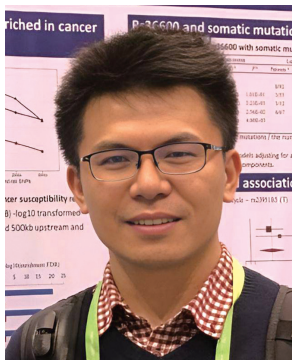




胡志斌, 二级教授, 博士生导师, 南京医科大学校长, 生殖医学与子代健康全国重点实验室主任, 中国医学科学院学部委员, 中国医师协会公共卫生医师分会副会长, 中华预防医学会流行病分会常务委员, 国家重点研发计划生育健康重点专项专家组成员。长期致力于出生队列研究及生殖障碍疾病、出生缺陷等复杂疾病的分子与遗传流行病学研究, 主持国家重点研发计划项目、基金委创新研究群体项目、国家杰出青年基金、国家自然科学基金重点项目等。在 *Nat Genet*、*Nat Med*、*Cancer Cell*、*Lancet Oncol* 等国际知名杂志上发表研究论文 200 余篇, 连续入选 Elsevier“化学、遗传与分子生物学”中国高被引学者(H指数 65, 总 SCI 他引超 18 000 次), 研究成果获国家自然科学基金二等奖、国家科学技术进步奖二等奖及多项省部级科技进步一等奖等。

<https://gwxy.njmu.edu.cn/2017/0718/c8947a98500/page.htm>



王铖, 副教授, 博士生导师, 入选江苏省“333工程”培养对象(第三层次)。主要从事基因组学、遗传学与人类健康的关系及相关生物信息学方法的研究, 主持国家自然科学基金优秀青年基金、面上、青年项目, 作为学术骨干参与科技部重点研发计划 2 项。近年来发表 70 余篇 SCI 论文, 其中包括以第一(含共同)或通讯作者身份在 *Nat Med*、*Cancer Cell*、*Lancet Oncol*、*Cell Res* 等高质量国际学术期刊上发表的 30 余篇论文, 获授权国家发明专利 3 项, 参与获得江苏省科学技术奖一等奖、中国抗癌协会科技奖一等奖。任中国抗癌协会肿瘤流行病学专业委员会委员、江苏省人类遗传资源管理专家委员会专家委员。

<https://gwxy.njmu.edu.cn/2024/0111/c8947a253626/page.htm>

人类新生突变的起源、特征及其对子代健康的影响

周俊^{1,2} 朱佳雯^{1,2} 王铖^{1,2,3*} 胡志斌^{1,2*}

(¹南京医科大学生殖医学与子代健康全国重点实验室, 南京 211166; ²南京医科大学公共卫生学院流行病学系, 南京 211166; ³南京医科大学生物医学工程与信息学院生物信息学系, 南京 211166)

摘要 除来自祖先的遗传变异外, 人类还携带了每一世代新产生的突变, 即新生突变。近年来, 基于“父-母-子代”核心家系设计的高通量测序研究初步解析了这些新生突变的起源和特征, 并且发现新生突变是多种疾病(例如神经发育障碍、先天性心脏病等)发生的重要原因。因此, 该文就人类新生突变的起源、特征以及对子代健康的影响进行综述, 以期对新生突变的进一步研究提供思路。

关键词 新生突变; 突变起源; 子代健康

收稿日期: 2023-12-31

接受日期: 2024-02-27

国家自然科学基金优秀青年科学基金项目(批准号: 82322032)、国家自然科学基金创新研究群体项目(批准号: 82221005)、国家重点研发计划(批准号: 2021YFC2700601)和江苏省杰出青年基金项目(批准号: BK20220050)资助的课题

*通信作者。Tel: 025-86869366, E-mail: cheng_wang29@njmu.edu.cn; Tel: 025-86868437, E-mail: zhibin_hu@njmu.edu.cn

Received: December 31, 2023

Accepted: February 27, 2024

This work was supported by the Science Fund Program for Distinguished Young Scholars of the National Natural Science Foundation of China (Grant No.82322032), the Science Fund Program for Creative Research Groups of the National Natural Science Foundation of China (Grant No.82221005), the National Key Research and Development Program of China (Grant No.2021YFC2700601), and the Outstanding Youth Foundation of Jiangsu Province (Grant No.BK20220050)

*Corresponding authors. Tel: +86-25-86869366, E-mail: cheng_wang29@njmu.edu.cn; Tel: +86-25-86868437, E-mail: zhibin_hu@njmu.edu.cn

The Origin and Characteristics of *De Novo* Mutations in Human and Its Impact on Offspring Health

ZHOU Jun^{1,2}, ZHU Jiawen^{1,2}, WANG Cheng^{1,2,3*}, HU Zhibin^{1,2*}

¹State Key Laboratory of Reproductive Medicine and Offspring Health, Nanjing Medical University, Nanjing 211166, China;

²Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing 211166, China; ³Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China)

Abstract In addition to genetic variations inheriting from ancestry, humans carry novel mutations that arise in each generation, known as *de novo* mutations. In recent years, high-throughput sequencing studies based on trios-design (“father-mother-offspring”) have preliminarily elucidated the origin and characteristics of *de novo* mutations. It has been discovered that *de novo* mutations play a significant role in various diseases, such as neurodevelopmental disorders and congenital heart diseases. Therefore, this article mainly reviews the origin, characteristics, and impact of human *de novo* mutations on offspring health, with a view to provide insight for the future study in this field.

Keywords *de novo* mutation; mutation origin; offspring health

人类新生突变(*de novo* mutations, DNMs), 特指在一代中新产生的突变。这些突变由于不受自然选择压力的影响, 可导致严重的疾病发生, 因而早前就被研究者们发现^[1]。但是受限于检测技术和极低的突变率, 这类突变一直缺乏系统研究。随着高通量测序技术在人群研究中的广泛应用, 研究者开始对“父-母-子代”典型核心家系的生物样本(通常是外周血样本)开展全基因组测序^[2], 人类DNMs的特征得以初见端倪。与经典的基因组突变相同, 广义的人类DNMs包括点突变、短缺失/插入和结构变异等, 但是由于二代测序的读长限制, 对于DNMs中大片段结构变异(>1 Kb)的研究仍然十分匮乏。因此, 本文中重点讨论的是狭义DNMs(即点突变)的起源、特征及其对子代健康的影响。

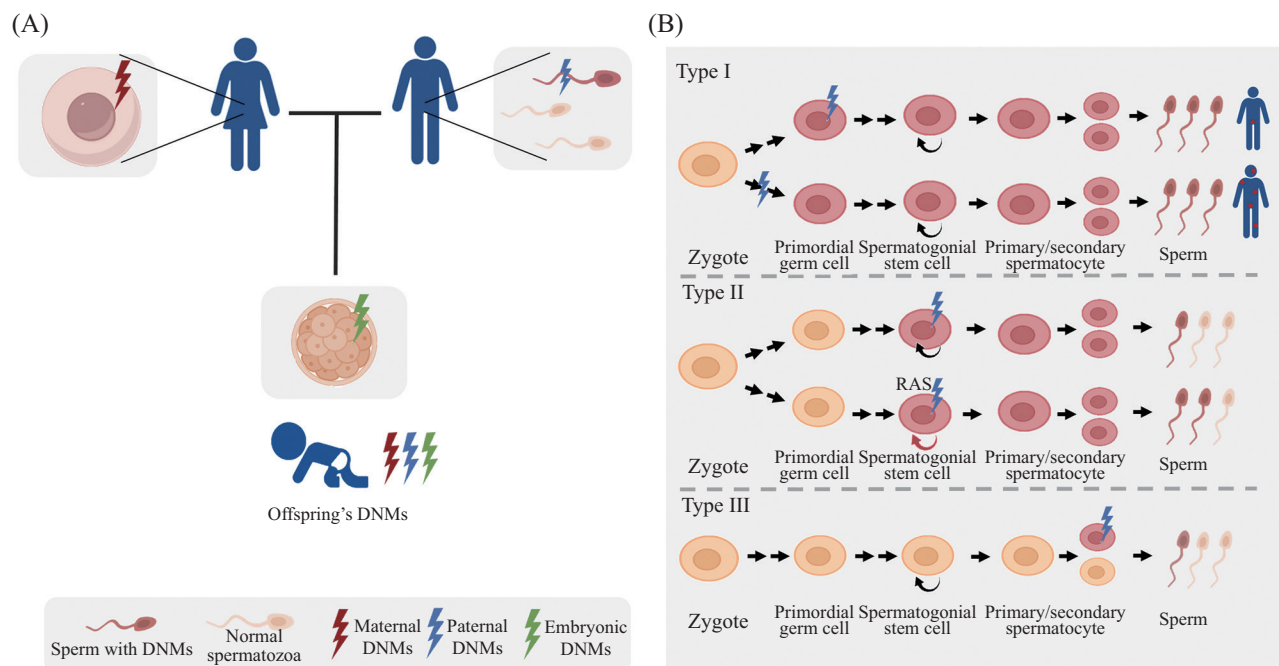
1 人类新生突变的起源、分类与影响因素

基于核心家系和全基因组测序检测的人类DNMs数量为38~77个^[3-6], 由于所获取的生物样本主要为外周血, 因此这些DNMs常具有不同来源。生殖细胞中产生的突变通常被认为是DNMs的来源, 但是随着研究的深入, 研究者们发现DNMs不仅仅来源于生殖细胞突变。按照DNMs来源, 可将其分为生殖系来源的突变和胚胎来源的突变(图1)。其中, 生殖系来源的DNMs可细分为父源突变和母源突变。胚胎来源的DNMs由于仅在部分细胞中发生,

因此也常被称为嵌合(mosaic)突变。目前新生突变检测多基于二代测序, 无法构建完整的父母源单倍型, 因此不能直接判断DNM的配子来源。研究者常利用核心家系信息和测序读段构建局部单倍型, 再根据单倍型上的遗传变异情况判断单倍型的来源, 最终进行DNMs的父母源推断。该方法的局限是仅能推断20%~30%的DNMs来源^[3,5,7], 且仅能追溯父母源突变, 因此目前对于人类DNMs来源的理解主要基于这些可推断突变的情况, 父母源突变的比例为2.8:1~3.9:1^[3-5,7]。近年来, 研究者发现依据“祖父母-父母-子代”三代家系特殊设计, 可以有效提升单倍型构建的长度和准确性, 同时家庭信息的整合也有助于追溯DNMs来源^[8-9]。根据一项包含33个三代家系研究的推断, 所有人类DNMs中, 约72%为父源突变, 18%为母源突变, 10%为胚胎源突变^[9]。

1.1 父源DNMs

父源DNMs, 特指起源于父亲生殖细胞的突变, 即精源突变, 是人类DNMs中最主要的突变来源。自2012年一项冰岛的家系研究首次发现父源DNMs与父亲的生育年龄高度相关以来^[4], 多项针对不同国家、不同种族的研究发现, 父亲生育年龄每增加一岁, 父源新生突变增加0.31~1.47个^[3,5,8,10], 这种现象被称为DNMs的父亲年龄效应。事实上, 早在1912年的流行病学研究中就观察到父亲生育年龄和子代疾病风险的关联^[11], 即父亲生育年龄的增加与多种子代疾



A: 子代新生突变的来源示意图。B: 精子中新发突变(父源新生突变)的起源及分类示意图。

A: the origin of offspring DNMs. B: the origin and classification of DNMs in sperm.

图1 新生突变的起源

Fig.1 The origin of *de novo* mutations

病的发病风险增加相关,并且父亲年龄的效应强于母亲年龄。DNMs父亲年龄效应的发现为解释上述现象提供了强有力的证据。该效应背后的生物学基础存在多种假设,其中最著名的是精原细胞的复制错误假说。精原细胞是产生精子的干细胞,位于生精上皮^[12-13]。精原细胞大量分裂增殖并自我更新,产生精母细胞,最终分化成精子细胞^[14]。在衰老的过程中,每个精原细胞的平均分裂次数在不断增加,而每次分裂的基因组复制过程都可能由于复制错误引入新的突变,因此精原细胞随着年龄增加积累突变被视为父亲年龄效应的一个最常见的假设,但是,也有研究认为DNA损伤的作用被低估了^[15-16]。

随着大规模测序研究的深入,人类DNMs的特征突变信号也逐渐被解析。突变信号特征模式首先在肿瘤体细胞突变中建立,包含30种常见的突变序列特征模式^[17]。通过特征模式比较发现,特征模式1和5的组合足以解释目前鉴定的人类DNMs突变特征谱^[6]。这两种模式出现在几乎所有肿瘤类型中,并以类似时钟的方式(即以恒定的速率)积累突变,因此可能代表衰老过程对基因组的影响^[18-19],这与DNMs的年龄效应相吻合。但是值得注意的是,DNMs的突变特征可随着父亲生育年龄发生变化:

高龄父亲的子代中可观察到显著增加的T>G替换和显著降低的非CpG位置上的C>T替换。这个现象提示,父源DNMs中可能存在一种与年龄有关的突变类型,这类突变具有独有的突变特征模式,随着父亲年龄的增加,其比例发生变化,从而影响了总的DNMs突变特征模式。

上述研究虽然对父源DNMs的特征进行了解读,但是这些突变并非在精子中被检测到的,而是通过单倍型推断出来的,因此并不能提供最直接的线索。为此,一些研究应用高深度且错误率优化的全基因组测序方案对精子中的突变进行检测,发现精子中存在丰富的嵌合现象。由于精子是单倍型细胞,因此这些精原突变嵌合的比例可以直接通过突变读段的比例来反映。结合疾病家系的传递模式和精源突变的嵌合比例,可将精子中的突变分为三类^[20-21]。

类型I: 发生在父亲的胚胎细胞或者原始生殖细胞(primordial germ cell, PGC)中的突变。根据这类突变发生的时间差异,又可以将其进一步分为Ia类和Ib类,其中Ia为仅在PGC中出现的突变,Ib则是除了在PGC中外,还可能在其他的胚层发育而来的组织或器官中出现的突变。这些突变会传递至多个精原细胞,最终以较高比例稳定地在精子中出现。因

此, 这些突变不存在父亲年龄效应, 在老年人和年轻人中的数量无显著差异^[21], 但是缺乏同一个体配对设计的研究, 无法确认其嵌合比例是否会发生变化, 缺乏证据证明其存在明确的选择优势。由于精子中携带突变的比例一般较高, 因此常可以通过Bulk测序发现该类突变。研究发现, 3%的DNMs可归类为此类突变, 这些突变的特征与其他精源突变存在差异, 表现为较高比例的CpG位置C>T替换和较低比例的T>C替换^[9,21]。值得注意的是, 三世代家系的研究发现这些突变可能以较高的比例传递给后代, 从而在同胞中被反复检测^[9], 这也意味着一旦携带的此类突变致病, 其子代当中再次出现该疾病的风险则会较高。该类型DNMs是临床上易于检测且具有较高临床意义的一类突变。

类型 II: 发生在精原细胞中的突变。这些突变被认为是随着父亲年龄积累的主要突变类型, 但是目前仍缺乏直接证据证明该结论。由于这些突变仅在个别精原细胞中出现, 因此其最终在精子中的携带率常较低, 需要依赖较高深度的测序进行鉴定。目前, 对此类突变占有所有精子突变的比例尚缺乏清晰的认识。在这些突变中, 绝大多数突变不会给精子细胞提供选择优势, 因此它们被归类为IIa类, 这类突变也是此前在子代中观察到的父亲年龄效应积累的主要突变; 但是也有一些突变可能对精原细胞本身的适应性产生影响, 从而形成选择优势导致克隆扩增, 携带此类突变的精子数量随着父亲衰老大量增加, 这类突变远超IIa类的父亲年龄效应, 可归类为IIb类。IIb类突变被选择的现象类似于癌前病变过程中特定突变被正向选择的过程。人类精子中的这类克隆扩增的现象被称为“自私的精子”模型^[22-23], 现阶段研究发现的此类突变主要通过影响表皮生长因子和RAS-MAPK通路的关键分子发挥作用, 从而形成明显的突变热点^[23-24]。这些热点突变的发现通常是因为其导致了子代疾病, 这些突变可在不同的家庭中导致相同的疾病发生^[25-27]。但是, 是否存在其他不导致疾病的IIb突变尚不清楚, 有待进一步研究。

类型 III: 发生在精子中的突变。这些突变理论上可发生在精子发生的终末期, 数量不随着父亲年龄的增长而积累, 在子代中检测出的DNMs中此类突变的比例可能较高, 尤其是年轻父亲生育的子代中。但是, 这类突变由于在精子中嵌合比例极低, 难

以被Bulk测序准确检测, 因此目前研究对此类突变的理解甚少。近期, 一项研究开发了一种单细胞长读段基因组测序方法及配套的相应分析流程, 其能够准确检测单个精子细胞基因组中的单核苷酸多态性和结构变异, 实现了整条染色体尺度的单倍型构建, 为此类突变的研究提供了契机^[28]。

1.2 母源DNMs

母源DNMs, 特指起源于母亲生殖细胞的突变, 即卵源突变, 母源DNMs的数量远低于父源DNMs^[3,5,7]。与父源DNMs相似, 母源DNMs也具有母亲年龄效应, 尽管突变积累的速率远低于父源DNMs: 据不同研究估算, 母亲生育年龄每增加一岁, 母源新生突变增加0.12~0.37个^[5,8,10]。一项研究发现, 母源DNMs在年老母亲中的增速可达到年轻母亲的2倍, 提示母源DNMs并非线性增加, 而是在年老母亲中加速。不同于精子发生的过程中细胞的大量分裂, 女性的生殖细胞被认为是相对静止的, 初级卵母细胞在出生前就停滞在减数分裂前期, 直到排卵^[29]。因此, 理论上母亲年龄效应的潜在机制可能与父亲年龄效应完全不同, 不涉及基因组复制过程, 但是其具体机制目前仍存在广泛的争议。越来越多的线索提示, DNA损伤可能是母亲年龄效应的主要原因。研究者们观察到母源DNMs存在热点基因组区域, 这些热点区域包括8号、9号和16号染色体上的数百万个碱基^[5,8,30]。这些区域中C>G突变的相对频率显著高于父源DNMs和其余区域的母源DNMs, 这些区域存在减数分裂基因交换的富集^[31]。同时, 随着母亲年龄的增加, 减数分裂过程中交叉干扰逃逸的几率也随之增加^[32]。这些现象都说明, 减数分裂重组事件可能在母源DNMs的产生过程中发挥重要作用。而近期一项研究系统探讨了减数分裂重组事件对DNMs的影响, 发现减数分裂过程中的双链断裂修复可能产生DNMs的热点区域, 是父母源DNMs的一个重要来源。不仅如此, 该机制产生的父源和母源DNMs比例与实际观察到的比例相近^[15]。尽管如此, 也不能完全排除其他可能的机制。例如, 有研究发现父源DNMs的数量与母亲年龄呈现正相关, 这提示母亲怀孕时的年龄和状态可能通过影响发育中的胚胎从而导致胚胎在分裂初期产生突变。这些突变常常难以被准确地与生殖系突变区分开, 在目前的父母源推断方法中, 也常常被错误地归类为父源或者母源突变。

1.3 胚胎源DNMs

在大部分基于全基因组测序的DNMs研究中, DNMs被假设来源于父亲或者母亲的生殖细胞。事实上, 在子代受精卵分裂早期也可能产生突变^[33-34]。尤其当突变发生在受精卵分裂的早期时, 该突变可能影响子代绝大多数的细胞, 与传递自祖先的遗传变异表现出类似的检测结果, 因此难以区分^[35]。随着突变发生的时间越晚, 其影响的细胞数量将逐渐降低, 在常规的二代测序数据中可与遗传变异区分开来, 其检测的灵敏度受到携带突变的细胞比例和二代测序深度的影响^[36]。

在大规模人群的基因组研究中, 外周血是最常采用的生物样本, 提取的主要是白细胞来源的DNA。但是, 在白细胞中存在一类特殊的突变, 这类突变可使造血干细胞获得竞争优势, 导致其分化产生的血细胞成为血液中的主要成分, 这类突变被称为克隆性造血(clonal hematopoiesis, CH)^[37]。在外周血的全基因组测序中, CH可表现为与胚胎源DNMs类似的测序结果, 从而难以区分。由于CH与血液系统的老化、炎症和肿瘤密切相关^[38], 越年轻的个体携带CH的可能性越小, 其生物样本更适合进行胚胎源DNMs的挖掘。需要注意的是, 造血细胞起源于胚胎的中胚层, 发生在其他胚层的胚胎源DNMs在外周血中常无法被检测到。因此, 采用多点采样的方式^[34,39-40], 在不同组织中进行测序, 寻找独有和共有的突变, 不仅有助于区分胚胎源DNMs和CH, 还可以发现更丰富的胚胎源DNMs。

1.4 DNMs的影响因素

父母生育年龄是DNMs最明确的影响因素(表1), 该关联证据来自冰岛的deCODE项目^[4], 结论在后续多项研究中被明确验证^[5,8,16,41]。但上述研究多为以基因组为出发点的研究, 即其重点关注的是基因组本身, 参与者的临床信息多缺失或者依赖医疗登记系统, 因此常缺乏用于探讨DNMs的影响因素的研究级数据。此外, 孕周与DNMs的关联也较为明确^[42], 但是机制尚不明确。

我们近期基于中国国家出生队列^[43-44]开展了一项小规模试点基因组学研究^[3]。依托队列完善的信息采集和随访系统, 我们得以全面系统地探索了DNMs的影响因素。除父母生育年龄外, 孕前的父母生活习惯也可以影响父母源DNMs的数量。例如, 怀孕前父亲饮酒行为可显著增加父源DNMs数量。此外, 患有

少弱畸形精子症的父亲通过辅助生殖技术生育的子代, 父源DNMs的数量显著增加, 而母亲的不孕因素则与母源DNMs无显著关联。此外, 我们发现辅助生殖子代DNMs数量显著高于自然妊娠子代, 其中父源和母源DNMs均在辅助生殖子代中显著增加。对于辅助生殖与DNMs的关联, 国外的研究结果存在差异, 一项较早的研究显示辅助生殖子代具有较多的DNMs^[10], 而近期的另一项小规模研究则未能发现这种关联^[45]。考虑到辅助生殖家庭具有较高的早产率^[46], 并且早产与DNMs增加有关^[42], 因此推测辅助生殖子代确实可能携带更多的DNMs。但是, 受限于样本量, 我们的研究未能全面解读导致辅助生殖子代DNMs增加的主要原因。由于父源和母源突变发生于辅助生殖干预之前, 因此其本身受到辅助生殖技术的影响可能有限, 但是尚不能排除辅助生殖技术的影响。一项在牛中开展的研究则提供了更多的线索, 该研究发现辅助生殖技术显著增加了DNMs, 并且由于这些突变发生在受精卵分裂的早期(一细胞和二细胞期), 因此可能无法将其与父母源DNMs区分开来^[47]。该结果提示辅助生殖有可能增加胚胎早期的突变, 但目前尚缺乏进一步的证据。

2 影响新生突变率的基因组特征

目前, 人群研究尚缺乏精良的设计和充足的样本量系统地揭示DNMs的影响因素。但是, 导致突变发生的细胞过程可在DNA序列中留下印记。因此, 探究不同基因组特征与突变率的关系可为突变机制的推断提供线索。

2.1 大规模染色质结构与复制时间

早在十余年前, 研究者们就发现了复制时间对突变率的影响, 即晚复制区域具有较高的突变率^[49]。游离核苷酸池的耗尽和S期后期频繁的复制叉停滞导致的复制准确率的下降可能是两者关联背后的机制^[50]。两者的关联还存在另一种解释: 大片的染色质结构与复制时间具有高度的相关性, 而这种染色质结构又与DNA损伤修复系统[如核苷酸切除修复(nucleotide excision repair, NER)和错配修复(mismatch repair, MMR)系统]的功能活性密切相关^[51-53]。目前, 已经有研究观察到TC到TA的突变在异染色质的晚复制区域显著富集^[54-55]。因此, 明确NER或者MMR异常所导致的突变的序列特征可以为上述机制的推断提供线索。

表1 新生突变的影响因素
Table 1 The influencing factor of *de novo* mutations

影响因素 Influence factor	突变类型 Mutation type	效应值(95%可信区间/ <i>P</i> 值) Effect value (95% CI/ <i>P</i> -value)	参考文献 Reference
Paternal conceived age	DNMs	1.51 (1.45, 1.57)	[8]
	DNMs	0.51 (0.29, 0.73)	[10]
	DNMs	1.44 (1.22, 1.77)	[9]
	DNMs	1.7 (<i>P</i> <0.05)	[48]
	DNMs	2.01 (1.68, 2.34)	[4]
	DNMs	1.02 (<i>P</i> <0.05)	[7]
	DNMs	2.87 (2.11, 3.64)	[6]
	DNMs	1.09 (0.91, 1.27)	[45]
	Paternal DNMs	1.47 (1.34, 1.59)	[8]
	Paternal DNMs	0.82 (0.52, 1.12)	[3]
	Paternal DNMs	0.31 (<i>P</i> <0.05)	[10]
	Paternal DNMs	0.91 (0.81, 1.02)	[5]
	Embryonic DNMs	Not applicable (<i>P</i> >0.05)	[9]
Paternal exercise habit	DNMs	-3.13 (-5.28, -0.98)	[3]
	Paternal DNMs	-2.20 (-4.37, -0.03)	[3]
Paternal drinking behavior	Paternal DNMs	1.88 (0.03, 3.73)	[3]
Oligoasthenoteratozoospermia	Paternal DNMs	2.91 (0.06, 5.76)	[3]
Maternal conceived age	DNMs	0.37 (0.32, 0.43)	[8]
	DNMs	0.64 (0.52, 0.77)	[10]
	DNMs	0.38 (0.21, 0.55)	[9]
	DNMs	0.34 (<i>P</i> <0.05)	[48]
	Paternal DNMs	0.33 (0.09, 0.57)	[3]
	Maternal DNMs	0.37 (0.30, 0.45)	[8]
	Maternal DNMs	0.12 (<i>P</i> <0.05)	[10]
	Maternal DNMs	0.23 (0.15, 0.34)	[5]
ART pregnancy	DNMs	4.59 (2.72, 6.47)	[3]
	Maternal DNMs	1.26 (0.07, 2.46)	[3]
	Paternal DNMs	3.32 (1.43, 5.22)	[3]
Preterm birth	DNMs	0.27 (<i>P</i> <0.05)	[42]

2.2 重组率

一系列直接和间接的证据提示,同源重组事件可增加突变率。一项研究基于“父-母-子代”的全基因组测序和芯片数据,系统地鉴定了复制重组事件和DNMs,发现无论是男性还是女性,重组交叉周围1 Kb范围内均表现出了5~50倍的DNMs富集^[56-57],这提示重组事件本身可能导致突变的发生。

2.3 核小体定位

核小体结构被认为和局部突变率有关。近期的一项研究发现,平移定位稳定性强的核小体周围突变率显著升高。低保真的复制、不充分的错配修复和增加的双链断裂是该现象的潜在机制^[58]。

2.4 转录因子结合

在肿瘤基因组中,一些线索指示转录因子结合和突变率相关,其机制存在两种可能,一种是转录因子结合阻断了NER系统的启动^[59-60],另一种则是转录因子结合导致DNA局部构象发生变化,从而导致序列对紫外线更敏感^[61]。有研究发现,在转录因子结合位点中突变率上升20%,但是这种变化可能是由于序列的偏倚所导致的^[62]。近期的一项研究则通过模型校正了序列的影响,发现转录因子结合位点的突变率显著升高,且这种现象仅出现在睾丸组织中特异的结合位点和位于启动子上的位点^[63]。

2.5 DNA甲基化

胞嘧啶和鸟嘌呤组成的两碱基序列组合, 常被称为CpG位点, 在人类基因组中常被甲基化。在此前的一系列研究中, 已经发现位于CpG位点的C>T突变率是其他位点上C>T突变率的15倍^[4,64-65]。基因组特征的研究也发现, 睾丸组织中的DNA甲基化水平与CpG位点的突变率显著相关^[58]。这些证据说明在CpG序列上, 甲基化与突变的发生密切相关。值得注意的是, CpG位点的C>T突变在复制叉的不同方向上差异巨大: 除ACG外, 所有携带CpG的三核苷酸序列均在滞后链上显示出显著升高的突变率^[66-68]。

2.6 复制与转录方向

DNA双链的对称性是其分子结构的关键特征, 但是突变在双链上却存在差异。复制和转录等过程是造成这种非对称性的主要原因。这种差异在实际的突变数据中表现为反向互补突变之间的不平衡, 而由复制和转录过程造成的这种差异分别被称为复制不对称和转录不对称^[69]。

有多项研究报道了复制不对称的现象^[54,66-67,69-70]。经典的机制是这些突变来源于复制过程, 而两种主要DNA聚合酶的保真度或修复错误的效率差异导致了突变率在不同链上的差异。事实上, 在DNA聚合酶突变的肿瘤中, 其突变信号即表现出了典型的复制不对称^[69,71-72]。

滞后链上的易错损伤是人类突变的一种主要来源。研究发现, 转录偶联的核苷酸切除修复可以修复转录链上的损伤, 而非转录链上的损伤则更容易转变为突变。因此, 在非转录链上的突变特征与复制过程中滞后链上的主要突变类型高度相似^[66]。

3 DNMs对子代健康的影响

在过去的几十年中, 疾病的遗传病因研究主要围绕遗传变异开展。这些研究通过在孟德尔疾病中开展基于家系的连锁分析, 或者在复杂疾病中开展全基因组关联研究, 发现了一系列影响疾病发生的易感变异。但是, 这些传统研究大多并未重点关注DNMs和疾病的关系。近年来, 在“父-母-子代”进行核心家系中的全外显子或者全基因组测序为理解DNMs的致病性提供了重要数据。

3.1 DNMs与神经相关疾病

DNMs与多种神经系统相关疾病密切相关, 包括神经发育性疾病如孤独症发育谱系障碍、智力障

碍等, 以及神经精神类疾病如精神分裂症、神经退行性疾病等。致病性DNMs往往造成关键基因的蛋白截断, 这些基因在神经系统发育、神经元分化与生成、突触的结构和功能、跨膜受体蛋白酪氨酸激酶信号通路、Wnt信号通路的调节以及代谢过程调控等通路中存在富集^[73-75]。

3.1.1 孤独症发育谱系障碍 (autism spectrum disorder, ASD) 目前, DNMs被视为诸多常见神经发育疾病的重要原因, 在ASD中研究最为广泛^[76-77]。受限于技术水平, 早期研究重点关注长度大于100 Kb的拷贝数变异(copy number variation, CNV)。2007年, SEBAT等^[78]采用比较基因组芯片技术对ASD患儿及其父母进行分析, 发现在ASD患儿中新生CNV的频率显著高于正常儿童, 随后多项研究均在更大的样本中证实了该关联^[79-81]。其中, SSC(Simons Simplex Collection)项目开展的研究重点针对ASD家系, 包括父母、1名ASD先证者以及1名正常子女, 该研究设计有利于控制关联研究中的一些混杂因素^[82-83]。研究发现与未受影响的兄弟姐妹相比, ASD先证者不仅携带更多新生CNV, 而且这些CNV范围更大并涉及更多基因。此外, 研究还发现了7q11.23、15q11.2-13.1、16p11.2等区域的新生CNV与ASD之间的关联^[80]。

除CNV外, 研究者也基于候选基因策略评估了神经发育相关基因上的DNMs对ASD的影响。2010年, AWADALLA等^[84]利用桑格测序技术对142名ASD患者的401个突触相关基因上的DNMs进行了研究, 共发现了6个DNMs, 其中包含4个可能对蛋白质功能/结构产生影响的突变。随着高通量测序技术的发展, 研究人员得以进一步全面评估DNMs对ASD的影响。一系列研究通过在多人家系中开展全外显子测序, 发现ASD先证者编码区的新生突变率高于一般人群或其正常的兄弟姐妹^[85-87]。研究发现非同义DNMs可解释不少于10%的病例^[87-88]。

此后, 全基因组测序技术的快速应用使得研究者可以在全基因组水平上探讨DNMs和疾病的关联。一项包含了1 902个ASD四人家系的全基因组研究发现ASD先证者在启动子区存在DNMs富集^[89]。然而, 在更大规模全基因组测序研究中却未能揭示该关联^[90-91]。这些不一致的结果提示目前对非编码区DNMs功能的解读仍存在巨大挑战。除单核苷酸变异(single nucleotide variation, SNV)外, 全基因组测序也可以检测部分大片段结构变异(>1 Kb)。有研究发

现ASD患儿与正常儿童新生结构变异率无显著差异,但是患儿的新生结构变异范围更大,且更倾向覆盖外显子区^[92];而另一项研究则发现父源的结构变异可通过影响顺式调控元件,影响ASD的发病风险^[93]。

3.1.2 智力障碍(intellectual disability, ID) 约0.5%的新生儿患有严重ID,该疾病与遗传因素密切相关^[94-95]。早期研究通过染色体芯片发现5%~10%的智力障碍患者携带大片段新生CNV^[96-97],并鉴定了包括17q21.31微缺失^[98]、15q13.3微缺失^[99]等与ID有关的新生CNV。2009年,一项研究对94名非综合征型ID患者的*SYNGAP1*基因进行测序,发现了3名患者携带可造成蛋白截断的DNMs,该基因编码的Ras GTP酶激活蛋白对认知和突触功能至关重要^[100]。同时期另一项研究对95名散发ID患者的197个候选基因进行了测序,在*SYNGAP1*等基因中发现了多个功能缺失型DNMs^[101]。

2010年后,外显子测序被应用于ID的家系研究中,一项研究在10个散发ID患儿中鉴定了9个非同义DNMs,这9个DNMs位于9个不同的基因,包含*RAB39B*和*SYNGAP1*等ID致病基因上^[102]。随后一项研究对100名ID患者及其父母进行外显子组测序,共鉴定了79个DNMs,除10个已知基因外,还鉴定了3个新的ID致病基因,累计解释了约13%的ID患者的致病原因^[103]。近期一项研究应用全基因组测序对21个ID先证者的非编码调控区DNMs进行分析^[104],发现与成人脑增强子相比,DNMs在脑发育期特异的增强子中存在富集。此外,该研究还发现重复突变的增强子簇主要调控神经系统发育相关基因,随后通过实验证明调控元件上的DNMs可影响潜在靶基因的表达,为非编码区DNMs与ID的关联提供了新的证据。

3.1.3 发育性癫痫性脑病(developmental and epileptic encephalopathy, DEE) DEE是一组神经发育性疾病,其主要临床综合征包括婴儿痉挛、大田原综合征(Ohtahara syndrome, OS)、Lennox Gastaut综合征(Lennox Gastaut syndrome, LGS)、肌阵挛-失张力癫痫(Doos综合征)、获得性癫痫性失语综合征(Landau-Kleffner syndrome, LKS)和儿童癫痫伴慢波睡眠期持续棘慢复合波(continuous spikes and waves during sleep, CSWS)等。癫痫性脑病相关DNMs在2001年被首次发现,在一项关于Dravet综合征的研究中,所有7名儿童在*SCN1A*基因上都发生了

DNMs^[105]。此后,一项研究报道了*STXBPI*基因上新生CNV(9q33.3-q34.11缺失)可导致早期婴儿癫痫性脑病的发生^[106]。2011年一项更大规模的研究发现3%的癫痫性脑病患者携带了致病新生CNV^[107]。随后,全外显子组测序研究证实Dravet综合征患者中约95%的*SCN1A*突变是DNMs^[108]。2013和2014年两项基于大规模家系的全外显子测序研究,发现了癫痫性脑病先证者的DNMs负荷显著增加,并且至少12%的先证者携带有致病DNMs^[109-110]。此外,其他的基于核心家系的外显子测序研究还报道了*KCNT1*^[111]、*CYFIP2*^[112]和*GABRG2*^[113]等基因上的DNMs与癫痫性脑病之间的关联。

3.1.4 精神分裂症(schizophrenia, SCZ) DNMs不仅与发育性疾病相关,也有研究报道DNMs可以引起成年期的神经精神类疾病。一项研究在359个SCZ先证者家系中发现患者携带新生CNV的概率是正常对照的8倍^[114],同期的一些研究发现22q11.2、3q29、15q11.2、15q13.3和16p11.2等区域的新生CNV与SCZ有关^[115]。一些研究基于候选基因策略,对突触相关基因上的DNMs进行了系统研究,发现了一些仅在SCZ患者中出现的非同义DNMs^[84,116]。GIRARD等^[117]开展了首个SCZ的全外显子组测序研究,该研究分析了14名SCZ先证者及其父母的外显子组,共报道了15个DNMs,其中包含4个无义突变,该结果被同期另一项研究所证实^[118]。XU等^[119]对231个SCZ核心家系和34个健康家系进行全外显子组测序,也发现了功能缺失型DNMs负荷显著增加,并且鉴定了4个受DNMs影响的关键基因(*LAMA2*、*DPYD*、*TRRAP*和*VPS39*)。此后,一项更大规模的测序研究纳入了623个SCZ核心家系,发现DNMs显著富集在谷氨酸能突触后蛋白相关基因中^[120]。GULSUNER等^[121]则通过转录组映射的方法,发现功能性DNMs破坏的基因主要在胎儿背外侧和腹外侧前额叶皮层发育过程中形成转录共表达网络和蛋白质相互作用网络,一些关键基因在神经元迁移、突触传递、信号转导、转录调节和运输等方面发挥作用。

3.1.5 神经退行性疾病 越来越多的证据表明, DNMs对多种神经退行性疾病也有影响。最早的报道是在一名37岁的阿尔茨海默病(Alzheimer's disease, AD)患者中鉴定的DNMs,该突变位于*PSEN1*基因上^[122];随后又有多项研究报道了携带该基因DNMs的其他病例^[123-124]。除了AD以外,也有多项

研究发现肌萎缩侧索硬化症(amyotrophic lateral sclerosis, ALS)先证者也携带 *SOD1*、*FUS* 等基因的 DNMs^[125-127]。一项全外显子测序研究对 14 名早发性 AD 患者及健康父母进行研究后,发现 6 名患者携带 12 个非同义 DNMs,其影响的基因显著富集在以 A β 肽为中心的基因网络中^[128]。ALS 的全外显子组研究则确定了 25 个被非同义 DNMs 影响的基因,并且这些 DNMs 富集在染色质调控相关通路中^[129]。但是这些研究的结果却未能在近期另一项研究中被验证^[130]。近期,有少量研究报道了 DNMs 对帕金森病(Parkinson disease, PD)的影响^[131-132]。例如,一项包括 21 名早发 PD 先证者家系的全外显子测序研究共鉴定了 19 个非同义 DNMs,发现 3 个受 DNMs 影响的基因(*PTEN*、*VAPB* 和 *ASNA1*)与已知 PD 基因具有互作^[132]。此外,也有一些研究报道了 DNMs 对 Bethlem 肌病^[133]、早发性小脑萎缩^[134]、腓骨肌萎缩症^[135]等神经退行性疾病的影响。

3.1.6 其他神经系统相关疾病 一些罕见的神经发育性疾病也与 DNMs 有关。例如,一项家系研究发现,先天性脑积水患儿中 *TRIM71*、*SMARCC1*、*PTCH1* 等神经管发育关键基因上的功能缺失型 DNMs 负荷显著升高,这些突变可解释约 10% 的病例^[136]。此外,编码区的 DNMs 也被发现与先天性巨结肠^[137]和局限性脑皮质发育不良和半侧巨脑畸形^[138]有关。此外,多项研究也报道了 DNMs 对强迫症^[139-140]、双相情感障碍^[141]等其他神经精神类疾病的影响。HALVORSEN 等^[142]开展了目前最大规模的强迫症全外显子组测序,涉及 587 个强迫症核心家系和 41 个强迫症四人家系,发现预测有害的错义 DNMs 在患者中显著增加,同时在功能缺失不耐受的基因上出现更多的功能缺失型 DNMs。另一项研究则对 53 个强迫症家系进行全基因组测序,报道了 4 个与强迫症有关的基因(*SETD5*、*KDM3B*、*ASXL3* 和 *FBL*),并且发现 DNMs 在部分非编码区域的负荷显著增加^[143]。双相情感障碍的首个外显子测序研究共纳入了 79 个先证者家系,也发现功能缺失型 DNMs 显著富集于功能缺失不耐受的基因中^[144]。此外,多项研究表明从双相情感障碍先证者中鉴定的 DNMs 显著富集在钙离子结合蛋白相关基因中^[144-145]。

3.2 DNMs与心脏相关疾病

DNMs 与心脏相关疾病的研究主要聚焦于先

天性心脏病(congenital heart disease, CHD)。大片段基因组 CNV 是导致 CHD 发生的明确病因^[146],这些 CNV 可以传递自祖先,也可以是新生 CNV。多项较大规模基于 CHD 家系的研究($n > 100$)均发现 CHD 患者新生 CNV 的频率显著增加^[147-148],据估算至少 10% 的 CHD 患者携带新生 CNV^[148]。2013 年,儿科心脏基因组学联盟(the Pediatric Cardiac Genomics Consortium, PCGC)对 362 名 CHD 先证者核心家系进行了全外显子测序,发现 CHD 病例的总体 DNMs 数量未显著升高,但在心脏发育关键基因中,功能缺失型 DNMs 数量显著增加,这些突变被认为解释了约 10% 的 CHD 病例^[149]。随后多个更大规模的研究也报道了相似结论^[150-151]。不仅如此,有研究进一步发现 CHD 相关的功能缺失型 DNMs 主要富集在染色质修饰相关基因^[152],揭示了 23 个可能导致 CHD 的新基因^[153]。近年来,两项基于全基因组测序的研究进一步揭示了非编码 DNMs 与 CHD 之间的关联。一项研究对 PCGC 中 749 个 CHD 先证者家系进行全基因组测序,发现 CHD 先证者携带更多功能性非编码 DNMs,同时 CHD 已知基因的增强子上 DNMs 数量也显著增多^[154]。我们近期的一项研究也发现,父源的 CpG 位置上 C>T 碱基替换 DNMs 和功能性调控区 DNMs 均可增加子代先天性心脏病的发病风险^[3]。有研究指出,约有 30% 与 CHD 相关的 DNMs 存在于参与组蛋白修饰和染色质重塑的基因中,可导致各种类型的 CHD 表型,包括左心室流出道梗阻、圆锥动脉干异常等^[155]。大动脉转位相关的 DNMs 等致病变异富集在纤毛基因中,为纤毛在 CHD 发病机制中的作用提供依据^[156]。此外,非编码区的 DNMs 也被认为可以通过破坏心脏发育的转录和转录后调控导致 CHD 的发生^[154]。

3.3 DNMs与生殖相关疾病

近期,一项研究对 298 个卵母细胞成熟障碍和 175 个胎停的不孕女性及其父母进行全外显子组测序,共鉴定了 481 个 DNMs,涉及 474 个基因,这些基因富集在减数分裂(减数分裂细胞周期、纺锤体组织、减数分裂染色体分离等)、胚胎发育(宫内胚胎发育、翻译起始调节、组蛋白修饰)等生殖过程相关的信号通路中^[157]。另一项研究也探讨了 DNMs 对男性不育的影响。该研究对 185 个不育男性及其父母进行了全外显子组测序,鉴定了 29 个无义 DNMs,不育男性的无义 DNMs 显著富集在功能缺失不耐受

的基因中^[158]。此外,一些研究在中枢性早熟^[159],低促性腺激素性腺功能减退^[160]等患者中报道了一些可能致病的DNMs。此外,有研究通过小鼠模型证实了*TUBA4A*基因上的DNMs通过破坏HeLa细胞微管网络的组装过程,导致小鼠卵母细胞成熟或胚胎发育异常,进一步揭示了这些DNMs致生殖相关疾病的潜在机制^[157]。

3.4 DNMs与免疫相关疾病

一些证据显示DNMs也可导致免疫系统疾病。例如,一项研究应用全基因组测序对8对同卵双胞胎(其中1个患有银屑病)进行分析,共鉴定了27个DNMs,发现这些DNMs附近5 Mbps的基因显著富集在免疫应答、先天免疫应答和NF- κ B信号通路等调控通路中^[161]。此外,也有研究报道了DNMs对自身炎症性疾病^[162]、系统性红斑狼疮^[163]、慢性肉芽肿^[164]等免疫性疾病的影响。

4 总结与展望

在大规模测序得以开展之前,通过细胞或者动物模型进行突变机制的研究可能是探索DNA损伤修复机制的唯一途径,但是这些研究却难以真实模拟人类DNMs的发生过程,无法为其发生机制提供直接证据。在过去十年中,测序技术的长足发展推动了群体遗传学研究,在人类DNMs的起源和致病性方面带来了全新的认知,但是目前对这些变异的研究仍存在诸多挑战。

首先,当前的人类DNMs研究仍主要基于二代测序技术,受限于读长较短的不足,DNMs的检测类型尚不完整:(1)大部分的研究主要聚焦在本文提及的点突变和短缺失插入,且常难以覆盖序列复杂度低的基因组区域;(2)对于大片段的结构变异,现有检测技术的分辨率不足,仅能对长度>100 Kb的变异实现检测,但是这些变异由于对基因组的影响较大,常导致严重的疾病,因此在人群中频率极低,而长度在50 bp~100 Kb间的结构变异更为丰富,虽然可以通过二代测序检测,但是仍不能准确识别其断点,并且在检测过程中难以与假阳性变异区分。考虑到配子发生中的双链断裂修复过程可能是DNMs发生的重要机制之一,突变类型不全可能是目前对人类DNMs发生机制理解缺失的重要原因。而第三代测序技术的逐渐成熟有望为解决上述问题提供契机:一方面,通过三代测序构建的参考基因组序列可以

有效提升序列复杂度低的区域变异检测的准确性;另一方面,随着三代测序变异检测准确性的提升,长读长测序可以准确覆盖基因组上的所有变异类型。需要注意的是,三代测序的高成本和高样本需求仍是制约其在人群研究中的应用的主要原因。

其次,人类DNMs的起源、影响因素及发生机制的研究仍然存在较多挑战。突变基因组特征的关联分析可以为DNMs的发生机制推断提供线索,但是这种联系仍然局限在关联层面,可能受到诸多基因组因素的干扰,因此此类研究需要良好的偏倚控制并结合分子生物学实验验证才能得到可靠的结论。人群关联研究是探讨DNMs发生机制更直接的方式,但是需要依赖特殊且严谨的流行病学研究设计:(1)大样本的“父-母-子代”核心家系是鉴定人类DNMs的基本设计,如何获取家系的生物样本并进行高质量基因组检测是该设计的难点之一;(2)前瞻性获取可能对突变发生产生影响的个体信息是进行因果推断的重要保障,也是该设计需面临的核心挑战。当前研究往往基于已有基因组数据的人群进行回顾性信息采集,因此常仅能获取父母生育年龄、孕周等这些易于无偏获取的基本信息。与之相比,建设高质量的出生队列并开展测序研究是全面解析人类DNMs影响因素的最有效途径,而这些影响因素的解析不仅可为DNMs的发生机制提供直接证据,也可为疾病预防向致病突变预防的“关口前移”新策略提供理论依据。

再次,核心家系结合二代测序的研究仅能推断20%~30%突变的父母起源,并且推断结果可能受到受精卵分裂早期突变的干扰。因此,人类DNMs起源的追溯需依赖更为精细的设计。例如,对于子代样本进行二代和三代测序有望实现父母源单倍型的准确构建,从而将突变父母源的推断率大幅提升;在三世代家庭中开展二代测序研究也有助于构建单倍型并提高父母源推断率,同时该设计还有可能提供胚胎源DNMs的准确识别;高精度深度测序技术和单细胞基因组技术的发展为检测配子和胚胎来源的DNMs提供了可能。这些特殊设计的研究不仅可为解析不同起源突变的发生机制差异提供关键信息,还可以为这些突变的筛查策略制定提供差异化的方案。

最后,人类DNMs的致病性评价仍然困难重重。一方面,DNMs主要发生在非编码区,而既往研究发

现的致病性DNMs仍然局限于经典的功能缺失型突变,对编码区的其他突变以及非编码突变的致病性评价尚缺乏成熟的体系,结合基于序列特征的深度学习模型和高通量的细胞/在体功能评价系统可能是解决该问题的可行方案。另一方面,致病性DNMs的发现为疾病的遗传模型提供了思路,但是越来越多的DNMs被发现并非完全外显,基于核心家系的病例-对照研究是发现这些不完全外显DNMs的最有效途径,依托良好的大规模队列构建完善的健康人群DNMs数据及共享体系可为上述研究提供优质对照,推动领域内致病性DNMs的高效发现。

参考文献 (References)

- [1] LEJEUNE J, GAUTIER M, TURPIN R. Study of somatic chromosomes from 9 mongoloid children [J]. C R Hebd Seances Acad Sci, 1959, 248(11): 1721-2.
- [2] ROACH J C, GLUSMAN G, SMIT A F, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing [J]. Science, 2010, 328(5978): 636-9.
- [3] WANG C, LÜ H, LING X, et al. Association of assisted reproductive technology, germline *de novo* mutations and congenital heart defects in a prospective birth cohort study [J]. Cell Res, 2021, 31(8): 919-28.
- [4] KONG A, FRIGGE M L, MASSON G, et al. Rate of *de novo* mutations and the importance of father's age to disease risk [J]. Nature, 2012, 488(7412): 471-5.
- [5] GOLDMANN J M, WONG W S, PINELLI M, et al. Parent-of-origin-specific signatures of *de novo* mutations [J]. Nat Genet, 2016, 48(8): 935-9.
- [6] RAHBARI R, WUSTER A, LINDSAY S J, et al. Timing, rates and spectra of human germline mutation [J]. Nat Genet, 2016, 48(2): 126-33.
- [7] MICHAELSON J J, SHI Y, GUJRAL M, et al. Whole-genome sequencing in autism identifies hot spots for *de novo* germline mutation [J]. Cell, 2012, 151(7): 1431-42.
- [8] JONSSON H, SULEM P, KEHR B, et al. Parental influence on human germline *de novo* mutations in 1,548 trios from Iceland [J]. Nature, 2017, 549(7673): 519-22.
- [9] SASANI T A, PEDERSEN B S, GAO Z, et al. Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation [J]. eLife, 2019, 8: e46922.
- [10] WONG W S, SOLOMON B D, BODIAN D L, et al. New observations on maternal age effect on germline *de novo* mutations [J]. Nat Commun, 2016, 7: 10486.
- [11] WEINBERG W. Zur vererbung des zwergwuchses [J]. Arch Rassen-u Gesel Biolog, 1912, 9: 710-8.
- [12] SOHNI A, TAN K, SONG H W, et al. The neonatal and adult human testis defined at the single-cell level [J]. Cell Rep, 2019, 26(6): 1501-17, e4.
- [13] SCHULZE W, REHDER U. Organization and morphogenesis of the human seminiferous epithelium [J]. Cell Tissue Res, 1984, 237(3): 395-407.
- [14] SHARMA S, WISTUBA J, POCK T, et al. Spermatogonial stem cells: updates from specification to clinical relevance [J]. Hum Reprod Update, 2019, 25(3): 275-97.
- [15] HINCH R, DONNELLY P, HINCH A G. Meiotic DNA breaks drive multifaceted mutagenesis in the human germ line [J]. Science, 2023, 382(6674): eadh2531.
- [16] GAO Z, MOORJANI P, SASANI T A, et al. Overlooked roles of DNA damage and maternal age in generating human germline mutations [J]. Proc Natl Acad Sci USA, 2019, 116(19): 9491-500.
- [17] ALEXANDROV L B, NIK-ZAINAL S, WEDGE D C, et al. Signatures of mutational processes in human cancer [J]. Nature, 2013, 500(7463): 415-21.
- [18] ALEXANDROV L B, KIM J, HARADHVALA N J, et al. The repertoire of mutational signatures in human cancer [J]. Nature, 2020, 578(7793): 94-101.
- [19] ALEXANDROV L B, JONES P H, WEDGE D C, et al. Clock-like mutational processes in human somatic cells [J]. Nat Genet, 2015, 47(12): 1402-7.
- [20] YANG X, BREUSS M W, XU X, et al. Developmental and temporal characteristics of clonal sperm mosaicism [J]. Cell, 2021, 184(18): 4772-83, e15.
- [21] BREUSS M W, ANTAKI D, GEORGE R D, et al. Autism risk in offspring can be assessed through quantification of male sperm mosaicism [J]. Nat Med, 2020, 26(1): 143-50.
- [22] GIANNOULATOU E, MCVEAN G, TAYLOR I B, et al. Contributions of intrinsic mutation rate and selfish selection to levels of *de novo* HRAS mutations in the paternal germline [J]. Proc Natl Acad Sci USA, 2013, 110(50): 20152-7.
- [23] GORIELY A, WILKIE A O. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease [J]. Am J Hum Genet, 2012, 90(2): 175-200.
- [24] MAHER G J, RALPH H K, DING Z, et al. Selfish mutations dysregulating RAS-MAPK signaling are pervasive in aged human testes [J]. Genome Res, 2018, 28(12): 1779-90.
- [25] GORIELY A, HANSEN R M, TAYLOR I B, et al. Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors [J]. Nat Genet, 2009, 41(11): 1247-52.
- [26] DAKOUANE GIUDICELLI M, SERAZIN V, LE SCIELLOUR C R, et al. Increased achondroplasia mutation frequency with advanced age and evidence for G1138A mosaicism in human testis biopsies [J]. Fertil Steril, 2008, 89(6): 1651-6.
- [27] CROW J F. The origins, patterns and implications of human spontaneous mutation [J]. Nat Rev Genet, 2000, 1(1): 40-7.
- [28] XIE H, LI W, GUO Y, et al. Long-read-based single sperm genome sequencing for chromosome-wide haplotype phasing of both SNPs and SVs [J]. Nucleic Acids Res, 2023, 51(15): 8020-34.
- [29] FULKA J JR, FIRST N L, MOOR R M. Nuclear and cytoplasmic determinants involved in the regulation of mammalian oocyte maturation [J]. Mol Hum Reprod, 1998, 4(1): 41-9.
- [30] GOLDMANN J M, SEPLYARSKIY V B, WONG W S W, et al. Germline *de novo* mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence [J]. Nat Genet, 2018, 50(4): 487-92.
- [31] HALLDORSSON B V, HARDARSON M T, KEHR B, et al. The

- rate of meiotic gene conversion varies by sex and age [J]. *Nat Genet*, 2016, 48(11): 1377-84.
- [32] CAMPBELL C L, FURLOTTE N A, ERIKSSON N, et al. Escape from crossover interference increases with maternal age [J]. *Nat Commun*, 2015, 6: 6260.
- [33] CURRIE C E, FORD E, BENHAM WHYTE L, et al. The first mitotic division of human embryos is highly error prone [J]. *Nat Commun*, 2022, 13(1): 6755.
- [34] COORENS T H H, MOORE L, ROBINSON P S, et al. Extensive phylogenies of human development inferred from somatic mutations [J]. *Nature*, 2021, 597(7876): 387-92.
- [35] ACUNA-HIDALGO R, BO T, KWINT M P, et al. Post-zygotic point mutations are an underrecognized source of *de novo* genomic variation [J]. *Am J Hum Genet*, 2015, 97(1): 67-74.
- [36] HA Y J, KANG S, KIM J, et al. Comprehensive benchmarking and guidelines of mosaic variant calling strategies [J]. *Nat Methods*, 2023, 20(12): 2058-67.
- [37] HEYDE A, ROHDE D, MCALPINE C S, et al. Increased stem cell proliferation in atherosclerosis accelerates clonal hematopoiesis [J]. *Cell*, 2021, 184(5): 1348-61.e22.
- [38] AVAGYAN S, HENNINGER J E, MANNHERZ W P, et al. Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis [J]. *Science*, 2021, 374(6568): 768-72.
- [39] LI R, DI L, LI J, et al. A body map of somatic mutagenesis in morphologically normal human tissues [J]. *Nature*, 2021, 597(7876): 398-403.
- [40] GORIELY A, LORD H, LIM J, et al. Germline and somatic mosaicism for FGFR2 mutation in the mother of a child with Crouzon syndrome: implications for genetic testing in "paternal age-effect" syndromes [J]. *Am J Med Genet A*, 2010, 152A(8): 2067-73.
- [41] YUEN R K, MERICO D, CAO H, et al. Genome-wide characteristics of *de novo* mutations in autism [J]. *NPJ Genom Med*, 2016, 1: 16027.
- [42] LI J, OEHLERT J, SNYDER M, et al. Fetal *de novo* mutations and preterm birth [J]. *PLoS Genet*, 2017, 13(4): e1006689.
- [43] DU J, LIN Y, XIA Y, et al. Cohort profile: the Jiangsu Birth Cohort [J]. *Int J Epidemiol*, 2023, 52(6): e354-e63.
- [44] LIN Y, LU Q, JIANG Y Q, et al. A sub-cohort study design of the maternal and infant microbes in China National Birth Cohort [J]. *Zhonghua Liu Xing Bing Xue Za Zhi*, 2021, 42(4): 597-601.
- [45] SMITS R M, XAVIER M J, OUD M S, et al. *De novo* mutations in children born after medical assisted reproduction [J]. *Hum Reprod*, 2022, 37(6): 1360-9.
- [46] WANG R, SHI Q, JIA B, et al. Association of preterm singleton birth with fertility treatment in the US [J]. *JAMA Netw Open*, 2022, 5(2): e2147782.
- [47] LEE Y L, BOUWMAN A C, HARLAND C, et al. The rate of *de novo* structural variation is increased in *in vitro*-produced offspring and preferentially affects the paternal genome [J]. *Genome Res*, 2023, 33(9): 1455-64.
- [48] BESENBACHER S, SULEM P, HELGASON A, et al. Multi-nucleotide *de novo* mutations in humans [J]. *PLoS Genet*, 2016, 12(11): e1006315.
- [49] STAMATOYANNOPOULOS J A, ADZHUBEI I, THURMAN R E, et al. Human mutation rate associated with DNA replication timing [J]. *Nat Genet*, 2009, 41(4): 393-5.
- [50] RHIND N, GILBERT D M. DNA replication timing [J]. *Cold Spring Harb Perspect Biol*, 2013, 5(8): a010132.
- [51] VOHRINGER H, HOECK A V, CUPPEN E, et al. Learning mutational signatures and their multidimensional genomic properties with TensorSignatures [J]. *Nat Commun*, 2021, 12(1): 3628.
- [52] SUPEK F, LEHNER B. Differential DNA mismatch repair underlies mutation rate variation across the human genome [J]. *Nature*, 2015, 521(7550): 81-4.
- [53] ZHENG C L, WANG N J, CHUNG J, et al. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes [J]. *Cell Rep*, 2014, 9(4): 1228-34.
- [54] SEPLYARSKIY V B, SOLDATOV R A, KOCH E, et al. Population sequencing data reveal a compendium of mutational processes in the human germ line [J]. *Science*, 2021, 373(6558): 1030-5.
- [55] AGARWAL I, PRZEWSKI M. Signatures of replication timing, recombination, and sex in the spectrum of rare variants on the human X chromosome and autosomes [J]. *Proc Natl Acad Sci USA*, 2019, 116(36): 17916-24.
- [56] HALLDORSSON B V, PALSSON G, STEFANSSON O A, et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map [J]. *Science*, 2019, 363(6425): eaau1043.
- [57] ARBEITHUBER B, BETANCOURT A J, EBNER T, et al. Cross-overs are associated with mutation and biased gene conversion at recombination hotspots [J]. *Proc Natl Acad Sci USA*, 2015, 112(7): 2109-14.
- [58] LI C, LUSCOMBE N M. Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome [J]. *Nat Commun*, 2020, 11(1): 1363.
- [59] SABARINATHAN R, MULARONI L, DEU-PONS J, et al. Nucleotide excision repair is impaired by binding of transcription factors to DNA [J]. *Nature*, 2016, 532(7598): 264-7.
- [60] PERERA D, POULOS R C, SHAH A, et al. Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes [J]. *Nature*, 2016, 532(7598): 259-63.
- [61] MAO P, BROWN A J, ESAKI S, et al. ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma [J]. *Nat Commun*, 2018, 9(1): 2626.
- [62] VIERSTRA J, LAZAR J, SANDSTROM R, et al. Global reference mapping of human transcription factor footprints [J]. *Nature*, 2020, 583(7818): 729-36.
- [63] SEPLYARSKIY V, KOCH E M, LEE D J, et al. A mutation rate model at the basepair resolution identifies the mutagenic effect of polymerase III transcription [J]. *Nat Genet*, 2023, 55(12): 2235-42.
- [64] FRANCIOLI L C, POLAK P P, KOREN A, et al. Genome-wide patterns and properties of *de novo* mutations in humans [J]. *Nat Genet*, 2015, 47(7): 822-6.
- [65] DUNCAN B K, MILLER J H. Mutagenic deamination of cytosine residues in DNA [J]. *Nature*, 1980, 287(5782): 560-1.
- [66] SEPLYARSKIY V B, AKKURATOV E E, AKKURATOVA N, et al. Error-prone bypass of DNA lesions during lagging-strand replication is a common source of germline and cancer mutations [J]. *Nat Genet*, 2019, 51(1): 36-41.
- [67] TOMKOVA M, TOMEK J, KRIAUCIONIS S, et al. Mutational signature distribution varies with DNA replication timing and strand asymmetry [J]. *Genome Biol*, 2018, 19(1): 129.

- [68] POULOS R C, OLIVIER J, WONG J W H. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes [J]. *Nucleic Acids Res*, 2017, 45(13): 7786-95.
- [69] HARADHVALA N J, POLAK P, STOJANOV P, et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair [J]. *Cell*, 2016, 164(3): 538-49.
- [70] CHEN C L, DUQUENNE L, AUDIT B, et al. Replication-associated mutational asymmetry in the human genome [J]. *Mol Biol Evol*, 2011, 28(8): 2327-37.
- [71] HARADHVALA N J, KIM J, MARUVKA Y E, et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair [J]. *Nat Commun*, 2018, 9(1): 1746.
- [72] SHINBROT E, HENNINGER E E, WEINHOLD N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication [J]. *Genome Res*, 2014, 24(11): 1740-50.
- [73] LI B, LI K, TIAN D, et al. *De novo* mutation of cancer-related genes associates with particular neurodevelopmental disorders [J]. *J Mol Med*, 2020, 98(12): 1701-12.
- [74] MCNEILL A, IOVINO E, MANSARD L, et al. SLC12A2 variants cause a neurodevelopmental disorder or cochleovestibular defect [J]. *Brain*, 2020, 143(8): 2380-7.
- [75] TURNER T N, COE B P, DICKEL D E, et al. Genomic patterns of *de novo* mutation in simplex autism [J]. *Cell*, 2017, 171(3): 710-22, e12.
- [76] LORD C, BRUGHA T S, CHARMAN T, et al. Autism spectrum disorder [J]. *Nat Rev Dis Primers*, 2020, 6(1): 5.
- [77] VISSERS L E, GILISSEN C, VELTMAN J A. Genetic studies in intellectual disability and related disorders [J]. *Nat Rev Genet*, 2016, 17(1): 9-18.
- [78] SEBAT J, LAKSHMI B, MALHOTRA D, et al. Strong association of *de novo* copy number mutations with autism [J]. *Science*, 2007, 316(5823): 445-9.
- [79] LEVY D, RONEMUS M, YAMROM B, et al. Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders [J]. *Neuron*, 2011, 70(5): 886-97.
- [80] SANDERS S J, ERCAN-SENCICEK A G, HUS V, et al. Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism [J]. *Neuron*, 2011, 70(5): 863-85.
- [81] PINTO D, PAGNAMENTA A T, KLEI L, et al. Functional impact of global rare copy number variation in autism spectrum disorders [J]. *Nature*, 2010, 466(7304): 368-72.
- [82] FISCHBACH G D, LORD C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors [J]. *Neuron*, 2010, 68(2): 192-5.
- [83] ALTSHULER D, DALY M J, LANDER E S. Genetic mapping in human disease [J]. *Science*, 2008, 322(5903): 881-8.
- [84] AWADALLA P, GAUTHIER J, MYERS R A, et al. Direct measure of the *de novo* mutation rate in autism and schizophrenia cohorts [J]. *Am J Hum Genet*, 2010, 87(3): 316-24.
- [85] NEALE B M, KOU Y, LIU L, et al. Patterns and rates of exonic *de novo* mutations in autism spectrum disorders [J]. *Nature*, 2012, 485(7397): 242-5.
- [86] O'ROAK B J, VIVES L, GIRIRAJAN S, et al. Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations [J]. *Nature*, 2012, 485(7397): 246-50.
- [87] SANDERS S J, MURTHA M T, GUPTA A R, et al. *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism [J]. *Nature*, 2012, 485(7397): 237-41.
- [88] RONEMUS M, IOSSIFOV I, LEVY D, et al. The role of *de novo* mutations in the genetics of autism spectrum disorders [J]. *Nat Rev Genet*, 2014, 15(2): 133-41.
- [89] AN J Y, LIN K, ZHU L, et al. Genome-wide *de novo* risk score implicates promoter variation in autism spectrum disorder [J]. *Science*, 2018, 362(6420): eaat6576.
- [90] TROST B, THIRUVAHINDRAPURAM B, CHAN A J S, et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation [J]. *Cell*, 2022, 185(23): 4409-27, e18.
- [91] WERLING D M, BRAND H, AN J Y, et al. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder [J]. *Nat Genet*, 2018, 50(5): 727-36.
- [92] BRANDLER W M, ANTAKI D, GUJRAL M, et al. Frequency and complexity of *de novo* structural mutation in autism [J]. *Am J Hum Genet*, 2016, 98(4): 667-79.
- [93] BRANDLER W M, ANTAKI D, GUJRAL M, et al. Paternally inherited cis-regulatory structural variants are associated with autism [J]. *Science*, 2018, 360(6386): 327-31.
- [94] GILISSEN C, HEHIR-KWA J Y, THUNG D T, et al. Genome sequencing identifies major causes of severe intellectual disability [J]. *Nature*, 2014, 511(7509): 344-7.
- [95] ROPERS H H. Genetics of early onset cognitive impairment [J]. *Annu Rev Genomics Hum Genet*, 2010, 11: 161-87.
- [96] VISSERS L E, DE VRIES B B, VELTMAN J A. Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis [J]. *J Med Genet*, 2010, 47(5): 289-97.
- [97] DE VRIES B B, PFUNDT R, LEISINK M, et al. Diagnostic genome profiling in mental retardation [J]. *Am J Hum Genet*, 2005, 77(4): 606-16.
- [98] SHAW-SMITH C, PITTMAN A M, WILLATT L, et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability [J]. *Nat Genet*, 2006, 38(9): 1032-7.
- [99] SHARP A J, MEFFORD H C, LI K, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures [J]. *Nat Genet*, 2008, 40(3): 322-8.
- [100] HAMDAN F F, GAUTHIER J, SPIEGELMAN D, et al. Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation [J]. *N Engl J Med*, 2009, 360(6): 599-605.
- [101] HAMDAN F F, GAUTHIER J, ARAKI Y, et al. Excess of *de novo* deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability [J]. *Am J Hum Genet*, 2011, 88(3): 306-16.
- [102] VISSERS L E, DE LIGT J, GILISSEN C, et al. A *de novo* paradigm for mental retardation [J]. *Nat Genet*, 2010, 42(12): 1109-12.
- [103] DE LIGT J, WILLEMSSEN M H, VAN BON B W, et al. Diagnostic exome sequencing in persons with severe intellectual disability [J]. *N Engl J Med*, 2012, 367(20): 1921-9.
- [104] DE VAS M G, BOULET F, JOSHI S S, et al. Regulatory *de novo*

- mutations underlying intellectual disability [J]. *Life Sci Alliance*, 2023, 6(5): e202201843.
- [105] CLAES L, DEL-FAVERO J, CEULEMANS B, et al. *De novo* mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy [J]. *Am J Hum Genet*, 2001, 68(6): 1327-32.
- [106] SAITSU H, KATO M, MIZUGUCHI T, et al. *De novo* mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy [J]. *Nat Genet*, 2008, 40(6): 782-8.
- [107] MEFFORD H C, YENDLE S C, HSU C, et al. Rare copy number variants are an important cause of epileptic encephalopathies [J]. *Ann Neurol*, 2011, 70(6): 974-85.
- [108] HARKIN L A, MCMAHON J M, IONA X, et al. The spectrum of SCN1A-related infantile epileptic encephalopathies [J]. *Brain*, 2007, 130(Pt 3): 843-52.
- [109] EPI K C, EPILEPSY PHENOME/GENOME P, ALLEN A S, et al. *De novo* mutations in epileptic encephalopathies [J]. *Nature*, 2013, 501(7466): 217-21.
- [110] EURO E-R E S C, EPILEPSY PHENOME/GENOME P, EPI K C. *De novo* mutations in synaptic transmission genes including DNMI cause epileptic encephalopathies [J]. *Am J Hum Genet*, 2014, 95(4): 360-70.
- [111] BARCIA G, FLEMING M R, DELIGNIERE A, et al. *De novo* gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy [J]. *Nat Genet*, 2012, 44(11): 1255-9.
- [112] NAKASHIMA M, KATO M, AOTO K, et al. *De novo* hotspot variants in CYFIP2 cause early-onset epileptic encephalopathy [J]. *Ann Neurol*, 2018, 83(4): 794-806.
- [113] SHEN D, HERNANDEZ C C, SHEN W, et al. *De novo* GABRG2 mutations associated with epileptic encephalopathies [J]. *Brain*, 2017, 140(1): 49-67.
- [114] XU B, ROOS J L, LEVY S, et al. Strong association of *de novo* copy number mutations with sporadic schizophrenia [J]. *Nat Genet*, 2008, 40(7): 880-5.
- [115] KIROV G, POCKLINGTON A J, HOLMANS P, et al. *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia [J]. *Mol Psychiatry*, 2012, 17(2): 142-53.
- [116] GAUTHIER J, CHAMPAGNE N, LAFRENIERE R G, et al. *De novo* mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia [J]. *Proc Natl Acad Sci USA*, 2010, 107(17): 7863-8.
- [117] GIRARD S L, GAUTHIER J, NOREAU A, et al. Increased exonic *de novo* mutation rate in individuals with schizophrenia [J]. *Nat Genet*, 2011, 43(9): 860-3.
- [118] XU B, ROOS J L, DEXHEIMER P, et al. Exome sequencing supports a *de novo* mutational paradigm for schizophrenia [J]. *Nat Genet*, 2011, 43(9): 864-8.
- [119] XU B, IONITA-LAZA I, ROOS J L, et al. *De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia [J]. *Nat Genet*, 2012, 44(12): 1365-9.
- [120] FROMER M, POCKLINGTON A J, KAVANAGH D H, et al. *De novo* mutations in schizophrenia implicate synaptic networks [J]. *Nature*, 2014, 506(7487): 179-84.
- [121] GULSUNER S, WALSH T, WATTS A C, et al. Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network [J]. *Cell*, 2013, 154(3): 518-29.
- [122] DUMANCHIN C, BRICE A, CAMPION D, et al. *De novo* presenilin 1 mutations are rare in clinically sporadic, early onset Alzheimer's disease cases. French Alzheimer's Disease Study Group [J]. *J Med Genet*, 1998, 35(8): 672-3.
- [123] GOLAN M P, STYCZYNSKA M, JOZWIAK K, et al. Early-onset Alzheimer's disease with a *de novo* mutation in the presenilin 1 gene [J]. *Exp Neurol*, 2007, 208(2): 264-8.
- [124] PORTET F, DAUVILLIERS Y, CAMPION D, et al. Very early onset AD with a *de novo* mutation in the presenilin 1 gene (Met 233 Leu) [J]. *Neurology*, 2003, 61(8): 1136-7.
- [125] MULLER K, OH K W, NORDIN A, et al. *De novo* mutations in SOD1 are a cause of ALS [J]. *J Neurol Neurosurg Psychiatry*, 2022, 93(2): 201-6.
- [126] DEJESUS-HERNANDEZ M, KOCERHA J, FINCH N, et al. *De novo* truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis [J]. *Hum Mutat*, 2010, 31(5): E1377-89.
- [127] ALEXANDER M D, TRAYNOR B J, MILLER N, et al. "True" sporadic ALS associated with a novel SOD-1 mutation [J]. *Ann Neurol*, 2002, 52(5): 680-3.
- [128] ROVELET-LECRUX A, CHARBONNIER C, WALLON D, et al. *De novo* deleterious genetic variations target a biological network centered on A β peptide in early-onset Alzheimer disease [J]. *Mol Psychiatry*, 2015, 20(9): 1046-56.
- [129] CHESI A, STAAHL B T, JOVICIC A, et al. Exome sequencing to identify *de novo* mutations in sporadic ALS trios [J]. *Nat Neurosci*, 2013, 16(7): 851-5.
- [130] VAN DOORMAAL P T C, TICOZZI N, WEISHAUPT J H, et al. The role of *de novo* mutations in the development of amyotrophic lateral sclerosis [J]. *Hum Mutat*, 2017, 38(11): 1534-41.
- [131] GUO J F, ZHANG L, LI K, et al. Coding mutations in NUS1 contribute to Parkinson's disease [J]. *Proc Natl Acad Sci USA*, 2018, 115(45): 11567-72.
- [132] KUN-RODRIGUES C, GANOS C, GUERREIRO R, et al. A systematic screening to identify *de novo* mutations causing sporadic early-onset Parkinson's disease [J]. *Hum Mol Genet*, 2015, 24(23): 6711-20.
- [133] DECONINCK N, RICHARD P, ALLAMAND V, et al. Bethlem myopathy: long-term follow-up identifies COL6 mutations predicting severe clinical evolution [J]. *J Neurol Neurosurg Psychiatry*, 2015, 86(12): 1337-46.
- [134] CHEMIN J, SIQUIER-PERNET K, NICOLEAU M, et al. *De novo* mutation screening in childhood-onset cerebellar atrophy identifies gain-of-function mutations in the CACNA1G calcium channel gene [J]. *Brain*, 2018, 141(7): 1998-2013.
- [135] LASSUTHOVA P, SAFKA BROZKOVA D, KRUTOVA M, et al. Severe axonal Charcot-Marie-Tooth disease with proximal weakness caused by *de novo* mutation in the MORC2 gene [J]. *Brain*, 2016, 139(Pt 4): e26.
- [136] FUREY C G, CHOI J, JIN S C, et al. *De novo* mutation in genes regulating neural stem cell fate in human congenital hydrocephalus [J]. *Neuron*, 2018, 99(2): 302-14.e4.
- [137] GUI H, SCHRIEMER D, CHENG W W, et al. Whole exome sequencing coupled with unbiased functional analysis reveals new Hirschsprung disease genes [J]. *Genome Biol*, 2017, 18(1): 48.
- [138] MIRZAA G M, CAMPBELL C D, SOLOVIEFF N, et al. Association of MTOR mutations with developmental brain disorders,

- including megalencephaly, focal cortical dysplasia, and pigmentary mosaicism [J]. *JAMA Neurol*, 2016, 73(7): 836-45.
- [139] CAPPI C, OLIPHANT M E, PETER Z, et al. *De novo* damaging DNA coding mutations are associated with obsessive-compulsive disorder and overlap with Tourette's disorder and autism [J]. *Biol Psychiatry*, 2020, 87(12): 1035-44.
- [140] CAPPI C, BRENTANI H, LIMA L, et al. Whole-exome sequencing in obsessive-compulsive disorder identifies rare mutations in immunological and neurodevelopmental pathways [J]. *Transl Psychiatry*, 2016, 6(3): e764.
- [141] MALHOTRA D, MCCARTHY S, MICHAELSON J J, et al. High frequencies of *de novo* CNVs in bipolar disorder and schizophrenia [J]. *Neuron*, 2011, 72(6): 951-63.
- [142] HALVORSEN M, SAMUELS J, WANG Y, et al. Exome sequencing in obsessive-compulsive disorder reveals a burden of rare damaging coding variants [J]. *Nat Neurosci*, 2021, 24(8): 1071-6.
- [143] LIN G N, SONG W, WANG W, et al. *De novo* mutations identified by whole-genome sequencing implicate chromatin modifications in obsessive-compulsive disorder [J]. *Sci Adv*, 2022, 8(2): eabi6180.
- [144] KATAOKA M, MATOBA N, SAWADA T, et al. Exome sequencing for bipolar disorder points to roles of *de novo* loss-of-function and protein-altering mutations [J]. *Mol Psychiatry*, 2016, 21(7): 885-93.
- [145] XY LING N, LANGENDORF C G, HOQUE A, et al. Functional analysis of an R311C variant of Ca²⁺-calmodulin-dependent protein kinase kinase-2 (CaMKK2) found as a *de novo* mutation in a patient with bipolar disorder [J]. *Bipolar Disord*, 2020, 22(8): 841-8.
- [146] FAHED A C, GELB B D, SEIDMAN J G, et al. Genetics of congenital heart disease: the glass half empty [J]. *Circ Res*, 2013, 112(4): 707-20.
- [147] GLESSNER J T, BICK A G, ITO K, et al. Increased frequency of *de novo* copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data [J]. *Circ Res*, 2014, 115(10): 884-96.
- [148] GREENWAY S C, PEREIRA A C, LIN J C, et al. *De novo* copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot [J]. *Nat Genet*, 2009, 41(8): 931-5.
- [149] ZAIDI S, CHOI M, WAKIMOTO H, et al. *De novo* mutations in histone-modifying genes in congenital heart disease [J]. *Nature*, 2013, 498(7453): 220-3.
- [150] SIFRIM A, HITZ M P, WILSDON A, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing [J]. *Nat Genet*, 2016, 48(9): 1060-5.
- [151] HOMSY J, ZAIDI S, SHEN Y, et al. *De novo* mutations in congenital heart disease with neurodevelopmental and other congenital anomalies [J]. *Science*, 2015, 350(6265): 1262-6.
- [152] WATKINS W S, HERNANDEZ E J, WESOŁOWSKI S, et al. *De novo* and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes [J]. *Nat Commun*, 2019, 10(1): 4722.
- [153] SEVIM BAYRAK C, ZHANG P, TRISTANI-FIROUZI M, et al. *De novo* variants in exomes of congenital heart disease patients identify risk genes and pathways [J]. *Genome Med*, 2020, 12(1): 9.
- [154] RICHTER F, MORTON S U, KIM S W, et al. Genomic analyses implicate noncoding *de novo* variants in congenital heart disease [J]. *Nat Genet*, 2020, 52(8): 769-77.
- [155] ZHU W, LO C W. Insights into the genetic architecture of congenital heart disease from animal modeling [J]. *Zool Res*, 2023, 44(3): 577-90.
- [156] LIU X, CHEN W, LI W, et al. Exome-based case-control analysis highlights the pathogenic role of ciliary genes in transposition of the great arteries [J]. *Circ Res*, 2020, 126(7): 811-21.
- [157] LI Q, ZHAO L, ZENG Y, et al. Large-scale analysis of *de novo* mutations identifies risk genes for female infertility characterized by oocyte and early embryo defects [J]. *Genome Biol*, 2023, 24(1): 68.
- [158] OUD M S, SMITS R M, SMITH H E, et al. A *de novo* paradigm for male infertility [J]. *Nat Commun*, 2022, 13(1): 154.
- [159] CANTON A P M, TINANO F R, GUASTI L, et al. Rare variants in the MECP2 gene in girls with central precocious puberty: a translational cohort study [J]. *Lancet Diabetes Endocrinol*, 2023, 11(8): 545-54.
- [160] XU H, LI Z, SUN T, et al. An isolated hypogonadotropic hypogonadism male with a novel *de novo* FGFR1 mutation fathered a normal son evidenced by prenatal genetic diagnosis [J]. *Andrologia*, 2020, 52(11): e13821.
- [161] LI J, LIN H, HOU R, et al. Multi-omics study in monozygotic twins confirm the contribution of *de novo* mutation to psoriasis [J]. *J Autoimmun*, 2020, 106: 102349.
- [162] MOGHADDAS F, ZENG P, ZHANG Y, et al. Autoinflammatory mutation in NLRC4 reveals a leucine-rich repeat (LRR)-LRR oligomerization interface [J]. *J Allergy Clin Immunol*, 2018, 142(6): 1956-67.e6.
- [163] DUAN R, LIU Q, LI J, et al. A *de novo* frameshift mutation in TNFAIP3 impairs A20 deubiquitination function to cause neuropsychiatric systemic lupus erythematosus [J]. *J Clin Immunol*, 2019, 39(8): 795-804.
- [164] WU E Y, KUEHN H S, ROSENZWEIG S D, et al. Expanding the clinical phenotype of chronic granulomatous disease: a female patient with a *de novo* mutation in CYBB [J]. *J Clin Immunol*, 2021, 41(1): 224-6.