中国科学 C辑:生命科学 2009 年 第 39 卷 第 3 期: 237 ~ 244
www.scichina.com life.scichina.com
《中国科学》杂志社
SCIENCE IN CHINA PRESS

人类基因组结构变异检测研究进展

武雪梅[①②], 肖华胜[①③*]

① 中国科学院上海生命科学研究院, 系统生物学重点实验室功能基因组中心, 上海 200031;
② 中国科学院研究生院, 上海 200031;
③ 生物芯片上海国家工程研究中心, 上海 201203
* 联系人, E-mail: huasheng_xiao@shbiochip.com

收稿日期: 2008-12-26; 接受日期: 2009-01-21
国家高技术研究发展计划(批准号: 2006AA020704)资助项目

摘要 高通量、高分辨率基因组学技术的出现推动了人类基因组中长度在 1 kb~3 Mb 的亚显微水平结构变异检测方法的发展, 这些结构变异主要包括基因拷贝数变异、倒置、插入、缺失、重复及其他基因重排. 而传统的细胞遗传学技术达不到如此高的分辨率. 本文介绍了目前主要的基因组结构变异的检测技术, 包括基于芯片的比较基因组杂交技术和代表性寡核苷酸芯片分析技术, 基于 PCR 的多重扩增探针杂交技术和依赖于连接反应的多重探针扩增技术, 配对末端图谱技术等. 还比较和分析了各种方法的优劣势并提出了目前结构变异数据库存在的问题. 最后讨论了这些变异对于人类表型多态性、疾病易感性、药物反应程度及群体遗传学的影响.

关键词
结构变异
细胞遗传学
基因芯片
PCR
下一代 DNA 测序
配对末端图谱

人类遗传变异被*Science*杂志评为2007年世界十大科技突破之一. 近年来的研究表明人类基因组存在大量变异, 研究这些变异不仅有助于揭示许多复杂疾病和个体性状的遗传学机制, 也加快了个性化用药的步伐.

根据发生突变的碱基数目, 遗传变异可分为单核苷酸多态性(single nucleotide polymorphisms, SNPs)和结构变异(structural variations, SVs). 单核苷酸多态性主要是指在基因组水平上由单个核苷酸的变异所引起的DNA序列多态性; 而结构变异则指 1 kb以上的DNA碱基的改变, 进一步分为大于 3 Mb的显微水平结构变异和大小在 1 kb~3 Mb之间的亚显微水平结构变异[1]. 最初的人类基因组分析揭示出SNP是导致遗传和表型多样性的主要原因[2], 在人类基因组中广泛存在, 总数可达1000~1500万个[3]. 截止到2007年, 二期人类基因组遗传整合图谱HapMap中已发现了310 万个SNP位点[4]. 与之相关的全基因组疾病关联研究也发展迅速, 在过去的几年里, 已经发现了与多发性疾病如老年痴呆症、帕金森症、阿尔茨海默氏症、糖尿病和心血管疾病相关的许多SNP位点[5,6].

SNP及低复杂性串联重复序列(tandem repeats)(一般小于 1 kb), 如微卫星变异(microsatellites)和小卫星变异(minisatellites)一直被视为人类遗传变异最主要的形式, 但这一概念在 2004 年之后发生了极大改变. Iafrate[7]和Sebat等人[8], 分别用BAC 芯片及ROMA技术证实了从数千碱基到数百万碱基长度的基因拷贝数变异广泛存在于健康个体中. 最广义的基因结构变异涉及所有非单碱基的基因突变[9], 包括DNA序列的插入、缺失、倒置、重复、转置及拷贝数变异(copy number variation, CNV)[10,11]. 此外, 根据基因数量改变与否, 结构变异还可分为平衡重排(balanced rearrangements)和不平衡重排(unbalanced

rearrangements)[12]. 目前, 某些变异形式在人类基因组中的粗略图谱已经公布, 如CNVs[13]和INDELs (insertion and deletion)[14]. 本研究着重对 1 kb~3 Mb 的亚显微水平结构变异的检测方法及其对人类表型多态性、疾病易感性、药物反应程度及群体遗传学方面的影响进行综述.

1 亚显微水平结构变异的检测方法

1.1 细胞遗传学方法

显微水平的结构变异长度在 3 Mb以上, 光学显微镜下即可观察到. 在高通量DNA测序技术出现以前, 研究者主要采用传统的细胞遗传学方法尤其是高分辨率的染色体带型技术在全基因组中寻找变异[15]. 通过这些方法, 染色体异型、转置、缺失、重复、插入和倒置均可检测到.

1969 年, 原位杂交技术首次被报道[16], 随之发展起来的荧光原位杂交技术(fluorescent *in situ* hybridization, FISH)[17]及DNA纤维荧光原位杂交技术(stretched-fiber FISH)[18]第一次使特定序列的变异情况得到高分辨率的分析, 但该技术费力费时, 不适用于全基因组变异扫描. 目前对靶序列的分辨率已经从最初的中期染色体水平(~5 Mb)或间期细胞核水平(50 kb~2 Mb)提高到纤维状染色质水平(5~500 kb)[19]. Raap等人[20]报道将FISH技术应用于解螺旋的裸露DNA纤维, 分辨率可达 1~400 kb. 这些都表明分子水平的细胞遗传学技术的发展显著提高了遗传变异分析的分辨率.

1.2 高分辨率分析方法

过去 10 年中, 分子生物学尤其是 DNA 测序技术的迅速发展使人类基因组结构变异的检测分辨率得到很大提高. 这些方法主要包括高通量分析和靶向性分析两种. 高通量分析是基于芯片技术和 DNA 测序技术, 靶向性分析是基于 PCR 技术.

(1) 基因芯片. 比较基因组杂交技术(CGH, comparative genomic hybridization)与基因芯片技术的结合使人们可以更快、更准确地检测基因的扩增或缺失. 在一张有成千上万特异DNA序列的芯片上, 用标记不同荧光素的测试样品和对照样品同时进行杂交, 从而快速直观地检测两样品之间基因拷贝数的差异[11], 这项技术就是基于芯片的比较基因组杂交技术(array-based CGH, array-CGH).

基因芯片上的探针可以是基因克隆(如BAC)、cDNAs、PCR产物及寡核苷酸. 其中, BAC和寡核苷酸芯片在全基因组变异扫描中的应用最为广泛. 由于BAC载体的插入片段长度一般在150 kb, 其分辨率只能达到 50 kb, 在大片段DNA变异的研究中应用广泛[21]. 在 2004 年启动的国际项目"拷贝数变异计划"中, BAC芯片技术用来全面鉴定 269 个样本中的基因拷贝数变异, 最终找到了 1447 个拷贝数变异区域(copy-number variant regions, CNRs), 占人类DNA序列的 12%[13]. 此外, SNP分型芯片也可用于拷贝数变异分析. 例如, Affymetrix人类基因组SNP 6.0 芯片拥有超过 180 万个遗传变异标志物, 包括超过 90 万个SNP和超过 94 万个用于检测拷贝数变化的探针[22]. Illumina Human1M芯片上也有 107 万个用于检测 1.4 万个拷贝数变异区域中的拷贝数变化情况的标志物[23].

其他类型的基因芯片也可用于高分辨率分析遗传结构变异, 如在单个外显子水平检测CNV的外显子芯片比较基因组杂交技术(exon array CGH)[24]和长度在 60~100 bp的寡核苷酸芯片技术[25]. 其中, 代表性寡核苷酸芯片分析技术(representational oligonucleotide microarray analysis, ROMA)在全基因组中的分辨率可达 30 kb[26]. ROMA技术中, 基因组DNA经限制性内切酶(如六碱基序列识别酶*Bgl* Ⅱ)切割后, 通过黏性末端与特定的接头序列相连, 然后用通用引物进行PCR扩增, 检测样品与对照样品的PCR产物经不同荧光素标记后与寡核苷酸芯片杂交(芯片上的探针是根据预测得到的限制性酶切片段设计). 对芯片扫描数据进行分析后得到的拷贝数变异图谱使研究者可以高精度地检测人类全基因组中的基因扩增和缺失[27].

基于基因芯片的各种方法可在全基因组水平扫描遗传变异, 这是其优点所在, 但也存在如下缺点: 不能检测拷贝数未发生变化的变异形式, 即平衡基因重排如倒置和平衡转置(balanced translocation); 不能准确描述基因重排发生的断裂点(breakpoints)和精细结构重排等. Array painting 技术克服了这一缺点, 对异常染色体显微切割后得到的片段进行扩增, 然

后与点有基因克隆的芯片杂交, 可分析转置等未涉及拷贝数变化的变异形式[28].

(2) 多重PCR. 为了精确验证变异区域, 基于定量PCR的新方法得到发展. 这些方法在一次实验中可同时检测 50~100 个区域的特异性短片段的突变[29]. 多重扩增探针杂交技术(multiplex amplifiable probe hybridization, MAPH)和依赖于连接的多重探针扩增技术(multiplex ligation-dependent probe amplification, MLPA)正是这样的技术.

MAPH 分析只需要小量的无需预处理或扩增的基因组DNA(0.5~1 μg). 将待分析的基因组DNA固定在膜上, 与连接有通用引物且长度各异的一系列探针杂交, 经洗液洗涤除去未结合探针, 然后将芯片上结合的特异性探针从膜上分离下来, 经通用引物扩增后得到的PCR产物进行电泳, 由于它们长度各异, 可很好的分离开来. 通过比较条带强度分析探针对应序列的拷贝数变化情况[30]. 尽管MAPH分析迅速且成本低廉, 但PCR多重水平及凝胶电泳分辨率的限制使MAPH一次只能分析 40 个序列[31], 所以低通量是它的最大劣势. 与芯片技术的结合成功解决了这个问题, 基于芯片技术的MAPH方法(microarray MAPH)将膜上分离下来的特异性结合探针与芯片上的寡核苷酸杂交, 从而大大提高了通量, 可同时分析更多区域的拷贝数变化[32]. Philippos等人[33]设计了特异性分析人类X染色体的MAPH array, 此芯片上包括 558 个探针.

MLPA 是一种灵敏度高、重复性好, 只需 20 ng DNA样品即可在单个外显子水平检测基因扩增或缺失的技术. 针对每个待检测靶基因需设计相互毗邻的两个探针, 且两个探针跨越的碱基长度各异[34]. 所有的探针对两侧都连接有通用引物, 与靶序列配对杂交后, 两个毗邻探针通过连接反应相连, 连接产物的量与完整靶基因的拷贝数成正比, 经PCR扩增后根据电泳结果分析基因的扩增与缺失, 其分析方法与MAPH类似. MLPA在肿瘤预诊断基因的剂量分析中应用广泛: 如遗传性非息肉病性大肠癌(HNPCC)相关的基因*hMLH1* 和 *hMSH2*; 软骨骨生成障碍相关的基因*SHOX*[35]. 该技术的优势是灵敏度高、特异性好、成本低廉.

2007 年, Isaksson[36]建立了一种依赖多重连接的基因组扩增技术(multiplex ligation dependent genome amplification, MLGA). 该技术使用两种探针, 针对酶切片段的特异性探针和用于酶切片段环化的探针. 基因组DNA经限制性酶切后环化, 环化产物的量与靶基因的拷贝数成正比. PCR扩增环化DNA后电泳分析拷贝数变化. 该技术的特点在于被扩增的是基因组DNA, 而非探针, 从一定程度上降低了背景噪音; 短长度的MLGA探针易于合成; 短时间内即可完成.

(3) 序列比对. DNA序列信息的公开化[37]和各种算法的优化为基因组结构变异的分析提供了另一途径. Tuzun等人[38]将从高密度fosmid 文库得到的 110 万个配对末端序列(paired-end sequences)与人类参考基因组(human genome reference assembly)进行比对, 在长度或方向上不一致的区域被确定为插入、缺失和倒置. 由于插入fosmid 质粒的片段长度限制在 40 kb以下, 该方法鉴定出的两个基因组间结构变异的分辨率局限在 8 kb. 由于该方法建立了fosmid文库, 所以可以对检测出来的变异进一步测序以更加精确地分析. 此外, 其优势还包括可以分析拷贝数不变的基因重排, 如倒置. 美国国家人类基因组研究所(NHGRI)于 2006 年提出 “人类基因组结构变异计划”(Human Genome Structural Variation Project), 该计划提出采用fosmid 配对末端测序方法对各种类型的变异进行鉴定、测序并最终实现基因型(genotyping)分析.

2007 年, Korbel等人[39]提出了一种新的大规模高通量的分析方法——配对末端图谱法(paired end mapping, PEM). 首先将基因组DNA剪切成长度约为 3 kb的片段, 片段两端与生物素标记的接头连接后环化, 对环化产物随机切割, 通过亲和素筛选带有生物素的剪切片段, 该片段包括了原来 3 kb片段的两个末端. 然后采用罗氏GS FLX 454 测序得到配对末端的序列信息, 此序列与人类参考基因组序列比对即可根据方向或长度的不一致找出存在的结构变异, 包括大于 3 kb的缺失、倒置、配对及非配对插入和长度在 2~3 kb的简单插入. 他们找到 1000 多个结构变异, 这表明实际上结构变异的数目要远多于起初的预测, 并且有些变异会影响基因功能. 2008 年, 罗氏 454 生

命科学公司进一步验证了利用高准确性、长读长的GS FLX系统可快速获得高质量的生物数据, 包括对未知基因组测序和基因组结构变异的检测[40].

事实上, 如果基因组序列完全解码, 序列比对是鉴定各种变异最直接、最简单的方法, 而且不受分辨率制约, 各种形式的变异都可检测到. Khaja等人[41]将人类Celera's R27c数据库中的序列与Build 35 reference序列进行比对, 找出了 13534 个非SNP的突变包括缺失、倒置等, 定量PCR, FISH及与其他数据库的比较进一步验证了其比对结果. 同样, 人类和黑猩猩基因组比对结果所揭示的某些种间结构变异在不同的人之间也存在[42]. 2007 年, Krzywinski 等人[43]提出了指纹图谱技术(fingerprint profiling, FPP): 用限制性内切酶作用于 493 个代表乳腺癌细胞系MCF7 基因组的BAC克隆, 得到的酶切片段与人类参考基因组(reference assembly)比对后揭示出包括 1~5 kb的微小缺失和平衡性基因重排在内的各种结构变异.

目前测序技术的发展日新月异, 自动化程度不断加强, 技术日渐成熟, 成本大大降低, 这使得个人基因组的解码成为可能. 第二代高通量测序技术包括罗氏 454 生命科学公司的 454 技术、 Illumina公司的Solexa 技术和ABI公司的Solid技术, 第三代测序技术包括Helicos BioSciences公司的纳米孔单分子测序技术. 与传统的双脱氧测序法相比, 这些技术具有高通量、自动化和低成本的优势, 加快了对人类和其他物种基因组序列的分析进程, 如伤寒沙门菌的基因变异和遗传进化规律在采用 454 技术和Solexa 技术实现全基因组测序后被揭示[44]. 现在已有一些公司如 23andMe推出了个人DNA测试服务, 所有这些进步无疑推动了包括结构变异在内的各种基因重排的发现.

2 结构变异的影响

无论其种类和大小如何, 基因变异都有可能与遗传疾病及个体间、人群间、物种间的表型差异相关, 并接受自然选择[45].

基因的结构变异可通过多种机制影响基因的表达. 非平衡基因重排如插入、缺失和串联重复等可导致基因剂量的改变, 从而影响携带者表型[46]. 非编码区的结构变异可通过位置效应(position effects)改变基因表达基本调控原件的位置或影响其调控作用, 从而改变基因的表达, 如遗传性疾病地中海贫血[47]; 也可通过改变局部染色质结构影响基因的表达[48]. 缺失突变还可通过另外一种机制导致表型改变: 两个等位基因中的显性基因发生缺失后, 隐形基因发挥显性抑制效应(dominant negative effect), 导致突变表型[49]. 此外, 有些突变虽然没有造成表型的改变, 但影响了个体的疾病易感性和药物反应程度[50]. 但由于技术的限制, 目前对结构变异的关联研究尚处于起步阶段.

2.1 结构变异与表型

果蝇X染色体上*Bar*基因的重复导致棒眼表型, 这是首次将大的遗传变异与表型相联系[51]. 到目前为止, 越来越多的人类疾病与基因变异之间的关系被揭示, 结构变异的检测已成为筛选疾病候选基因最迅速的方法之一[52].

大片段重复和缺失与表型的关系很早就有报道, 如剂量敏感型发育相关基因的拷贝数变化可导致遗传性疾病[53]. 某些最早报道的可遗传性状, 如色盲[54]和恒河因子(Rh因子)[55]后来被证实与基因重复和缺失相关. X染色体上神经胶质素(neuroligin)基因 *NLGN3* 和 *NLGN4* 的缺失与自闭症相关[56]; 70%的普拉德-威利症候群(Prader-Willi syndrome, PWS)病人来自父方的 15 号染色体 15q11~q13 区段发生缺失, 目前已有两种缺失亚型报道[57]. 人类基因组突变数据库显示, 与孟德尔遗传性疾病相关的突变中有 5%属于亚显微水平的插入或缺失[58]. 同时, 许多广泛存在于健康个体中的基因拷贝数变异是无害的, 因此区分致病性基因变异与无害的变异至关重要.

与其他类型的结构变异不同, 倒置未涉及基因的获得或缺失, 只是DNA序列在方向或位置上发生了改变. 针对平衡基因重排, 缺乏高通量低成本的检测方法, 因此目前对此类变异的认识相当有限. 但有些研究已经揭示了它们与疾病的关系. 例如, 40%的A型血友病人的凝血因子VIII基因存在 400 kb的缺失[59]; 自闭症病人10 号染色体10q21.3区段的臂内倒置破坏了基因*TRIP8* 和 *REEP3*, 这两个基因都是自闭症的候选相关基因[60]; 此外, 在某些情况下, 染色体上的DNA倒置未导致当代个体的表型变化, 但却

增加了下一代发生微小缺失或转置突变的几率从而导致疾病, 如威廉氏症候群、安格曼症候群、Sotos氏症和Wolf-Hirschhorn氏综合征[1]. 因此对普通健康人群倒置突变的检测具有重要意义.

2.2 结构变异与疾病易感性

结构变异除可导致表型改变之外, 还会影响个体对于疾病的易感性及对药物的反映程度[11]. 例如, 基因*UGT2B17* 存在缺失多态性, 个体对前列腺癌的易感程度和病人对睾丸激素的反应程度在种间和个体间的差异与之明显相关[61]. 基因*CCL3L1* 的拷贝数减少将导致个体更易感染HIV且更易于AIDS的发 展[62]. 细胞色素P450 基因如*CYP2D6* 参与药物代谢, 其拷贝数的改变会影响机体对三环抗抑郁药和安定药的代谢并且与喉癌和肺癌相关[63,64]. 谷胱甘肽S-转移酶基因*GSTT1* 和*GSTM1* 的纯合缺失可增加个体患多种癌症的几率[65].

2.3 结构变异与群体遗传学

结构变异在某些种类的基因, 尤其是参与分子间和环境间相互作用的基因中明显富集. 这类基因或参与机体对细菌感染和外界刺激的防御反应、或参与药物代谢、或调节细胞结构和生物合成, 或负责感官知觉, 都富含结构变异区[38,66]. 这表明, 基因的结构变异可能参与了人类对新环境压力的适应, 有助于分析不同人种的人口统计学历史和导致多样性产生的突变过程[11]. Jakobsson等人[67]正是通过分析世界上 29 个人群样本的基因型、单倍体型(haplotype)及拷贝数变异推测出人类种群结构, 这是遗传变异在群体遗传学上的又一应用. 此外, 通过位置效应(position effect)发挥作用的基因突变似乎在发育相关的基因中富集, 这一现象还未得到清楚的解释. 原因可能是发育相关基因中重复片段的大量存在增加了这类基因重排的几率; 也可能是其他基因对于基因突变产生的位置效应不敏感所致[29].

基因的结构变异对人类的影响不只局限于上述几点, 它们广泛存在, 影响普遍, 因此对于此类变异的研究不亚于 SNPs 研究的重要性.

3 展望

2004 年至今, 结构变异的信息海量增加. 不同技术检测到的突变中有 80%是不相互重叠的, 这预示着大量结构变异还未被揭示[68]. 根据人类基因变异数据库(the database of genomic variants, DGV)的最新公布, 目前已有 31615 个基因突变被提交, 其中包括 19792 个CNVs, 487 个倒置, 11336 个InDels (100 bp~1 kb的插入和缺失).

在结构变异的检测阶段, 有几方面问题值得关注. 首先, 由于结构变异的检测技术和数据处理方法各不相同, 不同研究机构获得的突变信息的数据质量及假阳性、假阴性率必然存在差异, 因此有必要建立一套标准规则, 以规范各种结构变异的特征和提交程序, 包括标准的命名体制和准确全面的注释信息等. 其次, 不同研究者采用的人类参考基因组(reference assembly)也有不同, 或采用不同的数据库, 或采用不同来源的DNA, 都最终导致数据的不一致并且增加了不同数据库整合在一起的难度. 而且由于目前的人类基因组序列仍未全部测通, gap的存在以及不同参考基因组的差异(如NCBI reference assembly 和Celera assembly)无疑都会影响获得数据的可靠性和准确性. 因此有必要建立一个标准通用且序列完整的参考基因组. 第三, 检测阶段一个最基本的问题就是缺乏能同时全面扫描基因组中包括平衡重排和非平衡重排在内的所有突变的高通量高分辨率的技术. 目前已有的方法如aCGH、定量PCR等在通量或分辨率上各有局限. 解决的关键即如何有效整合各技术平台, 尤其是新一代测序技术和基因芯片技术, 以实现同时分析各种突变(包括SNPs和结构变异)的目的[69,70].

结构变异的检测分析包括 3 个主要步骤: 除上面提到的检测即鉴定外, 还包括测序和基因型频率分析. 对于疾病关联研究而言, 在大量样本中进行充足的基因型分析以获得基因型频率和连锁不平衡规律至关重要. 在此阶段, 不可避免的问题就是如何建立准确、全面而又可信的人类全基因组水平的结构变异数据库. 目前, 已建立了一些有关基因变异的数据库如: 基因变异数据库(DGV)(http://projects.tcag.ca/variation/)、人类结构变异数据库(http://humanparalogy.gs.washington.edu/structuralvariation/)、数据来源于Ensemble的染色体不平衡和表型数据库(DECIPHER)

(www.sanger.ac.uk/postgenomics/decipher/)、人类基因组表观遗传学网络(HuGENet) (www.cdc.gov/genomics/hugenet/default.htm). 众多数据库的最大缺陷在于对提交的结构变异没有标准要求, 因此有必要建立数据整合处理分析的公共平台.

目前, 各类结构变异的鉴定尚处于起步阶段, 其信息量将会继续增加. 采用某一方法检测到的突变一定要通过其他不同方法加以验证, 以降低假阳性率. 只有这样, 得到的数据才有意义. 另外, 一个不可忽视的问题就是结构变异检测的发展速度远快于基因型分析和关联研究, 海量的变异数据只有部分被进一步研究, 如果仍是这种状况, 海量的突变信息将失去价值, 因此结构变异与疾病的关联研究同样要加快速度[50].

参考文献

1 Feuk L, Carson A R, Scherer S W. Structural variation in the human genome. Nat Rev Genet, 2006, 7(2): 85—97[DOI]

2 Sachidanandam R, Weissman D, Schmidt S C, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 2001, 409(6822): 928—933[DOI]

3 Kruglyak L, Nickerson D A. Variation is the spice of life. Nat Genet, 2001, 27(3): 234—236[DOI]

4 Frazer K A, Ballinger D G, Cox D R, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature, 2007, 449(7164): 851—861[DOI]

5 Li Y, Grupe A, Rowland C, et al. Evidence that common variation in NEDD9 is associated with susceptibility to late-onset Alzheimer's and Parkinson's disease. Hum Mol Genet, 2008, 17(5): 759—767[DOI]

6 Shastry B S. SNP alleles in human disease and evolution. J Hum Genet, 2002, 47(11): 561—566[DOI]

7 Iafrate A J, Feuk L, Rivera M N, et al. Detection of large-scale variation in the human genome. Nat Genet, 2004, 36(9): 949—951[DOI]

8 Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. Science, 2004, 305(5683): 525—528[DOI]

9 Check E. Human genome: patchwork people. Nature, 2005, 437(7062): 1084—1086[DOI]

10 Freeman J L, Perry G H, Feuk L, et al. Copy number variation: new insights in genome diversity. Genome Res, 2006, 16(8): 949—961[DOI]

11 Sharp A J, Cheng Z, Eichler E E. Structural variation of the human genome. Annu Rev Genomics Hum Genet, 2006, 7: 407—442[DOI]

12 Sebat J. Major changes in our DNA lead to major changes in our thinking. Nat Genet, 2007, 39(7 Suppl): S3—5[DOI]

13 Redon R, Ishikawa S, Fitch K R, et al. Global variation in copy number in the human genome. Nature, 2006, 444(7118): 444—454[DOI]

14 Mills R E, Luttig C T, Larkins C E, et al. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res, 2006, 16(9): 1182—1190[DOI]

15 Sperling K, Wiesner R. A rapid banding technique for routine use in human and comparative cytogenetics. Humangenetik, 1972, 15(4): 349—353[DOI]

16 Gall J G, Pardue M L. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci USA, 1969, 63(2): 378—383[DOI]

17 Bauman J G, Wiegant J, Borst P, et al. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Exp Cell Res, 1980, 128(2): 485—490[DOI]

18 Parra I, Windle B. High resolution visual mapping of stretched DNA by fluorescent hybridization. Nat Genet, 1993, 5(1): 17—21[DOI]

19 Speicher M R, Carter N P. The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet, 2005, 6(10): 782—792[DOI]

20 Raap A K, Florijn R J, Blonden L A J, et al. Fiber FISH as a DNA Mapping Tool. Methods, 1996, 9(1): 67—73[DOI]

21 de Stahl T D, Sandgren J, Piotrowski A, et al. Profiling of copy number variations (CNVs) in healthy individuals from three ethnic groups using a human genome 32 K BAC-clone-based array. Hum Mutat, 2008, 29(3): 398—408[DOI]

22 McCarroll S A, Kuruvilla F G, Korn J M, et al. Integrated detection and population-genetic analysis of SNPs and copy number varia-

tion. Nat Genet, 2008, 40(10): 1166—1174[DOI]
23 Butler H, Ragoussis J. BeadArray-based genotyping. Methods Mol Biol, 2008, 439: 53—74[DOI]
24 Dhami P, Coffey A J, Abbs S, et al. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. Am J Hum Genet, 2005, 76(5): 750—762[DOI]
25 Ijssel P, Ylstra B. Oligonucleotide array comparative genomic hybridization. Methods Mol Biol, 2007, 396: 207—221[DOI]
26 Lucito R, Healy J, Alexander J, et al. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res, 2003, 13(10): 2291—2305[DOI]
27 Jobanputra V, Sebat J, Troge J, et al. Application of ROMA(representational oligonucleotide microarray analysis) to patients with cytogenetic rearrangements. Genet Med, 2005, 7(2): 111—118
28 Backx L, Van Esch H, Melotte C, et al. Array painting using microdissected chromosomes to map chromosomal breakpoints. Cytogenet Genome Res, 2007, 116(3): 158—166[DOI]
29 Feuk L, Marshall C R, Wintle R F, et al. Structural variants: changing the landscape of chromosomes and design of disease studies. Hum Mol Genet, 2006, 15(1): R57—66[DOI]
30 Hollox E J, Atia T, Cross G, et al. High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH). J Med Genet, 2002, 39(11): 790—795[DOI]
31 Sellner L N, Taylor G R. MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat, 2004, 23(5): 413—419[DOI]
32 Gibbons B, Datta P, Wu Y, et al. Microarray MAPH: accurate array-based detection of relative copy number in genomic DNA. BMC Genomics, 2006, 7: 163[DOI]
33 Philippos P C, Kousoulidou L, Mannik K, et al. Detection of small genomic imbalances using microarray-based multiplex amplifiable probe hybridization. Eur J Hum Genet, 2007, 15(2): 162—172[DOI]
34 Schouten J P, McElgunn C J, Waaijer R, et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res, 2002, 30(12): e57[DOI]
35 Gatta V, Antonucci I, Morizio E, et al. Identification and characterization of different SHOX gene deletions in patients with Leri-Weill dyschondrosteosys by MLPA assay. J Hum Genet, 2007, 52(1): 21—27[DOI]
36 Isaksson M, Stenberg J, Dahl F, et al. MLGA–a rapid and cost-efficient assay for gene copy-number analysis. Nucleic Acids Res, 2007, 35(17): e115[DOI]
37 Abdellah Z, Ahmadi A, Ahmed S, et al. Finishing the euchromatic sequence of the human genome. Nature, 2004, 431(7011): 931—945[DOI]
38 Tuzun E, Sharp A J, Bailey J A, et al. Fine-scale structural variation of the human genome. Nat Genet, 2005, 37(7): 727—732[DOI]
39 Korbel J O, Urban A E, Affourtit J P, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science, 2007, 318(5849): 420—426[DOI]
40 Jarvie T, Harkins T. *De novo* assembly and genomic structural variation analysis with genome sequencer FLX 3K long-tag paired end reads. Biotechniques, 2008, 44(6): 829—831[DOI]
41 Khaja R, Zhang J, MacDonald J R, et al. Genome assembly comparison identifies structural variants in the human genome. Nat Genet, 2006, 38(12): 1413—1418[DOI]
42 Feuk L, MacDonald J R, Tang T, et al. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet, 2005, 1(4): e56[DOI]
43 Krzywinski M, Bosdet I, Mathewson C, et al. A BAC clone fingerprinting approach to the detection of human genome rearrangements. Genome Biol, 2007, 8(10): R224[DOI]
44 Holt K E, Parkhill J, Mazzoni C J, et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. Nat Genet, 2008, 40(8): 987—993[DOI]
45 Conrad D F, Hurles M E. The population genetics of structural variation. Nat Genet, 2007, 39(7 Suppl): S30—36[DOI]
46 Rodriguez-Revenga L, Mila M, Rosenberg C, et al. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. Genet Med, 2007, 9(9): 600—606
47 Barbour V M, Tufarelli C, Sharpe J A, et al. alpha-thalassemia resulting from a negative chromosomal position effect. Blood, 2000, 96(3): 800—807
48 Kleinjan D J, van Heyningen V. Position effect in human genetic disease. Hum Mol Genet, 1998, 7(10): 1611—1618[DOI]

49 Lupski J R. Structural variation in the human genome. N Engl J Med, 2007, 356(11): 1169—1171[DOI]

50 Sharp A J. Emerging themes and new challenges in defining the role of structural variation in human disease. Hum Mutat, 2009, 30(2): 135—144 [DOI]

51 Bridges C B. The Bar "Gene" a Duplication. Science, 1936, 83(2148): 210—211[DOI]

52 Hurles M E, Dermitzakis E T, Tyler-Smith C. The functional impact of structural variation in humans. Trends Genet, 2008, 24(5): 238—245[DOI]

53 McCarroll S A, Altsh uler D M. Copy-number variation and association studies of human disease. Nat Genet, 2007, 39(7 Suppl): S37—42[DOI]

54 Nathans J, Piantanida T P, Eddy R L, et al. Molecular genetics of inherited variation in human color vision. Science, 1986, 232(4747): 203—210[DOI]

55 Blunt T, Steers F, Daniels G, et al. Lack of RH C/E expression in the Rhesus D--phenotype is the result of a gene deletion. Ann Hum Genet, 1994, 58(Pt 1): 19—24[DOI]

56 Jamain S, Quach H, Betancur C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet, 2003, 34(1): 27—29[DOI]

57 Butler M G, Fischer W, Kibiryeva N, et al. Array comparative genomic hybridization (aCGH) analysis in Prader-Willi syndrome. Am J Med Genet A, 2008, 146(7): 854—860

58 Armour J A, Barton D E, Cockburn D J, et al. The detection of large deletions or duplications in genomic DNA. Hum Mutat, 2002, 20(5): 325—337[DOI]

59 Lakich D, Kazazian H H, Antonarakis S E, et al. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat Genet, 1993, 5(3): 236—241[DOI]

60 Castermans D, Vermeesch J R, Fryns J P, et al. Identification and characterization of the TRIP8 and REEP3 genes on chromosome 10q21.3 as novel candidate genes for autism. Eur J Hum Genet, 2007, 15(4): 422—431[DOI]

61 Jakobsson J, Ekstrom L, Inotsume N, et al. Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. J Clin Endocrinol Metab, 2006, 91(2): 687—693[DOI]

62 Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science, 2005, 307(5714): 1434—1440[DOI]

63 Buckland P R. Polymorphically duplicated genes: their relevance to phenotypic variation in humans. Ann Med, 2003, 35(5): 308—315[DOI]

64 Agundez J A, Gallardo L, Ledesma M C, et al. Functionally active duplications of the CYP2D6 gene are more prevalent among larynx and lung cancer patients. Oncology, 2001, 61(1): 59—63[DOI]

65 Garcia-Closas M, Malats N, Silverman D, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet, 2005, 366(9486): 649—659[DOI]

66 Conrad D F, Andrews T D, Carter N P, et al. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet, 2006, 38(1): 75—81[DOI]

67 Jakobsson M, Scholz S W, Scheet P, et al. Genotype, haplotype and copy-number variation in worldwide human populations. Nature, 2008, 451(7181): 998—1003[DOI]

68 Eichler E E. Widening the spectrum of human genetic variation. Nat Genet, 2006, 38(1): 9—11[DOI]

69 Scherer S W, Lee C, Birney E, et al. Challenges and standards in integrating surveys of structural variation. Nat Genet, 2007, 39(7 Suppl): S7—15[DOI]

70 Eichler E E, Nickerson D A, Altshuler D, et al. Completing the map of human genetic variation. Nature, 2007, 447(7141): 161—165[DOI]