

CNV及其与癌症间的关系

张慧莲 陈鲤群*

(福州大学生物科学与工程学院, 福州 350108)

摘要 拷贝数变异(copy number variation, CNV)是指基因组发生1 Kb 以上的DNA片段的增添、缺失或重排。癌症的早期诊断与治疗一直是本世纪亟待解决的难题。CNV的相关研究为人类健康和疾病的治疗提供了宝贵的见解。目前, CNV的研究引发了人们对疾病的新探索, 尤其体现在与遗传物质息息相关的疾病(例如, 癌症)的病因研究、临床诊断、新药研发和治疗。该文主要综述了CNV的研究方法、形成机制以及其与癌症间的联系, 以期推动癌症相关研究的发展。

关键词 CNV; 癌症; 诊断与治疗

CNV and Its Relation with Cancer

ZHANG Huilian, CHEN Liqun*

(College of Biological Science and Engineering, Fuzhou University, Fuzhou 350108, China)

Abstract CNV (copy number variation) refers to the addition, deletion or rearrangement of more than 1 Kb of DNA fragments in the genome. The early diagnosis and treatment of cancer have always been difficult problems to be solved in this century. CNV-related researches provide valuable insights for human health and disease treatment. At present, CNV research has triggered a new understanding of diseases, especially in the etiology research, clinical diagnosis, new drug development and treatment of diseases (such as cancer) that are closely related to genetic material. This paper reviews the research methods, causes of CNV, and the relationship between CNV and cancer in order to promote the development of cancer-related research.

Keywords CNV; cancer; diagnosis and treatment

CNV是指一类具有拷贝数变化的结构变异, 其通常涉及长度大于1 Kb的DNA片段。其中, CNV的结构变异主要包括基因组的获得(重复或插入转座)、缺失或复杂的重排^[1]。CNV存在于正常人的基因组中^[2]。平均每个个体的基因组中至少有1 000个CNV, 约有400万个碱基对^[3-4]。此外, 不同个体间的CNV约有1%的差异^[5]。研究表明, CNV可能为人类多样性作出重大贡献^[6]。但是, CNV也被确定为许多人类疾病的常见病因^[7]。

癌症(cancer)是一种以不正常的细胞生长和扩

散为特征的疾病。如果其扩散不受控制, 可能会导致患者死亡。癌症的发生是一个多因子、多步骤的复杂过程。癌症发生的根本原因是体细胞遗传物质的改变。研究表明, 癌症与体细胞拷贝数紧密相关^[8], 细胞可通过CNV获得无限增殖、迁移、侵袭等能力。CNV在许多癌症的发生中起重要作用, 例如卵巢癌(ovarian cancer)^[9]、肝细胞癌(hepatocellular carcinoma, HCC)^[10]和结直肠癌(colorectal cancer, CRC)^[11]。本文综述CNV与癌症之间的关系, 着重描述CNV在多种癌症中的作用, 例如CRC、乳腺癌(breast cancer)、

收稿日期: 2019-08-30 接受日期: 2020-02-17

国家自然科学基金(批准号: 31500616)资助的课题

*通讯作者。Tel: 18106973597, E-mail: lqchen@fzu.edu.cn

Received: August 30, 2019 Accepted: February 17, 2020

This work was supported by the National Natural Science Foundation of China (Grant No.31500616)

*Corresponding author. Tel: +86-18106973597, E-mail: lqchen@fzu.edu.cn

URL: <http://www.cjcb.org/arts.asp?id=5276>

胃癌(gastric cancer)和肺癌(lung cancer)等,希望从中找到疾病的早期诊断方法并为拟定最优治疗方案提供线索。

1 CNV的研究方法

1985年, HARRISON等^[12]在微观水平下观察到染色体结构的变化。20世纪90年代, 比较基因组杂交技术(comparative genomic hybridization, CGH)出现并被应用于检测CNV的研究^[13]。随后染色体微阵列(chromosomal microarray, CMA)也被证明是一种检测结构变异, 特别是杂合性丢失(loss of heterozygosity, LOH)和CNV的方法, 并被广泛应用于临床诊断^[14]。但是直到2004年, 研究人员才在微观水平观察到人类基因组中的CNV^[15]。

自从在微观水平下观察到染色体, 研究人员一直没有停止对其结构、序列的探究。最初, 人们主要通过细胞遗传学方法进行CNV检测。这些检测方法可以观察到染色体的物理结构。荧光原位杂交(fluorescent in situ hybridization, FISH)利用荧光探针定位基因位置。CGH通过荧光团可视化检测CNV, 然后比较染色体的长度。但是CGH只能检测不平衡的染色体异常。平衡的染色体异常, 例如染色体倒位、易位, 不会导致拷贝数变化, 则无法使用CGH技术检测。上述技术的主要缺点是基因组分辨率相对较低, 只能检测大片段重复, 例如全基因重复。

随着生物技术的发展, 研究人员已经研发出许多具有高分辨率的CNV检测技术。细菌人工染色体(bacterial artificial chromosome, BAC)阵列可以检测重排热点中的CNV。生物学家利用单核苷酸多态性(single nucleotide polymorphisms, SNP)的连锁不平衡概率判断重组热点区域, 以识别CNV。目前, 研究人员常用的CNV检测技术是染色体微阵列和新一代测序(next generation sequencing, NGS)。

染色体微阵列是一种能够在同一时间定量分析基因组情况的检测方法, 具有快速、精确、低成本等特点。实验中先将预先合成好的核酸探针按特定顺序排列于玻片载体上制成芯片, 再从样本中分离基因组DNA并加以标记, 随后基因组DNA与芯片进行反应, 最后检测芯片信号并进行结果分析。染色体微阵列已被推荐为针对神经发育障碍(neuro-developmental disorders, NDD)或多发性先天性畸形(multiple congenital anomalies, MCA)患者的一级临

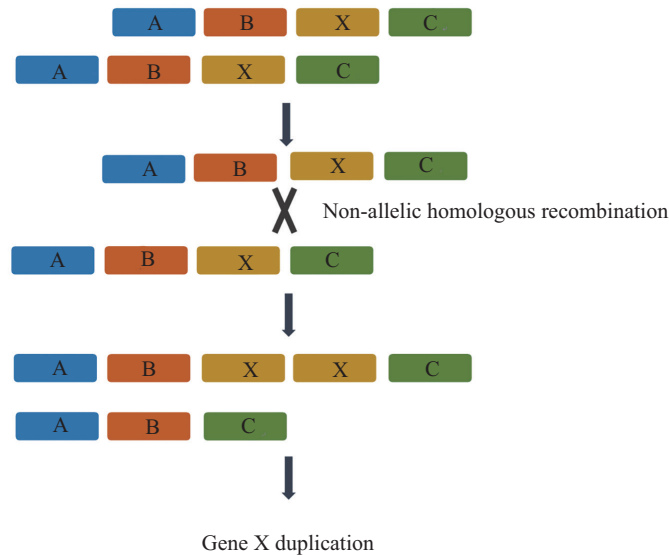
床诊断方法^[14], 其分辨率远高于传统的细胞遗传学方法, 可筛选基因组不平衡, 并可以更好地确定失衡基因的大小^[16]。在实验研究中, 染色体微阵列已经在很大程度上取代了荧光原位杂交分析, 用于鉴定基因组中的致病性CNV^[17]。

新一代测序又被称为高通量测序(high throughput sequencing), 是一种通过大量且同时进行的测序反应, 从而获得大量的测序数据的方法。实验中将测序读数映射到参考基因组, 然后从映射数据中得到DNA结构变异的数据, 即通过信息学方法从NGS数据中检测CNV^[4]。从NGS数据检测CNV的4种主要方法是: RP(read-pair)、SR(split read)、RD(read depth)和AS(assembly)^[1]。RP方法可以从NGS数据中检测中等大小片段的插入和缺失, 但对小片段的插入或缺失不敏感^[18]。SR方法可以精确检测缺失片段开始和结束的位置, 但无法识别大型结构变异^[1]。RD方法可以检测到CNV的确切数量, 并且可以识别大型CNV, 而RP和SR很难检测到它们^[19]。但是此方法对小型CNV(小于1 Kb)的灵敏度较低^[20]。AS方法不需要参考基因组, 但只能检测到纯合的结构变异^[21]。此外, RP、SR和AS方法不适用于检测重复区域中的变异^[18,20,22]。可见, 上述每种方法都有其自身的优缺点。为了结合每种方法的优点, 可以采用CA(combined approach)方法, 通过逐步合并来自两个或多个方法的数据, 从而可以克服只使用一种方法的局限性^[22]。目前常用的NGS平台有Illumina HiSeq/MiSeq、Life Technologies SOLiD和Roche 454^[4]。新一代测序技术已成为检测癌症细胞中CNV的重要手段^[23-24]。新一代测序技术推动癌症研究的同时, 被广泛应用于临床癌症易感综合征的评价、发育迟缓和癌症预后诊断等, 在胃癌、肺癌和乳腺癌的研究及临床检测应用较为突出^[4,25-28]。

目前, 测序技术主要依赖于大基因组的短读取, 获得的数据根据重叠区域以形成更长的序列, 最终形成整个基因组序列。但是由于CNV是重复次数不定的区域, 因此这些重叠区域的出现给CNV检测带来巨大的难题。

2 CNV的形成机制

如今, 研究人员已经提出许多CNV形成的机制。但是目前仍没有非常明确的证据可以说明CNV形成的原因, 因此这些理论仍然是推测。不过, 存在



在同源染色体配对时, 等位基因尚未对准便发生重组, 即非等位基因同源重组, 则可能导致基因重复。其中, A、B、X、C代表基因。

When homologous chromosomes are paired, recombination occurs before the alleles are aligned, that is, non-allelic homologous recombination, may result in gene duplication. Among them, A, B, X, C represent genes.

图1 非等位基因同源重组图解

Fig.1 Diagrammatic representation of non-allelic homologous recombination

2种被大众认可的理论: “同源性”理论和“非同源性”理论。

“同源性”理论是基于同源重组的形成理论。在减数分裂过程中, 同源染色体配对, 非姐妹染色单体发生交叉互换, 此过程称为同源重组。同源重组主要包括3个阶段, 即前联会体阶段、联会体形成和Holliday结构的拆分。在同源染色体配对时, 倘若等位基因尚未对准便发生重组, 即非等位基因同源重组(图1), 则可能导致基因重复。

同源性配对是指高度相似(约97%)且长度适宜的DNA链配对。非同源配对是仅依赖于两条DNA链中少数碱基对的相似性进行配对。非同源末端连接是引发CNV的可能原因之一。DNA双链发生断裂时, 断裂末端不需要同源性即可连接起来以此修复破损DNA。修复过程中的末端连接可能会导致缺失或增添。因此, 通过此方法修复的DNA必然存在错误^[29]。上述内容即为CNV形成的“非同源性”理论内容之一。

3 CNV在癌症发生发展中的作用

癌症发生的根本原因是体细胞遗传物质的改变, 其中包括单碱基取代、易位和CNV等。几乎所有的癌症均存在染色体结构不稳定性和CNV^[30]。基因组分析表明, CNV通过影响基因表达量(即转录和

翻译水平)从而使细胞发生癌变。因此, 基因组不稳定性, 尤其是CNV, 是癌症的一个标志, 在临床诊断上具有巨大的潜力。文中出现的基因、蛋白的名称列于表1。

现代系统生物学研究主要包含2个部分, 以细胞、分子实验等为主的“湿”研究方法和以计算机模拟、理论分析为主的“干”研究方法。目前, 癌症中CNV的研究一般采用是“干”、“湿”实验相结合的研究方法。常见的研究方法包括样品处理或培养、样本制备(基因组提取、文库构建等)、高通量测序或组织芯片(tissue chip)、生物信息学方法进行数据分析, 其中常常会涉及染色体z值分析(chromosome arm-level z score analysis, CAZA)以及TCGA(The Cancer Genome Atlas, <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>)、GEO(Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>)等数据库的运用。此外, 研究人员也会采用原位杂交、细胞增殖、迁移和侵袭等方法对分析结果作进一步验证。由此可见, “湿”研究方法侧重于问题探究、数据验证, 而“干”研究方法侧重于实验数据的整理与分析。

3.1 消化系统相关癌症

结直肠癌是常见的癌症之一。研究表明, 结直肠癌组织中存在特异性CNV, 并集中分布于染色

表1 文中的基因和蛋白
Table 1 Genes and proteins in this review

蛋白 Protein	编码基因 Coding gene	蛋白 Protein	编码基因 Coding gene
Polycystin 1 like 2 (gene/pseudogene)	<i>PKD1L2</i>	Cyclin dependent kinase inhibitor 1B	<i>CDKN1B</i>
Solute carrier family 18 member A1	<i>SLC18A1</i>	Microtubule affinity regulating kinase 4	<i>MARK4</i>
PLAG1 like zinc finger 2	<i>PLAGL2</i>	FAT atypical cadherin 4	<i>FAT4</i>
Erb-b2 receptor tyrosine kinase 2	<i>ERBB2</i>	Phosphodiesterase 5A	<i>PDE5A</i>
Small nucleolar RNA H/ACA box 18-like 5	<i>SNORA18L5</i>	Purkinje cell protein 4	<i>PCP4</i>
MYC proto-oncogene, bHLH transcription factor	<i>MYC</i>	Macrophage scavenger receptor 1	<i>MSR1</i>
Tumor protein p53	<i>TP53</i>	Exostosin glycosyltransferase 1	<i>EXT1</i>
RB transcriptional corepressor 1	<i>RB1</i>	Diaphanous related formin 3	<i>DIAPH3</i>
RPTOR independent companion of MTOR complex 2	<i>RICTOR</i>	MIA SH3 domain ER export factor 2	<i>MIA2</i>
GRB2 associated binding protein 2	<i>GAB2</i>	NADH: ubiquinone oxidoreductase subunit C2	<i>NDUFC2</i>
BRCA1 DNA repair associated	<i>BRCA1</i>	Eukaryotic translation initiation factor 4E binding protein 1	<i>EIF4EBP1</i>
BRCA2 DNA repair associated	<i>BRCA2</i>	Prostate stem cell antigen	<i>PSCA</i>
Phosphatase and tensin homolog	<i>PTEN</i>	MYB proto-oncogene like 2	<i>MYBL2</i>
Checkpoint kinase 2	<i>CHEK2</i>	ATP binding cassette subfamily C member 4	<i>ABCC4</i>
RAD51 paralog C	<i>RAD51C</i>	UDP glucuronosyltransferase family 2 member B28	<i>UGT2B28</i>
Damage specific DNA binding protein 1	<i>DDB1</i>	Plectin	<i>PLEC</i>
ERCC excision repair 1, endonuclease non-catalytic subunit	<i>ERCC1</i>	Cyclin D1	<i>CCND1</i>
ERCC excision repair 2, TFIIH core complex helicase subunit	<i>ERCC2</i>	Cyclin dependent kinase inhibitor 2A	<i>CDKN2A</i>
Pre-mRNA processing factor 19	<i>PRPF19</i>	Branched chain amino acid transaminase 1	<i>BCAT1</i>

数据来自HGNC数据库(<https://www.genenames.org/>)。

Data from HUGO Gene Nomenclature Committee database (<https://www.genenames.org/>).

体臂^[31]。这些特异性CNV与结直肠癌发生相关。PLAGL2作为细胞核内转录因子,其在结肠直肠中表达上调从而导致细胞癌变。其中,CNV是结直肠癌中PLAGL2过表达的原因之一^[32]。*PKD1L2*或*SLC18A1*基因的缺失均提高了结直肠癌发生的风险^[33-34]。由此可见,CNV是导致结直肠癌发生的重要原因。XU等^[30]通过结直肠癌患者血浆样品($n=32$)和健康人血浆样品($n=38$)的全基因组测序确定了结直肠癌中CNV的情况,并在此基础上设计了SVM(support vector machine)分类器。研究证明,用SVM分类器检测早期结直肠癌的灵敏度可达100%。此外,MOLPAR-IA等^[35]的研究提出了基于循环肿瘤DNA(circulating tumor DNA, ctDNA)衍生的CNV进行结直肠癌诊断的可行性。DRULINER等^[36]也提出可以将18q的遗传畸变用作散发性结直肠癌检测及风险评估的临床标志物。因此,我们认为,利用结直肠癌相关CNV设计癌症早期诊断方案是可行的。

胃癌是常见癌症类型之一,威胁着人类的健

康。大多数胃癌是胃腺癌(gastric adenocarcinoma, GAC),根据Lauren分类法进一步被分为肠型和弥漫型2种类型^[37]。研究显示,17号染色体臂存在较多CNV^[38],并且在染色体17q12-21上观察到拷贝数增加,其中包含*ERBB2*,这是一种驱动胃腺癌的致病性CNV^[26]。扩增区段中的许多基因可能协同促进癌症发生,但是大多数基因的影响仍然是未知的^[26],有待进一步研究。

肝细胞癌是死亡率极高的癌症^[39],研究其成因在肝细胞癌的诊断和治疗方面具有重要意义。JIANG等^[40]使用NGS和CAZA分析检测肝细胞癌血浆中1号和8号染色体的CNV,发现84.4%的患者血浆中至少有1个CNV,而对照组几乎没有CNV。在CNV的全基因组相关的研究中,CAO等^[41]发现,在染色体15q13.3发生拷贝数的重复,导致*SNORA18L5*过表达。过表达的*SNORA18L5*通过一系列反应致使细胞周期停滞,促进小鼠中肝细胞癌细胞增殖和肿瘤生长。ZHOU等^[42]在肝癌组织研究中发现,

CNV的存在导致138个基因的表达量发生变化(93个基因过表达, 45个基因表达下调)。这些基因参与翻译、细胞分裂、DNA损伤反应的正调节和代谢过程等与癌症形成密切相关的生化反应。因此, 肝细胞癌的发生与CNV密切相关, 可以此为突破口探索肝细胞癌的临床标志物。

食道癌(esophageal cancer)是常见的癌症之一, 可以分为2种主要类型: 食道鳞状细胞癌(esophageal squamous-cell carcinoma, ESCC)和食管腺癌(esophageal adenocarcinoma, EAC)。细胞周期调节因子*MYBL2*基因拷贝数增加, 并且*MYBL2*的过表达增加了食道鳞状细胞癌细胞系的增殖^[43]。因此, CNV与食道鳞状细胞癌发生存在间接关系。*SUN*等^[44]在染色体13q32.1处发现了CNV区域, 并且证实该区域中的*ABCC4*基因表达增加与食道鳞状细胞癌的形成显著相关。此外, *UGT2B28*的拷贝数丢失和*PLEC*的拷贝数增加提高了患食道鳞状细胞癌病率^[45]。上述研究结果表明, 食道癌的发生与CNV存在一定联系, 研究人员需要进一步探索以确定关键的致病性CNV。

3.2 呼吸系统相关癌症

肺癌根据组织病理学可以分为2种类型: 小细胞肺癌(small-cell lung cancer, SCLC)和非小细胞肺癌(non-small-cell lung carcinoma, NSCLC)。小细胞肺癌约占所有肺癌的15%, 是与吸烟最密切相关的癌症类型^[46]。*DU*等^[47]对小细胞肺癌患者细胞中的游离DNA(cell free DNA, cfDNA)进行拷贝数分析发现许多CNV, 包括*MYC*的扩增、*TP53*的缺失、*RBI*的缺失等。这些CNV涉及多个信号通路, 如Notch1信号传导途径等。因此*DU*等^[47]认为, 研究小细胞肺癌患者cfDNA中的CNV对于开发检测和治疗疾病方法具有重要的临床价值。*RICTOR*基因的拷贝数变化导致小细胞肺癌细胞中的*RICTOR*蛋白表达改变, 这与小细胞肺癌的发生密切相关。*SAKRE*等^[46]发现上述现象后, 预测可使用靶向药物雷帕霉素(rapamycin)治疗小细胞肺癌。可见, 生物学家可以通过研究致病性CNV, 预测治疗效果, 并在此基础上进行药物研发及临床试验。这为癌症药物治疗指明方向。

非小细胞肺癌是致死率较高的癌症之一, 主要有肺腺癌(lung adenocarcinoma, LUAD)和肺鳞状细胞癌(lung squamous cell carcinoma, LUSC)2种类型^[48]。此外, 在CNV方面肺腺癌和肺鳞状细胞癌

超过47%基因(11 000个基因)具有明显差异^[48]。由于3q26.2位点的扩增导致原始致癌环状RNA(proto-oncogenic circular RNA, circPRKCI)在肺腺癌组织中过表达, 促进了细胞的增殖和癌症的发生^[49]。*PARK*等^[50]研究癌症基因组CNV数据, 推测*GAB2*基因的扩增可能在非吸烟者的肺鳞状细胞癌发生中起重要作用, 并且其可作为非吸烟者肺鳞状细胞癌的潜在生物标志物。之后, *CHEN*等^[51]开发了一种监测转移性肺鳞状细胞癌患者的cfDNA拷贝数变异的方法, 此方法可为癌症早期诊断与后期治疗提供重要数据。

鼻咽癌(nasopharynx cancer)是在鼻咽黏膜中发生的癌症。*LOW*等^[52]认为, CNV与鼻咽癌易感性紧密关联, 推测CNV在鼻咽癌发生中有重要作用。在鼻咽癌小鼠模型中检测到细胞周期相关基因的拷贝数变异, 包括*CCND1*扩增和*CDKN2A*缺失。因此, *Hsu*等^[53]提出, 细胞周期蛋白依赖性激酶抑制剂帕博西林(palbociclib)可以作为鼻咽癌的新型候选药物。

3.3 泌尿生殖系统与妇科癌症

卵巢癌是最致命的妇科癌症, 全世界每年死亡人数超过150 000^[54]。研究人员对卵巢癌相关基因进行CNV检测, 鉴定出12种不同的致病性突变, 包括*BRCA1*、*BRCA2*、*TP53*、*PTEN*、*CHEK2*和*RAD51C*突变^[55], 也发现多个染色体区域发生拷贝数的增添或缺失, 确定了7个可以特异性靶向晚期卵巢癌的治疗和预后的基因, 包括*DDB1*、*ERCC1*、*ERCC2*、*PRPF19*、*BCAT1*、*CDKN1B*和*MARK4*^[54]。但是, 上述结论是否可行, 仍需临床验证。

前列腺癌(prostate cancer)是男性中最常被诊断出的癌症, 在美国每年约有220 800人确诊, 27 540人死亡^[56]。在前列腺癌中, DNA拷贝数的缺失和增加改变了癌症相关基因的表达, 导致癌症发生。研究人员通过比较前列腺癌样本与正常样本之间的基因表达确定至少93个癌症相关CNV^[57-58]。在这些CNV区域中, *HAN*等^[57]发现3个表达量显著降低的基因, 即*FAT4*、*PDE5A*和*PCP4*, 并且这3个基因表达的蛋白富集于特定信号通路, 如钙调蛋白信号传导途径。此外, *ROSE*等^[59]发现, 位于19号染色体的*MSR1*基因发生拷贝数变异, 从而致使前列腺癌发病率升高。

乳腺癌是女性常见的癌症之一^[60]。在临床上, 乳腺癌是基于3个基本乳腺癌的生物标志物——雌激素(estrogen rector, ER)、孕酮受体(progesterone re-

表2 CNV在癌症中的研究情况
Table 2 CNV research in cancer

癌症类型 Types of cancer	癌症中CNV的特点 Characteristics of CNV in cancer	典型CNV Typical CNV	参考文献 Reference
Colorectal cancer	CNV is concentrated in the chromosome arm	<i>PLAGL2, SLC18A1</i>	[31-34]
Gastric cancer	There are more CNV on chromosome 17 arm	<i>ERBB2</i>	[26,38]
Hepatocellular carcinoma	138 CNV-related genes were found in liver cancer tissue	<i>SNORA18L5</i>	[41-42]
Esophageal cancer		<i>MYBL2, ABCC4, UGT2B28, PLEC</i>	[43-45]
Small-cell lung cancer	Copy number analysis of cfDNA in small cell lung cancer patients reveals many CNV	<i>RICTOR</i>	[46-47]
Non-small-cell lung carcinoma		<i>GAB2</i>	[50]
Nasopharynx cancer		<i>CCND1, CDKN2A</i>	[52-53]
Ovarian cancer	Addition or deletion of copy number was found in multiple chromosomal regions	<i>BRCA1, BRCA2, TP53, PTEN, CHEK2, RAD51C</i>	[54-55]
Prostate cancer	At least 93 cancer-related CNVs have been identified	<i>FAT4, PDE5A, PCP4, MSRI</i>	[57-59]
Breast cancer	There are gene additions and deletions in breast cancer tissues The most frequently added gene is EXT1 The most frequently deleted gene is DIAPH3	<i>BRCA1, MSRI</i>	[25,59,62]

ceptor, PR)和人表皮生长因子受体2(human epidermal growth factor receptor 2, HER2)分类的^[61]。FATIMA等^[62]在59个乳腺癌细胞系和650个乳腺癌患者中进行了MammaPrint™基因的CNV分析,结果显示,乳腺癌细胞系和乳腺癌患者中均存在基因增添和缺失,其中最常增添的基因是*EXT1*,最常被删除的基因是*DIAPH3*。RIZZA等^[25]认为,*BRCA1*基因重排可能会导致*BRCA1*蛋白关键功能域的丧失,从而引发乳腺癌。ROSE等^[59]发现,*MSRI*作为调控基因表达的分子开关,由于其发生拷贝数变异而富集于基因组的调节区域可能是乳腺癌发生的原因之一。乳腺癌易感基因*BRCA1*重排检测和MammaPrint™基因的CNV谱可分别用于常规诊断测试与乳腺癌患者的预后预测^[62-63]。*MIA2*、*NDUFC2*、*EIF4EBP1*和*PSCA*基因同样可以应用于乳腺癌的早期诊断^[64]。此外,这些致病性CNV也可以作为疾病的药物靶标。因此,乳腺癌相关CNV的研究在癌症发生和患者的诊断、治疗及预后中起关键作用。

在临床试验中,SCHUTZ等^[65]发现,血清中cfDNA中存在的染色体不稳定性如CNV,可作为检测前列腺癌的生物标志物,准确率高达83%。XU等^[30]研究发现,CNV作为诊断早期CRC的标志物时检测灵敏度可达100%,并且在多种癌症CNV检测中均发现致病性CNV的存在,尤其体现在血浆CNV的检测。

上述结果表明,血浆中的CNV可能是具有高灵敏度、特异性的癌症生物标志物^[30]。此外,通过致病性CNV研究,可以准确预测靶向药物。例如,SAKRE等^[46]发现,*RICTOR*基因的拷贝数变化与小细胞肺癌发生密切相关后,预测可使用雷帕霉素治疗疾病。HSU等^[53]在鼻咽癌小鼠模型中检测到细胞周期相关基因的拷贝数变异,以此提出帕博西林可以作为鼻咽癌的新型候选药物。因此,对于多种癌症中CNV的研究,重在发现癌症相关的关键CNV,以此为癌症早期诊断、治疗(靶向药物)及预后提供方向(表2)。

4 结语与展望

CNV是大片段DNA(大于1 Kb)的缺失或重复,可能含有0个、1个甚至多个基因。CNV虽然为人类遗传多样性做出贡献,但是也与神经发育迟缓(neurodevelopmental delay)、自闭症谱系障碍(Autism spectrum disorders, ASD)等多种精神疾病有关。PINTO等^[66]分析2 446个存在自闭症谱系障碍病史的家庭,发现最常见的缺失区域是染色体16p11.2和2p16.3,最常见的增添区域是15q11-13。在精神分裂症患者中进行的全基因组研究发现染色体15q13.3和2p16.3的缺失以及16p11.2的重复^[3]。MEFFORD等^[67]使用全基因组阵列比较基因杂交检测517例混合型癫痫患者的拷贝数变化发现,大多数(69%)患者

存在基因缺失, 并且发现了5例患者各自在染色体15q11.2、15q13.3和16p13.11处有缺失。研究结果显示, 这些精神疾病中有许多相同的致病性CNV, 并且这类精神疾病的发生不能简单地归因于CNV内的单个驱动基因, 而是多因素作用的结果^[68]。此外, 精神分裂症仍然无法治愈, 只能控制其症状, 并且目前的治疗药物会对患者产生明显的副作用甚至出现不同的疗效^[69]。因此, 精神疾病治疗的研究迫在眉睫。而致病性CNV的研究直接暗示了精神疾病中的特定基因, 这为新药的研发提供了新的方向。

大量数据表明, CNV与癌症的发生发展息息相关。几乎所有癌症在染色体某些区域都存在拷贝数增添或缺失。越来越多的癌症相关CNV被发现, 但是其引发疾病的作用机制仍然需要探索。此外, 基于CNV精确定位致病基因的研究方案的实施仍然充满困难。因为大多数最常见的复发性CNV均非常大(大于500 Kb), 并且包含多个基因^[3]。不少致病性CNV中的部分基因的作用未知, 仍待研究。因此, 我们需要更先进的技术手段来研究CNV。癌症的早期诊断难和晚期转移性强导致其治疗难度大大增加。研究人员检测癌症患者基因组中的CNV, 通过数据比对找出致病基因, 将其作为诊断或预后的生物标志物, 开展新药研发。因此, 深入研究癌症的相关CNV对于其诊断、治疗及预后等具有重要的意义。

参考文献 (References)

- [1] PIROOZNIYA M, GOES F S, ZANDI P P. Whole-genome CNV analysis: advances in computational approaches [J]. *Front Genet*, 2015, 6: 138.
- [2] SEBAT J, LAKSHMI B, TROGE J, et al. Large-scale copy number polymorphism in the human genome [J]. *Science*, 2004, 305(5683): 525-8.
- [3] MALHOTRA D, SEBAT J. CNVs: harbingers of a rare variant revolution in psychiatric genetics [J]. *Cell*, 2012, 148(6): 1223-41.
- [4] ABEL H J, DUNCAVAGE E J. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches [J]. *Cancer Genet*, 2013, 206(12): 432-40.
- [5] PANG A W, MACDONALD J R, PINTO D, et al. Towards a comprehensive structural variation map of an individual human genome [J]. *Genome Biol*, 2010, 11(5): R52.
- [6] VILLELA D, SUEMOTO C K, LEITE R, et al. Increased DNA copy number variation mosaicism in elderly human brain [J]. *Neural Plast*, 2018, 2018: 2406170.
- [7] ZHAO X, LIU X, ZHANG A, et al. The correlation of copy number variations with longevity in a genome-wide association study of Han Chinese [J]. *Aging (Albany NY)*, 2018, 10(6): 1206-22.
- [8] SERVANT N, VAROQUAUX N, HEARD E, et al. Effective normalization for copy number variation in Hi-C data [J]. *BMC Bioinformatics*, 2018, 19(1): 313.
- [9] NAKABAYASHI M, KAWASHIMA A, YASUHARA R, et al. Massively parallel sequencing of cell-free DNA in plasma for detecting gynaecological tumour-associated copy number alteration [J]. *Sci Rep*, 2018, 8(1): 11205.
- [10] XU H, ZHU X, XU Z, et al. Non-invasive analysis of genomic copy number variation in patients with hepatocellular carcinoma by next generation DNA sequencing [J]. *J Cancer*, 2015, 6(3): 247-53.
- [11] HORPAOPAN S, SPIER I, ZINK A M, et al. Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel causative candidate genes for colorectal adenomatous polyposis [J]. *Int J Cancer*, 2015, 136(6): E578-89.
- [12] HARRISON C J, JACK E M, ALLEN T D, et al. Investigation of human chromosome polymorphisms by scanning electron microscopy [J]. *J Med Genet*, 1985, 22(1): 16-23.
- [13] WANG J, CHEN L, ZHOU C, et al. Identification of copy number variations among fetuses with ultrasound soft markers using next-generation sequencing [J]. *Sci Rep*, 2018, 8(1): 8134.
- [14] UDDIN M, THIRUVAHINDRAPURAM B, WALKER S, et al. A high-resolution copy-number variation resource for clinical and population genetics [J]. *Genet Med*, 2015, 17(9): 747-52.
- [15] RIGGS E R, LEDBETTER D H, MARTIN C L. Genomic variation: lessons learned from whole-genome CNV analysis [J]. *Curr Genet Med Rep*, 2014, 2: 146-50.
- [16] HU G, FAN Y, WANG L, et al. Copy number variations in 119 Chinese children with idiopathic short stature identified by the custom genome-wide microarray [J]. *Mol Cytogenet*, 2016, 9: 16.
- [17] WANG J C, BOYAR F Z. Chromosomal microarray analysis as the first-tier test for the identification of pathogenic copy number variants in chromosome 9 pericentric regions and its challenge [J]. *Mol Cytogenet*, 2016, 9: 64.
- [18] MEDVEDEV P, STANCIU M, BRUDNO M. Computational methods for discovering structural variation with next-generation sequencing [J]. *Nat Methods*, 2009, 6(11 Suppl): S13-20.
- [19] YOON S T, XUAN Z Y, MAKAROV V, et al. Sensitive and accurate detection of copy number variants using read depth of coverage [J]. *Genome research*, 2009, 19(9): 1586-92.
- [20] BELLOS E, JOHNSON M R, COIN L J. cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data [J]. *Genome Biol*, 2012, 13(12): R120.
- [21] XI R, LEE S, PARK P J. A survey of copy-number variation detection tools based on high-throughput sequencing data [J]. *Curr Protoc Hum Genet*, 2012, Chapter 7: Unit7. 19.
- [22] ZHAO M, WANG Q G, WANG Q, et al. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives [J]. *BMC Bioinformatics*, 2013, 14(11): S1.
- [23] REN T, SUO J, LIU S, et al. Using low-coverage whole genome sequencing technique to analyze the chromosomal copy number alterations in the exfoliative cells of cervical cancer [J]. *J*

- Gynecol Oncol, 2018, 29(5): e78.
- [24] KADER T, GOODE D L, WONG S Q, et al. Copy number analysis by low coverage whole genome sequencing using ultra low-input DNA from formalin-fixed paraffin embedded tumor tissue [J]. *Genome Med*, 2016, 8(1): 121.
- [25] RIZZA R, HACKMANN K, PARIS I, et al. Novel BRCA1 large genomic rearrangements in Italian Breast/ovarian cancer patients [J]. *Mol Diagn Ther*, 2019, 23(1): 121-6.
- [26] ZHANG J, QIU W, LIU H, et al. Genomic alterations in gastric cancers discovered via whole-exome sequencing [J]. *BMC Cancer*, 2018, 18(1): 1270.
- [27] PENG H, LU L, ZHOU Z, et al. CNV detection from circulating tumor DNA in late stage non-small cell lung cancer patients [J]. *Genes*, 2019, 10(11): 926.
- [28] LIAO Y, MA Z, ZHANG Y, et al. Targeted deep sequencing from multiple sources demonstrates increased NOTCH1 alterations in lung cancer patient plasma [J]. *Cancer Med*, 2019, 8(12): 5673-86.
- [29] ARLT M F, WILSON T E, GLOVER T W. Replication stress and mechanisms of CNV formation [J]. *Curr Opin Genet Dev*, 2012, 22(3): 204-10.
- [30] XU J F, KANG Q, MA X Y, et al. A novel method to detect early colorectal cancer based on chromosome copy number variation in plasma [J]. *Cell Physiol Biochem*, 2018, 45(4): 1444-54.
- [31] BIGAGLI E, DE FILIPPO C, CASTAGNINI C, et al. DNA copy number alterations, gene expression changes and disease-free survival in patients with colorectal cancer: a 10 year follow-up [J]. *Cell Oncol (Dordr)*, 2016, 39(6): 545-58.
- [32] SU C, LI D, LI N, et al. Studying the mechanism of PLAGL2 overexpression and its carcinogenic characteristics based on 3'-untranslated region in colorectal cancer [J]. *Int J Oncol*, 2018, 52(5): 1479-90.
- [33] ZHANG D, LI Z, XU X, et al. Deletions at SLC18A1 increased the risk of CRC and lower SLC18A1 expression associated with poor CRC outcome [J]. *Carcinogenesis*, 2017, 38(11): 1057-62.
- [34] PARK C, KIM J I, HONG S N, et al. A copy number variation in PKD1L2 is associated with colorectal cancer predisposition in korean population [J]. *Int J Cancer*, 2017, 140(1): 86-94.
- [35] MOLPARIA B, OLIVEIRA G, WAGNER J L, et al. A feasibility study of colorectal cancer diagnosis via circulating tumor DNA derived CNV detection [J]. *PLoS One*, 2018, 13(5): e0196826.
- [36] DRULINER B R, RUAN X, SICOTTE H, et al. Early genetic aberrations in patients with sporadic colorectal cancer [J]. *Mol Carcinog*, 2018, 57(1): 114-24.
- [37] ZHANG J, LIU X, YU G, et al. UBE2C is a potential biomarker of intestinal-type gastric cancer with chromosomal instability [J]. *Front Pharmacol*, 2018, 9: 847.
- [38] TAMILZHALAGAN S, MUTHUSWAMI M, GANESAN K. Juxtaposed genes in 7q21-22 amplicon contribute for two major gastric cancer sub-types by mutual exclusive expression [J]. *Mol Carcinog*, 2017, 56(4): 1239-50.
- [39] LIAO H, LIAO M, XU L, et al. Integrative analysis of h-prune as a potential therapeutic target for hepatocellular carcinoma [J]. *EBioMedicine*, 2019, 41:310-9.
- [40] JIANG P, CHAN C W, CHAN K C, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients [J]. *Proc Natl Acad Sci USA*, 2015, 112(11): E1317-25.
- [41] CAO P, YANG A, WANG R, et al. Germline duplication of SNORA18L5 increases risk for HBV-related hepatocellular carcinoma by altering localization of ribosomal proteins and decreasing levels of p53 [J]. *Gastroenterology*, 2018, 155(2): 542-56.
- [42] ZHOU C, ZHANG W, CHEN W, et al. Integrated analysis of copy number variations and gene expression profiling in hepatocellular carcinoma [J]. *Scientific reports*, 2017, 7(1): 10570.
- [43] CHEN Y B, JIA W H. A comprehensive genomic characterization of esophageal squamous cell carcinoma: from prognostic analysis to in vivo assay [J]. *Chin J Cancer*, 2016, 35(1): 76.
- [44] SUN Y, SHI N, LU H, et al. ABCC4 copy number variation is associated with susceptibility to esophageal squamous cell carcinoma [J]. *Carcinogenesis*, 2014, 35(9): 1941-50.
- [45] HU L, WU Y, GUAN X, et al. Germline copy number loss of UGT2B28 and gain of PLEC contribute to increased human esophageal squamous cell carcinoma risk in Southwest China [J]. *Am J Cancer Res*, 2015, 5(10): 3056-71.
- [46] SAKRE N, WILDEY G, BEHTAJ M, et al. RICTOR amplification identifies a subgroup in small cell lung cancer and predicts response to drugs targeting mTOR [J]. *Oncotarget*, 2017, 8(4): 5992-6002.
- [47] DU M, THOMPSON J, FISHER H, et al. Genomic alterations of plasma cell-free DNAs in small cell lung cancer and their clinical relevance [J]. *Lung Cancer*, 2018, 120: 113-21.
- [48] QIU Z W, BI J H, GAZDAR A F, et al. Genome-wide copy number variation pattern analysis and a classification signature for non-small cell lung cancer [J]. *Genes Chromosomes Cancer*, 2017, 56(7): 559-69.
- [49] QIU M, XIA W, CHEN R, et al. The circular RNA circPRKCI promotes tumor growth in lung adenocarcinoma [J]. *Cancer Res*, 2018, 78(11): 2839-51.
- [50] PARK Y R, BAE S H, JI W, et al. GAB2 amplification in squamous cell lung cancer of non-smokers [J]. *J Korean Med Sci*, 2017, 32(11): 1784-91.
- [51] CHEN X, CHANG C W, SPOERKE J M, et al. Low-pass whole-genome sequencing of circulating cell-free DNA demonstrates dynamic changes in genomic copy number in a squamous lung cancer clinical cohort [J]. *Clin Cancer Res*, 2019, 25(7): 2254-63.
- [52] LOW J S Y, CHIN Y M, MUSHIRODA T, et al. A genome wide study of copy number variation associated with nasopharyngeal carcinoma in malaysian chinese identifies CNVs at 11q14.3 and 6p21.3 as candidate loci [J]. *PLoS One*, 2016, 11(1): e0145774.
- [53] HSU C L, LUI K W, CHI L M, et al. Integrated genomic analyses in PDX model reveal a cyclin-dependent kinase inhibitor Palbociclib as a novel candidate drug for nasopharyngeal carcinoma [J]. *J Exp Clin Cancer Res*, 2018, 37(1): 233.
- [54] LI L, BAI H, YANG J, et al. Genome-wide DNA copy number analysis in clonally expanded human ovarian cancer cells with distinct invasive/migratory capacities [J]. *Oncotarget*, 2017, 8(9): 15136-48.
- [55] PELTTARI L M, SHIMELIS H, TOIMINEN H, et al. Gene-panel testing of breast and ovarian cancer patients identifies a recurrent RAD51C duplication [J]. *Clin Genet*, 2018, 93(3): 595-602.
- [56] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2015 [J]. *CA Cancer J Clin*, 2015, 65(1): 5-29.
- [57] HAN Y, JIN X, LI H, et al. Microarray analysis of copy-number

- variations and gene expression profiles in prostate cancer [J]. *Medicine (Baltimore)*, 2017, 96(28): e7264.
- [58] WU Y, CHEN H, JIANG G, et al. Genome-wide association study (GWAS) of germline copy number variations (CNVs) reveal genetic risks of prostate cancer in Chinese population [J]. *J Cancer*, 2018, 9(5): 923-8.
- [59] ROSE A M, KRISHAN A, CHAKAROVA C F, et al. MSR1 repeats modulate gene expression and affect risk of breast and prostate cancer [J]. *Anna Oncol*, 2018, 29(5): 1292-303.
- [60] KUMARAN M, CASS C E, GRAHAM K, et al. Germline copy number variations are associated with breast cancer risk and prognosis [J]. *Sci Rep*, 2017, 7(1): 14621.
- [61] DAI X, XIANG L, LI T, et al. Cancer hallmarks, biomarkers and breast cancer molecular subtypes [J]. *J Cancer*, 2016, 7(10): 1281-94.
- [62] FATIMA A, TARIQ F, MALIK M F A, et al. Copy number profiling of mammprint genes reveals association with the prognosis of breast cancer patients [J]. *J Breast Cancer*, 2017, 20(3): 246-53.
- [63] CONCOLINO P, RIZZA R, MIGNONE F, et al. A comprehensive BRCA1/2 NGS pipeline for an immediate Copy Number Variation (CNV) detection in breast and ovarian cancer molecular diagnosis [J]. *Clin Chim Acta*, 2018, 480: 173-9.
- [64] BEHRING M, SHRESTHA S, MANNE U, et al. Integrated landscape of copy number variation and RNA expression associated with nodal metastasis in invasive ductal breast carcinoma [J]. *Oncotarget*, 2018, 9(96): 36836-48.
- [65] SCHUTZ E, AKBARI M R, BECK J, et al. Chromosomal instability in cell-free DNA is a serum biomarker for prostate cancer [J]. *Clin Chem*, 2015, 61(1): 239-48.
- [66] PINTO D, DELABY E, MERICO D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders [J]. *Am J Hum Genet*, 2014, 94(5): 677-94.
- [67] MEFFORD H C, MUHLE H, OSTERTAG P, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies [J]. *PLoS genetics*, 2010, 6(5): e1000962.
- [68] TAKUMI T, TAMADA K. CNV biology in neurodevelopmental disorders [J]. *Curr Opin Neurobiol*, 2018, 48: 183-92.
- [69] AVRAMOPOULOS D. Recent advances in the genetics of schizophrenia [J]. *Mol Neuropsychiatry*, 2018, 4(1): 35-51.