

· 专论与综述 ·

基因组拷贝数变异研究进展*

梁燕 吴建新[△]

(北京大学首都儿科研究所教学医院 北京 100020)

摘要 基因组结构变异分为两个层次:显微水平(microscopic)和亚显微水平(submicroscopic)。显微水平的基因组结构变异主要是指显微镜下可见的染色体畸变,包括整倍体或非整倍体、缺失、插入、倒位、易位、脆性位点等结构变异。亚显微水平的基因组结构变异是指 DNA 片段长度在 1Kb-3Mb 的基因组结构变异,包括缺失、插入、重复、重排、倒位、DNA 拷贝数目变化(copy number variation, CNV) 这些统称为 CNV 或者 CNP(copy number polymorphisms, CNP)。对 CNV 的研究能够帮助研究者建立遗传检测假说,进而发现疾病易感基因,同时加深对表型变异的理解,为今后研究人类生物功能、进化、疾病奠定基础。本文主要从 CNV 的研究历史、分子机制、研究方法、研究意义等四个方面进行综述。

关键词 亚微观结构变异 拷贝数变异

中图分类号:Q75, Q78 文献标识码:A 文章编号:1673-6273(2012)05-964-04

Research Progress on Gene Copy Number Variations*

LIANG Yan, WU Jian-Xin[△]

(Capital Institute of Pediatrics, Peking University, Beijing, 100020, China)

ABSTRACT: Two types of gene structural variations are defined as microscopic variations and submicroscopic variations. The genomic structure of micro-level variations that can be seen under the microscope mainly refers to the chromosome aberrations, including euploid or aneuploid, deletions, insertions, inversions, translocations, fragile sites and other structural variations. Submicroscopic structural variations include deletions, insertions, duplications, rearrangements, inversions, translocations or copy number variations (CNV) in large DNA segments (1Kb-3Mb). These are collectively referred as CNV or CNP (copy number polymorphisms, CNP). Researches on CNV could establish the genetic testing hypothesis, detect the disease susceptible genes, and understand the phenotype variations better. Besides, CNV would provide the basis for determining how genomic diversity impacts the biological functions, the evolution and human diseases. Here, we review the CNV research history, molecular mechanisms, related research methods, and the research significance.

Key words: Structural variations; Copy number variations

Chinese Library Classification(CLC): Q75, Q78 **Document code:** A

Article ID:1673-6273(2012)05-964-04

基因组结构变异(structural variation)可分为两个层次:显微水平(microscopic)和亚显微水平(submicroscopic)。显微水平的基因组结构变异主要是指显微镜下可见的染色体畸变,包括整倍体或非整倍体、缺失、插入、倒位、易位、脆性位点等结构变异。亚显微水平的基因组结构变异是指 DNA 片段长度在 1Kb-3Mb 的基因组结构变异,包括缺失、插入、重复、重排、倒位、DNA 拷贝数目变化(copy number variation, CNV)等,这些统称为 CNV 或者 CNP (copy number polymorphisms, CNP)^[1]。CNV 是一组具有良性、致病性或未知临床意义的基因组结构改变。染色体非等位同源重排、非同源突变和非βDNA 结构是造成基因组拷贝数变异的重要原因。目前研究表明 CNV 偏向分布于基因超保守区域外的位置,多达 40%的 CNV 位于基因沙漠区(gene deserts)。存在 CNV 的基因经常参与人体对外界环境的反应的生理过程,进而在细胞连接、感观理解、化学刺激、

神经生理等过程中发挥重要作用。不存在 CNV 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,同时 CNV 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。

1 基因拷贝数研究历史

自 1998 年 Lupski 给出了基因病的定义之后^[2],目前已经发现大量的基因病是由基因组结构改变引起的,而非传统的 Watson-Crick 碱基配对变化所引起,其中一些基因病是由重组区域的基因拷贝数发生改变所致。

2000 年 6 月 26 日参加人类基因组计划 (human genome project, HGP) 6 个国家 (包括中国) 的科学家公布完成了人类基因组草图,随后人类基因组序列绘制成功,首次在分子层面

* 基金项目:国家重点基础发展规划项目(973 项目)(2007CB511903)

作者简介:梁燕(1986-)女,博士研究生,主要研究方向:儿科相关疾病的生物化学研究

△通讯作者:吴建新,男,博士生导师,电话:010-85695578, E-mail: wjianxin@gmail.com

(收稿日期:2011-08-02 接受日期:2011-08-23)

上为人类提供了一份生命“说明书”HGP从分子层面上为多种遗传疾病、癌症及神经退化症的治疗提供了基础。

高通量阵列比较基因组杂交技术(array based comparative genomic hybridization CGH)加速了CNV的探究。2004年Iafraite等人通过细菌人工染色体微阵列(bacterial artificial chromosome, BAC-based array)对39个非相关的健康人研究后发现255个变异位点,其中有24个位点出现的频率大于10%,有6个位点出现的频率大于20%,其平均间隔为1Mb^[3]。同年Sebat及同事通过代表性单核苷酸微阵列分析(representational oligonucleotide microarray analysis, ROMA)对20个健康人研究发现了221个CNV,代表着76种CNP, CNP间隔平均长为465kb^[4]。此外发现CNP间隔内70个不同基因的CNV,包括调节神经功能、细胞生长、新陈代谢的基因,以及几种已知疾病的相关基因。由此可以看出在正常人群中也可能存在一定数目的CNV。

伴随着研究者对CNV的研究越来越深入,Redon等人通过对270名具有欧洲、非洲或亚洲世系的4个群体研究,构建了人类基因组第一代拷贝数变异图谱^[5]。该研究表明:拷贝数变异非常复杂,类型多样。通过两种平台:WGTP platform, 500K EA platform, 及两种互补技术:单核苷酸多态性(SNP)基因型微阵列和基于克隆比较基因组杂交技术对这些个体DNA进行鉴定,结果显示有1447种拷贝数变异区(copy number variation regions, CNVR)涵盖了360万个碱基(占人类基因组12%),其中285种与孟德尔遗传疾病相关;并且指出CNV通常不编码发育相关的重要基因,而是编码与环境作用相关的基因,即“环境敏感性基因”,而这些基因通常参与细胞粘附、化学刺激、感官知觉、神经生理过程等活动。

2009年AnnaC等通过分析全基因组单核苷酸变异(Single Nucleotide Polymorphisms, SNP)和CNV遗传标记与精神分裂症患者的相关性,提出不常见的致病性CNV区域对于精神分裂症易感性方面发挥更重要的作用,而不支持共同变异(common variation)区域与精神分裂症的相关性^[6]。

2010年Christiaan等通过对95个血液肿瘤细胞系的高通量分析发现了一些共同发生的基因拷贝数变化位点,并对这些位点进行功能分析,绘制出基因拷贝数变化(获得或缺失)网络,从而发现了一些中心节点,进而提出:大规模低强度的拷贝数变化可能是肿瘤发生发展过程的重要特征^[7]。

目前已有几个数据库用来收集CNVs信息^[8]:健康人群CNV可到Genomic Variants (www.projects.tcag.ca/variation)查询,神经发育异常的患者CNVs可到DECIPHER(www.sanger.ac.uk/PostGenomic/decipher/)查询;染色体异常的患者CNVs数据库www.ukcad.org.uk/cocoon/ukcad、www.isca.genetics.emory.edu/;染色体非平衡变异的患者CNVs可到www.ecaruca.net查询。

2 基因拷贝数变异形成机制

基因的结构特征决定基因是否容易发生重组,进而影响基因拷贝数变化。重组主要发生在特定的重复序列区域,或者低拷贝重复区(low copy repeats, LCRs)。LCR中包含一个或多个基因、假基因、基因片段、逆转录病毒序列、基因调控区,通常分

布在端着丝粒和端粒区域,其大小、相对方向、各拷贝之间的距离及同源程度,均将影响到CNV的形成^[9]。然而目前CNV的确切机制仍不甚清楚,可能的机制主要包括非等位基因同源重组机制(non-allelic homologous recombination, NAHR),非同源末端连接机制(non-homologous end joining, NHEJ)。

NAHR机制一般发生在经常重组的区域,这些区域有如下特征:(1)片段大小>10kb,(2)序列同源程度>97%,(3)序列方向明确,(4)每个LCR大小控制在5Mb以内,(5)LCR在同一染色体上^[9];而NHEJ不需要重组断端之间的具有严格的DNA同源性,但是仍能够引发彼此毫不相干的DNA断端的连接,导致包括移位(移位)等在内的染色体之间的重排。不经常发生重组的LCR,或者各LCR区域大小不一致时,倾向于通过NHEJ机制引起基因拷贝数变化。尽管如此,很多遗传学家并不认可NHEJ机制。

3 基因拷贝数变化研究方法

3.1 比较基因组杂交(comparative genomic hybridization, CGH)技术

由荧光原位杂交技术结合消减杂交技术衍生而来的CGH,仅需少量DNA即可在整个基因组水平研究DNA拷贝数差异。将待测组织DNA和正常对照组DNA分别用不同颜色的荧光标记,在滴有正常染色体中期分裂相的玻片上进行原位杂交。根据荧光比率,得出待测DNA和对照DNA的相对拷贝数。此过程中需加入竞争性DNA(competitor DNA, Cot-1DNA)。鉴于该方法检测的缺失的分辨率为目标片段不小于10Mb,检测扩增区段不小于2Mb,故对于低水平的DNA扩增和小片段的丢失会漏检。

3.2 微阵列比较基因组杂交(Array-CGH)技术

将DNA克隆或cDNA做成微阵列,代替传统CGH法的中期染色体作为杂交靶。此方法不仅分辨率高,而且可以确定相关基因,从而确定最可能导致疾病和其他问题的基因。Array-CGH中微阵列探针通常来源于PCR扩增的BAC(细菌人工染色体)、YAC(酵母人工染色体)、PAC(噬菌体人工染色体)或cDNA分子,其中又以BAC-CGH芯片为主。虽然目前应用最多的是BAC Array-CGH,但还存在许多不足,包括BAC克隆文库处理不易、PCR污染、BAC克隆在人基因组上的定位还不是百分之百准确等。

3.3 代表性寡核苷酸阵列分析(representational oligonucleotide microarray analysis, ROMA)

ROMA原理是用限制性内切酶将基因组DNA消化成碎片,仅有相对较小的碎片PCR扩增,在固体表面集成已知序列的基因探针,待测生物细胞或组织中大量标记的核酸序列与上述探针阵列进行杂交,通过检测相应位置杂交探针,实现基因信息的快速检测。ROMA技术突出的特点是可一次性检测多种样品,获得多种基因的差别表达图谱,灵敏性和敏感性比Array-CGH更高,然而,基因组复杂性的降低可能导致部分染色体表达的差异,这可能被错误地解释为拷贝数改变。其次,不同个体可有不同的消化图谱,可能一些个体的探针比例与限制性片段的大小有关而不是真正的拷贝数的改变。

3.4 FISH(fluorescence in situ hybridization)技术

基本原理是: 如果被检测的染色体或 DNA 纤维切片上的靶 DNA 与所用的核酸探针是同源互补的,二者经变性-退火-复性,即可形成靶 DNA 与核酸探针的杂交体。将核酸探针的某一种核苷酸标记上报告分子如生物素、地高辛,利用该报告分子与荧光素标记的特异亲和素之间的免疫化学反应,经荧光检测体系在镜下对待测 DNA 进行定性、定量或定位分析。

鉴于上述技术只能定性或定量检测到 CNV,但是不能相对定位检测,所以平衡染色体结构变异,如平衡易位、倒置均不能检测。FISH 技术能唯一能够提供 CNV 在基因组中的分布情况,进而弥补其他技术的不足。用 FISH 技术则能检测到母源易位基因,因此在遗传咨询,FISH 技术能够提供准确的信息^[10]。

3.5 其他技术

DNA 小片段或特定目标基因的结构变异可用多重可扩增

探针杂交技术 MAPH(multiplex amplifiable probe hybridization, MAPH),多重连接探针扩增技术(multiplex ligation-dependent probe amplification, MLPA)。此外一些常规分子生物学方法,如 qPCR、DNA 印迹技术(Southern blotting)、同源序列比率实验等均可定量定性的检测 CNV。

4 基因拷贝数变异的研究意义及应用

4.1 基因拷贝数变化与表型和疾病易感性的关联分析

目前拷贝数变异疾病可分为三大类:生殖细胞遗传 CNVs,在父母中即存在,新生的获得性 CNVs,父母中未发现的;体细胞性 CNVs,在胚胎单细胞时期之后发生。其中生殖细胞性 CNVs 和新生 CNVs 与很多疾病相关(表 1)^[8]。

表 1 人类拷贝数变化主要相关疾病

Table 1 Diseases associated with human copy number variations

Disorder	CNV	Gene	Effect	Risk associated	Study type	Significance
Infectious disease						
HIV-1/AIDS susceptibility	Common	CCL3L1	Dosage	Low CNV	Case control	Varies in populations
Autoimmune disorder						
Systemic lupus erythematosus (SLE)	Common	FCGR3B	Dosage	Low CNV	Case control	P = 2.7 × 10 ⁻⁸
Psoriasis	Common	DEFB	Dosage	High CNV	Case control	P = 1.65 × 10 ⁻⁶
Crohn's disease	Common	HBD-2	Dosage	Low CNV	Case control	P = 0.002
Neurological disorders						
Autism Spectrum Disorders	Unknown	Multiple	Unknown	De novo CNVs; multiple CNVs	Familial	P = 0.043
Parkinson's disease	Rare	SA/CA	Dosage	Duplication/Triplication	Familial	NA
Bipolar disorder	Rare	GSK3b	Dosage	Deletions and duplications	Case control	P = 0.002
Schizophrenia	Rare	Multiple	Positional	Deletions and duplications; de novo CNVs	Case control	Varies in CNV types
Cancers						
Breast cancer	Rare	MTUS1 (exon 4)	Positional	Exon deletion (decrease risk)	Familial Case	P = 0.01, OR = 0.58
Prostate cancer	Common	UGT2B17	Positional	Gene deletion	Case control	Varies in study subjects
Neruoblastoma	Common	NBPF23	Dosage	Deletions and duplications	Case control	P = 1.7 × 10 ⁻¹¹ , OR = 2.23

CNV 从以下几个方面影响基因的功能,进而影响表型的变化:(1)一个或几个基因的缺失会导致基因产物的缺失,同时 CNV 会影响转录调控原件,产生新的基因融合产物;(2)通过位置效应、基因印记、等位基因表达差异性表达影响基因的功能;(3)携带者中一些平衡 CNV 没有功能效应,但是可能会引起后代基因组的不稳定性,增加疾病易感性。Lars Feuk 等提供了一些基因型与表型的网站,其数据库能够提供大量的基因型与表型的关联信息^[11]。

目前随着基因型与表型关联性研究的深入,此前一些曾被认为是良性或者未知临床意义的 CNV,现在也已证实与一些特定的基因病相关。Shaw-Smith 等人通过 array-CGH, FISH 等技术对三个发育迟缓、学习障碍病例的研究表明,17q21.3 存在大小为 500kb-650kb 的新生微缺失^[12]。Redo 等人利用 array-CGH 方法,通过对两个临床表现为巨头、过度生长、精神活动迟缓、过度活跃、特殊面部表情的儿童的研究指出:9q22.32-9q22.33 中间的微缺失是过度生长和智障的一个原因^[13]。Ewart 等在 WBS (Williams-Beuren Syndromes) 患儿中发现 7q11123 处 ELN 基因的微缺失,它引起弹性蛋白量或质的缺陷,造成血管病变、部分面部特征、疝气、声音沙哑等表现^[14];最近

在男性进行性神经发育综合征中发现 MECP2 基因(位于 Xq28,编码甲基化 CqG 结合蛋白)的重复,其临床表现包括重度智障,语言功能障碍、婴幼儿低张症、经常感染、癫痫、进行性强制性痉挛,此外有些患者表现出孤独症状^[15]。由此可见,无论是罕见疾病、单基因病、多基因病、感染性疾病几乎都与 CNV 有关。

4.2 基因拷贝数变化与生物进化的关系

基因组拷贝数引起基因组结构的进化,进而引起生物的进化。Cheng 等对人和黑猩猩的研究表明:基因的亚微观拷贝数变异比单核苷酸的改变更为重要,提示从进化的观点来研究 CNV 将会更有意义^[16]。Perry 等指出相对于单个核苷酸的替代而言,大片的 CNV 可能承受着更大的选择压力^[17]。Cooper 等认为 CNV 在基因沙漠区保持在较低的频率水平,因此 CNV 更有可能成为净化选择的目标;这类与环境应答相关联的基因,如某些感知功能和免疫功能的基因,其 DNA 序列及其周边序列存在丰富的 CNV,因此这类基因常被认为是哺乳动物基因组快速适应性进化的典型代表^[18],而伴随着适应性进化的基因数目的重复常常是基因产生新功能的原动力之一。未来在系统生物学、蛋白质进化相关的研究中,CNV 将发挥重要作用,它

将是基因组学和蛋白组学之间重要的媒介。

5 结语

随着高通量全基因组扫描技术的发展,对基因组亚微结构变异的研究会越来越深入。目前对人类遗传变异的研究已经将细胞遗传学、亚微观拷贝数变异和单核苷酸多态性结合在一起,进而为将来的人类遗传学研究提供新的框架。然而对于亚微结构的变异研究重要的挑战,在于如何将不同试验方法得出的数据进行比较;如何设立参照组的基因序列使其不断完善;如何纵向深入研究将亚微结构变异与表型信息充分结合起来,进而更好的理解其对表型、疾病易感性的影响,如何区分界定良性与致病性的拷贝数变异,更好的将拷贝数变异的检测应用到临床,为一些疾病的诊断治疗提供新的视角。

参考文献(References)

- [1] 何永蜀,张闻,杨照青.人类基因组结构变异[J].遗传,2009,8,31(8):771-778
HE Yong-Shu,ZHANG Wen,YANG Zhao-Qing.Structural variation in the human genome[J].HEREDITAS (Beijing), 2009,8,31(8):771-778
- [2] Lupski JR. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits [J]. Trends Genet,1998, 14: 417-422
- [3] A John Iafrate, Lars Feuk, Miguel N Rivera, et al. Detection of large-scale variation in the human genome [J]. Nature Genet,2004, 36, 949-951
- [4] Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome [J]. Nat Genet, 2004, 36(8): 949-951. Science 305,525-528
- [5] Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome[J]. Nature, 2006, Nov 23;444(7118):428-429
- [6] Need AC, Ge D, Weale ME, et al. A Genome-Wide Investigation of S-NPs and CNVs in Schizophrenia [J]. PLoS Genet, 2009, Mar;5(3):387-392
- [7] Christiaan Klijn, Jan Bot, David J.Adams, et al. Identification of Networks of Co-Occurring,Tumor-Related DNA Copy Number Changes Using a Genome-Wide Scoring Approach [J]. PLoS Comput Bio, 2010,6(1):e1000631
- [8] KW Choy, SR Setlur, C Lee, et al. The impact of human copynumber variation on a new era of genetic testing[J]. BJOG ,2010,117:391-398
- [9] W.Gu, J.R.Lupski. CNV and nervous system diseases-what's new[J]? Cytogenet Genome Res,2008, 123:54-64
- [10] Lars Feuk, Christian R.Marshall, Richard F, et al. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis[J]. Human Molecular Genetics, 2006, Vol.15, Review Issue1 R57-R66
- [11] Feuk L, Marshall CR, Wintle RF, et al. Structural variants: changing the landscape of chromosomes and design of disease studies[J]. Hum Mol Genet, 2006, 15(1): R57-R66
- [12] Charles Shaw-Smith, Alan M Pittman, Lionel Willatt, et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability[J]. Nat Genet ,2006,38: 1032-1037
- [13] Redon R, Baujat G, Sanlaville D, et al. Interstitial 9q22.3 microdeletion: clinical and molecular characterisation of a newly recognised overgrowth syndrome[J]. Eur J Hum Genet ,2006,14:759-767
- [14] Ewart AK, Morris CA , Ensing GJ, et al. A human vascular disorder , supravalvular aortic stenosis, maps to chromosome 7[J]. Proc Natl Acad Sci USA , 1993 , 90 :3226-3230
- [15] Bauters M, Van Esch H, Friez MJ, et al. Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair[J]. Genome Res ,2008 ,18: 847-858
- [16] Ze Cheng , Mario Ventura, Xinwei She, et al. A genome-wide comparison of recent chimpanzee and human segmental duplications[J]. Nature,2005, 437:88-93
- [17] George H.Perry, Joelle Tchinda, Sean D.McGrath, et al. Hotspots for copy number variation in chimpanzees and humans[J]. Proc Natl Acad Sci USA, 2006,103:8006-8010
- [18] Gregory M Cooper, Deborah A Nickerson , Evan E Eichler. Mutational and selective effects on copy-number variants in the human genome[J]. Nature Genetics,2007, 39, S22-29
- [27] Qu K, Chen CP, Halliwell B, et al. Hydrogen sulfide is a mediator of cerebral ischemic damage[J]. Stroke. 2006, 37(3):889-893
- [28] 任彩丽, 赵红岗, 蔡德亮, 等. 脑梗死患者血浆中内源性硫化氢含量的变化[J]. 中国神经精神疾病杂志, 2010, 36(1): 43-45
Ren Cai-li, Zhao Hong-gang, Cai De-liang, et al. Change in plasmic endogenous hydrogen sulfide of patients with cerebral infarction[J]. Chin J Nerv Ment Dis, 2010, 36(1): 43-45(In Chinese)
- [29] Ren C, Du A, Li D, et al. Dynamic change of hydrogen sulfide during global cerebral ischemia-reperfusion and its effect in rats [J]. Brain Research, 2010, 1345(23): 197-205
- [30] Kamoun P. Mental retardation in Down syndrome: a hydrogen sulfide hypothesis[J]. Med Hypotheses, 2001, 57(3): 389-392
- [31] Belardinelli MC, Chabli A, Chadeaux-Vekemans B, et al. Urinary sulfur compounds in Down syndrome [J]. Clin Chem, 2001, 47(8): 1500-1501

(上接第 978 页)

- [27] Qu K, Chen CP, Halliwell B, et al. Hydrogen sulfide is a mediator of cerebral ischemic damage[J]. Stroke. 2006, 37(3):889-893
- [28] 任彩丽, 赵红岗, 蔡德亮, 等. 脑梗死患者血浆中内源性硫化氢含量的变化[J]. 中国神经精神疾病杂志, 2010, 36(1): 43-45
Ren Cai-li, Zhao Hong-gang, Cai De-liang, et al. Change in plasmic endogenous hydrogen sulfide of patients with cerebral infarction[J]. Chin J Nerv Ment Dis, 2010, 36(1): 43-45(In Chinese)
- [29] Ren C, Du A, Li D, et al. Dynamic change of hydrogen sulfide during global cerebral ischemia-reperfusion and its effect in rats [J]. Brain Research, 2010, 1345(23): 197-205
- [30] Kamoun P. Mental retardation in Down syndrome: a hydrogen sulfide hypothesis[J]. Med Hypotheses, 2001, 57(3): 389-392
- [31] Belardinelli MC, Chabli A, Chadeaux-Vekemans B, et al. Urinary sulfur compounds in Down syndrome [J]. Clin Chem, 2001, 47(8): 1500-1501