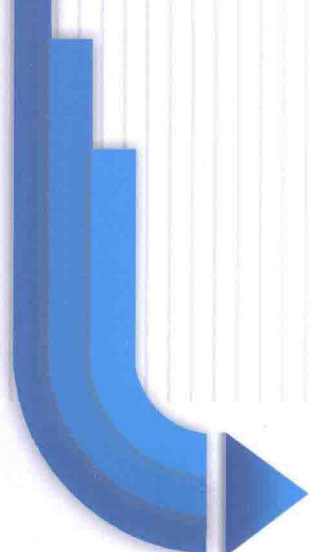
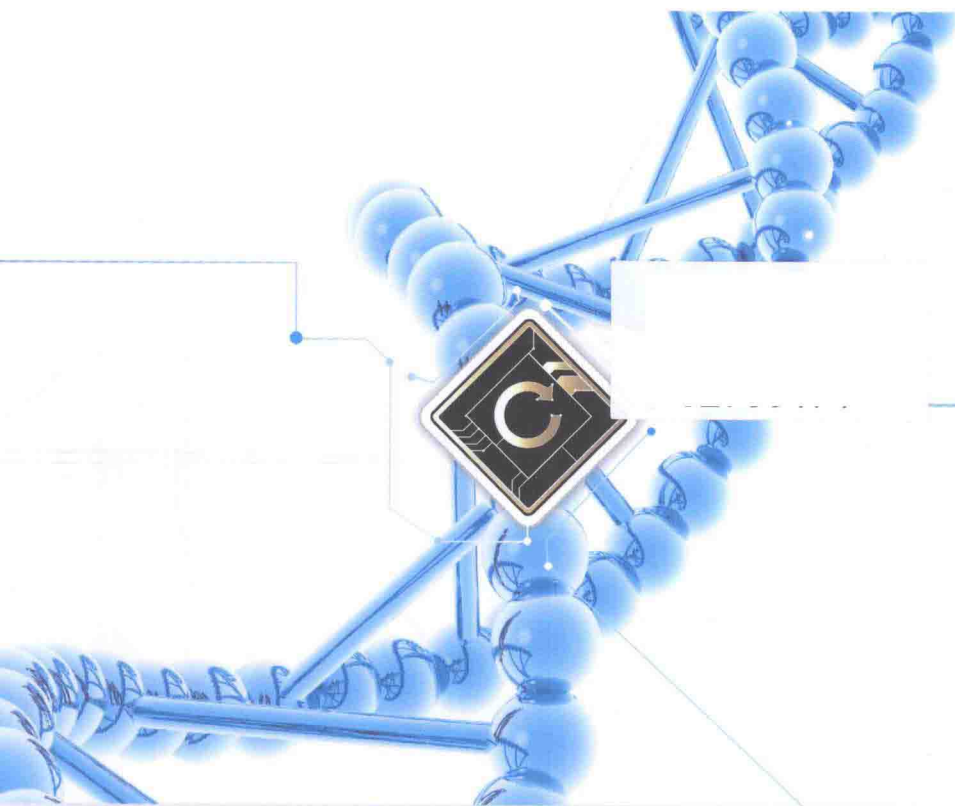




基因芯片制备 及数据分析技术

刘 军 冯艳君 宋 凯 著



西安电子科技大学出版社
<http://www.xduph.com>

基因芯片制备及数据分析技术

刘 军 冯艳君 宋 凯 著

西安电子科技大学出版社

内 容 简 介

本书是一本比较全面和系统研究基因芯片制备、图像处理和数据分析技术及其应用的学术专著。

本书在介绍基因芯片技术基本原理的基础上,简要阐述了基因芯片点样、核酸扩增、杂交、清洗等制备技术的实现方法。同时,重点针对核酸扩增热循环控制、荧光靶点聚焦成像、芯片图像扫描与拼接、基因芯片图像处理、数据分析等关键技术中存在的问题,给出了改进方法或提出了新型控制和计算模型,提高了控制品质 and 数据分析处理效率。

本书可供基因芯片分析设备设计开发人员、生物信息学领域科研人员使用,也可供测控技术与仪器、信号与信息处理和自动化等有关专业高年级本科生和研究生学习、研究和参考。

图书在版编目(CIP)数据

基因芯片制备及数据分析技术/刘军,冯艳君,宋凯著. —西安:西安电子科技大学出版社,2015.4
ISBN 978 - 7 - 5606 - 3633 - 7

I. ①基… II. ①刘… ②冯… ③宋 III. ①基因—芯片—数据—分析
IV. ①Q78

中国版本图书馆 CIP 数据核字(2015)第 060543 号

策划编辑 高 樱

责任编辑 阎 彬 董小兵

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfxb001@163.com

经 销 新华书店

印刷单位 北京京华虎彩印刷有限公司

版 次 2015年4月第1版 2015年4月第1次印刷

开 本 787毫米×1092毫米 1/16 印张 11.5

字 数 263千字

定 价 25.00元

ISBN 978 - 7 - 5606 - 3633 - 7/Q

XDUP 3925001 - 1

*** 如有印装问题可调换 ***

作者简介

刘军，男，汉族，辽宁省西丰人，1972年生。2010年毕业于中国科学院研究生院机械电子工程专业，获工学博士学位。目前，任教于沈阳理工大学信息科学与工程学院测控技术与仪器专业教研室，副教授，检测技术与自动化装置专业硕士生导师。主要研究方向为智能传感器与检测技术、图像与信号处理、现场总线仪表与控制技术。发表学术论文十余篇，其中EI收录3篇，辽宁省自然科学学术成果奖2项。软件著作权2项，国家发明专利1项。先后承担了辽宁省科学事业公益研究基金项目“桥梁裂缝扩展应力波检测关键技术研究”，辽宁省自然科学基金项目“基因工程关键技术研究”的子课题“核酸扩增杂交基因芯片检测分析系统研究和开发”，沈阳理工大学博士启动基金项目“基于多源图像融合的目标识别技术研究”，中国铁路科研院委托项目“基于计算机视觉的铁路罐车测量系统”，沈阳市科技局中小企业创新基金项目“基于计算机视觉的波纹管芯自动氩弧焊接系统联合开发”，中国科学院沈阳自动化研究所委托项目“激光冲击强化质量在线监测系统研究开发”、“激光成套设备激光光路转换控制及激光器控制系统开发”，沈阳恺米迪科技有限公司委托项目“可吸入颗粒物分析仪电气控制系统开发”，负责总体方案设计、控制系统硬件软件开发、算法设计等工作。电子邮箱：li-mail-sut@163.com。

冯艳君，女，汉族，辽宁省沈阳市人，1975年生，研究生学历，硕士学位。现于沈阳农业大学农业电气化与自动化专业攻读博士学位。同时，任教于沈阳理工大学信息科学与工程学院计算机科学与技术专业教研室，讲师。主要研究方向为计算机应用技术、服务计算等。发表学术论文十余篇，其中EI收录1篇。获国防科学技术进步奖1项，软件著作权2项。参与多项省部科技攻关项目及横向课题研究。电子邮箱：braverfyj@126.com。

宋凯，男，汉族，辽宁省辽中人，1964年生，教授/博导，博士学位。主要研究方向为计算机视觉、智能检测与控制。目前，在沈阳理工大学通信与信息系统专业从事教学、科研工作，沈阳农业大学农业电气化与自动化专业兼职博士研究生导师，辽宁省“百千万人才工程”百人层次人选。沈阳市第八届优秀科技工作者。辽宁省高新技术企业评审认定专家。校精品课“计算机网络”负责人，图像信息处理技术方向学术带头人。作为一线教师指导博士研究生4人、硕士研究生十余人，主讲本科主干课程“计算机网络”。积极进行教研、科研、学术等各项业务研究，近三年发表教学研究论文2篇，发表学术论文30余篇，其中EI检索11篇；主编出版教材3部，拥有国家发明专利1项。先后主持、参与国家级、省级课题7项。教学成果获省“教学成果奖”三等奖，科研成果获省“农业科技贡献奖”二等奖。先后承担多项辽宁省科学技术计划项目、辽宁省科学事业公益研究基金项目以及多项横向课题。电子邮箱：ap9351@163.com。

前 言

随着生物信息技术的飞速发展,基因芯片技术在诸多领域发挥着不可替代的作用:生物信息编码与存储、核酸检测与分析、遗传基因组学、食品药品检验、环境监测等方面的发展都离不开基因芯片技术。基因芯片技术是机械工程学、计算数学、分子生物学、信息科学等多学科交叉的产物,需要不同学科研究人员的通力合作才能实现基因芯片技术的深入发展和推广。

基因芯片技术是对微观生物遗传物质进行提取、处理、检测和分析的一种新型分子生物学试验方法。基因芯片设计、加工、制备与检测分析是实现基因芯片分析设备微型化、集成化、自动化、智能化的四项关键技术;基因芯片技术的发展必将对揭示生命本质和发展规律,提高临床医疗水平,保证药物、食品质量和安全做出重大贡献。

在基因芯片技术中,基因芯片 PCR(Polymerase Chain Reaction,聚合酶链式反应)热循环控制、微阵列图像分析以及基因芯片数据分析是基因芯片检测分析设备中的三个关键技术,决定着基因芯片检测分析系统的性能。

本书根据自动控制、数字信号处理、图像处理与模式识别、生物信息学等基础理论,对基因芯片 PCR 热循环控制和微阵列图像分析两项关键技术进行了深入研究,构建了基因芯片 PCR 热循环控制、微阵列图像扫描采集、微阵列图像处理和分析以及基因芯片数据抽取分析的理论框架。

根据 PCR 温度均匀性、精确性、快变性的控制特点和要求,本书提出一种基于环形管道“风浴”炉的 PCR 方案。该方案针对 PCR 温度升降速度慢、恒温精度低和系统不稳定等问题,采取一系列技术措施对 PCR 温控方案以及加热结构进行改进。书中采用 MATLAB 仿真工具对温度控制算法进行仿真实验研究,并对不同控制算法的控制仿真实验性能进行了分析和比较,提出了一种基于常规 PID,采用风门开关控制、温度误差死区、风门开度调节联合控制策略的 PCR 热循环控制算法,实现了 PCR 温度的快速高精度控制,有效改善了温度控制的动态和稳态性能。

针对基因芯片微阵列图像 CCD 扫描采集系统存在的显微成像人工调焦困难、荧光靶点需人工识别、图像分辨率低、扫描拼接速度慢等技术问题,提出了一种基于荧光靶点检测识别、显微自动调焦、芯片扫描路径规划、显微图像拼接的基因芯片微阵列图像 CCD 自动扫描采集系统,实现了基因芯片靶点阵列图像的自动扫描和采集。

本书在基因芯片图像预处理中,针对图像信噪比低、靶点阵列倾斜等问题,采用人工交互方式和图像旋转变换,实现了微阵列图像倾斜校正;采用数学形态学算子进行微阵列图像去噪,改善了微阵列图像分割质量,提高了基因芯片网格定位分析的准确性;采用基于热传导方程、中值曲率驱动方程、AMSS 方程的图像滤波算子,有效提高了基因芯片阵列及靶点图像的信噪比。

网格定位是确定和表达各个靶点在整个基因芯片微阵列图像上的二维空间几何位置的

一种图像分析技术,然而图像噪声和计算误差常常导致网格定位分析存在误划分问题。本书分别采用基于投影变换功率谱分析和峰值搜索的一维数字信号分析方法以及基于固定圆形模板匹配的方法来实现基因芯片微阵列图像的网格定位,提出一种基于投影变换差分序列分析和局部极值搜索的网格定位方法,提高了定位分析效率,降低了算法的复杂度。

基因芯片靶点图像分割是在网格定位确定的靶点方格区域内提取出单个靶点目标的图像分析技术。靶点图像分割质量决定了靶点数据抽取的准确性,但由于图像噪声的影响,一些靶点图像分割方法存在着过分割、误分割等问题。本书采用基于阈值、基于区域、基于边缘、基于 C-V 模型变分水平集等四种不同的分割方法进行靶点图像分割试验,通过后续靶点数据抽取对不同靶点图像分割方法的分割质量进行分析和比较。针对基因芯片靶点数据抽取问题,对常用的荧光靶点特征参数进行抽取并进行修正处理。在基因芯片数据表达方法方面,本书进行了数据数值表达方法和图形化表达方法的试验研究。本书指出的理论和方法为进一步的基因芯片数据分析和数据挖掘研究奠定了技术基础。

本书综合应用上述基础理论和关键技术,针对基因芯片制备(核酸扩增、杂交)技术、芯片靶点阵列图像检测、图像处理、数据分析等问题展开了研究,其研究范围包括基因芯片 PCR 热循环及杂交温度控制、微阵列图像 CCD 检测分析,并通过连续的基因芯片制备、芯片荧光成像检测、图像采集处理和图像数据分析过程,自动获取镶嵌于基因芯片中的物理和生物数据信息。

本书是对基因芯片制备、图像检测和处理、数据分析等技术研究和仪器设备研制工作的总结。相关研究和应用情况表明,上述研究成果、关键技术基因芯片分析工程中具有较好的实用价值和应用前景。

本书由刘军副教授、冯艳君博士和宋凯教授合作编写,其中,刘军撰写第 1~3 章,冯艳君撰写第 4~5 章,宋凯撰写第 6~8 章以及参考文献。本书全部算例和算法设计是由冯艳君完成的。

作 者

2014 年 6 月

目 录

第 1 章 引言	1
1.1 研究背景和选题意义	1
1.2 基因芯片设计、制备、检测及分析关键技术综述	4
1.2.1 基因芯片设计及制备关键技术	4
1.2.2 基因芯片杂交信号图像检测与采集关键技术	5
1.2.3 基因芯片图像和数据分析关键技术	6
1.3 研究课题在国内外发展现状	7
1.3.1 基因芯片制备检测分析设备的研究现状与进展	7
1.3.2 基因芯片 PCR 热循环控制技术的研究现状与进展	9
1.3.3 基因芯片杂交信号图像检测技术的研究现状与进展	11
1.3.4 基因芯片图像和数据分析技术的研究现状与进展	13
1.4 本书主要内容	18
第 2 章 基因芯片 PCR 热循环温度控制技术研究	20
2.1 基因芯片制备 PCR 方案的确定	20
2.1.1 PCR 的基本特点和控制性能要求	20
2.1.2 常用 PCR 方法及各自特点	21
2.1.3 PCR 方案比较和确定	21
2.2 PCR 热循环温控方案及风浴炉加热结构改进研究	22
2.2.1 基于温控试验和性能比较分析的 PCR 温控方案改进	22
2.2.2 基于温控试验的 PCR 炉加热结构改进	24
2.3 基于 MATLAB 的 PCR 热循环温度控制算法仿真研究	26
2.3.1 PCR 热循环温度常规 PID 控制算法仿真	26
2.3.2 PCR 热循环温度预测控制算法仿真	28
2.3.3 不同 PCR 热循环温度控制算法仿真实验性能比较	31
2.4 PCR 热循环温度控制算法试验研究	31
2.4.1 基于常规 PID 和风门开关控制温度的快速跟踪算法	31
2.4.2 基于常规 PID 和风门开度调节温度的快速跟踪算法	35
2.4.3 基于模糊自整定 PID 的 PCR 温度控制算法	36
2.4.4 不同 PCR 热循环温度跟踪控制算法温控试验性能比较	39
2.5 本章小结	39
第 3 章 基因芯片靶点阵列图像 CCD 自动扫描采集系统研究及算法实现	40
3.1 引言	40
3.2 基因芯片靶点阵列图像 CCD 自动扫描采集系统概述	40

3.3 基因芯片荧光靶点检测识别	42
3.3.1 荧光靶点图像获取、预处理及分割	42
3.3.2 改进前的荧光靶点检测识别算法	43
3.3.3 改进后的荧光靶点检测识别算法	44
3.3.4 荧光靶点识别试验结果分析	44
3.4 基于图像处理和粗精结合原则的显微自动调焦方法	45
3.4.1 显微调焦自动控制系统	45
3.4.2 聚焦清晰度评价函数及计算公式	46
3.4.3 基于粗精结合原则和清晰度评价函数的调焦方法	46
3.4.4 显微成像自动调焦试验	47
3.5 芯片图像预扫描	48
3.6 基因芯片靶点阵列图像扫描与图像拼接	48
3.6.1 图像扫描	48
3.6.2 图像拼接	48
3.7 本章小结	50
第4章 基因芯片图像预处理和网格定位方法研究及其算法实现	51
4.1 引言	51
4.2 基因芯片图像预处理方法及试验研究	52
4.2.1 常规图像预处理方法简介	52
4.2.2 基于数学形态学的图像预处理方法及试验研究	56
4.2.3 基于偏微分方程的图像预处理方法及试验研究	60
4.3 基于投影变换和功率谱分析的网格定位方法	63
4.3.1 方法原理	63
4.3.2 算法实现	64
4.3.3 网格定位分析试验及结果分析	65
4.4 基于投影差分序列分析和局部极值搜索的网格定位方法	68
4.4.1 方法原理	68
4.4.2 算法实现	68
4.4.3 网格定位分析试验及结果分析	69
4.5 基于固定半径圆模板匹配的网格定位方法	71
4.5.1 方法原理	71
4.5.2 算法实现	71
4.5.3 网格定位分析试验及结果分析	72
4.6 基于荧光靶点图像分割、标记和识别的网格定位方法	73
4.6.1 算法基本思想	73
4.6.2 算法实现及处理结果	74
4.6.3 网格定位试验结果	74
4.7 网格定位算法性能比较和选择	75
4.8 本章小结	75

第 5 章 基因芯片靶点图像分割及数据抽取方法研究	76
5.1 基因芯片靶点图像分割及数据抽取技术概述	76
5.2 基于阈值的图像分割方法及试验	76
5.2.1 全局阈值图像分割方法	76
5.2.2 最佳阈值图像分割方法	79
5.2.3 动态阈值图像分割方法	79
5.2.4 基于一维最大熵的图像分割方法	81
5.3 基于区域生长的图像分割方法	81
5.4 基于边缘检测算子的图像分割方法	83
5.4.1 图像边缘检测算子	83
5.4.2 边界跟踪与边界链码	84
5.5 基于偏微分方程的图像分割方法	85
5.5.1 变分原理与梯度下降流	85
5.5.2 变分水平集方法	88
5.5.3 基于 Chan - Vese 模型的变分水平集图像分割方法	90
5.6 曲线演化的水平集方法	93
5.6.1 曲线几何演化的一般方程式	93
5.6.2 求解演化方程的质点标注法	94
5.6.3 水平集方法	94
5.6.4 水平集方法的优点	97
5.7 图像分割方法对靶点数据抽取质量影响试验及分析	97
5.8 基因芯片荧光靶点数据抽取方法研究	98
5.8.1 荧光靶点数据抽取	98
5.8.2 数据修正处理方法简介	98
5.9 基因芯片数据表达方法	101
5.9.1 数据数值表达方法	101
5.9.2 数据可视化表达方法	101
5.10 基因芯片表达分析技术研究展望	102
5.11 本章小结	103
第 6 章 核酸扩增杂交基因芯片检测分析系统实现	104
6.1 核酸扩增杂交基因芯片检测系统的基本功能和性能指标	104
6.1.1 核酸扩增杂交基因芯片检测仪的基本功能	104
6.1.2 核酸扩增杂交基因芯片检测仪的性能指标	105
6.2 系统的机械结构	105
6.3 系统的工作流程	107
6.4 系统电气控制原理	108
6.5 系统软件设计	109
6.6 与国内外同类系统的比较	111
6.7 本章小结	112

第 7 章 基因芯片数据分析技术	113
7.1 基因表达数据的获取	113
7.1.1 cDNA 微阵列	114
7.1.2 寡核苷酸芯片	116
7.1.3 基因表达数据的网络资源	116
7.2 基因表达数据预处理	118
7.3 基因表达差异的显著性分析	121
7.3.1 倍数分析	121
7.3.2 t 检验	121
7.3.3 贝叶斯分析	122
7.4 基因表达谱聚类分析	123
7.4.1 相似性度量函数	124
7.4.2 聚类方法	125
7.4.3 基于模型的聚类方法	131
7.4.4 支持向量机	131
7.4.5 聚类结果的可视化	133
7.4.6 聚类结果的定量评价	135
7.5 基因表达数据的分类分析	137
7.5.1 朴素贝叶斯分类法	138
7.5.2 k -近邻法	138
7.5.3 其他分类法	138
7.6 主成分分析 PCA	139
7.7 基于基因表达谱的基因调控网络研究	147
7.7.1 布尔网络模型	148
7.7.2 线性组合模型	149
7.7.3 加权矩阵模型	150
7.7.4 数据整合分析	150
第 8 章 结论与展望	152
8.1 结论	152
8.2 展望	153
8.2.1 基因芯片技术领域的发展趋势和发展前景	153
8.2.2 未来的研究工作	154
参考文献	156
与本专著研究工作相关的学术论文.....	172
致 谢	173

第 1 章 引 言

1.1 研究背景和选题意义

20 世纪 90 年代中期至 21 世纪初,生命科学研究进入高速发展时期,在生物化学、分子生物学领域涌现的大量研究成果和先进技术快速向生命科学领域渗透。随着生命科学研究中新基因的发现、基因测序等工作的开展以及未来基因组学研究的深入,生命科学研究工作者面临着大量生物遗传物质处理、生物信号产生与获取、生物数据抽取、生物信息处理与分析等繁复工作的困扰。为加快生物遗传物质并行处理速度,提高生物试验以及生物信息获取与处理效率,降低科研、医疗等工作者的实验工作强度,生物芯片技术应运而生。

生物芯片(biochip)是指采用光导原位合成或微量点样等方法,将大量生物大分子,比如核酸片段、多肽分子甚至组织切片、细胞等生物样品有序地固化于支持物(如玻璃片、硅片、聚丙烯酰胺凝胶、尼龙膜等)的表面,组成密集的二维探针分子阵列,然后将其与已标记的待测生物样品中的靶分子杂交,通过特定的仪器,如激光共聚焦扫描仪或 CCD 扫描仪对杂交信号的强度进行快速、并行、高效的检测分析,从而判断样品中靶分子的有无或数量。

生物芯片试验设备是利用微机械、微电子、生物化学、物理技术、控制技术、计算机技术构建于固体芯片表面或内部的微流体或微阵列试验分析单元或系统,实现了生命科学研究所涉及的不连续的试验过程,如样品制备、化学反应和检测分析等的连续化、集成化、自动化和微型化。

生物芯片技术是一种高通量、高灵敏度和高特异性检测技术,包括四项关键技术:芯片微阵列构建、芯片样品制备、生物分子反应和信号检测分析。目前常见的生物芯片主要分为三大类:基因芯片(Gene chip)、蛋白质芯片(Protein chip)、芯片实验室(LOC, Lab-On-Chip)。

基因芯片又称为 DNA 芯片(DNA chip)、DNA 阵列(DNA array),此外,因 DNA 是一种寡核苷酸,故基因芯片也被称为寡核苷酸阵列(oligonucleotide array),它是在基因探针的基础上研制出来的,其基本原理是将大量的探针分子固定在特定的支持物上,然后利用基因探针与特异寡核苷酸的碱基互补原理,与经过荧光染料标记的样品分子进行杂交,通过检测和分析杂交反应荧光信号的强度和分布情况来获取样品分子数量、基因序列等有关信息。基因芯片由 Affymetrix 公司率先研制出来并最早实现了商品化。其突出优点是检测过程的快速高效,可以同时大量的基因甚至基因组进行分析。通常,基因芯片试验过程包括样品制备、聚合酶链式反应(Polymerase Chain Reaction, PCR, 包括核酸扩增反应和杂交(Hybridization)反应)、杂交信号检测分析三个基本环节。

蛋白质芯片(Protein chip)是构建在抗体与抗原结合特异性反应(免疫反应)基础上的

生物芯片。它以蛋白质代替 DNA 作为检测目的物,在适当的固相载体上以预先设定的方式结合大量的特定蛋白质分子(抗原或抗体),形成蛋白质的微阵列,然后加入与之特异性结合的标记蛋白分子,通过对标记物的检测来实现蛋白质的检测。该技术能够高效检测蛋白质的存在和运动变化,并且所需蛋白质样品量极少,检测结果稳定而灵敏,在医药和临床方面具有极为广阔的应用前景。

芯片实验室是能够完成生化分析全部操作过程的微分析系统,是目前生物芯片技术发展的最高阶段,它采用微加工技术在固相基片上加工出不同的功能微区,集样品制备、基因扩增、核酸标记及检测为一体,将各个试验步骤微缩于单个芯片上,从而将复杂而不精确的生物分析过程自动化、精确化。芯片实验室技术具有精确、灵活、使用方便等优点,特别是集中式的实验室分析和即时检验,在临床诊断等领域中具有非常广阔的商业应用前景。

生物芯片是生物科学与技术、制造科学与技术、信息科学与技术高度集成的产物,生物芯片技术的发展将对我国生命科学研究、医学诊断、新药筛选具有革命性的推动作用,也将对我国人口素质、农业发展、环境保护等作出巨大的贡献,同时带动我国科学技术的整体进步,为各相关高科技产业创造发展机会。

目前,生物芯片及其分析技术已成为 21 世纪生命科学和社会经济领域的前沿技术,生物芯片产业也在不断发展壮大,并已取得良好的社会效益。在我国,生物芯片检测分析技术及其系统设备已经开始在生命科学研究、临床疾病诊断、新药筛选、个性化给药、预防医学、食品安全检测、转基因产品开发、环境治理等领域得到广泛应用。

研究与掌握生物芯片检测分析核心技术是我国生命科学研究和社会经济发展的迫切要求。长期以来,我国的生物芯片制备、检测、分析仪器或设备市场一直被国外专业厂商占据,国内相关企业缺乏自有核心技术和自主知识产权,生物芯片设备关键技术研究长期被忽视,导致国产生物芯片技术与设备处于相对落后的状况。因此,研究生物芯片设计、加工、制备、检测、分析等共性关键技术,具有良好的应用价值和社会经济意义。

基因芯片技术采用体外核酸扩增检测分析方法实现基因或基因组分析,具有灵敏度高、表达特异性强、试验通量大和效率高的特点,在遗传工程等诸多领域得到广泛应用。在基因芯片检测分析系统中,高性能的 PCR 扩增杂交控制技术研究不同个体基因变异,不同组织、不同时间、不同生理状态等基因表达差异的关键技术前提,而高性能的基因芯片检测分析系统对于高效、快速地检测和分析大量遗传信息具有十分重要的意义。

基因芯片设计、制备、检测及分析技术的研究工作是从 20 世纪 90 年代中期开始的,现已形成完整的技术体系。它是建立在微光机电系统设计集成及制造、传感器与检测技术、控制理论、信号与信息处理、计算机数字图像处理、模式识别、生物信息学、数据挖掘等诸多理论及技术基础上的生物遗传信息检测分析技术,伴随着相关学科及研究的进展而得到迅速发展。

随着基因芯片设计、制备、检测和分析技术的发展,如何高效率、高精度、高可靠地提取基因芯片携带的生物遗传信息,已成为基因芯片分析技术中的首要任务。只有准确获取镶嵌在基因芯片中的生物数据和信息,才能实现基因芯片的特定应用。典型的基因芯片分析系统包括三个基本子系统:芯片设计制备子系统、信号检测子系统、图像和数据分析子系统。图 1-1 所示为采用自动分区点样制备方法的基因芯片设计、制备、检测、分析流程。

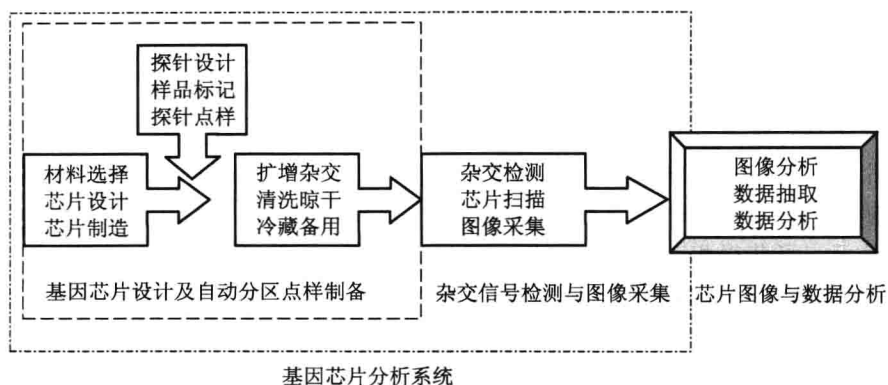


图 1-1 基因芯片设计、制备、检测、分析流程

在基因芯片设计制造过程中,为保证芯片试验可靠和高效,必须根据基因芯片生物动力学特性以及生物反应效率和质量要求,合理选择芯片材料、合理设计加工芯片及其反应体系结构。

基因芯片制备过程十分复杂,制备方法主要包括原位合成制备和自动分区点样制备两种方法,其中基因芯片自动分区点样制备方法因分区点样速度快、试验通量大、核酸扩增量显著、试验反应速度快、基因表达差异性显著而得到广泛采用。在基因芯片制备过程中,生物科学工作者必须根据特定应用要求合理设计与靶基因匹配的探针,并利用荧光染色剂对靶 DNA 分子进行荧光标记,然后再利用生物芯片点样设备将探针分子基团固定在芯片基片上。为提高检测灵敏度,首先必须对待测定的靶基因进行核酸扩增反应,之后可获得上百到上万倍的靶基因拷贝;然后令扩增后的靶分子与芯片基片上的探针分子在基因芯片微反应池中的液体环境及恒温条件(通常为 $45^{\circ}\text{C}\sim 50^{\circ}\text{C}$)下进行杂交反应(如果靶分子和探针分子碱基配对,则二者进行杂交反应,否则不进行杂交反应)。因此,杂交反应结束后,有反应的样点位置上有荧光发出,否则无荧光发出。为避免噪声干扰后续杂交荧光信号检测,芯片杂交后需利用专用液体对芯片进行清洗,然后将芯片晾干冷藏备用,至此完成基因芯片制备过程。

基因芯片制备完成后,为进一步提取镶嵌在芯片中的生物遗传信息,需要采用光电检测技术对芯片上的探针杂交信号进行检测、定位和成像,以便获得基因芯片微阵列杂交图谱。首先利用激光或荧光激发光源、精密光学显微镜(如荧光显微镜或共焦显微镜)及专用光电探测器(光电倍增管或 CCD 相机)对基因芯片表面经过荧光标记的呈矩形阵列排列的探针杂交信号(荧光信号)进行检测,由于探针点样密度高、样点小、杂交信号微弱,故采用的光电传感器应具有高空间分辨率、高灵敏度、低噪声和很好的线性响应。然后采用 X-Y 二维扫描平台对芯片表面进行连续扫描和杂交信号检测,得到呈阵列排布的探针杂交信号分布图。为获取基因芯片微阵列图像,需要采用杂交信号前置放大、模/数转换、视频编码、数/模转换、视频放大输出等电路进行处理。对于基因芯片微阵列图像 CCD 检测、扫描、采集系统,由于采用荧光显微镜进行芯片成像且成像视场小,因此一次成像不能获取芯片图像全貌。为通过多次成像来获得芯片清晰的高分辨率全图,还有必要对荧光显微镜进行自动调焦控制,以及对获取的芯片局部扫描图像再次进行图像拼接。至此,完成了基因芯片杂交信号检测和图像扫描采集过程。

在获取芯片基因图谱的全图后,为了进一步提取基因芯片试验结果,必须进行芯片图像和数据分析过程。芯片图像数据分析与常规图像数据分析过程相比,在具有相同点的同时也具有其自身的特点。为了削弱芯片制备和检测过程中引入的噪声,首先对芯片数字图像进行图像滤波、增强等预处理;然后为确定芯片上探针阵列中各个靶点在图像中的位置,对芯片进行网格定位分析,网格定位分析结果可以用靶点中心坐标或网格划分直线坐标数据来表示,也可以用芯片阵列图像上的网格划分直线可视化地表示出来;芯片网格定位分析完成后,可获得相互分离的芯片微阵列中的各个探针靶点图像;为进一步提取各个探针靶点所携带的生物数据信息,必须采用图像分割技术将各个探针靶点准确地与背景分离开来,图像分割的准确性、可靠性、稳定性极大地影响着后续数据抽取的准确性;通过图像分割,将目标(探针靶点)与背景严格区分开来,此时再对目标和背景的亮度信息进行检测,这就是基因芯片数据抽取过程;抽取芯片探针阵列中所有靶点数据后,再结合网格定位分析结果以及芯片探针阵列的行列定义信息,就能获得大量的关于不同基因不同特异性表达的数据矢量样本,对这些数据进行分析 and 映射即可获得试验样品(如血清)中所携带的生物遗传(如抗原抗体或致病基因)信息。

综上所述,基因芯片通过在芯片表面构建呈微阵列排布的生物探针,实现了生物遗传物质并行制备、生物遗传信息的高通量检测分析,提高了生物试验以及生物信息获取与处理效率,降低了科研医疗实验工作强度。但是,现存的基因芯片检测分析系统普遍存在下列技术问题:

- (1) 非集成化。功能单一,试验环境开放,样品易受污染。
- (2) 非连续化。试验过程不连续,试验环境和试验条件不易控制。
- (3) 非自动化。手工方式进行芯片扩增杂交、注液、清洗、图像扫描采集等工作,试验工作强度大。
- (4) 图像数据处理和分析效率低。

因此,研制基因芯片自动检测分析系统是解决系统集成化、测试连续化、试验自动化和数据处理分析智能化问题的根本途径。

本书通过对基因芯片 PCR 热循环控制方法以及 PCR 温度跟踪控制算法、基因芯片图像检测扫描采集技术、基因芯片图像预处理及网格定位分析方法、基因芯片靶点图像分割及数据抽取和分析方法等内容的理论和技术研究,为基因芯片分析技术和系统设备的推广应用奠定了良好的技术基础。

1.2 基因芯片设计、制备、检测及分析关键技术综述

1.2.1 基因芯片设计及制备关键技术

基因芯片设计及制备包括四个关键技术:芯片设计和制造、样品设计和制备、芯片分区点样和核酸扩增杂交反应控制。

芯片设计和制造是一项微纳米设计和制造技术,芯片结构是决定生物遗传物质生化反应动力学性能的首要因素之一。根据基因芯片的基本结构,基因芯片的设计及制造包括探针支持物(玻璃、尼龙膜等薄片)、微流道、微反应池、微结构或器件连接结构以及微电极

键合结构等的设计和制造工作。

样品设计和制备是生物遗传学和生物信息学相结合的一项技术,同样对生化反应动力学性能有重要影响,是实现基因芯片特定应用目的不可缺少的关键生物遗传技术。样品设计和制备主要包括探针或引物设计与制备、标记物选择、样品标记和清洗缓冲液选择与制备等环节。

芯片分区点样技术是将经荧光标记的靶分子液体反应物样品以分区分块的矩形阵列的形式固定在探针支持物如玻璃、尼龙膜等薄片上的技术,由于这些薄片面积很小(只有1平方厘米到几平方厘米)、样点微小(几微米到几十微米)、样品量少且要求精确定量控制,因此,必须采用专用的芯片点样设备来完成芯片分区点样操作,以达到样点大小和间距均匀以及自动分区分块形成矩形样点阵列的分区点样目标。

核酸扩增杂交反应控制即 PCR 热循环温度控制,是基因芯片制备过程中的最后关键技术环节,包括 PCR 控制对象分析和 PCR 控制方案设计两方面的研究内容。PCR 控制对象分析及 PCR 控制方案设计必须结合芯片设计制造技术和 PCR 反应控制要求来进行,并且以提高 PCR 反应效率、动态/稳态性能、可靠性等作为设计标准。PCR 是一个生物链式反应过程,对反应过程的温度控制性能有着极高的要求,要求温度升降速度足够快、过冲量尽可能小、温度稳态精度高、温度场均一性好。对核酸扩增反应的温度-时间模式曲线要求可以有很多种,比较典型的单次 PCR 温度循环的温度-时间模式曲线如图 1-2 所示,在单次 PCR 循环中,主要分为三个阶段:

(1) $t_0 \sim t_2$ 时刻之间,即温度从常温跳变到 95°C 并持续一段时间 $t_2 - t_1$,称为高温变性阶段。

(2) $t_2 \sim t_4$ 时刻之间,即温度从高温 95°C 跳变到 55°C 并持续一段时间 $t_4 - t_3$,称为低温(复性)退火阶段。

(3) $t_4 \sim t_6$ 时刻之间,即温度从低温 55°C 跳变到中温 72°C 并持续一段时间 $t_6 - t_5$,称为中温延伸阶段。对于已报道的特殊设计的 PCR 系统,温度升降速度最高达 $20^{\circ}\text{C}/\text{s}$,恒温阶段稳态精度达 $\pm 0.5^{\circ}\text{C}$ 。实用的商用 PCR 系统,温度升降速度达 $2^{\circ}\text{C}/\text{s}$ 以上,恒温阶段稳态精度为 $\pm 1^{\circ}\text{C}$ 。

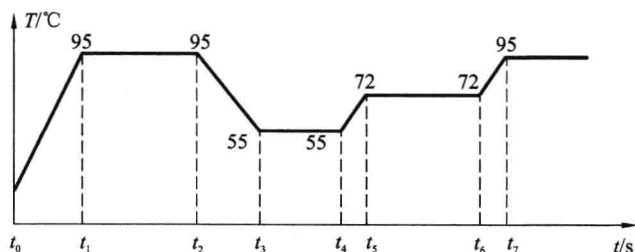


图 1-2 核酸扩增杂交反应系统的温度-时间模式曲线

1.2.2 基因芯片杂交信号图像检测与采集关键技术

基因芯片杂交信号检测是基因芯片分析技术中的重要组成部分,是在基因芯片制备过程完成后,用于提取镶嵌于基因芯片中的生物遗传信息的首要步骤。

杂交信号图像检测技术主要包括杂交信号产生、探测、传输和变换以及杂交信号成像

两个部分。

产生杂交信号的最常用方法是杂交反应标记物法,因使用的反应标记物不同,杂交信号检测方法大致可分为两类:荧光标记物杂交信号检测和生物素标记杂交信号检测。利用荧光标记物进行杂交信号检测的研究者最多,因此,相应的检测方法也最多、最成熟。无论采用何种杂交信号检测方法,都是采用基因芯片 DNA 分子发光,通过检测标记物发光信号来获取基因芯片杂交图谱^[9]的。

探测分子杂交的方法很多,多数方法都是采用光学成像检测技术。因此,杂交信号光学成像探测技术是基因芯片杂交信号图像检测系统中的一项关键技术,杂交信号光学探测系统的性能决定着杂交信号图像质量。高密度基因芯片具有体积小,密度大,点样量少、杂交信号微弱的特点,需要采用高稳定激发光源、高分辨率光学系统和高灵敏度弱光信号探测装置,因而对探测方法以及光学装置的灵敏度以及线性响应性能要求很高。

杂交信号传输通过专用光学成像系统实现。因荧光显微镜可以有选择性地激发和探测样品中的混合荧光标记物,并具有很好的空间分辨率和热分辨率,特别是当荧光显微镜中使用了共焦激光扫描技术时,分辨能力在实际应用中可接近由数值孔径和光波波长决定的空间分辨率(普通光学显微镜很难做到),从而为基因芯片微型化提供了重要的检测技术基础。多数杂交信号传输和成像方法都是在入射照明式荧光显微镜(epifluorescence microscope)基础上发展起来的,包括激光扫描荧光显微镜、激光共焦扫描荧光显微镜、采用 CCD 相机和改进荧光显微镜以及将基因芯片直接制作在光纤束切面上并结合荧光显微镜的光纤传感器微阵列等。

在杂交信号成像技术中,杂交信号的光电转换普遍采用光电倍增管或制冷 CCD 相机实现,前者通过光电倍增管获取的是微弱的模拟电信号,而后者直接获得杂交分子的数字荧光图像。对于采用光电倍增管作为探测器的系统,必须进行模拟光电信号放大、信号调理、模数转换、视频编码、数模转换以及芯片全图扫描等处理后,才能获得基因芯片靶点阵列的整幅数字图像。杂交信号 CCD 成像技术是一种基于半导体制冷 CCD 相机和专用荧光显微镜的基因芯片杂交信号成像检测方法。在此类基因芯片 CCD 扫描系统中,半导体制冷 CCD 相机和专用荧光显微镜是关键光电检测仪器,显微镜自动调焦、图像扫描和采集、显微图像拼接是实现高分辨率基因芯片图像采集的三个关键技术。基因芯片 CCD 扫描仪是在生命基础科学研究以及临床医疗中广泛采用的芯片检测扫描设备,可以对基因芯片进行荧光靶点阵列图像扫描检测,具有成本低、分辨率高、检测扫描速度快等特点。基因芯片荧光靶点阵列图像 CCD 检测扫描采集技术是基因芯片杂交图谱图像 CCD 扫描采集系统中的关键技术,是影响扫描仪性能的关键因素。

1.2.3 基因芯片图像和数据分析关键技术

基因芯片图像和数据分析关键技术主要包括图像分析和数据分析两个方面的内容。

1. 基因芯片图像分析关键技术

基因芯片图像分析技术包括芯片图像预处理、芯片靶点阵列图像网格定位、靶点图像分割、靶点图像表示与描述等。

基因芯片图像预处理包括图像增强、图像倾斜校正等内容,由于基因芯片图像含有大量噪声点、瑕疵等干扰信号,而增强图像有用信息、提高靶点图像的信噪比对于可靠提取

图像中携带的生物数据至关重要。因此,芯片图像预处理一直是基因芯片图像分析研究热点之一。

图像增强是增强芯片图像中 useful 信号、抑制噪声影响以便提高芯片图像信噪比的必要手段。图像增强包括空域增强、频域增强、变换域增强、图像锐化、图像滤波等方法。图像滤波是芯片图像分析的首要步骤,能否在保留有用信号的前提下滤除噪声干扰,成为影响芯片分析结果的重要因素。常用的图像消噪方法包括均值滤波、中值滤波、数学形态学滤波、小波变换滤波和基于偏微分方程的图像滤波等。芯片图像消噪方法具有其特殊性,适用于其他图像消噪处理的方法不一定适用于芯片图像。

芯片图像倾斜校正是在芯片靶点阵列相对于图像栅格出现旋转倾斜时采用的图像校正技术。图像倾斜校正可以以自动方式完成,也可采用人工交互方式进行。

芯片靶点阵列网格定位是为确定芯片上靶点阵列中各个靶点在图像中位置而进行的图像处理和分析过程。网格定位分析结果可以用靶点中心坐标或网格划分直线坐标数据来表示,也可以用芯片阵列图像上的网格划分直线可视化地表示出来。

靶点图像分割是为进一步提取各个探针靶点所携带的生物数据信息,将各个探针靶点准确地与背景分离开来的图像处理步骤。适用于靶点图像分割的方法主要包括基于阈值的方法、基于区域的方法、基于边缘的方法以及基于变分水平集的方法。

靶点图像表示与描述是为实现基因芯片靶点图像分割、识别、定位、数据抽取等分析目标的必要的图像分析技术。用于基因芯片图像分析的图像描述与表示方法主要包括靶点区域灰度直方图特征描述、靶点区域的统计矩特征描述、封闭边界 Freeman 链码描述等。

2. 基因芯片数据分析关键技术

基因芯片数据分析技术包括数据抽取、数据表达和表达分析三个内容。

数据抽取包括目标及背景亮度提取、亮度比率计算、背景修正计算、比率修正计算、数据归一化处理等,数据抽取是实现以像素表达或几何描述的靶点图像转换为以数据或曲线描述的物理数据的必要步骤。

数据表达包括数值表达和可视化表达两种方式,其目的是使芯片数据表达定量化、更加直观以及建立起物理数据与生物特异性表达之间的映射关系。

表达分析根据基因芯片应用目的和分析目标的不同分为判定分析和聚类分析两种技术,其中判定分析技术主要用于需要进行定性判别的场合,如某种生物特异性表达的有无;而聚类分析技术主要用来实现某些分类应用,如基因片段的分类或者生物意义上在相对定量测定如不同特异表达的相对水平。

1.3 研究课题在国内外发展现状

1.3.1 基因芯片制备检测分析设备的研究现状与进展

自 1996 年美国 Affymetrix 公司成功研制出世界上首批用于药物筛选和实验室试验用的生物芯片以及生物芯片系统以来,世界各国在生物芯片研究方面快速前进,不断取得新的突破。美国的 Hyseq 公司、Syntexi 公司、Nanogen 公司、Ineyte 公司及日本、欧洲各国都积极开展 DNA 芯片技术研究工作,摩托罗拉、惠普、IBM 等跨国公司更是相继投入巨