LATHAM (1), *Asuragen, Inc., Austin, Texas*

FRAN LEWITTER (22), *Bioinformatics and Research Computing, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts*

JIANWEI LI (9), *The Institute for Genomic Research, Rockville, Maryland*

WEI LIANG (9), *The Institute for Genomic Research, Rockville, Maryland*

NICHOLAS M. LUSCOMBE (15), *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

LAKSHMI V. MADABUSI (1), *Asuragen Discovery Services, Asuragen, Inc., Austin, Texas*

JAMES M. MINOR (12), *Agilent Technologies, Inc., Santa Clara, California*

DANIEL Q. NAIMAN (16), *Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, Maryland*

HELEN PARKINSON (17, 20), *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

JOHN QUACKENBUSH (9), *Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute Boston, Massachusetts*

MARK REIMERS (8), *National Cancer Institute, Bethesda, Maryland*

THOMAS E. ROYCE (15), *Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut*

JOEL S. ROZOWSKY (15), *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut*

LAO H. SAAL (7), *Institute for Cancer Genetics, College of Physicians and Surgeons, Columbia University, New York, New York*

ALEXANDER I. SAEED (9), *Department of Bioinformatics, The Institute for Genomic Research, Rockville, Maryland*

MARC SALIT (5), *Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland*

UGIS SARKANS (20), *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

PETER C. SCACHERI (14), *Department of Genetics, Case Western Reserve University, Cleveland, Ohio*

VASILY SHAROV (9), *The Institute for Genomic Research, Rockville, Maryland*

GAVIN SHERLOCK (10), *Department of Genetics, Stanford University Medical School, Stanford, California*

**Current affiliation: Inconix Pharmaceuticals, Inc., Mountain View, California.*

MOHAMMAD SHOJATALAB (20), *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

MICHAEL SNYDER (15), *Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut*

JOHN R. STEVENS (23), *Department of Mathematics and Statistics, Utah State University, Logan, Utah*

CHRISTIAN J. STOECKERT, JR. (17), *Department of Genetics, Center for Bioinformatics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania*

MATHANGI THIAGARAJAN (9), *Department of Bioinformatics, The Institute for Genomic Research, Rockville, Maryland*

JERILYN A. TIMLIN (6), *Biomolecular Analysis and Imaging, Sandia National Laboratories, Albuquerque, New Mexico*

CARL TROEIN (7), *Computational Biology and Biological Physics, Department of Theoretical Physics, Lund University, Lund, Sweden*

JOHAN VALLON-CHRISTERSSON (7), *Department of Oncology, Lund University Hospital, Lund, Sweden*

PATRICIA L. WHETZEL (17), *Department of Genetics, Center for Bioinformatics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania*

JOSEPH A. WHITE (9), *Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts*

IVANA V. YANG (4), *Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina*

HAIYUAN YU (15), *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut*

XIAOWEI ZHU (15), *Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut*

目　录

Table of Contents

[1] RNA Extraction for Arrays

By LAKSHMI V. MADABUSI, GARY J. LATHAM, and BERNARD F. ANDRUSS

Abstract

DNA microarrays enable insights into global gene expression by capturing a snapshot of cellular expression levels at the time of sample collection. Careful RNA handling and extraction are required to preserve this information properly, ensure sample-to-sample reproducibility, and limit unwanted technical variation in experimental data. This chapter discusses important considerations for "array-friendly" sample handling and processing from biosamples such as blood, formalin-fixed, paraffin-embedded samples, and fresh or flash-frozen tissues and cells. It also provides guidelines on RNA quality assessments, which can be used to validate sample preparation and maximize recovery of relevant biological information.

Introduction

DNA microarrays have enabled biologists to move from the realm of studying one gene at a time to understanding genome-wide changes in gene expression. The value of microarray studies has been vetted through numerous studies that have linked abnormal transcript levels with many different diseases (Archacki and Wang, 2004; Blalock *et al.*, 2004; Borovecki *et al.*, 2005; Dhanasekaran *et al.*, 2001; Glatt *et al.*, 2005; West *et al.*, 2001). Because these types of studies will be used increasingly to create and validate diagnostic and prognostic expression signatures and to support toxicological and functional studies that underlie the regulatory filings for new drug submissions, it will become increasingly important to create standardized and robust methods for sample procurement, sample processing, and data analysis. The goal of any RNA isolation procedure is to recover an RNA population that faithfully mirrors the biology of the sample at the time of collection. Problems associated with the extraction of biologically representative RNA primarily arise from the susceptibility of RNA to degradation by ubiquitous and catalytically potent RNases. For tissues and cells, protection of RNA has traditionally been accomplished by immediate lysis using high concentrations of detergents and/or chaotropic agents and organic solvents (such as TRI reagent). These methods,

0076-6879/06 $35.00
DOI: 10.1016/S0076-6879(06)11001-0

while effective, are complex to use at point of care and suffer from low sample throughput and poor stabilization of cellular RNA for long periods. Flash freezing of the sample in liquid nitrogen and subsequent transportation on dry ice, although effective, are impractical in most clinical settings. Finally, disease specimens can present biohazard risks to the operator and constrain sample collection, thus limiting the use of best sample handling and processing practices and compromising RNA quality.

The practicality and efficacy of RNA stabilization agents such as RNA*later* to preserve the RNA in tissues, cells, and blood are gaining broad acceptance. Procedures used for collection of samples with RNA*later* are simple and can be carried out in a hospital setting with minimal training. This reagent is aqueous and nontoxic and allows convenient transportation of samples at ambient temperature. However, RNA*later* does not remove the biohazard risks associated with biosamples, and, as a result, all proper safety precautions should be observed. It is beyond the scope of this chapter to provide details on the risks associated and preventive measures to be taken when dealing with samples considered to be a biohazard. Several regulatory agencies offer guidelines on the safety issues and precautions that need to be addressed with such samples.

In addition to the handling of biological material, limitations can be imposed by the large amounts of RNA necessary for microarray experiments. As a result, samples such as tumor biopsies, formalin-fixed, paraffin-embedded (FFPE) sections, or laser microdissected samples require RNA amplification to generate adequate amounts of labeled material for microarray hybridization. The most popular and best validated approaches for amplifying RNA are based on the linear RNA amplification method developed by Eberwine (Van Gelder, 1990). This technique has been widely accepted for microarray applications and is known to preserve the original transcript ratios in the sample (Feldman *et al.*, 2002; Polacek *et al.*, 2003). In terms of RNA quality, parameters such as A260:280 measurements and Agilent RNA integrity number (RIN) are often used to gauge the quality of samples and predict their suitability for microarray studies. The minimum A260:280 or RIN number suitable for analysis varies by the array platform, number of replicates, and the experimental questions to be answered in the study.

Blood as a Biological Specimen

Blood is a highly desirable biosample for research and clinical studies for several reasons. First, blood is highly accessible and can be collected using relatively simple methods. Second, limited infrastructure is required to draw blood from a large number of patients. Finally, blood circulates throughout the entire body and thus is a vast reservoir of host biological information and an ideal specimen for experiments that aim to understand human physiology and disease. As a source of RNA, however, blood poses a number of unique challenges. The ratio of total protein to RNA in blood is roughly 100-fold greater than the ratio for most solid tissues, complicating the isolation of pure, high quality RNA. The presence of multiple cellular components in blood, each at different maturation stages in their life cycle, can lead to variation between patients. Among these various cellular components, only white blood cells (WBC) or leukocytes are nucleated and thus transcriptionally active (Fan and Hegde, 2005). However, WBC constitute only about 0.1% of the total blood cellular composition. In contrast, red blood cells (RBC) comprise ~95% of the total cell count but do not contribute to the blood gene expression program in their mature form. Immature red blood cells, known as reticulocytes, comprise only about 1% of the RBC population, yet contain significant levels of nucleic acids, particularly globin mRNA, that can contribute to the background noise in a microarray experiment. This noise can be substantial and can reduce the number of genes that are called present on microarrays.

Collection and Preservation of Blood Samples

Whole blood samples can be collected in the presence of anticoagulants such as sodium citrate, EDTA, or heparin. However, these chemicals are not effective RNA preservatives because they do not readily inhibit the RNases in blood that are the primary threat to RNA intactness and do not maintain cellular homeostasis in the sample. Indeed, the gene expression levels of many transcripts in blood stored in EDTA can change by an order of magnitude or more within a few hours (Rainen *et al.*, 2002). Rapid changes in gene expression of cytokines and transcription factors have been observed during storage for 1 to 4 h with interleukin-8 expression increasing 100-fold by 4 h (Tanner *et al.*, 2002). Additional genes such as transcription factors and pro- and anti-inflammatory genes show large changes in gene expression within a few hours to 1 day after collection (Pahl and Brune, 2002; Tanner *et al.*, 2002). The stresses caused by handling and centrifugation can alter gene expression rapidly (Haskill *et al.*, 1988). It is important to note that the purity of the RNA as measured by A260/280 is very consistent even after the extended storage of whole blood at ambient temperature, and often the intactness of ribosomal RNA bands is also well maintained although the underlying representation of many genes may have changed dramatically.

Other commonly used methods of blood collection and preservation include use of commercial products such as PAXgene tubes (PreAnalytiX GmbH, Switzerland) and the CPT tube (Becton Dickinson, NJ) for peripheral blood mononuclear cell (PBMC) collection. A brief summary of the advantages and impact of such sample collection methods on gene expression profiling has been reviewed by Fan and Hegde (2005). We find that use of RNA*later* as a preservative in conjunction with an optimized RNA isolation protocol produces excellent RNA yields and stable and reproducible expression profiles of human whole blood (Fig. 1). A concomitant increase in RNA yield and a more consistent level of percentage present calls were observed with the use of RNA*later*.

Methodologies for Globin Transcript Removal from Whole Blood

The presence of high levels of globin transcripts in RNA isolated from whole blood can affect the quality of data generated by reducing the number of present calls, decreasing call concordance, and increasing the variation in signal between samples. To circumvent the problems associated with the presence of globin transcripts, protocols have been described that reduce globin mRNA levels by either depleting these transcripts in purified RNA or fractionating blood cells to reduce the red blood

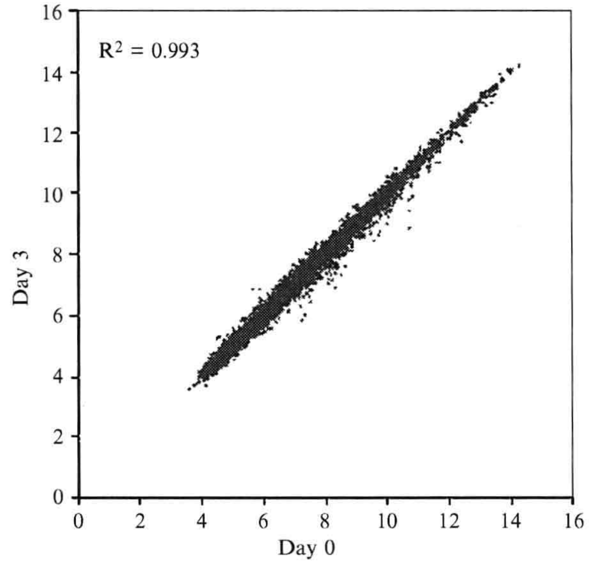

FIG. 1. RNA*later* provides room temperature stabilization of the global expression profile in human whole blood. Biological replicates of samples processed with RiboPure-Blood (Ambion) immediately after blood collection or after 3 days of storage at room temperature. The global expression level was assessed using Affymetrix human focus arrays with 10 μg aRNA input without globin reduction. Plots were constructed from signal-normalized data.

cell population, particularly reticulocytes, which are the primary reservoir of globin transcripts.

Depletion of Globin Transcripts from Whole Blood

To selectively deplete the globin transcripts from whole blood, commercial protocols were initially developed using enzymatic procedures to selectively degrade the globin transcripts. One of the first protocols was suggested by Affymetrix, Inc. This procedure described the hybridization of complementary DNA oligonucleotides to the various globin transcripts in blood followed by digestion of the RNA:DNA hybrid with RNase H. More recently, Affymetrix has launched the GeneChip Blood RNA Concentration Kit, which utilizes globin-specific peptide nucleic acid (PNA) oligomers as blocking molecules to prevent the amplification of these transcripts during T7 RNA polymerase-based linear amplification. An alternative strategy provide by Ambion is the GLOBINclear kit, which relies on the binding of biotinylated capture oligonucleotides to the RNA and uses biotin–streptavidin binding to deplete the globin transcript complex from the mixture. This method results in a dramatic reduction of the globin gene transcripts from whole blood RNA while substantially increasing the percentage present calls on Affymetrix Genechip arrays with human blood samples (Fig. 2). Thus, globin transcript reduction prevents

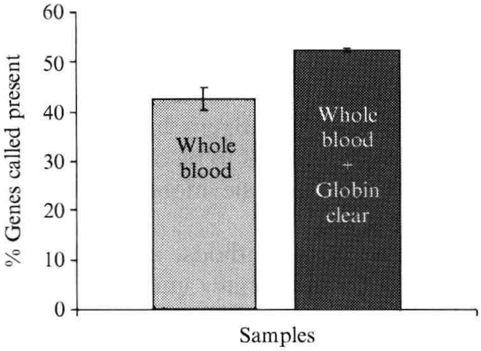

FIG. 2. GLOBINclear processing increases the sensitivity of microarrays. Quadruplicate GLOBINclear reactions were performed with pooled total RNA samples from human whole blood (from healthy donors under an IRB-approved protocol). The processed RNA was then amplified with MessageAmp II-96 to synthesize biotinylated aRNA for Affymetrix GeneChip array analysis. Quadruplicate untreated whole blood RNA samples were also amplified in parallel. Biotinylated aRNA was hybridized to Affymetrix human focus arrays. Present calls were determined using Affymetrix GCOS software with default settings. GLOBINclear processing resulted in an increase in genes called present.