

### 新闻动态

- Partners HealthCare的转译研究合作
- IPSOGEN签署Powered by Affymetrix协议
- 昂飞芯片应用的最新发现和最新项目
- Mouse and Rat Gene 1.0 ST Arrays
- 水稻tiling芯片和犬科全基因组SNP芯片

### 专题综述

遗传关联性研究的实验设计

### 技术介绍

GeneChip®DNA芯片设计

### 新技术与新启示

人类遗传变异, 拷贝数变异 (CNVs) 和疾病研究及检测

### 研究指南

DNA芯片实验和数据分析流程

### 应用案例

- WTCCC全基因组关联研究七种主要疾病
- HapMap人群中发现大量CNVs
- MIP技术与I型糖尿病易感性位点发现
- CustomSeq®与线粒体的体细胞性突变

### 本期相关技术要点

昂飞DNA分析产品: 从全基因组关联研究到定向基因分型和重测序

### 新闻动态

#### 昂飞公司和 Partners HealthCare的转译研究合作

7月23日, 昂飞公司和美国的医疗集团医护伙伴 (Partners HealthCare) 延续在转译研究 (Translational Research) 方面的合作并签署了一个新的合同芯片生产 (CAM) 的供应协议, 这使得医护伙伴的研究者能够将近期的芯片结果转换成得到充分验证的分子诊断试验。这种基于芯片的检测方法将有助于医生更为有效地对病人进行诊断并制定个性化治疗方案。这个刚刚签订的协议是建立在双方已经成功建立转译研究合作的基础之上的。去年, 哈佛医学院 (HMS)、医护伙伴以及哈佛医学院一医护伙伴遗传学和基因组学研究中心 (HPCGG) 的研究者采用昂飞技术发现了与一些复杂疾病相关的基因序列和标记。这些与临床相关的资料将用于开发分子诊断的检测技术。

#### 昂飞公司和IPSOGEN签署Powered by Affymetrix协议

8月28日, 昂飞公司和IPSOGEN宣布他们已经签署了一个Powered by Affymetrix (TM) (PbA) 协议, IPSOGEN是一家著名的设在法国和美国康涅狄格的分子诊断公司。根据该协议, IPSOGEN获得了昂飞芯片技术的非独占使用, 将在全球范围内开发并销售体外诊断 (IVD) 检测技术, 最初将用于乳腺癌诊断。昂飞公司诊断产品有GeneChip®3000Dx系统 (GCS3000Dx), 这是第一种用于分子诊断实验室的芯片仪器系统。美国食品和药物管理局已经对GCS3000Dx给予了批准, GCS3000Dx还通过了欧盟关于体外诊断使用的CE认证。美国昂飞公司及其PbA (Powered by Affymetrix (TM)) 合作伙伴正在利用该系统开发基于芯片的体外诊断试验。这种用户定制设计技术在很多方面都有应用, **Powered by AFFYMETRIX** 包括临床诊断、法医学、动物、工业和食品检测等等。

#### Empire Genomics获得一系列昂飞专利的非独占的全球许可以用于研究或诊断开发

9月7日, Empire Genomics获得了核酸芯片制造、使用和销售以及比较基因组杂交技术 (CGH) 相关产品和服务的一系列昂飞专利的非独占的全球许可。此类芯片和服务可以用于研究或诊断目的。位于美国纽约州布法罗市的 Empire Genomics公司, 活跃于分子诊断和基于基因组的检测研究领域, 它提供一系列一流的分子染色体组型平台技术的产品与服务, 用于对染色体畸变进行定性和定量分析。

#### Almac公司将昂飞芯片用于DSA® 诊断技术

Almac公司的癌症DSA® (Disease Specific Array) 研究工具应用昂飞芯片加快在发现、开发、核证并最终转向临床应用等方面研究的速度。作为第一种基于一种疾病的转录组的产生的更多有重要意义的信息的信息的芯片, 比起在一般芯片上获得的信息, 这些信息更为可靠, 且与所感兴趣的疾病更为相关。

## 昂飞芯片应用的最新发现

• 7月19日,英国Wellcome Trust Case Control Consortium (WTCCC)和心脏病协会 (Cardiogenics Consortium) 的研究者采用昂飞芯片技术发现了几个新的遗传变异,这些变异可能会提高冠心病 (CAD) 患病风险率达36%左右。该研究的结果发表在2007年7月18日的New England Journal of Medicine上。

• 9月20日,研究人员完成了一组来自于Framingham心脏研究 (Framingham Heart Study) 的两代参与者的全基因组扫描分析。研究者利用100K昂飞人类图谱分析芯片评估了1,345个FHS志愿者的基因型。在该项工作基础上,已有17篇论文同时发表了包括心血管疾病、乳腺癌和前列腺癌、与睡眠和生理节律相关的疾病、糖尿病等疾病的。该研究发表在9月19日的在线BMC医学遗传学杂志上 (BMC Medical Genetics)。

• 10月2日,一个由哈佛和麻省理工学院的Broad研究院、波士顿儿童医院、英国牛津大学和埃克赛特半岛医学院的研究者组成的研究团队,通过运用500K芯片和约5,000个样本的群体,确定了HMG2基因的变异与人的身高有很强的相关性。先前对于一般人群中身高差异的研究缺乏全面的统计学能力以检测真正的遗传关联性。该项目的研究结果发表在今年10月的Nature Genetics上。

## 昂飞芯片应用的最新项目

• 7月9日,欧洲细胞遗传学研究计划正式启动,它是昂飞公司和荷兰Nijmegen大学、德国Tuebingen大学以及英国伯明翰的NHS区域遗传学实验室共同合作的一个项目。合作方正在利用昂飞芯片技术确定那些在精神迟滞儿童中引起此类疾病的特异性突变。

• 7月25日,费城儿童医院将采用昂飞的人类全基因组SNP6.0芯片进行一系列的全基因组关联研究。该儿童医院的研究者正在与选择使用昂飞基因定型平台的研究者拓展合作,确定与一些复杂疾病相关的基因,如心血管、代谢和中枢神经系统失调等疾病。这些研究的结果有助于其他科学家开发更为有效的个性化治疗方案。费城儿童医院的研究小组在早期研究中将对7,000个个体的DNA样本进行全基因组关联研究。

## 美国国家环境卫生科学研究院 (NIEHS), Perlegen公司公布了小鼠Haplotype图谱

美国国家环境卫生科学研究院 (NIEHS) 和Perlegen公司公布了来自于小鼠基因组重测序和基因组分析研究的结果,利用这些研究结果可以将基因和与环境相关的疾病联系起来。在此项研究中,Perlegen公司的科学家采用了昂飞芯片对四种野生型和11种“经典”的小鼠品系进行重测序,在这15种品系中发现了827万个SNP。

## 美国国家人类基因组研究协会授予了昂飞实验室ENCODE研究经费

10月9日,美国国家人类基因组研究协会 (NHGRI) 授予了昂飞实验室共计1020万美元的研究经费,用于今后四年的研究,并将此作为ENCyclopedia Of DNA Elements (ENCODE) 计划拓展的一部分。该项批准是一个由16个国家的研究中心参与的,总经费超过8000万美元的计划的一部分。ENCODE计划的目的是识别人类基因组中各个不同的功能区域,并开发与本项目相关的新型高通量技术。

## 小鼠和大鼠基因1.0 ST芯片 ( Mouse and Rat Gene 1.0 ST Arrays )

9月10日,美国昂飞公司宣布GeneChip®小鼠和大鼠基因1.0 ST芯片 ( Mouse and Rat Gene 1.0 ST Arrays) 进入市场,它是3'端小鼠、大鼠表达谱芯片的最新补充。该芯片为研究者提供了一个更加精确和更大信息量的选择,可用于小鼠和大鼠基因表达检测,其成本低且易于使用。关于Gene1.0 ST芯片系统的介绍详见昂飞通讯1期。

## 水稻tiling芯片和犬科全基因组SNP芯片

• 水稻tiling芯片进入中国市场,一组共三张芯片,每张芯片包括大约六百五十万探针。该芯片是昂飞公司根据日本科学家需求而定制的。昂飞可提供各种各样的定制内容选择,包括筒式和片状结构。通过GeneChip®定制生物芯片项目,昂飞定制生物芯片设计组将与您紧密合作,共同确定最有效的设计方案。虽然生物芯片将按照您的特定规格进行设计,但是CustomExpress®芯片仍然采用GeneChip®芯片产品的严格探针选择方法进行设计。

• Broad研究院的科学家在Nature Genetics9月30日期上发表了关于用昂飞定制的犬科全基因组SNP芯片 (与500K、5.0/6.0的生产线的设计原理相同) 来绘制疾病基因图谱的研究结果。

## 专题综述

本期介绍遗传关联性研究的实验设计: **传统的两步研究法和新的联合分析模式**

遗传关联性研究成功的可能性决定于遗传力 (Genetics Power)。遗传力的大小受多方面因素影响。包括总样本数量,效应度 (Effect size /Odds Ratio), 等位基因频率 (Allele frequency), 连锁不平衡的程度 (R<sup>2</sup>), 种群结构和选择偏见等。理想的遗传关联性研究需要检测含100万个以上SNP的全基因组系列来得到关联性。由于潜在的遗传关联性通常是微弱的,因此还需要筛选尽可能多的样本,大多数情况下,遗传力直接与样本数量成比例

(图1)。在所有可用样本中最大限度地筛选全部SNP，这一理想的设计费用是相当昂贵的。长期以来，在实际研究过程中，研究者通常采用传统的两步法进行关联研究(图2)。

第一步，在样本子集中筛选全部的SNP系列，确定出符合有统计学重要性意义阈值的SNP子集。第二步，在更多的样本中筛选显示有潜在相关性的SNP子集。第二步的目的在于重复验证第一步的结果。为了更加准确地确定与疾病相关的等位基因的作用，保证有充足的遗传力确证第一步的结果，第二步通常需要更多的样本量。

这种分两步的方法有三个主要的缺点。第一，由于总样本集被分成了样本子集，使得初筛通量不够，因此遗漏真实关联性的可能性就提高了。第二，因为所有平台所固有的技术原因，使得有些SNP未能转入第二步的样本子集的新系列，从而引起真实关联性的丢失。第三，第一步和第二步之间定制系列会耽误6个星期到6个月的时间。尽管有以上缺点，两步法一直以来都是进行关联研究的明智选择。它在研究经费有限的情况下最大限度地发挥了遗传力。按单个样本算，进行全部SNP系列筛选的成本远远高于定制子集系列的成本。

直到最近，合适的芯片价格使得扩大全基因组系列分析的样本量成为可能，从而不需要像以前那样分成两步进行，全基因组关联研究实现一步到位的方法，降低了低通量的第一步中的假阴性风险，同时，还加快了研究的进程。

Skol等人(2006)在Nature Genetics 的研究表明这种整合的一步法比两步法的单独分析和联合分析的遗传力都要更为强大，即使是还要考虑到进行更为严格的多假设检测校正也依然如此。凭借昂飞技术提高了对更多样本进行检测的能力，遗传学研究领域已经开始采用这种更为强大的一步法方法。在今年七月Science发表的几篇关于确定与II型糖尿病相关变异的文章中，研究者利用昂飞芯片，在两步整合中使用相同的全基因组标记SNP集合，将全基因组筛选(第一步)和统计学重复(第二步)合二为一。

如今，传统的研究设计模式正在被基于全基因组标记集合如SNP6.0芯片这样的更大样本数据集合的联合分析所取代。SNP6.0芯片上的高密度标记还有助于研究者直接确定相关基因。该方法将大大加快从全基因组筛选到基因发现的进程。

## 技术介绍

本期技术介绍是DNA芯片设计。首先介绍昂飞的人类基因组图谱SNP芯片。如图3所示，每个SNP有六个探针四件套检测，每个探针四件套由每组等位基因的一个完全匹配和一个错配组成。每个SNP总共有24个不同的探针。SNP 5.0芯片和SNP 6.0芯片只取最佳SNP位点和每组等位基因的一个完全匹配，并有四次(SNP5.0)或三次(SNP6.0)重复检测，所以总共只有八个或六个探针。

人类基因组图谱芯片的所有SNP都经过严格的筛选和验证流程。被选择并铺载在芯片上的SNP是以准确度、callrate、连锁不平衡分析或基因组物理分布状况为基础的。500K芯片组有两张芯片组成，每张芯片平均能对250,000个SNPs进行基因分型。一张芯片使用NspI限制性内切酶(~262,000SNPs)，另一张芯片使用StyI(~238,000SNPs)。SNP5.0芯片是一种替代500K芯片组进行全基因组关联研究的单张芯片基因分型产品。除了500,568个SNPs，SNP5.0还含有420,000个用于检测拷贝数变异的探针。

最新的SNP6.0芯片含有超过906,600个SNPs以及超过946,000个用于检测拷贝数变异的探针。大约有482,000个SNPs来自于前代产品500K和SNP 5.0芯片。剩下的424,000个SNP包括了来源于国际HapMap计划中的标签SNPs，X，

图1. 说明增大样本量影响研究遗传力的例子

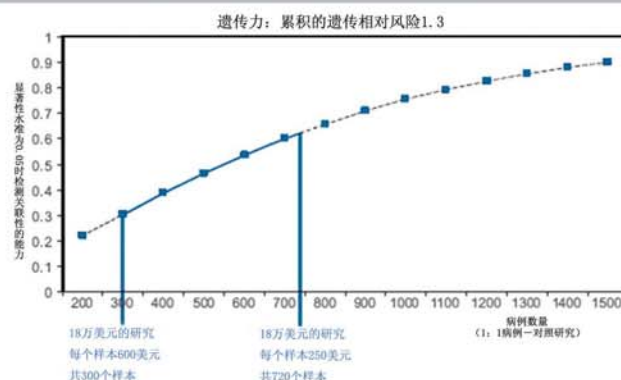


图2. 关联研究设计的新模型  
用于重复验证的全基因组方法

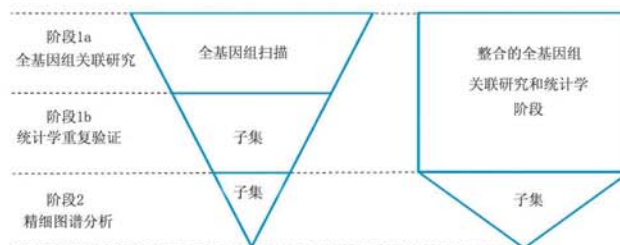
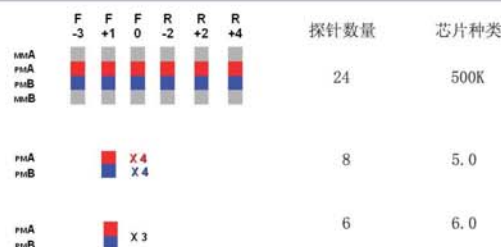


图3. 人类全基因组SNP芯片的Tiling策略—SNP探针设计的进化



Y染色体和线粒体上更具代表性的SNPs，以及来自于重组热点区域和500K芯片设计完成后新加入dbSNP数据库的SNPs。SNP6.0还包括了202,000个用于检测5,677个已知拷贝数变异区域的探针，这些区域来源于多伦多基因组变异体数据库。该数据库中的每3,182个非重叠片断区域分别平均用61个探针来检测。除了检测这些已知的拷贝数多态区域，还有超过744,000个探针平均分配到整个基因组上，用来发现未知的拷贝数变异区域。

第二类DNA芯片，定向基因分型芯片。通用标记物芯片上的原位合成的探针和每个分子倒置探针(MIP: Molecular Inversion Probe, 图4和图8示其基本原理)所含有的通用标记序列是互补的。所以芯片上的每个探针所具有的特征互补地代表了一段与感兴趣的独特SNP相应的通用标记序列。

第三类DNA芯片，CustomSeq®重测序芯片。使用扫描法(tiling, 每个碱基位用八个独特的25-mer探针)进行构建(如图5所示)，每个25-mer探针在中心位碱基互不相同——A、G、C或T——以便同时探测已知的和新的SNP。CustomSeq®重测序芯片能在一块芯片上分析多至30万双链序列碱基(总共600,000碱基)，提供在一次实验中大量测序的最有效和最经济的方法。

图4. 分子倒置探针 (MIP: Molecular Inversion Probe) 的基本原理

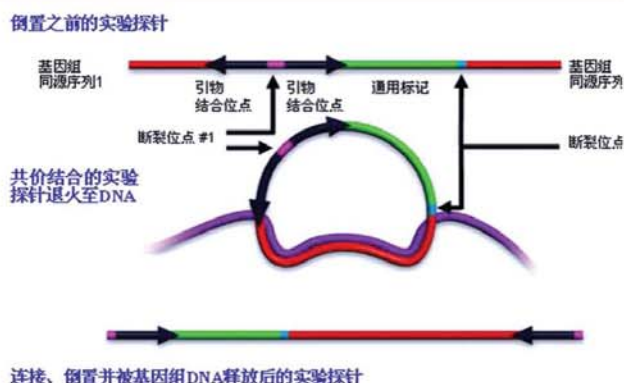
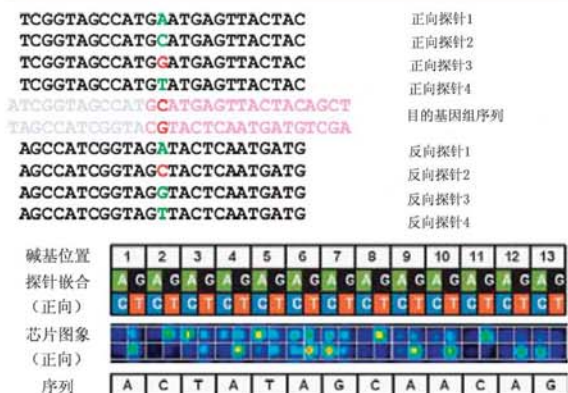


图5. CustomSeq®重测序芯片的基本原理



## 新技术与新启示

### 人类遗传变异，拷贝数变异 (CNVs) 和疾病研究及检测

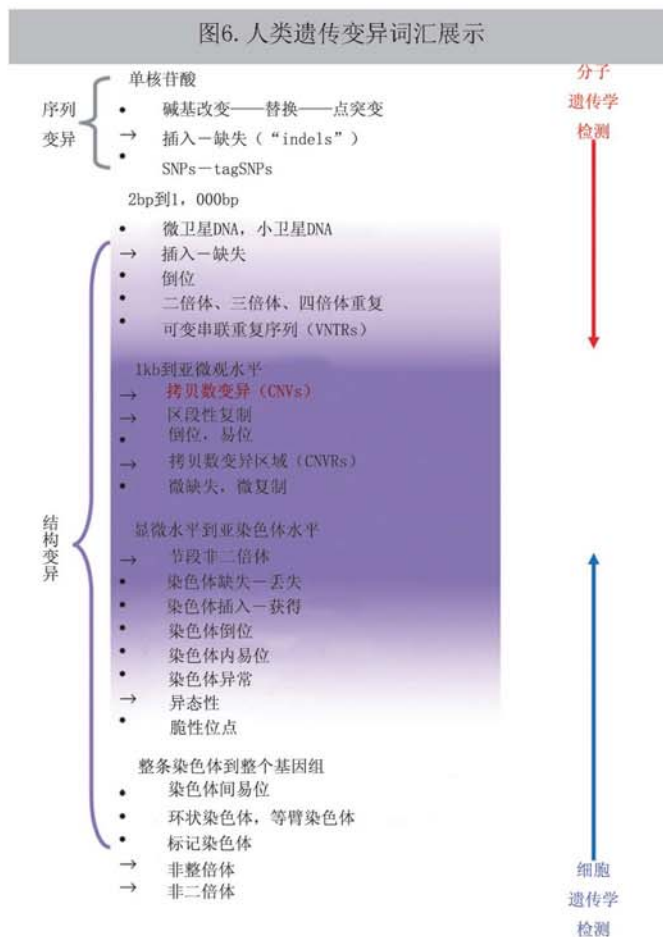
在人类细胞遗传学研究的早期，人们在显微镜下研究染色体，发现了染色体的拷贝数、重排和结构方面存在变异，而且在很多情况下这些变异可能与疾病相关。在分辨率频谱的另一端即高分辨率区域，DNA短片段的分析和测序方法的发展导致了短串联重复序列和单核苷酸多态性 (SNPs) 的发现。显而易见，人类遗传变异范围包括从序列水平上单一碱基对的变化到用显微镜检测到的几兆碱基长度的染色体差异。最近，通过观测亚微观DNA片段中广泛分布的拷贝数变异，我们对于人类遗传变异的认识又进一步得到了拓展。全基因组扫描方法的实行大大推动了这种关于人类变异的新认识，这些方法给我们提供了一个在显微镜细胞遗传学 (>5-10Mb) 和DNA序列分析 (1-700bp) 之间的对基因组中间范围遗传变异进行解读的强有力工具。正如图6所示的结构变异中的中等分辨率范围内的测亚微观部分。

现在已经发展了很多方法来检测这类中等大小范围内的DNA遗传变异，DNA生物芯片技术可能是其中最为有效的方法。拷贝数变异 (CNV) 鉴定的主要方法是比较基因组杂交 (CGH)，而商业化的标准CGH芯片在人类基因组的每1Mb长度范围有一个细菌人工染色体 (BAC) 克隆，这样就很难精确鉴定小于50kb的单拷贝数差异。昂飞的人类基因组图谱SNP芯片500K和SNP 5.0芯片的标记间距离中位数为2.5kb，最近推出的SNP 6.0的中位数则少于700个碱基对。这类基因型芯片通过将测试样本所获取的信息强度与其他个体的进行比较来确定每个位点相对基因组拷贝数。同时，拷贝数检测运算法中将探针的长度和GC含量考虑到其中，从而进一步降低了基因型芯片检测噪音。另外一个优点是，基因型芯片对拷贝数变异区域进行全面检测，并通过在连续的几个探针中要有重大的比率变化来确认。所以说，这样的工具明显提高了检测的精确度。除了提供拷贝数信息，SNP基因型芯片提供的基因型信息不但可以用于遗传关联性研究，还可以用于检测杂合性丢失，这为缺失的存在提供支持证据，还可能揭示片段性单亲二体。

近年通过拷贝数变异 (CNVs) 的研究，我们知道人类群体中的任何两个个体基因组结构上的差异比核苷酸序列水平上的差异更大 (请参阅应用案例2)。保守的估计显示个体之间CNVs总计有4Mb (相当于每800bp就有1个不同)，不保守估计则认为有多达5-24Mb范围内存在CNVs。无论是哪种估计，平均来说CNVs中的核苷酸变异数量比SNPs还要多，后者总数大约是2.5Mb (相当于每1,200个bp中有1个SNP)。因此人类个体之间的所有基因组差异性要远远大于先前所认为的，至少存在0.2%的差异：结构水平上有0.12%以上的差异，核苷酸水平上有0.08%的差异。

昂飞芯片技术革新不但为之前未被发现的人类健康人群中存在的基因组变异的基础研究敞开了大门，也为研究疾病的遗传基础打开了一扇新的窗户。致癌基因的扩增和/或肿瘤抑制基因的缺失是癌症起始和发展的特点，这一特点近来被认为可用来暗示癌症对治疗剂的反应。因此在细胞系和肿瘤样品中对这些改变进行定位是癌症研究的重要目标。近几年来，科学家应用了昂飞芯片技术对基因组总拷贝数改变和杂合性丢失进行评估，发表了众多的有关文章，并探讨其研究结果在癌症鉴定和分类中的应用。另外，针对目前最广泛采用的可贮存临床组织的形式之一——福尔马林固定、石蜡包埋（FFPE）样品，昂飞最近提供了总结推荐的质量控制（QC）步骤和实验方案，从而使500K芯片可以分析来源于FFPE的合格DNA样品，全面了解基因型、杂合性丢失以及拷贝数。

除了癌症研究以外，关于CNVs在遗传疾病中的作用，之前我们所知的大量信息都来源于关于“基因组疾病”的文献资料。基因组疾病是一大类由DNA拷贝数变化引起的遗传疾病。这些突变可以是相当大的，能在显微镜下观察，即可见的失衡区域；也可以是非常小的，需要更高分辨率的手段才能检测。而以前我们由可利用的方法所了解到的关于基因组疾病的信息主要是局限在：那些有明显临床症状的疾病，或是通过细胞遗传学方法可以检测到基因组不平衡性的疾病，以及性状以显性方式遗传的疾病。应用高分辨率全基因组扫描的昂飞芯片技术来研究遗传疾病可以大大提高检测与疾病相关的CNVs的能力。此外，越来越多的证据证明这些遗传学发现为医生提供了非常有用的与疾病临床特点有关的信息。



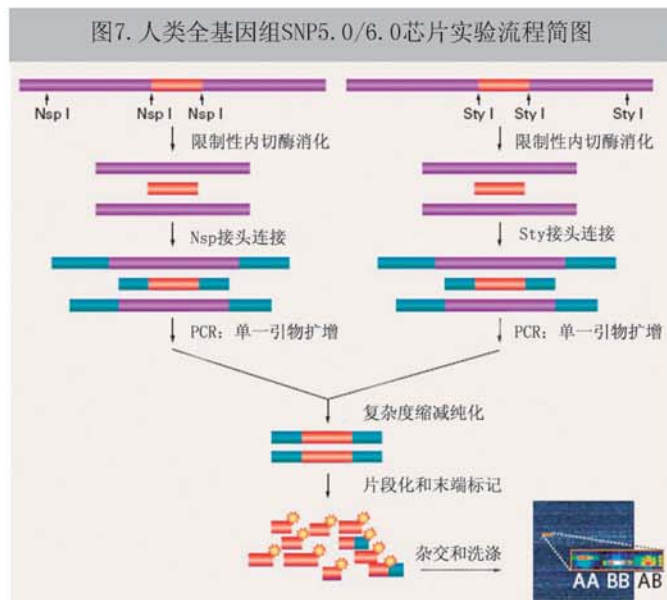
上图及此文中的大量信息来源于 Nature Genetics Supplement Vol 39, July 2007

## 研究指南

本期研究指南是DNA芯片实验和数据分析流程方面的内容。首先，昂飞500K, SNP 5.0和SNP 6.0人类基因组图谱芯片的策略分析采用了减少基因组复杂性的可靠而简单的方式。具体如图7所示，总基因组DNA（每个酶用250ng）由限制性内切酶内切（NspI和StyI）并连接到可以识别粘性4bp末端的接头上。所有的片段均由限制性内切酶消化而来，无论其大小都是接头连接的底物。然后用一个可识别接头序列的通用引物来扩增与接头连接的DNA片段。将PCR条件设定在扩增某一长度的最佳状态（200~1100bp）。这些扩增的DNA之后被合在一起并打碎，标记并杂交到芯片（500K操作中两种酶扩增的DNA不合并，后续分别杂交到NspI和StyI两张芯片）。关于详细信息，请参考昂飞公司操作手册：Affymetrix Genome-Wide Human SNP Nsp/Sty 5.0 Assay 5.0 or 6.0 Manual 和 GeneChip® Mapping 500K Assay Manual。

图8则简单示意了定向基因分型芯片的系统概览，在此系统中，定制试剂盒可用于3,000到50,000的SNP配置。通量为每天可以对48个样本（包括对照）中最多50,000个SNPs进行基因分型，每周就是192个样本。所以说，每天可做240万基因分型，而且在使用额外设备时通量可以更大。关于详细信息，请参考昂飞公司操作手册：Affymetrix®GeneChip® Scanner3000 Targeted Genotyping System User Guide

CustomSeq®重测序具体实验如图9所示。就芯片通量来说，通过长范围（Long-range）-PCR，所需的扩增反应大大减少，也减少了PCR所需的资金和时间。芯片



在48小时内就能以最小的线性排列、注解和手工编辑的方式测出整个序列。这样，可以将时间和资金投入研究者的实验中，而不是相应的试剂。详细信息请参考操作手册：GeneChip® CustomSeq® Resequencing Array Protocol。

图8. 基于分子倒置探针实验的定向基因分型系统

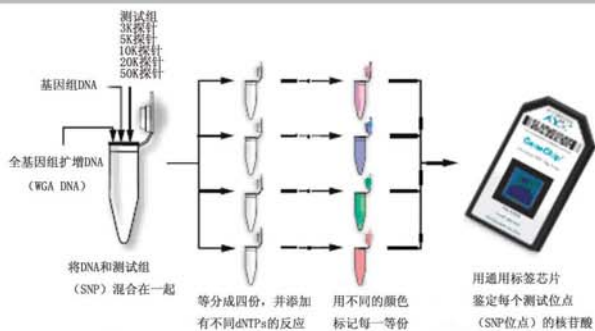
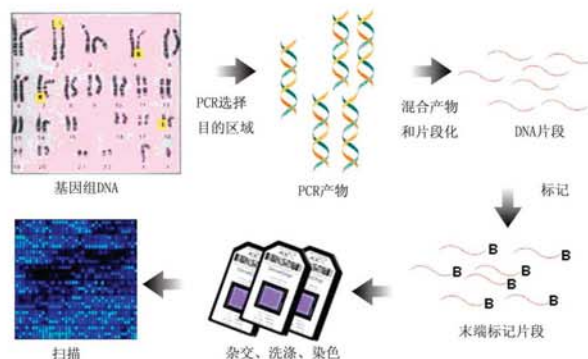


图9. 重测序实验流程图

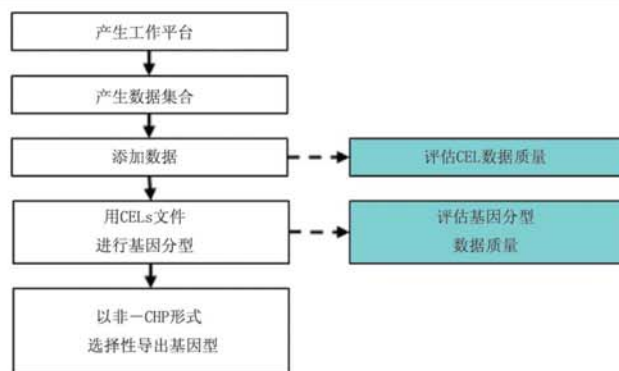


DNA芯片数据分析方面，昂飞基因分析软件（GTYPE）是为了给人类基因组图谱芯片提供高准确率、自动化的SNP allele calls而特别设计的。BRLMM Analysis Tool (BAT) 2.0是可用于500K 和SNP5.0芯片的包括BRLMM and BRLMM-P算法的软件工具。

新软件Genotyping Console则用来分析 SNP 5.0/6.0芯片数据。该软件整合了由BRLMM分型算法演化而来的新分型算法（Birdseed）。与SNP5.0芯片的BRLMM-P算法相似，Birdseed用多张芯片对每个SNP的等位基因估算出一个信号值，利用探针-特异性的效果来提高准确度。从而使得分型结果在两维A-信号对B-信号空间中符合Gaussian mixture model，并通过SNP-特异性模型来提高其精确性。

如图10所示，Genotyping Console的设计目的是提高分型以及分型质量控制的效率。除了算法特征外，还包括自动化QC，根据QC结果对样品进行分类，对所有样品的各种QC指标图形显示以及SNP群的可视化和跟踪指定的SNP列表。通过对72个指定SNP分型结果与不同技术或者其它已知的参考信息进行比较，可以对样品进行验证。同时可变的SNP筛选和输出使得下游分析能够进行。除了Birdseed算法，Genotyping Console也支持SNP 5.0芯片的BRLMM-P算法。

图10. Genotyping Console的基本流程



昂飞的Copy Number Analysis Tool (CNAT) 支持500K芯片数据拷贝数和杂合性缺失(LOH)分析。整合基因组浏览器(IGB)能实现基因中染色体变化可视化，并方便生物标记的辨认。另外，可免费下载拷贝数分析软件：哈佛大学Cheng Li博士的dChip SNP (<http://biosun1.harvard.edu/complab/dchip/>) 和东京大学SeishiOgawa博士的CNAG (<http://www.genome.umin.jp/>)。用于支持SNP5.0/6.0芯片拷贝数数据分析的软件包括在中国有销售商业软件Partek® Genomics Suite™和昂飞即将开发出来的相应分析软件([www.affymetrix.com/products/software/compatible/snp.affx](http://www.affymetrix.com/products/software/compatible/snp.affx))。

在定向基因分型芯片方面，TG软件包（GTGS）功能全面，使研究者能够基于一组质量控制（QC）度量体系，建立和跟踪实验的进程、分析原始数据并进行基因分型。具体包括：输入样本相关信息作为初始项目建立的一部分；条形码读取器可以跟踪反应板，并促进样本信息与获取的基因型鉴定的集成；聚类算法能够从原始生物芯片中进行基因型鉴定；如有必要，研究员可通过用户可编程的QC度量体系为给定的研究定制其鉴定参数；数据文件可以标记界定格式，轻松地输出给第三方分析工具或研究者自己的分析软件。

重测序芯片的序列分析软件(GSEQ)整合了先进的算法——重测序算法2.0版，对测得的杂交强度数据进行自动测序和基因分型。所有以上提到的昂飞分析软件可以在([www.affymetrix.com/products/software/](http://www.affymetrix.com/products/software/))查询。

## 应用案例

全基因组关联（GWA）研究是一种鉴定与常见人类疾病相关基因的有效方法。WTCCC采用了500K芯片在英国人群中进行联合GWA研究，检测了七种主要疾病（每种疾病2,000个病人）和3,000个共享正常对照个体。病例-对照对

比后鉴定了24个独立的关联性信号( $P < 5 \times 10^{-7}$ ): 双极抑郁症1个, 冠心病1个, 克罗恩病9个, 风湿性关节炎3个, I型糖尿病7个, II型糖尿病3个。基于先前的研究发现和目前已经完成的重复性研究, 显示几乎所有的这些信号都反映了真实的易感性效应。在大多数已鉴定的位点处观察到的适度效应大小说明拥有一批适当规模的样品是非常重要的。该研究是至今完成的最大的GWA研究和对GWA方法的一种全面验证。

The Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007 July 7; 447 661-683

Redon等用两种不同的芯片平台(昂飞500K芯片和Whole Genome TilePath)分析了270个正常个体内的拷贝数变异, 研究发现CNV在人类多样化和进化中非常重要。CNV区域比SNP包含更多的核苷酸内容。在HapMap人群中已识别1447个CNV区域, 这些区域覆盖了12%的人类基因组(360MB)。CNV在人类生物学和疾病中非常重要。CNV区域包含成百上千的基因、功能要素和分段重复序列。CNV在遗传研究中非常有价值, 可作为SNP的补充信息。HapMap人群的许多CNV中存在大量的CNV型变异和明显的连锁不平衡。

Redon, R., et al. Global variation in copy number in the human genome. *Nature*. 2006 Nov 23; 444: 444-454.

Smyth等采用了全基因组非同义SNP(nsSNP)扫描方法研究了剩下的I型糖尿病(T1D)易感性位点, 该方法使用了高度多元化、高通量的分子倒置探针(MIP)技术。研究以有说服力的统计学支持, 表明在染色体2q24.3先天免疫病毒RNA受体基因区域IFIH1(又被称为mda-5或Helicard)存在第六个I型糖尿病(T1D)相关位点。在对全基因组非同义SNP(nsSNP)扫描的期中分析中, 研究发现了这一相关性, 并在病例-对照群体中对其进行了确证, 最后还在一个独立的家系群体中进行了重复性实验。对4,253个病例, 5,842个对照和2,134个父母-孩子的三口之家的基因型进行检测发现, nsSNP rs1990760 A-G(A946T)的次要等位基因的风险率是0.86(95%置信区间=0.82-0.90),  $P = 1.42 \times 10^{-10}$ 。

Smyth, D. J., et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nature Genetics* 38(60) : 617-619 (2006).

线粒体的体细胞性突变在人类的癌症中较为常见, 所以可以将它用作检测早期癌症的工具。Maitra等开发了一种线粒体客户定制的Reseq™芯片, 将其作为基于芯片的测序平台, 快速且高通量地分析线粒体DNA。该研究对多个个体总计超过200万碱基对的线粒体DNA进行测序, 成功地读取了96.0%的核苷位置, 重复实验显示了>99.99%的重现性。在5个膀胱癌患者和4个胰腺癌患者的配对流体取样样品中(分别用尿液和胰液), MitoChip在9个样本的6个中(66%)检测到了至少一个癌症相关的线粒体突变。MitoChip是一款高通量的测序工具, 可以从一些临床样本的主要肿瘤中可靠地鉴定线粒体DNA突变。

Maitra, A. et al. The Human MitoChip: A High-Throughput Sequencing Microarray for Mitochondrial Mutation Detection *Genome Research*, 14:812 - 819 .2004.

## 本期相关技术要点

### 昂飞的人类基因组图谱SNP芯片

- 500K是一高分辨率产品, 可用于癌症(包括FFPE样品)和细胞遗传学中拷贝数的检测。不仅提供拷贝数信息, 还提供基因型信息(杂合性丢失, LOH)。

- SNP6.0是一高性能产品, 为用于大规模关联性研究的在固定经费基础上提供最高的遗传力(genetic power)。芯片全面发挥了光刻技术的能力, 可以在一张芯片上同时检测180万个分子标记。犹如作为全基因组表达谱研究标准的昂飞表达谱芯片。SNP6.0芯片将有望作为全基因组关联研究的标准, 使得研究人员用一个标准化的产品进行数据交流和整合。

- SNP5.0是介于500K芯片和6.0芯片之间的一种中间产品, 可以为那些经费有限的研究人员提供一种极有吸引力的选择。

表1. 各种昂飞人类基因组图谱SNP芯片的参数

	500K	SNP 5.0	SNP 6.0
SNPs	500,000	500,000	900,000
CNVs		420,000	900,000
芯片数	2	1	1
探针/SNP	24	8	6
算法	DM/BRLMM	BRLMM-P	Birdseed
重复性	>99.8%	>99.8%	>99.8%
Concordance	>99%	>99%	>99%
探针种类	PM/MM	PM	PM
Feature Size	5 micron	5 micron	5 micron

表2. 各种基因组图谱SNP芯片的分析软件

SNP 6.0	Genotyping Console	Partek® Genomics
SNP 5.0	Genotyping Console	Partek® Genomics
500K 2 chip sets	BAT 2.0 + SNP Signal Tool GTYPE	Partek® Genomics

### 定向基因分型和重测序

定向基因分型通过为每个芯片提供3,000到50,000个SNP灵活多样的配置，行业领先的将数据库SNP转化为功能芯片的高转换率及其SNP内容。不论SNP是常见的还是稀有的，抑或是不同义的，昂飞CustomSeq®重测序项目让研究者用自行设计定制芯片来快速鉴别一个特定区域的所有SNP。另外，CustomSeq®重测序使研究者能够比以往更有效、更经济地进行大规模单倍体生物和微生物基因组重测序。

从种群基因到致病基因再到药物基因组学，GeneChip®DNA分析产品提供您所期望的通量、精确度和可靠性。DNA分析产品以标准和定制的生物芯片、检测方法以及软件，将完整的基因工作流程用到了从全基因组关联研究到定向基因分型和重测序的各个方面。同时，为了进一步将生物功能与遗传发现联系起来，研究者可迅速结合GeneChip®系统上的表达分析与已获得的基因分型和重测序的数据。

### 美国昂飞公司上海代表处

中国上海陆家嘴环路1233号汇亚大厦1508室 200120  
 电话: 86-21-68881980 传真: 86-21-68881981  
 saleschina@affymetrix.com supportchina@affymetrix.com  
 www.affymetrix.com

Let Professionals Serve Professionals

香港: 852-28966283	北京: 010-51665161	上海: 021-64951899
广州: 020-85524840	天津: 022-88293136	成都: 028-85431195
武汉: 027-87166462	昆明: 0871-5349992	沈阳: 024-23341315
济南: 0531-86560825	南京: 025-83248692	长沙: 0731-4476562
西安: 029-82501170	杭州: 0571-87229824	重庆: 023-68614842



**Gene Company Limited**  
 基因有限公司

A Gene Group Company