

DOI: 10.3724/SP.J.1005.2009.00339

拷贝数变异: 基因组多样性的新形式

吴志俊, 金玮

上海交通大学医学院附属瑞金医院心内科, 上海 200025

摘要: 基因拷贝数变异是指 DNA 片段大小范围从 kb 到 Mb 的亚微观突变, 是一可能具有致病性、良性或未知临床意义的基因组改变。Fosmid 末端配对序列比较策略、比较基因组杂交芯片是当前较多使用的检测手段。染色体非等位的同源重排、非同源突变和非 β DNA 结构是造成基因组拷贝数变异的重要原因。拷贝数变异可导致不同程度的基因表达差异, 对正常表型的构成及疾病的发生发展具有一定作用。文章在总结基因拷贝数变异的认识过程和研究策略的基础上, 分析了拷贝数变异的形成和作用机制, 介绍了第一代人类基因组拷贝数变异图谱, 阐述了拷贝数变异研究的临床意义, 提示在探索疾病相关的遗传变异时不能错失拷贝数变异这一基因组多样性的新形式。

关键词: 拷贝数变异; 单核苷酸多态; 国际单倍体计划

Copy-number variation: a new pattern of structural diversity in genome

WU Zhi-Jun, JIN Wei

Department of Cardiology, Rui Jin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200025, China

Abstract: Copy number variation (CNV) is increasingly recognized as a source of inter-individual differences in genome sequence and has been proposed as a driving force for genome evolution and phenotypic variation. Many CNVs resulted in different levels of gene expression, which may account for a significant proportion of normal phenotypic variation and human diseases. This review unveiled the research process and study strategy of CNVs. Subsequently, the potential mechanisms of CNV formation and its clinical implications were discussed. In addition, the first-generation copy number variation map of the human genome was introduced, which demonstrated that DNA copy number variation was associated with specific chromosomal rearrangements and genomic disorders.

Keywords: copy number variation; single nucleotide polymorphisms; the International HapMap Project

2001 年 2 月, 6 国科学共同体和 Celera 公司同时宣告人类基因组测序计划基本完成^[1], 这意味着人类基因组学取得了“结构基因组学”研究的突破, 进入了以研究基因组功能为主体内容的“后基因组时

代”。2004 年, 全球数个“人类基因组计划”研究意外发现, 正常个体间部分基因的拷贝数存在差异, 其中 2 个科研小组率先公布了正常个体人类全基因组的拷贝数变异情况^[2, 3]。基因拷贝数变异 (Copy

收稿日期: 2008-08-12; 修回日期: 2008-08-27

基金项目: 国家自然科学基金项目 (编号: 30500576) 和上海市科委“上海青年科技启明星计划”项目 (编号: 06QA14033) 资助

作者简介: 吴志俊 (1982-), 女, 硕士, 住院医师, 研究方向: 心血管疾病基因组学和药物基因组学。Tel: 021-64370045;

E-mail: totito19822005@hotmail.com

通讯作者: 金玮 (1971-), 女, 博士, 副主任医师, 硕士生导师, 研究方向: 疾病基因组学和药物基因组学。Tel: 13022108899;

E-mail: jinwei_ivy@hotmail.com

number variations, CNVs)^[4-6]是指较之于参照基因组, DNA片段缺失或复制大于 1 kb至Mb的结构变异, 由于CNVs本身蕴涵着丰富的遗传信息, 昭示着全面了解和系统评估CNVs的重要性。本文就CNVs的研究进程作一全面综述。

1 基因拷贝数变异的认识过程

人类基因组存在着广泛的变异, 其中包括: 单核苷酸多态(Single nucleotide polymorphisms, SNPs); 可变数目串联重复序列多态(Variable number of tandem repeats, VNTRs), 例微卫星突变, 转座因子位置改变(例Alu因子), 以及其他基因结构改变等。1959年, Lejeune等^[7]首先运用细胞遗传学分析手段, 发现了人类基因组中存在着大于 5 Mb的微观变异, 并观测到人类基因组遗传物质的复制、缺失及染色体重排。随着分子生物学研究技术的突飞猛进, 人类对基因组变异的认识也逐步深入^[8]。荧光原位杂交(Fluorescence *in situ* hybridization, FISH)技术可以在更高分辨率的水平准确探究DNA序列的改变, 应用特定的探针准确定位已知位点, 快速评估多位点的基因连锁不平衡。光谱基因分型技术可探测到更小的DNA变异。近几年来, 第三代遗传标记SNPs被认为是最重要的人类基因组变异形式, 人类基因组中至少存在 1 千万个SNPs, 平均每 300 个核苷酸内就有一个SNP, 是造成正常表型个体间差异的主要原因^[1, 9]。

随着高通量比较基因组杂交芯片(Array-based comparative genomic hybridization, CGH)技术的发展, 更大范围内的基因组微观变异被发现^[2, 3]。2004年, 全球范围内数个“人类基因组计划”研究意外发现, 正常个体间部分基因的拷贝数存在差异^[4, 5], DNA片段存在大小范围从kb至Mb的亚微观突变^[10], 主要包括缺失(Deletions)、嵌入(Insertions)、复制(Duplications)和复合多位点变异(Complex multisite variants), 统称为拷贝数变异(Copy number variations, CNVs)或拷贝数多态(Copy number polymorphisms, CNPs)。CNVs一度被认为可能具有良性的或未知的临床意义^[11], 但随着临床表型-基因分型关联研究的深入, 目前已明确, CNVs与多基因疾病密切相关^[12-14]。因此, 基于当前对于人类基因组结构

和功能的认识, 现将CNVs定义为一可能具有致病性、良性或未知临床意义的基因组改变。

Redon等^[15]将CNVs定义为一个长为 1 kb或者更长的, 比较于参考基因组不同拷贝数的DNA片段。CNVs的平均长度为 465 kb, 其中半数以上的CNVs在多个个体中重复出现, 并经常定位于其他类型的染色体重排附近。排列结构也较简单, 例如, 顺接重复, 即重复顺序所携带遗传信息的顺序及方向与染色体上原有的相同。CNVs常位于富含可重复序列的位置, 例如: 端粒、着丝点、异染色质等^[10]。

2 基因拷贝数变异的研究策略

2.1 CNVs 的方法学研究

CNVs数据库的构建理论上可以借鉴SNPs数据库的构建方式。SNPs数据库的建立经过了若干阶段。首先, 运用快速发展的检测手段对候选SNPs位点进行研究; 其次, 对所有的检测手段标准化以提高灵敏度, 降低假阳性率; 第三, 对已获得的大量SNPs数据准确分型, 统计人群分布频率^[3, 16, 17], 构建SNPs图谱。最后, 集合图谱上所有达到的一定灵敏度的数据进行系统重排。这些结果将有助于研究者和技术公司设计特定的基因芯片进行全基因扫描或者探究未知的SNPs。建立全面的人类基因组CNVs数据库亦遵循以上的步骤。

当前, CNVs的研究方法各有优缺点。Fosmid末端配对序列比较策略是一种比较优良的CNVs探究方法^[18], 但受限于DNA序列数据的有效性。美国国立卫生研究院国际人类基因组研究中心最近公布了一个研究意向, 他们将从 48 个HapMap个体中建立多个DNA文库(<http://www.genome.gov/18016538>), 每个DNA文库对 1 百万个克隆末端测序, 用于Fosmid末端配对序列比较。这一工作将为那些有代表性的个体提供结构变异的目录, 包括CNVs和染色体平衡重排。由于Fosmid文库的插入片段最长不超过 40 kb, 这项策略的理想检测范围是 8 kb至 40 kb。比较基因组杂交芯片(Array-based CGH)因能在特定的平台上对多个基因组同时进行快速扫描而凸显优势, 但其对分辨率的要求较高。目前, 嵌合芯片(Tiling arrays)和超高密度寡核苷酸芯片(High density oligonucleotide arrays)已经广泛用于CNVs的

研究^[19~22]。Redon等^[15]构建CNVs图谱时采用了两种互补的技术手段:高分辨率全基因组SNPs芯片(Affymetrix 500 K EA基因芯片)^[23]和WGTP(Whole genome tilepath)细菌人工染色体(Bacterial artificial chromosome, BAC)基因芯片^[24]。SNPs芯片适用于探测较小的探针覆盖区域的CNVs,而WGTP平台适用于检测较大、较复杂的CNVs。典型的比较基因组杂交芯片虽可检测出较大片段的拷贝数,但不能提供等位基因特异性CNVs信息,却可通过Fosmid配对末端序列比较策略推论出相关信息。

在CNVs的研究中,应尽量排除假阳性或假阴性的影响^[20]。对相同的CNVs区域运用不同的实验方法(Array CGH、Fosmid末端序列配对、SNPs数据分析等)和验证手段(定量PCR、直接测序法、荧光原位杂交等),可全面了解同一区域DNA片段的变异情况,提高CNVs检测的准确性。

2.2 CNVs的人群研究

随着分子生物学技术的发展,基因组研究已从探索基因的特殊位点发展至全基因组扫描。Iafrate等^[2]运用BAC基因芯片和比较基因组杂交技术,对人类全基因组大小间隔1 Mb的DNA片段进行克隆。研究在5个人群39名无血缘健康个体中共证实了255个DNA片段位点存在连锁不平衡遗传模式,大多数的CNVs存在于1p13.3,进一步的高通量荧光原位杂交技术证实此区域变异大小介于150 kb至425 kb间。Sebat等^[3]在9个人群20个无血缘个体中扩增*Bgl*I-片段,杂交后运用微阵列技术(Representational oligonucleotide microarray analysis, ROMA)^[25],分析人类全基因组大小间隔35 bp DNA片段,研究运用3个连续性寡核苷酸截止标准检测DNA平均长度,共发现76个222 kb以及1个465 kb的DNA片段。随后,Tuzun等^[18]运用电子策略比较两组不同种族人类基因组DNA序列,共发现241组CNVs,其片段大小介于8~40 kb,其中大多数CNVs是在先前的研究中未曾探测到的DNA片段。这项电子策略与先前的CNVs探索方法相比,其主要优势在于,既可发现其他形式的基因组结构变异,亦可发现多种配对末端序列排列方向的一致性差异。研究共检测到56个倒位断裂点和241个CNVs。上述3项研究均探明,CNVs常位于已知的染色体复制片段的附近,偶见

孤立存在。Sharp等^[26]运用包含人类基因组已知复制片段区域的DNA克隆定制芯片,对7个人群47名无血缘个体分析后发现119个CNVs,其中的39%与先前的研究所检测到的CNVs重叠。Liu等^[27]利用来自于4个前列腺癌家系23个个体近116 000个SNPs基因型数据,运用Affymetrix 100 K基因芯片检测胚系细胞CNVs的类别和频率,结果发现,在基因组中>100 kb的CNVs少见,但大小介于100~1 000 bp的CNVs普遍存在。

2.3 CNVs与SNPs的交叉性研究

许多CNVs位于基因结构变异的复杂区域,一些CNVs与邻近SNPs存在着连锁不平衡,可通过检验SNPs基因型推测邻近的CNVs^[16, 28, 29],但另一些CNVs独立分布于基因组内^[10, 30]不能直接通过探测SNPs而得到。而且,位于同一区域的SNPs、STRs基因型和CNVs会相互影响。这些基因组的多点变异会影响基因的正确分型与评分^[31],故通过标记SNPs对CNVs分型较为困难。

近期,又有三项关于CNVs的全基因组研究公布^[10, 16, 28]。其中,Conrad等^[10]和McCarroll等^[28]的研究是基于“国际单倍体计划”(The International HapMap Project)^[17]公布的SNPs数据。由于“国际单倍体计划”的SNPs数据信息丰富,具有较高的空间分辨率,较适合用于研究影响SNPs基因型分布的CNVs。McCarroll等^[28]猜想拷贝数缺失变异会以3种形式的“足迹(Footprints)”表现在SNPs数据上:(1)特定个体携带的无效基因型;(2)邻近区域SNP等位基因频率背离Hardy-Weinberg平衡原理预期估计值;(3)SNPs基因型运算结果不符合孟德尔遗传模式。McCarroll等^[28]发现了541个拷贝数缺失变异,大小介于1 kb至745 kb。其中,120个为纯合子缺失(即两条染色体上拷贝数均缺失)。在这些纯合子缺失中,10个CNVs表现为有表达活性的多个外显子缺失,是个体在脂质代谢、嗅觉和药物反应性等方面产生差异的原因。Conrad等^[10]共发现了586个拷贝数缺失变异,大小介于300 bp至1.2 Mb,同时指出大多数缺失的CNVs区域是相对的乏基因(Gene-poor)区域,揭示了基因的拷贝数缺失服从净化选择,尽管如此,人类基因或多或少还是受到拷贝数缺失的影响。研究发现92个基因存在完全性缺失,109个基因的编码序列存在部

分性缺失。Hinds等^[16]利用高通量寡核苷酸芯片杂交 24 个无血缘个体的DNA样本, 共发现 215 个潜在的缺失片段, 大小介于 70 bp至 10 kb之间。随后对其 100 个CNVs进行PCR验证时发现, 41 个缺失变异相关等位基因频率 10%。再次对 100 个CNVs在 71 个无血缘个体中进行重复性检验, 同时用相同人群构成的 1.6 M的SNPs图谱进行基因分型研究。结果发现, CNVs普遍与周围相关的SNPs具有连锁不平衡的遗传关联, 证明CNVs与SNPs具有相似的进化过程。

3 基因拷贝数变异的形成和作用机制

3.1 CNVs 潜在的形成机制

CNVs常发生在同源重复序列或DNA重复片段之内或之间的区域, 染色体结构重排造成非同位的同源重组(Non-allelic homologous recombination, NAHR), 可引起CNVs^[13]。NAHR可造成大的结构多态和染色体重排, 直接造成基因组不稳定或早发性高频外显紊乱(Penetrant disorders)^[32]。但并非所有的CNVs都与DNA重复片段有关, 非同源突变(Non-homology-based mutation)也可造成CNVs。另一些CNVs则与非βDNA结构相关, 即DNA的结构有别于经典的右旋β双螺旋结构, 包括左旋Z型DNA和十字型DNA。这些非βDNA结构促进染色体重排和CNVs形成^[33, 34]。通过针对CNVs的大规模基因组扫描, 可进一步探索DNA序列脆性差异、DNA双链断裂非同源末端的连接修复机制。Lee等^[35]利用高分辨率寡核苷酸比较基因组杂交芯片和断点序列分析技术, 分析了 17 个Pelizaeus-Merzbacher (PMD)患者蛋白脂质蛋白质基因(Proteolipid protein gene, *PLP1*)周围的低CNVs序列, 提出复制叉停滞和模板转换(Replication Fork Stalling and Template Switching, FoSTeS)的假想, 认为是PMD患者染色体重排的主要机制。但FoSTeS是否是细胞在正常情况下控制模板转换的机制仍需进一步探索^[36]。

CNVs的大小亦和突变的机制相关。尽管机理未完全探明, 但较之于小片段的CNVs, 大片段的CNVs与DNA重复片段更为密切^[13]。此外, 由于基因组遗传物质获得或丢失的方式不同, 使DNA片段复制或缺失的进化压力有所不同^[31]。不过, 非同源突

变机制是造成小片段CNVs的主要原因^[31]。基因组的不平衡造成的分子表型结果包括特定的CNVs、邻近的DNAs、及其他有影响的基因组区域(增强子、阻遏物等)等位基因所组成的基因型。随着位点特异性和等位基因特异性定量芯片的发展, CNVs的遗传模式将日趋明朗。

3.2 CNVs 对基因表达的影响机制

当前研究表明, CNVs主要通过两个方面的机制影响基因的表达: (1)直接通过剂量效应改变特定基因的表达量, 造成基因微小缺失或微小复制, 引起功能紊乱^[32]。例如: Angelman 综合征, DiGeorge综合征, Charcot-Marie-Tooth综合征等。(2)影响基因转录调控因子, 间接影响基因表达的量^[28, 37, 38]。例如: 阻遏子的片段缺失可导致基因转录活性的增强; 启动子及邻近区域的片段重复可导致次稳定重排, 降低基因表达的活性。Stranger等^[37]将HapMap发现的CNVs数据与基因表达活性数据联系起来, 发现CNVs与 17.7%的基因表达活性相关。大多数呈正相关, 即拷贝数增加使覆盖区域或临近区域的基因表达活性增强; 但也有 15%的关联呈负相关, 例如阻遏子区域的片段拷贝数增多可使相关的基因表达活性减弱。例如, *PLP1* 基因下游小于 150 kb的片段重复可抑制基因的表达, 不产生PLP1 蛋白, 导致 2 型遗传痉挛性截瘫的发生^[39]。CNVs甚至可以间隔 6 Mb的远距离来调控基因的表达^[37]。

尽管某些 CNVs 具有高效的干扰作用, 但不必过分担忧这种遗传性基因不平衡。正常个体单一基因拷贝数缺失不一定会导致相关表型的疾病发生, 因为正常个体间可能存在其他形式的基因组变异缓冲了 CNVs 这种不平衡作用。但由于多基因改变和环境等综合因素, 就有可能影响多基因背景的慢性复杂性疾病, 包括精神分裂症、牛皮癣、动脉硬化症、先天性白内障等。

4 基因拷贝数变异的临床意义

4.1 人类基因组 CNVs 图谱

2006 年 11 月 23 日, 由Redon等多国研究人员组成的研究小组在《Nature》杂志上, 公布了人类基因组第一代CNVs图谱^[15]。这是首张鉴别影响人类基因

活性的最重要的DNA变异片段的图谱,首次在人类基因组范围内观测个体独特的遗传变异,并揭示了其独特的基因活动形态机制。这张遗传图谱是在4个人群270个个体的单倍体图谱(HapMap)基础上,检测培养细胞中14 000多个基因的活性,随后根据HapMap的定义、SNPs的索引和CNVs的新索引,运用两个互补的技术(SNPs基因分型技术和以克隆为基础的比较基因组杂交技术),进行CNVs筛选,根据基因序列的重复及缺失对CNVs进行分类^[24,40]。为优化数据的有效性,及CNVs和SNPs的综合效应,使用了国际HapMap DNA和来自与正常人群的细胞采集比对^[41],共获得了1 447个CNVs区域,其中包含了成百上千的基因、疾病位点、功能性因子和部分重复序列,估计有1 000多个基因的活性受序列变化的影响。将每个基因的活性与相邻的遗传变异联系起来,共检测了210个无血缘个体的14 000个基因的活性。该研究发现,至少10%~20%的基因活性遗传变异是由CNVs引起的。这张图谱也描述了许多CNVs的连锁不平衡模式,揭示了人类拷贝数标记突变。研究同时发现SNPs和CNVs的变异分别与900个和240个基因的活性变化有关。人类基因组间个体差异至少大于20 Mb,大约有3 000个基因与CNVs相关。与CNVs有关的彗星变异中,只有10%亦与SNPs相关,故提示在研究疾病相关的遗传变异时必须考虑CNVs,否则会错过重要的遗传信息^[23]。

4.2 CNVs研究与复杂性疾病的关系

基因组多位点的CNVs可以引起基因组和分子表型的异质性,影响复杂性疾病的发生发展。CNVs偏向分布于基因和超保守区域外的位置,多达40%的CNVs位于基因沙漠区(Gene deserts)^[15,42]。尽管如此,仍有大量的基因位于CNVs区域。Redon等^[15]在1 447个经HapMap检测到的CNVs区域中,发现了2 908个NCBI参考序列基因和285个OMIM基因,提示CNVs与复杂性疾病或孟德尔遗传紊乱相关。存在CNVs的基因通常不编码与生长发育相关的蛋白,但是,存在CNVs的基因常影响人体对外界环境的反应,在细胞连接、感知理解、化学刺激、神经生理过程中发挥重要作用。不存在CNVs的基因往往是剂量敏感性基因,参与维持细胞的生长发育,包括细胞信号传导、增殖、激酶化和磷酸化等过程。此

外,一些CNVs区域覆盖非编码的RNAs区域,包括miRNAs^[15,43]。miRNAs调节基因表达转录后加工的能力,对生长发育、正常生理过程及包括癌症在内的复杂性疾病的形成起重要作用。有文献报道,人13q14区域CNVs可干扰miRNAs表达,导致血液系统恶性肿瘤的发生^[44]。

CNVs不会直接造成中性基因早发性高频遗传紊乱,但在迟发性复杂性疾病中扮演重要的角色^[45]。CNVs的功能研究显示,CNVs在分子-环境相互作用以及人体对外界环境刺激的特殊反应中发挥重要作用^[3,18,45,46],主要包括(1)药物代谢反应,例如:谷胱甘肽还原酶基因、细胞色素P450基因、羧酸酯酶基因家族等;(2)炎症反应,例如:白细胞免疫球蛋白样受体、防卫素、APOBEC基因家族等;(3)膜表面完整性,例如:表皮角化包膜和粘蛋白基因家族等;(4)膜表面抗体,例如:半乳凝素、黑色素瘤抗原基因、Rh抗体家族等。

同样,CNVs可引起药物反应性和疾病易感性方面的差异^[47],例如:CCL3L1是HIV感染辅助受体CCR5的抑制性趋化因子和配体^[47],低拷贝数的CCL3L1基因可降低血浆CCR5-CCL3L1复合体水平,使血浆游离CCR5增多,增加HIV/AIDS感染细胞的机率。Aitman等^[48]发现老鼠*Fcgr3-rs*基因,其人类的直系同源基因*FCGR3B*拷贝数缺失可引起血管球性肾炎。*Fcgr3*基因编码巨噬细胞膜表面跨膜受体,故*Fcgr3*基因拷贝数缺失可导致免疫应答过度活跃或启动自身免疫反应,是血管球性肾炎的独立危险因素。

由于复杂性疾病染色体水平的分子机制未明,在基因组水平仅凭CNVs来解释复杂疾病的发生发展言之尚早^[6]。理论上,复杂性疾病的基因组变异,更为“弹性(Soft)”,例如,非编码区序列拷贝数的变异,尽管改变了基因的剂量但并不使基因的功能完全消失^[49]。由于CNVs存在多种作用机制,对于分子表型和基因表达的效果也不同,因此对CNVs的临床意义和遗传方式的解释必须倍加小心,需要基于全面的基因组变异评估^[45]。

5 结语

人类基因组研究的核心内容是了解分子表型变

异的遗传机制^[45], 探索基因变异在生物进化、生理过程及疾病进程的作用, 当前开展的CNVs研究亦将遵循上述原则。未来, 新的高通量基因扫描技术将推动更完善的基因组CNVs图谱的构建, 通过验证实验结果, 结合各种数据来源, 检测CNVs在不同群体中的分布频率, 构建全面的CNVs数据库。必须注意, 对于CNVs的认识与研究应着眼于全基因组变异的角度, 将各种大小、机制不同的基因组结构和功能改变综合分析, 才能全面揭示基因变异对于生物学功能、进化和疾病的影响。

参考文献(References):

- [1] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome, 2001, 1304–1351.
- [2] Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet*, 2004, 36(9): 949–951. [\[DOI\]](#)
- [3] Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. *Science*, 2004, 305(5683): 525–528. [\[DOI\]](#)
- [4] Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. Copy number variation: new insights in genome diversity. *Genome Res*, 2006, 16(8): 949–961. [\[DOI\]](#)
- [5] Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*, 2006, 7(2): 85–97. [\[DOI\]](#)
- [6] Goidts V, Cooper DN, Armengol L, Schempp W, Conroy J, Estivill X, Nowak N, Hameister H, Kehrer-Sawatzki H. Complex patterns of copy number variation at sites of

- segmental duplications: an important category of structural variation in the human genome. *Hum Genet*, 2006, 120(2): 270–284. [DOI](#)
- [7] Lejeune J, Turpin R, Gautier M. Mongolism; a chromosomal disease (trisomy). *Bull Acad Natl Med*, 1959, 143(11–12): 256–265.
- [8] Hollox EJ, Armour JA, Barber JC. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am J Hum Genet*, 2003, 73(3): 591–600. [DOI](#)
- [9] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chisoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrino A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409(6822): 860–921. [DOI](#)
- [10] Conrad DF, Andrews TD, Carter NP, Hurler ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet*, 2006, 38(1): 75–81. [DOI](#)
- [11] Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, 2006, 52(1): 103–121. [DOI](#)
- [12] Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid BM, Schoumans J, Sistermans EA, Veltman JA, Brunner HG, de Vries BB. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet*, 2006, 38(9): 999–1001. [DOI](#)
- [13] Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, Bobrow M, Carter NP. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet*, 2004, 41(4): 241–248. [DOI](#)
- [14] Redon R, Baujat G, Sanlaville D, Le Merrer M, Vekemans M, Munnich A, Carter NP, Cormier-Daire V, Colleaux L. Interstitial 9q22.3 microdeletion: clinical and molecular characterisation of a newly recognised overgrowth

- syndrome. *Eur J Hum Genet*, 2006, 14(6): 759–767. [\[DOI\]](#)
- [15] Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature*, 2006, 444(7118): 444–454. [\[DOI\]](#)
- [16] Hinds DA, Kloeck AP, Jen M, Chen X, Frazer KA. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet*, 2006, 38(1): 82–85. [\[DOI\]](#)
- [17] Consortium TIH. The international HapMap project. *Nature*, 2003, 426(6968): 789–796. [\[DOI\]](#)
- [18] Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. *Nat Genet*, 2005, 37(7): 727–732. [\[DOI\]](#)
- [19] Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet*, 2004, 36(3): 299–303. [\[DOI\]](#)
- [20] Dhami P, Coffey AJ, Abbs S, Vermeesch JR, Dumanski JP, Woodward KJ, Andrews RM, Langford C, Vetrie D. 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. [\[DOI\]](#)
- [21] Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer*, 2005, 44(3): 305–319. [\[DOI\]](#)
- [22] Shabana W, Peeters E, De Maeseneer M. Measuring thyroid gland volume: should we change the correction factor? *AJR Am J Roentgenol*, 2006, 186(1): 234–236. [\[DOI\]](#)
- [23] Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J, Liu G, Ihara S, Nakamura H, Hurles ME, Lee C, Scherer SW, Jones KW, Shapero MH, Huang J, Aburatani H. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res*, 2006, 16(12): 1575–1584. [\[DOI\]](#)
- [24] Fiegler H, Redon R, Andrews D, Scott C, Andrews R, Carder C, Clark R, Dovey O, Ellis P, Feuk L, French L, Hunt P, Kalaitzopoulos D, Larkin J, Montgomery L, Perry GH, Plumb BW, Porter K, Rigby RE, Rigler D, Valsesia A, Langford C, Humphray SJ, Scherer SW, Lee C, Hurles ME, Carter NP. Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res*, 2006, 16(12): 1566–1574. [\[DOI\]](#)
- [25] Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res*, 2003, 13(10): 2291–2305. [\[DOI\]](#)
- [26] Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*, 2005, 77(1): 78–88. [\[DOI\]](#)
- [27] Liu W, Chang B, Li T, Dimitrov L, Kim S, Kim JW, Turner AR, Meyers DA, Trent JM, Zheng SL, Isaacs WB, Xu J. Germline copy number polymorphisms involving larger than 100 kb are uncommon in normal subjects. *Prostate*, 2007, 67(3): 227–233. [\[DOI\]](#)
- [28] McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM. Common deletion polymorphisms in the human genome. *Nat Genet*, 2006, 38(1): 86–92. [\[DOI\]](#)
- [29] Newman TL, Rieder MJ, Morrison VA, Sharp AJ, Smith JD, Sprague LJ, Kaul R, Carlson CS, Olson MV, Nickerson DA, Eichler EE. High-throughput genotyping of intermediate-size structural variation. *Hum Mol Genet*, 2006, 15(7): 1159–1167. [\[DOI\]](#)
- [30] Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszałek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet*, 2006, 38(4): 463–467. [\[DOI\]](#)

- [31] Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ. Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet*, 2004, 36(8): 861–866.[\[DOI\]](#)
- [32] Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet*, 2005, 1(6): e49.[\[DOI\]](#)
- [33] Kurahashi H, Emanuel BS. Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum Mol Genet*, 2001, 10(23): 2605–2617.[\[DOI\]](#)
- [34] Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeyasinghe SS, O'Connell CD, Cooper DN, Wells RD. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci USA*, 2004, 101(39): 14162–14167.[\[DOI\]](#)
- [35] Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 2007, 131(7): 1235–1247.[\[DOI\]](#)
- [36] Branzei D, Foiani M. Template switching: from replication fork repair to genome rearrangements. *Cell*, 2007, 131(7): 1228–1230.[\[DOI\]](#)
- [37] Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 2007, 315(5813): 848–853.[\[DOI\]](#)
- [38] Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*, 2005, 76(1): 8–32.[\[DOI\]](#)
- [39] Lee JA, Madrid RE, Sperle K, Ritterson CM, Hobson GM, Garbern J, Lupski JR, Inoue K. Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann Neurol*, 2006, 59(2): 398–403.[\[DOI\]](#)
- [40] Khaja R, Zhang J, MacDonald JR, He Y, Joseph-George AM, Wei J, Rafiq MA, Qian C, Shago M, Pantano L, Aburatani H, Jones K, Redon R, Hurles M, Armengol L, Estivill X, Mural RJ, Lee C, Scherer SW, Feuk L. Genome assembly comparison identifies structural variants in the human genome. *Nat Genet*, 2006, 38(12): 1413–1418.[\[DOI\]](#)
- [41] International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, 437(7063): 1299–1320.[\[DOI\]](#)
- [42] Derti A, Roth FP, Church GM, Wu CT. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. *Nat Genet*, 2006, 38(10): 1216–1220.[\[DOI\]](#)
- [43] Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet*, 2007, 80(1): 91–104.[\[DOI\]](#)
- [44] Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA*, 2002, 99(24): 15524–15529.[\[DOI\]](#)
- [45] Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet*, 2006, 15 Spec No 1: R57–66.
- [46] Nguyen DQ, Webber C, Ponting CP. Bias of selection on human copy-number variants. *PLoS Genet*, 2006, 2(2): e20.[\[DOI\]](#)
- [47] Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, 2005, 307(5714): 1434–1440.[\[DOI\]](#)
- [48] Ouahchi K, Lindeman N, Lee C. Copy number variants and pharmacogenomics. *Pharmacogenomics*, 2006, 7(1): 25–29.[\[DOI\]](#)
- [49] de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N, Smeets D, Sistermans EA, Feuth T, van Ravenswaaij-Arts CM, van Kessel AG, Schoenmakers EF, Brunner HG, Veltman JA. Diagnostic genome profiling in mental retardation. *Am J Hum Genet*, 2005, 77(4): 606–616.[\[DOI\]](#)