何嘉玲,王天奇,暴国,等. 斑马鱼小头震颤突变体 *mise* 的鉴定与基因定位［J］. 中国比较医学杂志, 2022, 32(8): 48-56.
He JL, Wang TQ, Bao G, et al. Identification and gene localization of a microcephaly and microcephaly and suspected epilepsy mutant *mise* in zebrafish［J］. Chin J Comp Med, 2022, 32(8): 48-56.
doi: 10.3969/j.issn.1671-7856.2022.08.007

斑马鱼小头震颤突变体 *mise* 的鉴定与基因定位

何嘉玲,王天奇,暴　国,张长勇,李　楠,孙德明*

(国家卫生健康委科学技术研究所实验动物中心,北京　100081)

【摘要】　目的　以自发突变导致的小头、震颤(microcephaly and seizure,*mise*)斑马鱼为模型,研究其遗传方式和突变基因定位。**方法**　通过比较正常和突变体斑马鱼的头宽和眼球面积,鉴定突变体的小头小眼性状;通过构建家系确定突变遗传方式。对极端表型群体进行基因组重测序,随后利用分离群体分组分析(bulked segregation analysis,BSA)和竞争性等位基因特异性 PCR(kompetitive allele specific PCR,KASP),一代测序等方法获得突变基因定位。**结果**　与野生型相比,突变体 *mise* 受精后 3 d(3 days post fertilization,3 dpf)表现为小头小眼震颤表型。经基因定位及验证明确与表型相关的突变基因为 *terfa*。**结论**　本研究通过全基因组重测序结合连锁分析的方法快速定位突变基因,可为斑马鱼突变体的基因定位提供参考,为神经系统发育异常的相关基因功能分析奠定基础。

【关键词】　斑马鱼;小头畸形;基因定位
【中图分类号】R-33　　**【文献标识码】**A　　**【文章编号】**1671-7856 (2022) 08-0048-09

Identification and gene localization of a microcephaly and microcephaly and suspected epilepsy mutant *mise* in zebrafish

HE Jialing, WANG Tianqi, BAO Guo, ZHANG Changyong, LI Nan, SUN Deming*
(Experimental Animal Center of National Research Institute for Family Planning, Beijing 100081, China)

【Abstract】　Objective　Taking the congenital microcephaly and suspected epilepsy (*mise*) zebrafish mutant caused by spontaneous as a model, to study its inheritance pattern and phenotype-related mutation gene location. **Methods**　By comparing the head width and eye area of normal and mutant zebrafish, the microcephaly and ommatidia traits of mutants were identified; the inheritance pattern of the mutation was determined by constructing a family. Genome resequencing combined with bulked segregation analysis (BSA) and kompetitive allele specific PCR (kompetitive allele specific PCR, KASP) and Sanger sequencing were used to obtain the location of mutant genes. **Results**　Compared with the wild type, the head and eye were significantly reduced at 3dpf, accompanied by convulsions and tremors. The mutant gene associated with phenotype was identified as *terfa* by gene mapping and verification. **Conclusions**　This study uses whole-genome resequencing combined with linkage analysis to quickly locate mutant genes. It laid the foundation for the locational cloning of zebrafish mutants and the functional analysis of genes related to abnormal development of the nervous system.

【Keywords】　zebrafish; microcephaly; gene mapping

原发性小头畸形(microcephaly primary hereditary,MCPH)是一种大脑发育障碍,导致头围比年龄和性别的平均值低 3 个以上的标准差[1]。患者可出现一定程度的智力障碍,或伴发多动症、

[基金项目]2019 年度国家卫生健康委科研所中央级公益性科研院所基本科研业务费专项(2019GJM03);2018 年度国家卫计委科研所科技创新重点项目(2018GJ07)。
[作者简介]何嘉玲(1980—),女,助理研究员,研究方向:实验动物科学。E-mail: missu329@aliyun.com
[通信作者]孙德明(1958—),男,研究员,研究方向:实验动物科学。E-mail: sundemingnet@163.com

表达性语言障碍和癫痫等症状,因此也常称为小头畸形综合征。原发性小头畸形多数基于常染色体隐性遗传模式,目前已经报道过的与小头畸形相关的基因有 18 个[2],是小头畸形的病因研究和诊疗的切入点。小头畸形致病机理复杂,利用单基因突变动物模型开展研究是表型发生机制良好的切入点。基因突变动物模型是基因功能生物学研究的重要工具。通过研究突变体的表现型和鉴定突变基因,能获得关于基因功能和异常发育相关的大量信息。

斑马鱼作为一种重要的模式生物,具有与人类高度保守的大脑组织区域和发育过程[3],加之早期胚胎透明易于观察,是新兴的研究中枢神经系统发育性疾病的重要工具[4]。本文作者在斑马鱼繁育过程中发现了一种可以遗传的,表现为小头、小眼、震颤的斑马鱼突变体,是典型的小头畸形综合征。在人群中,虽然遗传性小头畸形综合征相对罕见,但小头畸形致病基因编码的一系列蛋白在细胞有丝分裂的关键步骤中起作用[1],因此,建立 MCPH 突变的动物模型,不仅能够加深人们对该疾病本身的理解,也能提高我们对大脑发育过程的认识。目前,MCPH 部分相关基因的动物模型研究十分缺乏,导致很多相关基因功能和作用机制仍无法明确。本研究通过建立家系,研究新发现的小头畸形突变体斑马鱼的遗传方式。通过基因组重测序结合 BSA 连锁分析以及 KASP 技术,研究表型相关突变基因定位,以期获得特定基因相关的小头畸形斑马鱼模型,为小头畸形发病机制研究和相关基因功能研究奠定基础。同时摸索出一套适用于斑马鱼突变基因定位的方法,为斑马鱼突变体基因定位提供可借鉴的参考。

1 材料和方法

1.1 实验动物

本研究使用的野生型 TU 品系斑马鱼成鱼,由北京大学生命科学学院细胞增殖与分化教育部重点实验室张博教授馈赠。India 品系野生型斑马鱼成鱼由清华大学生命学院孟安明实验室馈赠。成鱼体重约 350~390 mg,均为 20~24 周龄,雌雄各 4 尾。按照 Monte[5] 述及的方法饲养成鱼并繁育、收集受精卵。受精卵去除杂质后,放入含有 1‰(w/v)亚甲基蓝的培养液中,于 28.5℃培养箱中培养 4 d,期间每天更换新鲜培养液。以上斑马鱼繁育及实验均在国家卫生健康委科学技术研究所实验动物中心[SYXK(京)2018-0010]完成。饲养温度为(28±1)℃,每天光照 14 h,黑暗 10 h。本文所述及的斑马鱼相关实验方案均通过国家卫生健康委科学技术研究所实验动物福利伦理委员会审批(NRIFH21-2104-1),所有实验均符合动物伦理学标准,严格遵循 3R 原则进行动物饲养和实验操作。

1.2 主要试剂与仪器

海洋动物组织基因组 DNA 提取试剂盒(DP324,天根生化科技有限公司,中国)。Illumina 测序平台(Illumina,Inc.,San Diego,CA,USA);体视镜(Stemi 305, ZEISS, Germany); Nanodrop2000 (Thermo,USA);OMEGA F Reader 96 and 384 well compatible,OMEGA F SNP 分型检测仪(LGC,UK)。

1.3 实验方法

1.3.1 软件与数据库

本研究中用到的软件及数据库名称和链接如表 1 所示。

1.3.2 突变体生物学特性观察

突变体呈现小头,小眼表型,为了解头部和眼睛的改变情况,利用 ZEN 软件随机测量 8 尾 2 dpf 和 3 dpf *mise* 斑马鱼和无表型同胞的最大头宽和眼睛的面积。利用 GraphaPad Prism 软件对测量数据进行统计。

1.3.3 遗传分析

由于 *mise* 突变体 5 d 死亡,为研究其遗传方式,

表 1 主要实验软件与数据库链接

Table 1 Main experimental softwares and database links

软件 Softwares	版本名称 Version	相关链接 Related links
Picard	version1.119	https://sourceforge.net/projects/picard/
GATK	version4.0.11.0	https://software.broadinstitute.org/gatk/
ZEN	version2.3	https://www.zeiss.com.cn/microscopy/products/microscope-software/zen-lite.html
GraphaPad Prism	version6	https://www.graphpad.com/
LGC KlusterCaller	version 3.4.1.36	https://www.lgcstandards.com/CN/en/
斑马鱼参考基因组 Danio rerio reference genome	GRCz11	http://asia.ensembl.org/Danio_rerio/Info/Annotation

利用 *mise* 的亲本与野生 TU 进行外交产生 F1 代。随后待 F1 性成熟后进行自交,分析其子代(F2)的表型情况,计算发生表型分离的 F2 中有表型鱼和无表型同胞的比例。

1.3.4 突变基因初步定位

(1)DNA 样品制备及重测序

为获得丰富的多态性标记,含杂合突变基因的 *mise* 亲本(P0)与野生型 India 品系鱼外交,获得 F1。待 F1 培养至性成熟,通过自交获得 F2。选取 F2 发生性状分离的两个群体:即 *mise* 突变体和无表型同胞各 40 尾构建极端混池。使用海洋动物组织基因组 DNA 提取试剂盒,分别提取 *mise* 亲本(P0),野生型 India 亲本,F2 *mise* 混池和 F2 正常表型混池的总基因组 DNA,并用 Nanodrop 2000 定量后送百迈客有限公司用于文库构建和全基因组重测序。具体流程包括:使用 S2/E210 超声波发生器将提取的 DNA 产物进行超声波处理以产生 350 个碱基的片段(Covaris,Woburn,MA,USA);随后进行末端修复产生核苷酸突出物。然后,使用 T4 脱氧核糖核酸连接酶连接测序接头,并进行聚合酶链反应。最后将聚合酶链反应产物纯化并加载到 Illumina 测序平台(Illumina, Inc., San Diego, CA, USA)上进行双端测序。

为了简化描述,我们将杂合亲本 P0 命名为 R01,将野生型 India 命名为 R02,将 F2 *mise* 混池命名为 R03,将 F2 正常表型混池命名为 R04。

(2)突变位点关联分析

主要采取对极端混池的突变位点进行批量分离分析的方法。首先为获得 clean read,使用百迈客有限公司(中国北京)提供的内部 Perl 脚本,对原始数据进行过滤,去除低质量的 reads,包括带接头 reads,同时比对到多条染色体上的 reads,N 含量超过 10%的 reads,以及质量值低于 10 的碱基超过 50%的 reads。然后,使用 Burrows-Wheeler Aligner[6] 将这些高质量数据与斑马鱼基因组序列进行比对。使用 Picard 软件中的 Mark Duplicate 工具消除 PCR 重复,提高 SNP/InDel-calling 的准确性。利用 GATK 软件检测 SNP 和小的 InDel 变异,之后过滤数据,排除其中有多个基因型、read 深度小于 4、混池间基因型相同、隐性混池基因并非来源于隐性亲本的 4 类 SNP 位点。对于混池间的 Indel 关联分析,采用同样的标准进行过滤。关联分析分别采用欧氏距离(euclidean distance, ED)算法[7] 和 SNP-index[7-8] 方法进行。将上述两种方法取交集。最后对 SNP 和 InDel 对应的关联区域取交集作为小头小眼表型相关的候选区域。

1.3.5 突变基因精细定位

在初步定位候选区域中,筛选可信的 SNP 位点。根据 SNP 位点和侧翼序列设计 PCR 扩增引物,根据标记的距离合理分区段设计 KASP 标记。每个标记各设计 F1 与 F2 两条特异性 SNP 引物和一条通用引物(R)。特异性引物尾部分别添加能够与 FAM 荧光和 VIC 荧光结合的特异性序列。如 1.3.4 所述,分别提取、定量 F2 代 *mise* 个体,无表型同胞个体和 *mise* 亲本(P0),野生型 India 亲本全基因组为模板,进行 KASP 标记有效性检验,选取能有效区分野生亲本,隐性亲本和突变体的 KASP 标记。随后,使用能有效分型的 KASP 标记,对 192 枚 F2 *mise* 胚胎进行 KASP 基因分型实验。按照 LGC 基因组公司(英国米德尔塞克斯,LGC)的方案,在 384 孔板中进行 KASP 分析。PCR 反应体系为 2×Taq DNA Polymerase Mix,2 μL,SNP Primer Mix(4×),1 μL;DNA 样品,2 μL。PCR 扩增体系:预变性,94°C,10 min,94°C,变性,20 s,61°C ~ 55°C(drop 0.6°C per cycle),退火/延伸 45 s(10 cycle);94°C 变性 20 s,55°C,退火/延伸 45 s(35 cycle)。PCR 扩增循环结束后,在低于 40°C 的环境下,利用 SNP 分型检测仪(LGC)读取荧光值。被动参比染料 ROX(passive reference dye ROX)用于校正孔与孔之间由于反应体积误差导致的信号差异。结果数据分析软件为设备自带基因型读取软件 LGC KlusterCaller software。

1.3.6 转录本序列分析

TRIzol 法提取突变体 *mise* 和野生 TU 的总 RNA,取 1 μg 反转录成 cDNA,利用 CDS 全长引物 PCR 扩增候选基因的 CDS,并送测序。最后将测序结果与参考转录本序列进行比对。引物序列见表 2。

1.3.7 Morpholino 合成、注射及表型观察

委托 genetool 公司合成靶向 *terfa* 2 号外显子的剪接抑制 Morpholino 及相应的错配对照,terfa-sMO:5'-GCGAACGTCACTGGAAAATATTACT-3' 及 terfa-scMO:5'-GCCAACCTGACTGCAAAATATTAGT-3',以每个胚胎 6 ng 注射量分别注射单细胞期斑马鱼胚胎,并观察表型变化。

1.4 统计学方法

实验数据均用平均数±标准误差($\bar{x}\pm s\bar{x}$)表示。

统计学分析采用方差分析和 Dunnett's *T*-检验，如 $P<0.05$ 表明具有显著性差异，$P<0.01$ 表明差异极显著。各实验组数据利用 GraphPad Prism 进行 ANOVA 检验比较。

表 2　候选基因 CDS 扩增引物序列

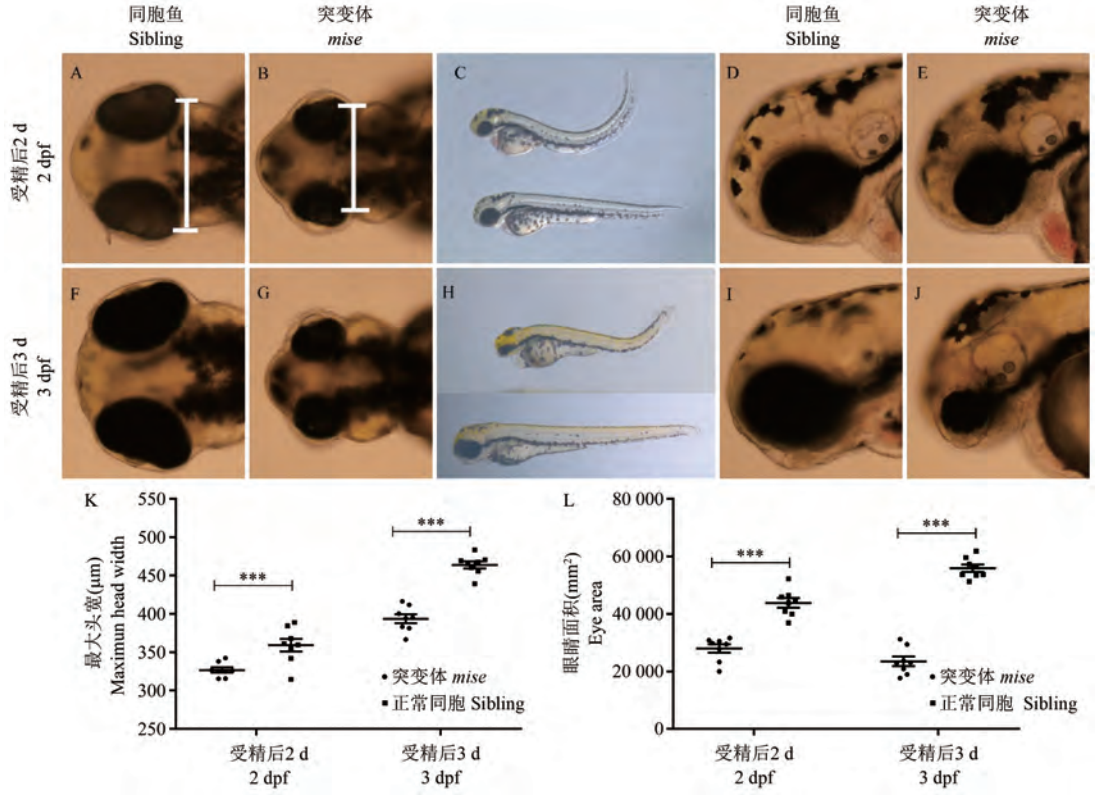
Table 2　CDS amplification primers for candidate genes

引物名称 Primer name	序列(5'-3') Sequence
*arfgap*3F	TTGTGGAGCGTTTGTCTGTC
*arfgap*3R	TCAGCGCTCGACACACTACT
*ikbip*F	TCTCACCGCTAAGGCAGAGT
*ikbip*R	AGCAGTCGACCAATCAGAAGA
*terfa*F	GCCGAGGCTCACATTAAATC
*terfa*R	TCGCAAACACACACATCAGA
*ckap*4F	CGCACTCACTCACGGTAAGA
*ckap*4R	TTCCACCACCCGGATATAAA

2　结果

2.1　突变体 *mise* 的表型分析

2 dpf 时 *mise* 斑马鱼突变体出现明显的小头，小眼表型，部分胚胎体轴向背部卷曲。通过对 2 dpf 和 3 dpf 的幼鱼进行测量发现，*mise* 的头（图 1A 和 1E）和眼（1B 和 1F）的尺寸明显小于正常表型同胞（1C、1G 和 1D、1H）。*mise* 平均最大头宽为（326.67±3.36）μm 和（393.63±5.79）μm，正常同胞平均头宽度为（359.11±8.33）μm（$n=8$，$P=0.0028$）和（463.99±4.55）μm（$n=8$，$P<0.0001$），*mise* 斑马鱼的头宽远小于正常同胞，差异极显著，出现了明显的发育迟滞。*mise* 斑马鱼眼睛面积平均值分别为：（28013±1454）μm^2 和（23562±1670）μm^2，都远小于无表型同胞（43921±1660）μm^2（$n=8$，$P<0.0001$）和



注：A、B、D、E 和 F、G、I、J 分别显示受精后 2 d 和 3 d 同胞鱼和突变体局部放大的头部外观；C，H 分别显示整体外观，图中上面为突变体，下面为同胞鱼；K 和 L 分别为受精后 2 d 和 3 d 突变体和同胞鱼头宽和眼睛面积的数值。与同胞鱼对照比较，*** $P<0.001$。

图 1　*mise* 胚胎的异常表型观察

Note. A, B, D, E and F, G, I, J show the head appearance of the sibling fish and mutant with local amplification at 2 and 3 days after fertilization, respectively. C and H respectively show the overall appearance while the mutant at the top and the sibling fish at the bottom. K and L are the values of maximun head width and eye area of mutant and sibling fish at 2 and 3 days after fertilization, respectively. Compared with the control fish, *** $P<0.001$.

Figure 1　Observation of abnormal phenotype of *mise* embryo

(55956±1258) μm^2 ($n=8$, $P<0.0001$),差异极显著。比较 2 dpf 和 3 dpf,眼睛不仅没有随着发育时间的延长增大,反而出现了一定程度的萎缩(图 1B、1F)。

2.2 遗传规律分析

mise 的亲本与野生 TU 进行外交产生 F1 代,表型分析表明,F1 代所有斑马鱼表型正常。F1 代斑马鱼自交,结果部分 F2 代表型正常,部分 F2 代出现表型分离。突变体 2 dpf 开始出现小头、小眼、震颤表型。统计 3 对产生 F2 代表型分离的 F1 代斑马鱼发现:第一对中,*mise* 突变体 25 尾,正常表型 78 尾;第二对中,*mise* 突变体 35 尾,正常表型 105 尾;第三对中,*mise* 突变体 44 尾,正常表型 124 尾(表 3)。经卡方 χ^2 测验,该表型符合隐性基因 1(隐性):3(显性)的分离比。表明 F1 代鱼发生了单基因隐性纯合致死突变。

2.3 斑马鱼重测序和 SNP、InDel 检测

利用 Illumina HiSeq 平台分别对 R01~R04 4 个样本进行重测序分析。测序共获得 211.42 Gbp 数据量,过滤后得到的 clean bases 为 209.43 Gbp。个样本测序数据 Q30 为 89.47%~90.25%,GC 含量 37.72%~38.45%。样品与斑马鱼基因组(GRCz11)平均比对效率为 98.27%。两亲本(F0)平均覆盖深度为 17×,子代混池(F2)的平均覆盖深度为 42×。总基因组覆盖度 95.48%(至少覆盖 1×)。过滤后得到的 clean reads 数分别为 97471743、106703350、238971525、255944524(表 4)。

根据过滤条件最终筛选获得高质量 SNP 3364571 个,高质量的可信 InDel 位点 778134 个。具体筛选结果统计如表 5、表 6。

2.4 关联分析

基于 SNP 的连锁分析,按照 ED 方法,计算得到 SNP 关联阈值为 0.07。关联值分布如图 2A。通过对此阈值的数据分析,可在 4 号染色体筛选到 1 个性状关联区域,长度 10.46 Mb,共包含 564 个基因,其中包含非同义突变 SNP 位点基因 241 个。

运用 SNP-index 关联分析方法进行计算机模拟[9]。结果表明,置信度为 0.99 时,在 4 号染色体获得 1 个性状关联区域,长度 5.67 Mb,包含 353 个基因,其中非同义突变基因 175 个 SNP-index 及 ΔSNP-index 的分布如图 2B 所示。

对于两种算法取交集,与突变关联的 SNP 位于 Chr4:72420000~78090000,包含 175 个非同义突变的基因。

基于 InDel 的关联分析,按照 ED 方法,通过计算将 InDel 关联的显著性阈值设置为 0.08。关联值分布如图 3A。根据该阈值判定,在 4 号和 5 号染色体共获得 2 个性状关联区域,长度 10.88 Mb,包含 586 个基因,其中包含移码突变 InDel 位点的基因共 39 个。

运用 InDel-index 方法进行计算机模拟[9]。结果表明,置信度为 0.99 时,在 4 号染色体获得 1 个性状关联区域,长度 5.61 Mb,包含 353 个基因,其中包含非同义突变的基因共 175 个。关联值分布如图 3B。

对于两种算法取交集,与突变关联的 InDel 位于 Chr4:72450000~78060000,包含 30 个移码突变的基因。对 SNP 和 InDel 对应的关联区域取交集,得到 1 个区域,总长度为 5.61 Mb,位于 Chr4:72450000~78060000,共包含 353 个基因,其中包含非同义突变的基因共 174 个,移码突变的基因共 30 个。

2.5 突变基因精细定位

由于 5.61 Mb 区域仍然包含大量的基因,我们开发了 22 个 KASP 标记,筛选出能有效区分亲本(P0),F2 纯合突变体和无表型同胞的 6 个 KASP 标记对 192 个 F2 突变个体进行基因分型。通过 KASP 标记进行连锁分析,最后将与 *mise* 突变体相

表 3 F2 表型分离情况统计

Table 3 Statistics of F2 phenotype separation

组别 Groups	有表型鱼/正常表型同胞 *Mise* phenotype zebrafish/sibling	有表型鱼/正常表型同胞 *Mise* phenotype zebrafish/sibling	卡方 χ^2	P 值 P value
1	25/78	1:3.12	2.15	0.143
2	35/105	1:3	0	1
3	44/124	1:2.82	0.127	0.722

表 4　测序数据质控及与参考基因组比对结果

Table 4　Sequencing data quality control and results of comparison with reference genome

编号 Atom number	清洁读值 Clean_reads	质量值 30(%) Q30	GC 含量(%) GC	正确比对率(%) Properly_mapped	平均深度 Ave depth
R01	97,471,743	90.25	38.45	90.30	17×
R02	106,703,350	89.80	37.72	92.73	17×
R03	238,971,525	89.47	37.72	92.15	42×
R04	255,944,524	89.67	38.08	92.41	42×

注:Q30(%):质量值大于等于 30 的碱基占总碱基数比率;正确比对率:比对到参考基因组且方向正确的 reads 数占总 reads 数的百分比。

Note. Q30(%), Ratio of bases number with quality value greater or equal to 30/total bases number. Properly_mapped, Ratio of reads match to the reference genome and correctly oriented to total reads.

表 5　SNPs 过滤统计数据

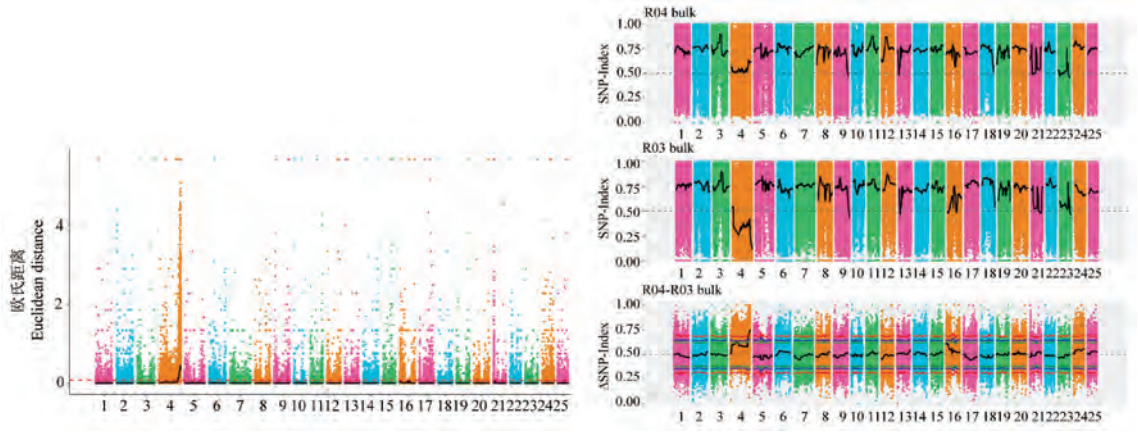
Table 5　Statistical datas of filter SNPs

SNP 数量 Total number	多个等位基因的位点 Multi-allelic loci	测序深度小于 4 的位点 Sites with coverage depth less than 4	混池基因型一致的位点 Sites with same genotype in two pools	与亲本来源不一致的位点 Non-parental origin sites	高质量 SNPs High quality SNPs
12950890	51244	237933	2421782	6875360	3364571

表 6　InDels 过滤情况统计

Table 6　Statistics of filter InDels

InDel 总数 Total InDels	多个等位基因的位点 Multi-allelic loc	测序深度小于 4 的位点 Sites with coverage depth less than 4	混池基因型一致的位点 Sites with same genotype in two pools	与亲本来源不一致的位点 Non-parental origin sites	高质量 InDels High quality InDels
3497763	216050	110455	581790	1811334	778134



注:A:横坐标为染色体名称,彩色点为 SNP 位点 ED 值,黑色线为拟合后 ED 值,红色虚线为关联阈值的显著性,ED 值越高关联效果越好;B:横坐标为染色体名称,彩色点为计算后的 SNP-index 值(或 ΔSNP-index 值),黑色线为经过拟合的 SNP-index 值(或 ΔSNP-index 值)。从上至下分别为隐性混池 SNP-index 值分布图、显性混池 SNP-index 值分布图、ΔSNP-index 值分布图,下图中绿色阈值线置信度为 0.90、蓝色阈值线置信度为 0.95、红色阈值线置信度为 0.99。

图 2　SNP 关联分析 ED 和 SNP 关联值在染色体上的分布情况

Note. A, Abscissa is the chromosome name, the colored dots are the ED values of the SNP sites, the black line is the ED value after fitting, and the red dotted line is the significance of the association threshold. The higher the ED value, the better the association effect. B, Abscissa is the chromosome name, the colored dots are the calculated SNP-index (or ΔSNP-index) values, and the black line is the fitted SNP-index (or ΔSNP-index) value. From top to bottom are the distribution of SNP-index values of recessive pools, the distribution of SNP-index values of dominant pools, and the distribution of ΔSNP-index values. Green threshold line in the figure below has a confidence level of 0.90, and the blue threshold line the confidence level is 0.95 and the red threshold line confidence level is 0.99.

Figure 2　Distribution of the associated values of ED and SNP on chromosomes in SNP correlation analysis

关的突变位点定位在 Chr4：77940617～78006239，一个 65.6 kb 的区域。

根据斑马鱼基因组数据库提供的基因注释信息，该区间内包含 6 个注释基因（表 7），可能为该小头小眼震颤突变体斑马鱼的候选基因。将这些基因与 BSA 结果中的 174 个非同义突变的基因和 30 个移码突变基因比对，发现这里有两个 *LOC*108183534 和 *zgc*：194336 不包含在内。因此，发生非同义突变的有 4 个，分别为 *arfgap*3，*zgc*：113921（*ikbip*），*terfa* 和 *zgc*：85975（*ckap*4）。

2.6 候选基因序列比对

为获取表型相关的候选基因，首先对候选基因的编码序列进行测序。利用扩增候选基因 CDS 全长的引物，扩增编码序列。通过与 Ensemble 中公布的参考序列比对发现，*arfgap*3、*ikbip* 和 *ckap*4 发生同义突变的位置均属于 SNP，因此不可能与该致死表型相关。而在比对 *terfa* 的序列时发现，*mise* 在第二外显子发生 c.236 G>A 突变，密码子 TGG 突变为 TAG 终止密码子，导致翻译提前终止（图 4）。

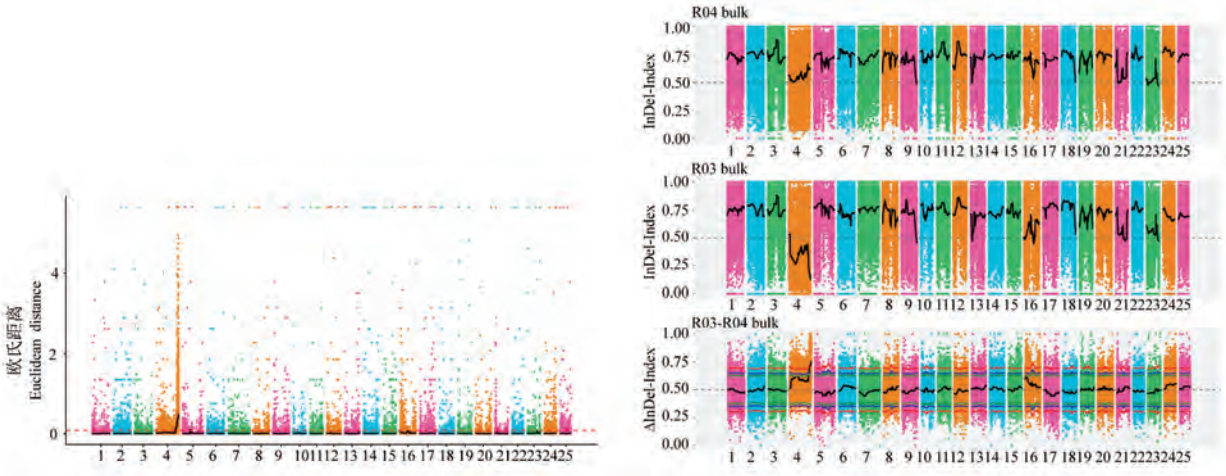
2.7 morpholino 验证候选基因

基于突变体的测序结果，合成靶向 *terfa* 基因剪接抑制 morpholino，注射单细胞期斑马鱼。注射 3 d 后观察表型发现，注射 *terfa*-sMO 的斑马鱼出现类似 *terfa* 上述点突变的小头小眼表型（图 5A），成功复制了 *terfa* 缺失的表型。而注射错配对照 *terfa*-scMO 的斑马鱼，没有类似的表型改变（图 5B）。说明 *terfa* 的剪接抑制的基因型变化是导致小头小眼表型的原因。

表 7 定位区间内的候选基因注释

Table 7 Annotation of candidate gene in the positioning interval

基因简称 Gene symbol	基因全称 Gene name
zgc：194336	zgc：194336
*arfgap*3	ADP-ribosylation factor GTPase activating protein 3
zgc：113921	IKBKB interacting protein
terfa	Telomeric repeat binding factor a
zgc：85975	Cytoskeleton associated protein 4
*LOC*108183534	Uncharacterized LOC108183534

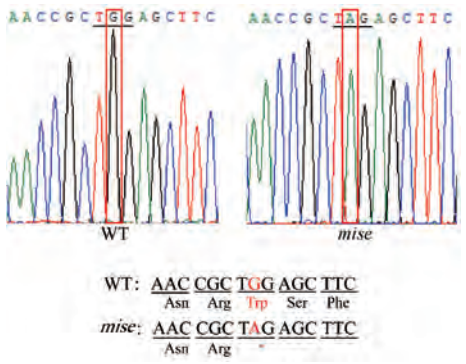


注：A：横坐标为染色体名称，彩色点为 InDel 位点 ED 值，黑色线为拟合后 ED 值，红色虚线为关联阈值的显著性，ED 值越高关联效果越好。B：横坐标为染色体名称，彩色点为计算后的 InDel-index 值（或 ΔInDel-index 值），黑色的线为经过拟合的 InDel-index 值（或 ΔInDel-index 值）。从上至下分别为隐性混池 InDel-index 值分布图、显性混池 InDel-index 值分布图、ΔInDel-index 值分布图，下图中绿色阈值线置信度为 0.90、蓝色阈值线置信度为 0.95、红色阈值线置信度为 0.99。

图 3 InDel 关联分析中，ED 和 InDel 关联值在染色体上的分布情况

Note. A, Abscissa is the chromosome name, the colored dots are the ED value of the InDel site, the black line is the ED value fter fitting, and the red dotted line is the significance of the association threshold. Higher the ED value, the better the association effect. B, Abscissa is the chromosome name, the colored dots are the calculated InDel-index (or ΔInDel-index) value, and the black line is the fitted InDel-index (or ΔInDel-index) value. From top to bottom are the distribution map of InDel-index value of recessive mixed pool, the distribution map of InDel-index value of dominant mixed pool, and the distribution map of ΔInDel-index value. Green threshold line in the figure below has a confidence level of 0.90 and the blue threshold line the confidence level is 0.95 and the red threshold line confidence level is 0.99.

Figure 3 Distribution of the associated values of ED and InDel on chromosomes in InDel correlation analysis

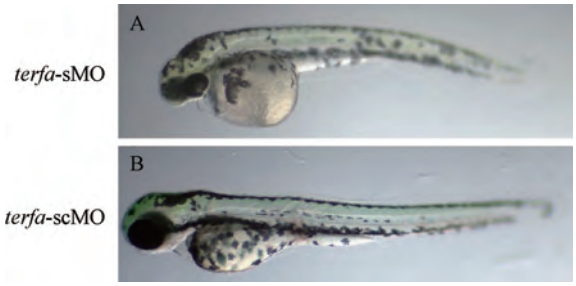


注：A：WT 与 *mise* 突变体的测序峰图，红框分别标注了 *mise* 的突变碱基和 WT 相应位置的碱基；B：WT 与 *mise* 相应的密码子和氨基酸，红字标注了碱基和氨基酸的变化。

图 4　*mise* 突变体的密码子及氨基酸变化。

Note. A, Sequencing peak diagrams of WT and *mise* mutants respectively, and red boxes indicate the mutated bases of *mise* and corresponding bases of WT. B, Codon and amino acid corresponding to WT and *mise*, and the changes of bases and amino acids are marked in red.

Figure 4　Codon and amino acid changes in *mise* mutant



注：A：注射 *terfa* 对照 morpholino（*terfa*-scMO）的表型；B：注射 *terfa* 剪接抑制 morpholino（*terfa*-sMO）的表型。

图 5　WT 斑马鱼注射 *terfa* 剪接抑制 morpholino 及对照 3 d 后的表型观察

Note. A, Phenotype after *terfa*-scMO injection. B, Phenotype after *terfa*-sMO injection.

Figure 5　Phenotype of WT zebrafish injected by *terfa* splicing morpholino and corresponding control 3 days after injection

3　讨论

原发性小头畸形是一种大脑发育障碍，目前已知的 MCPH 的致病基因参与了多种细胞过程，超过一半的 MCPH 的致病基因编码参与中心粒生物发生的中心体蛋白[10]，部分参与 DNA 复制和修复[11-12]，胞质分裂[13]、着丝粒功能[14]、跨膜或细胞内转运[15]以及 Wnt 信号自噬[16]等。可见虽然遗传性小头畸形综合征相对罕见，但对综合征的研究可以揭示对神经祖细胞、大脑大小和人脑进化的调节至关重要的分子机制。

本文作者在斑马鱼繁育过程中，发现了一个能稳定遗传的小头畸形突变体，为研究小头相关分子机制，首先要明确导致突变表型的基因。对于突变基因的初步定位，本研究采用了基因组重测序结合 BSA 的分析方法，快速确定了突变所在的染色体。自 2013 年 4 月英国桑格研究所（Wellcome Trust Sanger Institute）完成了斑马鱼的参考基因组的测序和全基因组分析[17]以来，已发布了 11 个斑马鱼基因组拼接版本，是继人类和小鼠以后，第三大高质量的有参基因组。BSA 连锁分析技术，不需要构建复杂定位群体，只需要选择群体中两极端性状个体构建混池，对照参考基因组，通过计算变异（SNP，In/Del，CNV 等）及其频率，即可快速定位与目的基因紧密连锁的分子标记。相较于限制性长度多态性 RFLP，简单序列重复数 SSR 和短串联重复序列 STR 等分子标记方法，SNP 具有数量多、覆盖密度大、遗传稳定性强、可实行批量化等优点。斑马鱼的多态性远高于大小鼠等实验动物，不同品系间的斑马鱼的 SNPs 高达 29.9%～60%[18]。本研究通过二代测序就获得了 3364571 个高质量的 SNP 位点。为基因定位提供了丰富的分子标记。由于高通量测序的获得的变异信息量大，因此基于 SNP 的 BSA 连锁分析方法较传统的通过构建遗传图谱进行连锁分析获得的定位区间，具有关联区间范围小，定位准确，成本低，耗时短的优点。

在对目的基因进行精细定位中，本研究基于上述的高通量测序结果中 SNP 的二等位信息以及连锁区间，开发 KASP 标记对 SNP 进行基因分型。通过计算交换率获得与突变基因连锁的区间。KASP 基因分型技术具有稳定性高、准确性高和成本低廉等优势，日益在高通量 SNP 分型中被广泛应用[19]。目前这种基因定位的策略主要广泛应用于农作物相关性状的基因定位[20-21]。在斑马鱼突变基因定位中也略有涉及[22]，大大加速了基因定位的效率。斑马鱼突变基因的定位之所以可以套用类似的思路，与斑马鱼的基因组特点和特殊的生物学特性有关。一是，如前所述，斑马鱼存在丰富的 SNP[18, 23]，密集的分子标记将获得更小的连锁区间；其次，斑马鱼子代数量大，在精确定位时，可以通过获得足够数量的子代分析个体的基因型计算交换率获得与性状连锁的位点。

本研究通过定位一个跟小头畸形相关的基因座，摸索了一套适用于斑马鱼表型相关基因定位的方法，为斑马鱼突变体相关基因定位提供有益的参考。也为 *mise* 基因的克隆和功能研究奠定了基础。

参考文献：

[1] Jayaraman D, Bae BI, Walsh CA. The genetics of primary microcephaly [J]. Annu Rev Genomics Hum Genet, 2018, 19: 177-200.

[2] Kleber de Oliveira W, Cortez-Escalante J, De Oliveira WT, et al. Increase in reported prevalence of microcephaly in infants born to women living in areas with confirmed Zika virus transmission during the first trimester of pregnancy-brazil, 2015 [J]. Morb Mortal Wkly Rep, 2016, 65(9): 242-247.

[3] Cañestro C, Yokoi H, Postlethwait JH. Evolutionary developmental biology and genomics [J]. Nat Rev Genet, 2007, 8(12): 932-942.

[4] 黄春念, 张晶晶. 模式动物斑马鱼在中枢神经系统疾病研究中的应用 [J]. 中国实验动物学报, 2018, 26(3): 392-397.

[5] Monte W. The Zebrafish Book: a guide for the laboratory use of zebrafish (*Danio rerio*) [Z]. Oregon: University of Oregon Press, 1995.

[6] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. Bioinformatics, 2009, 25(14): 1754-1760.

[7] Hill JT, Demarest BL, Bisgrove BW, et al. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq [J]. Genome Res, 2013, 23(4): 687-697.

[8] Fekih R, Takagi H, Tamiru M, et al. MutMap$^+$: genetic mapping and mutant identification without crossing in rice [J]. PLoS One, 2013, 8(7): e68529.

[9] Takagi H, Abe A, Yoshida K, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations [J]. Plant J, 2013, 74(1): 174-183.

[10] Megraw TL, Sharkey JT, Nowakowski RS. Cdk5rap2 exposes the centrosomal root of microcephaly syndromes [J]. Trends Cell Biol, 2011, 21(8): 470-480.

[11] Rai R, Dai H, Multani AS, et al. BRIT1 regulates early DNA damage response, chromosomal integrity, and cancer [J]. Cancer Cell, 2006, 10(2): 145-157.

[12] Shen J, Gilmore EC, Marshall CA, et al. Mutations in PNKP cause microcephaly, seizures and defects in DNA repair [J]. Nat Genet, 2010, 42(3): 245-249.

[13] Harding BN, Moccia A, Drunat S, et al. Mutations in citron kinase cause recessive microlissencephaly with multinucleated neurons [J]. Am J Hum Genet, 2016, 99(2): 511-520.

[14] Genin A, Desir J, Lambert N, et al. Kinetochore KMN network gene CASC5 mutated in primary microcephaly [J]. Hum Mol Genet, 2012, 21(24): 5306-5317.

[15] Sheen VL, Ganesh VS, Topcu M, et al. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex [J]. Nat Genet, 2004, 36(1): 69-76.

[16] Kadir R, Harel T, Markus B, et al. ALFY-controlled DVL3 autophagy regulates Wnt signaling, determining human brain size [J]. PLoS Genet, 2016, 12(3): e1005919.

[17] Howe K, Clark MD, Torroja CF, et al. The zebrafish reference genome sequence and its relationship to the human genome [J]. Nature, 2013, 496(7446): 498-503.

[18] Stickney HL, Schmutz J, Woods IG, et al. Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays [J]. Genome Res, 2002, 12(12): 1929-1934.

[19] Semagn K, Babu R, Hearne S, et al. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement [J]. Mol Breeding, 2014, 33(1): 1-14.

[20] Amitha Mithra SV, Kar MK, Mohapatra T, et al. DBT propelled national effort in creating mutant resource for functional genomics in rice [J]. Curr Sci India, 2016, 110(4): 543-548.

[21] Lun Y, Wang X, Zhang C, et al. A CsYcf54 variant conferring light green coloration in cucumber [J]. Euphytica, 2016, 208(3): 509-517.

[22] Mackay EW, Schulte-Merker S. A statistical approach to mutation detection in zebrafish with next-generation sequencing [J]. J Appl Ichthyol, 2014, 30(4): 696-700.

[23] Stickney HL, Schmutz J, Woods IG, et al. Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays [J]. Genome Res, 2002, 12(12): 1929-1934.

〔收稿日期〕2021-12-15