

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，人为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 片段中广泛颁布的拷贝数变异，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为变异的新认识，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，还可能提

示片段性单亲二体。

近年来通过拷贝数变异 (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,即可见的不平衡区域;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。