

DOI: 10.13376/j.cblls/2017049

文章编号: 1004-0374(2017)04-0371-09

基因拷贝数变异与人类疾病

胡力文, 杨康*

(中国人民解放军南京总医院, 南京 210001)

摘要: 拷贝数变异 (copy number variation, CNV) 是人类遗传多样性的一类重要形式。在前期的研究中, 人们通过寡核苷酸分型、比较基因组杂交以及测序等技术手段, 在人类基因组中鉴定出了大量拷贝数变异位点。这些变异可能是由于基因组重组或复制过程中的差错而产生。CNV 在人群中的覆盖率远远高于寡核苷酸多态性 (single nucleotide polymorphism, SNP), 它们可以通过多种机制改变基因的表达水平, 如基因剂量效应、基因断裂 - 融合效应, 以及远距调控效应, 进而引起多种人类复杂疾病。认识基因组中的拷贝数变异对于我们更好地认识基因与疾病的关系、遗传 - 环境因素的相互作用, 以及基因组变异与物种进化的关系具有重要的意义。

关键词: 拷贝数变异; 系统发育; 感染与免疫; 神经与精神疾病; 肿瘤

中图分类号: R394.3 文献标志码: A

Copy number variation and human diseases

HU Li-Wen, YANG Kang*

(Nanjing General Hospital, Nanjing 210002, China)

Abstract: Copy number variation (CNV) is an important form of human genetic polymorphisms. In previous studies, researchers have identified amounts of copy number variation by technologies such as single nucleotide genotyping, comparative genomic hybridization, and genome sequencing. Copy number variation may be derived from errors during genome recombination or replication, which covers much more genome range than single nucleotide polymorphisms (SNP). CNV may affect gene expression level by multiple ways such as gene dosage, gene fusion or disruption, and long-range regulation effects. Research on gene copy number variation may facilitate the understanding about the heredity-environment interaction on human pathogenesis, and be valuable to uncover the association between genomic variation and species evolution.

Key words: copy number variation; system development; infection and immunity; neural and mental diseases; cancer

随着测序技术及相关分子生物学技术的飞速发展, 人们对基因组的认识也越来越深刻。前期的研究中, 人们发现人类基因组中存在着大量的遗传变异, 这些变异对个体的性状产生不同的影响。在人群中比例大于 1% 的单个碱基位点变异被称之为寡核苷酸多态性 (single nucleotide polymorphisms, SNP)。前期大量的研究报道已揭示这种多态性广泛存在于人类基因组中, 且某些位点的变异与特殊的疾病发生密切相关^[1-4]。迄今为止, 数以千万计的寡核苷酸多态性已被人们揭示, 其功能意义也已经得到广

泛的阐释。SNP 的发现与功能鉴定解释了某些疾病的遗传学基础, 但对于大部分复杂疾病来说, 当前的解释仍远远不足。

相比于单个碱基的寡核苷酸多态性, 人们对于大片段基因组变异认识则相对较晚。直到 20 世纪 90 年代, 人们才发现, 由基因组扩增或丢失引

收稿日期: 2016-08-23; 修回日期: 2016-11-08

基金项目: 江苏省基础研究计划(自然科学基金)青年项目 (BK20160606)

*通信作者: E-mail: yangkangxwk@163.com

起的拷贝数变异同样可能引起一些符合孟德尔遗传定律的单基因疾病。

拷贝数变异是指相对于常见的二倍体基因组来说,发生的较大规模(>1 kb)的基因组结构变异,包括拷贝数的重复、丢失、倒位及易位,而我们所常提及的狭义的拷贝数变异指的是基因拷贝数目的改变^[5]。2006年,Redon等^[6]绘制了人类拷贝数变异图谱,发现CNV广泛存在于人类基因组中,其涉及范围之广超出了人们前期的预期,自此拉开了CNV研究的新时代。近年来,随着测序技术的快速发展,被鉴定发现的CNV数量越来越多,已有超过数十万个CNV位点被记录在数据库(Database of Genomic Variation),这些CNV所覆盖的染色体范围约占人类全基因组的20%以上。

1 产生机制

CNV产生的机制中最为人熟知的是非等位基因的重组,以及非同源末端连接^[7-8]。其中,非等位基因的重组是CNV产生的最主要的机制。该机制认为,在细胞分裂时,两条相似但位于基因组不同位置上的序列之间可发生配对及交换。如果发生交换的序列同向且位于同一条染色体,那么将会产生DNA拷贝数的扩增或丢失。非同源末端连接认为,当DNA受到电离辐射及氧化应激刺激时,将发生DNA损伤及修复事件,非同源末端连接即发生于该过程之中^[9]。发生非同源末端连接的DNA序列,并不需要用于重组的正向/反向重复序列,且在末端连接完成时,可能在发生连接处发生碱基的插入,这种特征使其不同于其他机制。

2007年, Lee等^[10]针对CNV的产生原理提出了一个新的模型:复制叉停滞与模板交换(fork stalling and template switching, FoSTeS)。该模型认为, DNA在复制过程中,其复制叉可以发生停滞。复制叉上的滞后链可以从引导链上解离,并通过一种名为微同源序列的元件转移到其他的复制叉上完成后续的复制。新的模板链与原来复制叉中的模板链不一定高度相似,但它们在空间上彼此靠近,模板转换的结果可以导致缺失或者重复。这种与原来引导链解离并在新的引导链上继续复制的过程可以连续多次重复,最终导致更加复杂的基因组重排事件。

此外,长间隔元件L1(long interspersed element-1, L1)介导生成的新序列是拷贝数变异发生的一类新机制。L1是目前已知的有转座活性的自发性转座子^[11]。全长的L1元件包含两个完整的开

放阅读框(open reading frame, ORF): ORF1编码RNA结合蛋白,而ORF2编码同时具备核酸内切酶及逆转录酶活性的蛋白。ORF1蛋白可以结合L1转录产物,并携带它们返回细胞核重新融入宿主基因组。此过程主要在ORF2帮助下完成,ORF2具有内切酶和逆转录酶的活性。逆转座时ORF2切开基因组DNA,并利用游离的3'-羟基基团起始逆转座的转录过程^[12]。L1转座是通过由RNA聚合酶II转录而成的RNA中间体完成,这种RNA中间体可以被逆转录然后插入到基因组新的位置,插入位置的两端是一对重复序列^[13]。

2 CNV与疾病关联分析的研究策略

近年来,随着测序技术以及基因组杂交技术的迅速发展,CNV的研究方法有了新的突破和进步,不再局限于既往的显微水平。目前关于拷贝数变异与疾病关联分析的方法主要有全基因组的关联分析以及基于候选基因的策略,其技术手段有:基于全基因组水平的比较基因组杂交^[14]、全基因组SNP分型芯片扫描^[15],以及下一代测序技术^[16]。基于候选位点策略的主要技术,包括实时定量PCR、多重连接的探针扩增(multiplex ligation-dependent probe amplification, MLPA)、多重扩增探针杂交技术(multiple amplifiable probe hybridization, MAPH)^[17]及旁系同源基因比例检测(paralogue ratio test)等^[18]。全基因组策略适用于大规模地筛查疾病相关CNV位点,代价高昂,而基于候选位点的策略适用于大样本量验证。

3 CNV与疾病关系

拷贝数变异可通过基因剂量效应直接改变所在基因的表达水平,或是通过位置效应调控远处基因表达,以及通过基因融合或断裂效应阻碍基因的表达,因此,在人类疾病的发生中具有重要意义。在相关检测技术快速发展的推动下,人们已经发现了诸多与各类疾病发生密切相关的拷贝数变异。当前人们对拷贝数变异与疾病关系的认识主要集中于发育疾病、感染与免疫疾病、神经系统疾病以及各类肿瘤。

3.1 与人类发育及先天性疾病的关系

很早以前,人们便已经认识到染色体结构异常与人类发育缺陷之间的关联,如唐氏综合征是由于21号染色体呈三倍体状态所造成的。这种大规模的染色体畸变可以在显微镜下观察到;而更多的染色体畸变的范围更小,无法在镜下被发现,但这些小

规模的拷贝数变异同样与人类的发育以及一些复杂疾病的发生密切相关。拷贝数变异与先天性疾病的关系见于先天性心脏病(GATA4、NKX2-5基因)^[19]、白化病(OCA2基因)^[20]、血友病(F8基因)^[21]等相关报道。

3.2 与感染及免疫性疾病之间的关系

该类疾病的发生往往与免疫系统功能相关的基因家族的扩增或丢失相关, 其中较为常见的包括趋化因子配体(C-C motif chemokine ligand, CCL)家族、 β -防御素(β -defensin, DEFB)家族、免疫球蛋白Fc片段受体家族(Fc fragment of IgG receptor, FCGR)、补体C4基因家族、白介素基因家族、Toll样受体基因家族等。

趋化因子配体家族成员:CCL3L1基因区域是人类基因组中的高度可变区, 极易发生拷贝数变异。CCL3L1编码产物可与趋化因子受体结合, 在人类感染与免疫疾病中发挥着重要的作用, 如CCL3L1与趋化因子受体CCR5结合可阻止HIV病毒进入细胞, 并阻止HIV病毒的复制^[22]。CCL3L1拷贝数丢失的个体, 其感染HIV的可能性更高, 且感染者的病情进展更快, 该观点已得到多个研究报道证实^[23]。CCL3L1的拷贝数丢失同样与丙型肝炎^[24]、肺结核^[25]的感染风险相关; 在自身免疫性疾病中, CCL3L1的低拷贝数同样与类风湿性关节炎、川崎病的发病相关^[26-27]。此外, Lee等^[28]报道, CCL3L1的拷贝数丢失与韩国人群哮喘发病风险相关。

β -防御素: DEFB基因家族编码抗菌或细胞毒性肽, 该基因功能缺陷可能导致细菌对黏膜上皮的黏附力增强, 进而增强细菌对黏膜的侵袭, 促进肠道黏膜炎症。研究表明, 该基因的拷贝数异常与感染性及炎症性疾病相关, 如DEFB2拷贝数丢失导致人群罹患克隆恩病可能性增加^[29]。

FCGR基因家族: 编码免疫球蛋白G Fc片段受体。其中, FCGR3B是最常见的免疫疾病相关基因, 主要表达于多形核白细胞表面。该基因家族主要与循环免疫复合物的清除有关, 其功能缺失可加重人类免疫性疾病。FCGR3B的拷贝数丢失与系统性红斑狼疮患者的易感性密切相关, 该论点已得到多个病例-对照研究的证实^[30-31]。来自于荷兰的病例-对照研究表明, 该位点的拷贝数丢失还与白种人的类风湿性关节炎相关^[31]。此外, 抗中性粒细胞胞浆抗体相关的脉管炎(ANCA-associated vasculitis)、银屑病^[32]、干燥综合征^[33]、硬化综合征^[34]、结节病等发病均与FCGR3B的拷贝数丢失相关^[35]。

补体C4: 补体系统可与免疫复合物结合, 该系统的激活可形成攻膜复合体, 降解靶细胞, 在促进病原吞噬、防止免疫复合物沉着以及中和病毒方面发挥着重要作用。C4是补体经典激活途径的一个重要组分, 定位于染色体6q21.3区, 该家族的拷贝数丢失可导致青少年系统性红斑狼疮及系统性红斑狼疮心包炎的发生风险增加^[36]。此外, C4基因的C4L亚基拷贝数丢失以及C4S的拷贝数增加可导致克隆恩病发病风险增加^[37], 与青少年皮炎发生密切相关^[38], 而高拷贝可降低中国汉族人群的原田综合征发病风险^[39]。此外, Liu等^[40]的研究表明, C4基因成员的拷贝数减少可降低人群患Grave's眼病的风险, 但当C4A基因低于两拷贝时, 可增加Grave's眼病患者的白癜风发病风险。

除了以上一些常见的拷贝数变异以外, 近年来, 一些疾病特异性的CNV位点也不断被鉴定发现, 如白细胞免疫球蛋白样受体(leukocyte immunoglobulin-like receptor A6, LILRA6)的单拷贝与过敏性皮炎的发病相关^[41]。这些基因往往与免疫系统的发育及功能发挥密切相关, 暗示了基因-环境因素的共同作用在疾病发生中的作用。

3.3 与神经系统疾病之间的关系

CNV与常见神经精神疾病的关系是当前的重点研究领域, 研究者在此领域获得了诸多重要发现。其中, 最常见的是关于自闭症的报道。自闭症在青少年中的发病率较高, 据报道, 每1万个青少年中, 可能有15~60个孤独症患者。在5%~10%的孤独症患者中, 都能检测到明显的染色体结构异常。其中, 在大约3%的孤独症儿童中发现了染色体15q11-q13区的拷贝数扩增。在随后的孤独症基因组计划中(autism genome project), 研究者们还发现, 除了15q11-q13、2p16、1q21以及17q12、16p11.2、22q11.2的拷贝数异常均与自闭症相关, 16p11.2区域也可同时出现拷贝数的扩增或丢失, 该区域的异常发现于大约1%的孤独症患者。此外, 研究人员还发现, 孤独症相关的拷贝数丢失常出现于一些与神经突触形成及突触蛋白合成相关的基因位点, 包括SHANK3、轴突蛋白neurexin 1, 以及神经连接蛋白(neurologin, NLGN4)等^[42-43]。这些位点的发现为人们认识孤独症的遗传机制提供了新的证据。

与自闭症相类似, 精神分裂症也已被证明与多个位点的拷贝数变异相关。前期的报道中, 精神分裂症相关的拷贝数变异见于ERBB4、SLC1A3、RAPGEF4、CIT、CHRNA7等基因的拷贝数丢失^[44-45],

这些基因的功能往往与神经系统的发育及递质传递功能密切相关,如 ERBB4 的编码产物可与神经调节蛋白相结合,并被其激活; CHRNA7 基因编码尼古丁乙酰胆碱受体,可调节神经元突触之间的快速递质传递。CIT 产物参与细胞的分裂,主要与神经系统的发育相关。

除此之外,研究还表明,CNV 与帕金森病、阿尔茨海默病、癫痫以及智障等神经系统疾病的发病有关。突触核蛋白 α (α -synuclein, SNCA) 高表达于中枢神经系统,可以衔接神经递质的突触前传递及跨膜转运。在帕金森患者的脑组织中,SNCA 的表达明显高于健康人群,其编码产物可形成淀粉样蛋白斑,是帕金森患者脑组织的结构特征之一。研究表明,帕金森患者脑组织中 SNCA 高表达可能与其拷贝数扩增有关,在一定程度上解释了帕金森的发病机理^[46]。此外,PTEN 诱导激酶 1 (PTEN induced putative kinase, PINK1)、泛素蛋白连接酶 (parkin RBR E3 ubiquitin protein ligase, PARK2) 等基因拷贝数变异也与帕金森患者的易感性相关^[47-48]。在阿尔茨海默病患者中,淀粉样前体蛋白 (amyloid beta precursor protein, APP) 的拷贝数扩增与 AD 的发病相关^[49]。在癫痫病患者中,同样发现了与精神分裂症相同的 CHRNA7 拷贝数丢失^[50],表明某些基因或通路在神经系统的发育中扮演着不可或缺的角色。综合文献报道,发现与癫痫相关的染色体丢失主要集中于 15q13.3、15q11.2、16p13.11 等位点。此外,钠离子通道 α 亚基 (sodium voltage-gated channel alpha subunit, SCN) SCN2A、SCN3A 基因拷贝数变异在癫痫发生中的作用也被揭示^[51]。在智障患者家系中,Moey 等^[52]发现 Xp11.22 染色体区的拷贝数扩增是最为常见的变异,其中涉及的基因包括 SMC1A、RIBC1、HSD17B10、HUWE1 等。人们还发现染色体 5q14.3 上肌细胞增强子家族成员 (myocyte enhancer factor, MEF2C)^[53] 基因及 9q33.1 上的星形肌动蛋白 (astrotactin, ASTN2)^[54] 的拷贝数异常与智力发育障碍、注意缺陷多动等智力行为相关。

3.4 CNV 与人类肿瘤之间的关系

拷贝数变异与人类肿瘤易感性的关联是近年来的研究热点,检测特定位点拷贝数可为人群肿瘤的早期预防提供预见性的指导。各类肿瘤发生相关的拷贝数变异已见于大量文献报道。根据 CNV 在人群中出现的频率,可将 CNV 分为常见 CNV (common CNV, MAF > 5%) 和稀有 CNV (rare CNV)。这些 CNV 发挥作用的特点各异。

位于编码基因位点的拷贝数异常,可通过基因剂量效应直接改变其编码产物的表达水平,进而改变肿瘤发生风险,如代谢基因家族谷胱甘肽—S 转移酶家族成员 GSTT1、GSTM1,以及 UDP 葡萄糖醛酸基转移酶基因家族的 UGT2B17、细胞色素 P450 基因家族成员 CYP2D6、磺基转移酶基因家族 SULT1A1。这些基因拷贝数变异导致代谢基因的异常表达,对外源毒力物质的代谢能力降低,促进肿瘤发生,其拷贝数变异与多种人类肿瘤的发生密切相关^[55-56]。此外,与细胞分裂、生长、染色体复制、修复等信号途径相关的基因拷贝数变异则可能导致细胞的异常生长,促使恶性转变,如促分裂原活化蛋白激酶激活的蛋白激酶 -2 (mitogen-activated protein kinase-activated protein kinase 2, MAPKAPK2) 的拷贝数增加则与肺癌、鼻咽癌的易感性相关^[57-58]。前期的研究表明,MAPKAPK2 可激活细胞周期蛋白 Cdc25B/C 并参与 DNA 修复及细胞凋亡过程,因此具有原癌基因的功能,其拷贝数增加与肺癌发生相关。该基因可被 EB 病毒直接激活,因此该拷贝数扩增与 EB 病毒感染联合作用可明显增加人群罹患鼻咽癌风险;此外,含 WW 域的氧化还原酶 (WW domain containing oxidoreductase, WWOX) 是一个被广泛认可的抑癌基因,可抑制癌基因 *c-Jun*、*erbB4*、*MDM2*,以及激活抑癌基因 *p53* 的表达,该基因拷贝数丢失与肺癌和胶质瘤的易感性相关^[59-60]。

而位于非编码区的拷贝数变异,则因其远距或近距离调控功能,可发挥增强子、抑制子样元件作用,间接对下游基因的表达水平产生影响,如位于染色体 2q35 基因荒漠区的一段增强子样元件丢失可通过改变胰岛素样生长因子结合蛋白 5 (insulin like growth factor binding protein, IGFBP5) 的表达促进乳腺癌发生^[61]。

4 常见肿瘤易感性相关 CNV 位点及其功能特点

4.1 乳腺癌

Long 等^[62]通过全基因组关联分析的方式,自中国人群中筛选到了与乳腺癌易感性密切相关的载脂蛋白 B mRNA 编辑酶 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3, APOBEC3) 基因家族丢失,该基因家族编码产物与 RNA 编辑功能相关,在控制细胞周期与生长中发挥着重要的作用。Xuan 等^[63]进一步在欧洲人群中证实了该基因的丢失与乳腺癌发病风险之间的关系。此外,2q35 上的拷贝数丢失与乳腺癌的关系亦得到了初步证实^[61]。

Frank 等^[64]报道, 线粒体上的抑癌基因 (mitochondrial tumor suppressor, MTUS1) 外显子 4 的拷贝数丢失可降低家族乳腺癌发病风险。作者认为, 该外显子的缺失可能产生高活性的 MTUS1 异构体, 丰富了对 CNV 作用机制的认识。Krepischi 等^[65]利用比较基因组杂交的方法, 对 BRCA 突变阴性的巴西人群进行了易感 CNV 筛查, 结果发现 26 个罕见 CNV 位点与乳腺癌发病相关, 其中包括唾液酸转移酶 ST6GALNAC5。研究表明, 该基因可能与乳腺癌的脑转移有关。9p13 基因组区的 KIA1797 及 MIR491 的缺失基因也与乳腺癌风险相关, 既往的研究还表明了该位点在结直肠癌发病中的作用。

4.2 肺癌

Yang 等^[59]在中国地区 4 个不同人群的 2 942 个肺癌患者以及 3 074 个健康对照中, 对位于抑癌基因 WWOX 第五内含子位点处的 CNV-67048 拷贝数进行了检测, 结果发现, CNV-67048 拷贝数丢失的个体, 其患肺癌的可能性明显增加, 且 CNV-67048 的丢失可致外显子的缺失。与正常的两拷贝的肺组织相比, 拷贝数丢失的样本其 WWOX 表达水平降低, WWOX 是一种抑癌基因, 该基因的低表达可促进多种肿瘤的发生。Gruber 等^[66]对 398 个肺癌及 697 个对照的研究表明, UGT2B17 的拷贝数丢失可导致女性患肺腺癌可能性增加, 与男性的肺癌易感性无明显关系。广州医科大学吕嘉春研究团队在对三个病例 - 对照组, 共计 2 332 肺癌患者及 2 457 健康对照的关联分析中发现, 丝裂原激活蛋白激酶激活的蛋白激酶 2 (MAPKAPK2) 启动子区的 CNV-30450 的拷贝数扩增可导致人群患肺癌的几率增加, 尤其是当 MAPKAPK2 的拷贝数达到 4 个时, 不仅患病几率大大增加, 且患者预后更差。研究还发现, 生殖系中的拷贝数与组织样本中的拷贝数存在明显正关联, 表明组织中的拷贝数变异可能来源于生殖系中的变异, 为解释组织中基因组变异提供了重要的证据。此外, 与两拷贝个体的组织样本相比, 4 拷贝样本中 MAPKAPK2 的表达量更高, 表明生殖系拷贝数的扩增可能通过改变组织中的拷贝数最终改变组织中的表达水平, 促进肿瘤的发生。Li 等^[67]通过对 EAGLE 和 PLCO 数据库分析发现, 位于 8q23.3、13q21.1、18q22.1 的拷贝数扩增, 以及 5q35.2 的拷贝数丢失与肺癌的易感性相关, 其中的某些位点在肺癌组织中也出现明显的拷贝数异常, 如 8q23.3 区域。这也在一定程度上表明了生殖系与实体组织变异之间的关系。Hu 等^[68]发现, 位

于 6q13 上一段基因荒漠区的 CNVR2966.1 的拷贝数丢失可增加中国人群肺癌风险, 值得注意的是, 该片段的丢失与胰腺癌的易感风险也相关, 表明两者在发病机制上的某些共同因素。

肺癌是一类遗传 - 环境因素共同作用的疾病, 以下两项研究进一步证明了该观点。Lee 等^[69]发现, 17p13.1 区 KCTD11 基因丢失的吸烟人群罹患肺癌的风险明显增加, KCTD11 是一类抑癌基因, 可负调控 Hedgehog 通路。在一项基于 7 880 个对象的大规模研究中, Yang 等^[70]发现, 胆碱能受体 (cholinergic receptor nicotinic $\alpha 7$, CHRNA7) 上的 CNV-3956 可显著增加肺癌风险, 并提示预后不良。该位点拷贝数扩增可导致 CHRNA7 的表达水平增加, 且该拷贝数扩增与吸烟的联合作用可导致肺癌风险的大幅增加。CHRNA7 可被尼古丁激活, 通过 Akt 及 NF- κ B 依赖的信号途径诱导凋亡抵抗, 进而促进肿瘤生长。

4.3 前列腺癌

Liu 等^[71]通过全基因组范围 SNP 扫描的方式, 发现染色体 2p24.3 上一段长约 6 kb 的缺失可提高前列腺癌的患病几率, 该区域内无基因座, 附近 100 kb 范围内只存在 AY928977 和 FAM84A 两个基因。既往的研究表明, FAM84A 在前列腺癌组织中高表达, 可促进细胞的迁移, 研究者推测该 CNV 区域可能对 FAM84A 的表达起到远距负向调控作用, 其丢失使这种抑制作用减弱, FAM84A 的表达则升高, 促进前列腺癌的发生。Jin 等^[72]在该基础上, 进而通过全基因组关联分析对 CNV 位点与前列腺癌的侵袭风险进行了分析, 结果发现, 20p13 区的一段长 32.3 kb 的片段 CNP2454 可能与前列腺癌的侵袭性相关, 该片段位于 SIRPB1 基因内部, 其丢失可能改变 SIRPB1 的可变剪切, 产生不同的异构体。Brueau 等^[73]发现, 在加勒比地区非洲人群中, GSTT1 和 GSTM1 的拷贝数扩增与男性前列腺癌的发病相关。Laitinen 等^[74]发现酪氨酸蛋白激酶受体 EPHA3 编码区内一段内含子的丢失增加芬兰人群前列腺癌患病风险。

4.4 肝癌

Clifford 等^[75]利用 SNP 芯片技术筛查到六个染色体区域的拷贝数变异与肝癌的发生相关, 其中 1p36.33 的 CNV 区域不含任何基因, 但其与线粒体基因组具有同源性, 可能是线粒体基因组变异的指标。其他五个区域分别有 KNG1、C4orf29、LARP2、LARP2、ALDH7A1、PHAX、C5orf48、LMNB1、

SRPK2、PUS7、TMPO。其中大多数基因均已报道与乙肝病毒的致病及肝肿瘤的发生相关，如 SRPK2 基因可抑制乙肝病毒的体内复制，TMPO 基因编码产物调控 Rb 基因的核内定位，TMPO 基因功能敲出可干扰细胞周期进展。该研究最重要的发现在于免疫相关基因与肝癌易感性之间的关系。例如，KNG1 基因调控 T 细胞衰老，其拷贝数丢失对于人群的保护作用明显。除此之外，T 细胞受体基因 TRA 与 TRG 的拷贝数丢失在健康对照中明显更多，说明这两个基因的丢失对人群具有保护作用，该结果表明，对于人类来说，拷贝数变异起到的未必是不良的影响，也可能导致良性的结果。同时，该研究也证实了免疫调控异常在肝癌发生中的作用。

4.5 肾癌

Park 等^[76]通过对 SNP 芯片数据分析发现，8 个拷贝数丢失（例如 6q22.2 的 SLC35F1、4q24 区 SLC39A8 的拷贝数丢失等，以及 G 蛋白偶联受体 GPR133）及 3 个拷贝数扩增（包括 3p11.2 上的 EPHA3 以及 17q12 的 ERBB2）与肾癌易感性相关，其机制尚不明确。

4.6 结直肠癌

Yang 等^[77]通过 GWAS 及定量验证分析，发现在癌症患者中出现了 12p12.3 区内一段长约 94.4 kb 的缺失，该缺失在健康对照中却并未发现，且在数据库中并未被收录，是一个新发现的罕见 CNV。RERGL (RAS like estrogen regulated growth inhibitor like) 基因是该区域所包含的唯一基因。对该区域两侧序列分析时发现非同源末端连接事件，表明这一段序列的丢失可能是在 DNA 修复过程中完成的。Bi 等^[78]对 E3 泛素连接酶家族的五个成员进行拷贝数检测后发现，家族成员 MDM2 的拷贝数扩增增加结直肠癌的患病风险，SKP2 的拷贝数丢失降低结直肠癌风险，SKP2 与鱼类、水果食用习惯的相互作用可明显降低结直肠癌的患病风险，既往的报道已显示这两类生活习惯可降低结直肠癌的风险，而该研究则进一步证明了遗传-环境相互作用在结直肠癌发病中的重要作用；Weren 等^[79]报道抑癌基因 FOCAD (focadhesin) 拷贝数丢失与家族性结直肠多发息肉及结直肠癌的发病相关，该基因主要表达于结肠憩室上皮细胞，是结直肠癌常见的起始部位，结直肠隐窝上皮的异常增殖被视为癌变的起始步骤。Fernandez-Rozadilla 等^[80]通过全基因组关联分析的方式发现 2p22.2、11p11、15q11.3 及 16p11.2 区的拷贝数变异与结直肠癌的发病相关。其中，

11p11 区的可重复性最好，该区域内编码嗅觉受体基因 OR4A15，此前尚无证据表明，OR4A15 与结直肠癌的发病直接相关，但相关研究表明，11p11 的拷贝数变异与肥胖相关。既往的研究表明，15q11.3 和 16p11.2 的变异与体重指数明显相关，肥胖与结直肠癌的关系已得到流行病学证实。因此，本研究为结直肠癌的发病机制提供了新的证据。

在结直肠癌发病的高危因素中，家族性多发息肉扮演着重要的角色。Masson 等^[81]在 56 例病例及 40 例对照进行 SNP 芯片筛查发现，18p11.32 区的拷贝数丢失与结直肠多发息肉的发病风险相关，该区域内有长链非编码 RNA TCONS_00026231 的编码位点，长链非编码 RNA 在肿瘤发生中的作用已得到广泛证实，但长链非编码 RNA 拷贝数变异与肿瘤发病风险的报道极为罕见，有待于进一步研究揭示其功能与机制。

4.7 食管癌

Sun 等^[82]研究发现，ATP 结合蛋白家族成员 C4 (ATP-binding cassette, sub-family C, ABCC4) 的拷贝数扩增与食管癌易感风险密切相关，功能试验结果表明，ABCC4 表达抑制可降低食管癌细胞的增殖与侵袭能力，其促进食管癌发生的机制尚不明确。Hu 等^[83]通过候选位点策略，发现了 UGT2B28 基因的拷贝数丢失导致中国西南人群食管癌易感风险增加，推测其机制可能与非甾体类药物代谢异常相关。

4.8 其他

Fidalgo 等^[84]报道 ANGPT1、IDH1、PDE5A、HIST1H1B 和 GCNT2 等罕见 CNV 位点与黑色素瘤的发病风险相关。Conde 等^[84]报道，11q25 染色体上 B3GAT1 上游 84 kb 区域的一段长约 340 kb 的扩增与弥漫性 B 细胞淋巴瘤的发生风险相关，长链非编码 RNA LOC283177 是该区域唯一的基因座，13q14、11q22-23、14q32 及 22q11.22 区的丢失则与慢性淋巴细胞白血病发病相关，ST13P4、IR16-1、MIR15 及 ADAM6 基因定位于这几个染色体区域，这些基因与肿瘤生长等途径密切相关。

线粒体 DNA (mitochondrial DNA, mtDNA) 独立于染色体 DNA 之外，其拷贝数在不同组织、不同人群中的数量差异较大。由于线粒体在能量代谢及细胞凋亡中的重要作用，其拷贝数的异常可能通过改变细胞正常代谢途径等方式促进肿瘤发生。mtDNA 拷贝数变异与肿瘤发生的关系较为复杂，其高拷贝及低拷贝均可能导致肿瘤的发生，在肿瘤

中的作用尚无定论, 其与种族、环境因素的交互作用在肿瘤发生中的机制值得深入探讨。目前认为, 线粒体 DNA 的拷贝数丢失将导致正常细胞的有氧代谢受到抑制, 而无氧代谢增强, 引发毒性代谢产物增加进而提高人群罹患肿瘤的风险。而线粒体拷贝数增加是机体处于氧化应激状态的标志, 细胞长期的氧化应激状态是肿瘤发生的前兆。

5 展望

CNV 是遗传变异的一种重要形式, 由于其覆盖范围广, 突变频率高, 可造成人群中巨大的遗传差异, 进而导致不同的性状。然而, 受制于目前的研究手段, 人们对于 CNV 的认识尚有很大不足, 随着新的高通量芯片技术的发展, 我们将获得更加完善的全基因组 CNV 图谱, 必将有力地推动我们从更深入的层面认识 CNV 与人类疾病发病机制之间的关系, 为疾病的早期干预提供有力的理论依据。

[参 考 文 献]

- [1] Schwartz S. Clinical utility of single nucleotide polymorphism arrays. *Clin Lab Med*, 2011, 31: 581-94
- [2] Johnson AD. Single-nucleotide polymorphism bioinformatics: a comprehensive review of resources. *Circ Cardiovasc Genet*, 2009, 2: 530-6
- [3] Bacolod MD, Schemmann GS, Giardina SF, et al. Emerging paradigms in cancer genetics: some important findings from high-density single nucleotide polymorphism array studies. *Cancer Res*, 2009, 69: 723-7
- [4] Dutt A, Beroukhi R. Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol*, 2007, 19: 43-9
- [5] Zarrei M, MacDonald JR, Merico D, et al. A copy number variation map of the human genome. *Nat Rev Genet*, 2015, 16: 172-8
- [6] Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*, 2006, 444: 444-54
- [7] Qi J, Chen Y, Copenhaver GP, et al. Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *Proc Natl Acad Sci USA*, 2014, 111: 10007-12
- [8] Arlt MF, Rajendran S, Birkeland SR, et al. *De novo* CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. *PLoS Genet*, 2012, 8: e1002981
- [9] Lieber MR, Lu H, Gu J, et al. Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system. *Cell Res*, 2008, 18: 125-33
- [10] Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 2007, 131: 1235-47
- [11] Goodier JL, Kazazian HH Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, 2008, 135: 23-35
- [12] Feng Q, Moran JV, Kazazian HH Jr, et al. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, 1996, 87: 905-16
- [13] Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet*, 2001, 35: 501-38
- [14] Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 1992, 258: 818-21
- [15] Zahir FR, Marra MA. Use of affymetrix arrays in the diagnosis of gene copy-number variation. *Curr Protoc Hum Genet*, 2015, 85: 1311-3
- [16] Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 2008, 452: 872-6
- [17] Sellner LN, Taylor GR. MLPA and MAPH: new techniques for detection of gene deletions. *Hum Mutat*, 2004, 23: 413-9
- [18] Armour JA, Palla R, Zeeuwen PL, et al. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res*, 2007, 35: e19
- [19] Liu Z, Wang J, Liu S, et al. Copy number variation of *GATA4* and *NKX2-5* in Chinese fetuses with congenital heart disease. *Pediatr Int*, 2015, 57: 234-8
- [20] Rooryck C, Morice-Picard F, Lasseaux E, et al. High resolution mapping of *OCA2* intragenic rearrangements and identification of a founder effect associated with a deletion in Polish albino patients. *Hum Genet*, 2011, 129: 199-208
- [21] Green PM, Bagnall RD, Waseem NH, et al. Haemophilia A mutations in the UK: results of screening one-third of the population. *Br J Haematol*, 2008, 143: 115-8
- [22] Nakajima T, Kaur G, Mehra N, et al. HIV-1/AIDS susceptibility and copy number variation in *CCL3L1*, a gene encoding a natural ligand for HIV-1 co-receptor CCR5. *Cytogenet Genome Res*, 2008, 123: 156-60
- [23] Liu S, Yao L, Ding D, et al. *CCL3L1* copy number variation and susceptibility to HIV-1 infection: a meta-analysis. *PLoS One*, 2010, 5: e15778
- [24] Grunhage F, Nattermann J, Gressner OA, et al. Lower copy numbers of the chemokine *CCL3L1* gene in patients with chronic hepatitis C. *J Hepatol*, 2010, 52: 153-9
- [25] Mamtani M, Mummidi S, Ramsuran V, et al. Influence of variations in *CCL3L1* and *CCR5* on tuberculosis in a northwestern Colombian population. *J Infect Dis*, 2011, 203: 1590-94
- [26] Ben Kilani MS, Achour Y, Perea J, et al. Characterization of copy number variants for *CCL3L1* gene in rheumatoid arthritis for French trio families and Tunisian cases and controls. *Clin Rheumatol*, 2016, 35: 1917-22
- [27] Burns JC, Shimizu C, Gonzalez E, et al. Genetic variations in the receptor-ligand pair *CCR5* and *CCL3L1* are important determinants of susceptibility to Kawasaki

- disease. *J Infect Dis*, 2005, 192: 344-9
- [28] Lee H, Bae S, Choi BW, et al. Copy number variation of *CCL3L1* influences asthma risk by modulating IL-10 expression. *Clin Chim Acta*, 2011, 412: 2100-4
- [29] Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human β -defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet*, 2006, 79: 439-48
- [30] Yuan J, Zhao D, Wu L, et al. *FCGR3B* copy number loss rather than gain is a risk factor for systemic lupus erythematosus and lupus nephritis: a meta-analysis. *Int J Rheum Dis*, 2015, 18: 392-7
- [31] Chen JY, Wang CM, Chang SW, et al. Association of *FCGR3A* and *FCGR3B* copy number variations with systemic lupus erythematosus and rheumatoid arthritis in Taiwanese patients. *Arthritis Rheumatol*, 2014, 66: 3113-21
- [32] Wu Y, Zhang Z, Tao L, et al. A high copy number of *FCGR3B* is associated with psoriasis vulgaris in Han Chinese. *Dermatology*, 2014, 229: 70-5
- [33] Nossent JC, Rischmueller M, Lester S. Low copy number of the Fc- γ receptor 3B gene *FCGR3B* is a risk factor for primary Sjogren's syndrome. *J Rheumatol*, 2012, 39: 2142-7
- [34] McKinney C, Broen JC, Vonk MC, et al. Evidence that deletion at *FCGR3B* is a risk factor for systemic sclerosis. *Genes Immun*, 2012, 13: 458-60
- [35] Wu J, Li Y, Guan W, et al. *FCGR3A* and *FCGR3B* copy number variations are risk factors for sarcoidosis. *Hum Genet*, 2016, 135: 715-25
- [36] Pereira KM, Faria AG, Liphaut BL, et al. Low *C4*, *C4A* and *C4B* gene copy numbers are stronger risk factors for juvenile-onset than for adult-onset systemic lupus erythematosus. *Rheumatology : Oxford*, 2016, 55: 869-3
- [37] Cleynen I, Konings P, Robberecht C, et al. Genome-wide copy number variation scan identifies complement component C4 as novel susceptibility gene for crohn's disease. *Inflamm Bowel Dis*, 2016, 22: 505-5
- [38] Lintner KE, Patwardhan A, Rider LG, et al. Gene copy-number variations (CNVs) of complement *C4* and *C4A* deficiency in genetic risk and pathogenesis of juvenile dermatomyositis. *Ann Rheum Dis*, 2016, 75: 1599-606
- [39] Hou S, Qi J, Liao D, et al. High *C4* gene copy numbers protects against Vogt-Koyanagi-Harada syndrome in Chinese Han. *Br J Ophthalmol*, 2014, 98: 1733-7
- [40] Liu YH, Wan L, Chang CT, et al. Association between copy number variation of complement component *C4* and Graves' disease. *J Biomed Sci*, 2011, 18: 71
- [41] Lopez-Alvarez MR, Jiang W, Jones DC, et al. *LILRA6* copy number variation correlates with susceptibility to atopic dermatitis. *Immunogenetics*, 2016, 68: 73-7
- [42] Qin J, Jia M, Wang L, et al. Association study of *SHANK3* gene polymorphisms with autism in Chinese Han population. *BMC Med Genet*, 2009, 10: 61
- [43] Liu Y, Du Y, Liu W, et al. Lack of association between *NLGN3*, *NLGN4*, *SHANK2* and *SHANK3* gene variants and autism spectrum disorder in a Chinese population. *PLoS ONE*, 2013, 8: e56639
- [44] Silberberg G, Darvasi A, Pinkas-Kramarski R, et al. The involvement of ErbB4 with schizophrenia: association and expression studies. *Am J Med Genet B: Neuropsychiatr Genet*, 2006, 141B: 142-8
- [45] Deng X, Shibata H, Takeuchi N, et al. Association study of polymorphisms in the glutamate transporter genes *SLC1A1*, *SLC1A3*, and *SLC1A6* with schizophrenia. *Am J Med Genet B: Neuropsychiatr Genet*, 2007, 144B: 271-8
- [46] Moura KC, Junior MC, de Rosso AL, et al. Exon dosage variations in Brazilian patients with Parkinson's disease: analysis of *SNCA*, *PARKIN*, *PINK1* and *DJ-1* genes. *Dis Markers*, 2012, 32: 173-8
- [47] Tan EK, Refai FS, Siddique M, et al. Clinically reported heterozygous mutations in the PINK1 kinase domain exert a gene dosage effect. *Hum Mutat*, 2009, 30: 1551-7
- [48] Nuytemans K, Theuns J, Cruts M, et al. Genetic etiology of Parkinson disease associated with mutations in the *SNCA*, *PARK2*, *PINK1*, *PARK7*, and *LRRK2* genes: a mutation update. *Hum Mutat*, 2010, 31: 763-80
- [49] Bushman DM, Kaeser GE, Siddoway B, et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *elife*, 2015, 4: 1-26
- [50] Sinkus ML, Graw S, Freedman R, et al. The human *CHRNA7* and *CHRFAM7A* genes: a review of the genetics, regulation, and function. *Neuropharmacology*, 2015, 96: 274-88
- [51] Thureson AC, Van Buggenhout G, Sheth F, et al. Whole-gene duplication of *SCN2A* and *SCN3A* is associated with neonatal seizures and a normal intellectual development. *Clin Genet*, 2017, 91: 106-10
- [52] Moey C, Hinze SJ, Brueton L, et al. Xp11.2 microduplications including *IQSEC2*, *TSPYL2* and *KDM5C* genes in patients with neurodevelopmental disorders. *Eur J Hum Genet*, 2016, 24: 373-80
- [53] Carr CW, Zimmerman HH, Martin CL, et al. 5q14.3 neurocutaneous syndrome: a novel contiguous gene syndrome caused by simultaneous deletion of *RASAI* and *MEF2C*. *Am J Med Genet A*, 2011, 155A: 1640-5
- [54] Freitag CM, Lempp T, Nguyen TT, et al. The role of *ASTN2* variants in childhood and adult ADHD, comorbid disorders and associated personality traits. *J Neural Transm : Vienna*, 2016, 123: 849-58
- [55] Jiang XY, Chang FH, Bai TY, et al. Susceptibility of lung cancer with polymorphisms of *CYP1A1*, *GSTM1*, *GSTM3*, *GSTT1* and *GSTP1* genotypes in the population of Inner Mongolia region. *Asian Pac J Cancer Prev*, 2014, 15: 5207-4
- [56] Pan ZJ, Huang WJ, Zou ZH, et al. The *GSTT1* null genotype contributes to increased risk of prostate cancer in Asians: a meta-analysis. *Asian Pac J Cancer Prev*, 2012, 13: 2635-8
- [57] Liu B, Yang L, Huang B, et al. A functional copy-number variation in *MAPKAPK2* predicts risk and prognosis of lung cancer. *Am J Hum Genet*, 2012, 91: 384-90
- [58] Yang L, Liu B, Qiu F, et al. The effect of functional *MAP-*

- KAPK2* copy number variation CNV-30450 on elevating nasopharyngeal carcinoma risk is modulated by EBV infection. *Carcinogenesis*, 2014, 35: 46-52
- [59] Yang L, Liu B, Huang B, et al. A functional copy number variation in the *WWOX* gene is associated with lung cancer risk in Chinese. *Hum Mol Genet*, 2013, 22: 1886-94
- [60] Yu K, Fan J, Ding X, et al. Association study of a functional copy number variation in the *WWOX* gene with risk of gliomas among Chinese people. *Int J Cancer*, 2014, 135: 1687-91
- [61] Wyszynski A, Hong CC, Lam K, et al. An intergenic risk locus containing an enhancer deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5 expression. *Hum Mol Genet*, 2016, 25: 3863-76
- [62] Long J, Delahanty RJ, Li G, et al. A common deletion in the *APOBEC3* genes and breast cancer risk. *J Natl Cancer Inst*, 2013, 105: 573-9
- [63] Xuan D, Li G, Cai Q, et al. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis*, 2013, 34: 2240-3
- [64] Frank B, Bermejo JL, Hemminki K, et al. Copy number variant in the candidate tumor suppressor gene *MTUS1* and familial breast cancer risk. *Carcinogenesis*, 2007, 28: 1442-5
- [65] Krepischi AC, Achatz MI, Santos EM, et al. Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res*, 2012, 14: R24
- [66] Gruber M, Le T, Filipits M, et al. UDP-glucuronosyltransferase 2B17 genotype and the risk of lung cancer among Austrian Caucasians. *Cancer Epidemiol*, 2013, 37: 625-8
- [67] Li X, Chen X, Hu G, et al. Combined analysis with copy number variation identifies risk loci in lung cancer. *Biomed Res Int*, 2014, 2014: 469103
- [68] Hu XY, Bai XM, Qiao X, et al. Copy number variation at 6q13 is associated with lung cancer risk in a Han Chinese population. *Exp Lung Res*, 2013, 39: 427-33
- [69] Lee M, Lee Y, Cho HJ, et al. Copy number variations of chromosome 17p13.1 might be linked to high risk of lung cancer in heavy smokers. *Mol Biol Rep*, 2011, 38: 5211-7
- [70] Yang L, Lu X, Qiu F, et al. Duplicated copy of *CHRNA7* increases risk and worsens prognosis of COPD and lung cancer. *Eur J Hum Genet*, 2015, 23: 1019-24
- [71] Liu W, Sun J, Li G, et al. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res*, 2009, 69: 2176-9
- [72] Jin G, Sun J, Liu W, et al. Genome-wide copy-number variation analysis identifies common genetic variants at 20p13 associated with aggressiveness of prostate cancer. *Carcinogenesis*, 2011, 32: 1057-62
- [73] Brureau L, Emeville E, Broquere C, et al. Copy number variation of *GSTT1* and *GSTM1* and the risk of prostate cancer in a Caribbean population of African descent. *PLoS One*, 2014, 9: e107275
- [74] Laitinen VH, Akinrinade O, Rantapero T, et al. Germline copy number variation analysis in Finnish families with hereditary prostate cancer. *Prostate*, 2016, 76: 316-24
- [75] Clifford RJ, Zhang J, Meerzaman DM, et al. Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma. *Hepatology*, 2010, 52: 2034-43
- [76] Park RW, Kim TM, Kasif S, et al. Identification of rare germline copy number variations over-represented in five human cancer types. *Mol Cancer*, 2015, 14: 25
- [77] Yang R, Chen B, Pflutze K, et al. Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. *Carcinogenesis*, 2014, 35: 315-23
- [78] Bi H, Tian T, Zhu L, et al. Copy number variation of E3 ubiquitin ligase genes in peripheral blood leukocyte and colorectal cancer. *Sci Rep*, 2016, 6: 29869
- [79] Weren RD, Venkatachalam R, Cazier JB, et al. Germline deletions in the tumour suppressor gene *FOCAD* are associated with polyposis and colorectal cancer development. *J Pathol*, 2015, 236: 155-64
- [80] Fernandez-Rozadilla C, Cazier JB, Tomlinson I, et al. A genome-wide association study on copy-number variation identifies a 11q11 loss as a candidate susceptibility variant for colorectal cancer. *Hum Genet*, 2014, 133: 525-34
- [81] Masson AL, Talseth-Palmer BA, Evans TJ, et al. Copy number variants associated with 18p11.32, DCC and the promoter 1B region of APC in colorectal polyposis patients. *Meta Gene*, 2016, 7: 95-104
- [82] Sun Y, Shi N, Lu H, et al. *ABCC4* copy number variation is associated with susceptibility to esophageal squamous cell carcinoma. *Carcinogenesis*, 2014, 35: 1941-50
- [83] Hu LW, Wu YY, Guan XY, et al. Germline copy number loss of *UGT2B28* and gain of *PLEC* contribute to increased human esophageal squamous cell carcinoma risk in Southwest China. *Am J Cancer Res*, 2015, 5: 3056-71
- [84] Fidalgo F, Rodrigues TC, Silva AG, et al. Role of rare germline copy number variation in melanoma-prone patients. *Future Oncol*, 2016, 12: 1345-57
- [85] Conde L, Riby J, Zhang J, et al. Copy number variation analysis on a non-Hodgkin lymphoma case-control study identifies an 11q25 duplication associated with diffuse large B-cell lymphoma. *PLoS One*, 2014, 9: e105382