

· 论 著 · DOI:10. 3969/j. issn. 1672-9455. 2019. 08. 005

已知位点突变检测中测序引物的规范化设计

谢嘉玲, 申 霞, 张 蕾, 顾怡瑾, 董 磊[△]

(上海交通大学医学院附属瑞金医院病理科 200021)

摘要:目的 作为突变检测的经典方法,第 1 代测序法(Sanger 测序法)可以清晰地读取基因中碱基置换、颠换、缺失和插入等改变,临床已知位点突变引物的设计涉及数据库查询、转录本及位点的核查等环节,过程繁琐。本研究尝试提出一套实用、易操作的已知位点突变引物设计流程。**方法** 随机抽取 2018 年 7 月在该院病理科分子病理实验室进行 BRAF(V600E)基因检测和 FOXL2(C134W)基因检测的甲状腺乳头状癌患者和颗粒细胞瘤患者检测结果各 1 例。对于 LRG 网站中已录入基因组、氨基酸、蛋白质 3 种序列的基因,进入 LRG 网站下载 3 种序列,使用 Lasergene(DNASTAR, USA)软件通过 3 种序列的人工比对,确定基因版本号无误,根据突变位点前后 400 个序列,设计引物;针对 LRG 网页还未收入 3 种序列的基因,先人工通过 NCBI 各个基因库搜索相应的基因组、氨基酸、蛋白质 3 种序列,并核实 3 种序列的版本匹配,再根据上述流程设计引物。**结果** 引物设计完成后,经上海生物工程公司订购,使用已知突变结果的 FFPE 样本作检测,用 ABI3500Dx (Thermo Fisher, Scientific, America)测序,观察测序结果,确认所验证基因突变点包含在该序列内。**结论** 目前临床治疗及报告习惯使用氨基酸位置提示突变,而引物设计需要通过核酸位点,因此确认 3 种序列基因版本号很重要;对于部分已收入 LRG 网站的基因,一些版本号也会与临床报告中的突变位点有出入,所以即使是有 LRG 号的基因也要核查基因组、氨基酸、蛋白质 3 种序列。

关键词:已知位点突变; Sanger 测序法; 引物设计**中图法分类号:**R331**文献标志码:**A**文章编号:**1672-9455(2019)08-1023-04

Standardized primer designing process for point mutation detection

XIE Jialing, SHEN Xia, ZHANG Lei, GU Yijing, DONG Lei[△]

(Department of Pathology, Ruijin Hospital, Medical School of Shanghai Jiaotong University, Shanghai 200021, China)

Abstract: **Objective** Sanger sequencing is a classical method to detect the gene substitution, deletion and insertion. Designing primer for point mutation detection is a multistep process, including database query, primer creation, sequence and position confirmation. In this study, we present a set of easy-to-use process for primer design. **Methods** Two FFPE samples, a BRAF (V600E) mutated papillary carcinoma of thyroid gland and a FOXL2 (C134W) wild type ovarian granulosa cell tumor, were selected from Department of Pathology, Ruijin Hospital of Shanghai Jiao Tong University School of Medicine. For the genes included by Locus Reference Genomic (LRG) database, three sequences (NG, NM, NP) of them were downloaded. Software was applied for position confirmation. For genes were not included in LRG, their sequences were downloaded from NCBI database. Primers were designed according to the results of the above process. **Results** The primers were verified using Sanger sequencing on the above FFPE samples. The results showed that the primers were correctly designed. **Conclusion** Confirmation the version of NG, NP, NM sequences version is important step for accurate primer design. The position of sequence of some genes in LRG database may not fit with the position which was used in clinical situation. So, NG, NP, NM sequences need to be checked even for those LRG included genes.

Key words: point mutation; Sanger sequencing; primer design

作为突变检测的经典方法, Sanger 测序法可以清晰地读取基因中碱基置换、颠换、缺失和和插入等改变,还可以对待检测 DNA 标本的某一个片段的碱基逐一精确检测,然后与已知的标准样本序列比较,以

低误读率、低成本的优势寻找新的突变位点。在一定的灵敏度范围内,它还可作为聚合酶链反应(PCR)和下一代测序法(NGS)检测突变的可靠验证平台^[1-5]。

成功的引物设计是完成 Sanger 测序的重要步

骤。在临床分子病理的实践中,常常要求对需要检测的定点突变进行引物设计。当实验室需要开展对已知点突变或小片段插入缺失(本文中统称为定点突变)的检测时,就需要针对该位点进行引物设计;用其他方法,如下一代测序得到的突变位点也常常需要用 Sanger 测序进行验证^[6]。

临床应用,点突变或小片段插入缺失必须做到规范化、可溯源^[7-8]。同一基因可有多个不同的转录本,该基因的同一位点会在不同的转录本里有不同位置。所以只有使用统一的、或可溯源的转录本设计引物,才能确保点突变引物的准确性^[9]。临床定点突变引物设计也要遵循引物设计的一般规则,如引物的长度、鸟嘌呤和胞嘧啶所占的比例(CG 含量)、Tm 值、ΔG 值、二聚体及错配修复等^[10-13]。此外,临床定点突变引物的设计还涉及数据库查询、转录本及位点核查等环节,过程比较繁琐^[14]。目前较少见报道提供相关的简明、易行的操作流程^[15]。本研究尝试提出一套临床定点突变引物设计工作流程,以利于分子病理地工作者更方便、准确地开展 Sanger 测序的定点突变检测。

1 资料与方法

1.1 一般资料 随机抽取 2018 年 7 月在本院病理科分子病理实验室进行 BRAF(V600E)基因检测和 FOXL2(C134W)基因检测的甲状腺乳头状癌患者和颗粒细胞瘤患者检测结果各 1 例。经检测,BRAF(V600E)基因检测到该位点突变,FOXL2(C134W)基因检测未能检测到该位点突变。

1.2 方法

1.2.1 拥有明确 LRG 号的基因(BRAF V600E) 当明确所需引物的基因名和氨基酸位点后,在 LRG 网站(<https://www.lrg-sequence.org/>)查找其 LRG 号(LRG_299),查询并下载该基因的 FASTA 文件——LRG_299[(其中包含 NG_007873、NM_004333 和 NP_001341538)3 种序列(图 1),分别代表人类基因组序列、mRNA 序列和氨基酸序列]。用软件 Lasergene(DNASTAR,USA)中的 Editseq 软件打开下载的 3 种序列,在 mRNA 序列中查找第 1 个 ATG,即 BRAF NM 序列中的第 226 位,切去这个 ATG 以前的序列。再用 Lasergene 中的 Goodies 选项中的 Translate DNA 功能将剩余的碱基序列转为氨基酸序列。将 mRNA 序列与 NP 序列进行比对,确定两者一致。然后将氨基酸位点(V600E)即 600×3,得到相应的核酸位点 1800,并选取这一位点前 10 个碱基至后 10 个碱基之间的片段(TAGCTACAGT-GAAATCTCGA)。

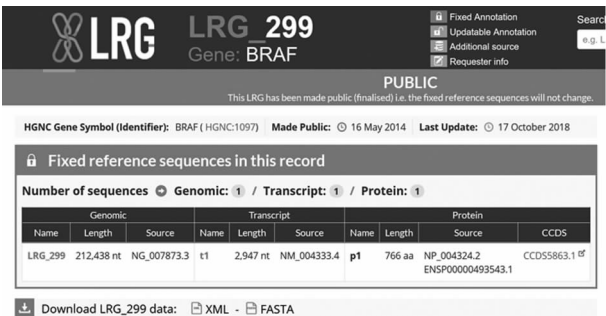


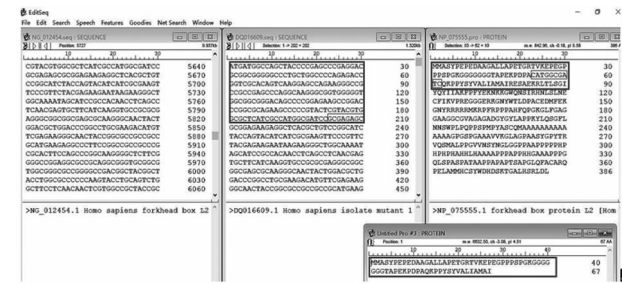
图 1 LRG 官网上 LRG_299 即 BRAF(V600E)信息

找出目的片段,并对其特异性核查,将这 20 个序列放入 NG 序列中,找到序列相同的位置。在 DNA 序列上选取前后各 200 位的碱基序列,含该 20 个碱基序列(即该基因序列 LRG_299g 的第 176 221~176 620 位),共 400 个碱基的长度。用 BLAST(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)比对该片段,确保其特异性。

针对以上序列进行引物设计,用在线软件 primer3(<http://bioinfo.ut.ee/primer3-0.4.0/>)将复制好的 400 bp 长的序列复制进去,选择长度为 250~350 bp 长度的引物,得到引物一对。序列为:3'-TCA TAA TGC TTG CTC TGA TAG GA,5'-TTT CTA GTA ACT CAG CAG CAT CT。最后进入 BLAST PRIMER(<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>)对引物的特异性进行验证。

1.2.2 LRG 未收录的基因 LRG 未收录或虽已收录,但转录位点与已知信息有出入的基因(FOXL2 C134W)找出可溯源、准确的 mRNA、基因组、蛋白质 3 个序列号;以 FOXL2 为例,目前 LRG 网页中还没录入该基因相关信息。在明确所需引物的基因名和氨基酸位点后,如 FOXL2 基因突变点为 C134W,在 NCBI(National Center for Biotechnology Information,<https://www.ncbi.nlm.nih.gov/>)上输入该基因,在侧拉选项“Results by taxon”下方选择“Homo sapiens”,首先显示了 mRNA 序列(NM_023067)及蛋白质序列(NP_075555),在 NM 序列中查找第 1 个 ATG,删除 ATG 之前的序列,用 EditSeq 中的“Goodies”,选择“Translate DNA”功能将 mRNA 序列转为氨基酸序列,并与旁边的 NP 序列人工比对,发现没有完全匹配。

此时,需重新查找其他匹配的转录本版本。进入 NCBI 中的 GenBank,发现该库中 FOXL2 对应转录本为 DQ016609。在该序列中查找第 1 个 ATG(第 147 位),删除 ATG 之前的 146 位序列,用 Goodies 中的 Translate DNA 功能将 RNA 序列转为氨基酸序列,并与旁边的 NP 序列人工比对。经验证,该序列为 NP_075555 号对应的转录本核苷酸序列(图 2)。



注:框内为氨基酸序列去除第 1 个 ATG 之前所有序列后,氨基酸翻译为蛋白后与原蛋白序列的对比

图 2 FOXL2 基因的 3 种序列

确认 NM、NP 序列对应无误后,将氨基酸位点(C134W)即 134X3,得到相应的核酸位点 402,并选取这一位点前 10 个碱基至后 10 个碱基之间的片段(ACCCGGCCTGCGAAGACATG)。将这 20 个序列放入 NG 序列中。对该序列,笔者团队适当延长目标序列,将目标序列扩展为 DNA 前后共 446 位序列(即该基因序列 NG_012454.1 的第 5 630~6 075 位),根据上述方法进行引物设计,得到引物序列分别为:(1) 3'-GCC TCA ACG AGT GCT TCA TC, 5'-GGA AGC CAG ACT GCA GGT A;(2) 3'-ATC GCG AAG TTC CCG TTC TA, 5'-TTG AGG AAG CCA

GAC TGC AG;(3) 3'-AAC CTC AGC CTC AAC GAG TG, 5'-GTT GTT GAG GAA GCC AGA CTG;(4) 3'-ACA ACC TCA GCC TCA ACG A, 5'-GGC CAC GAG TTG TTG AGG AA。最后进入 BLAST PRIMER 对引物的特异性验证,经核查,4 对引物均可使用。

2 结 果

2.1 选取 BRAF (V600E) 及 FOXL2 (C134W) 引物 经过上述操作,针对 BRAF(V600E)基因,得到引物序列 1 对: 3'-TCATAATGCTTGCTCTGAT-AGGA, 5'-TTTCTAGTAACTCAGCAGCATCT。3'、5'端引物长度均为 23 bp,3'端 Tm 值为 57.19,5'端 Tm 值为 57.33,GC 百分比均在 39.15,没有发卡结构(Hairpin),二聚体(Dimer)ΔG 值为-5.6,错配修复(False Priming)ΔG 值为-9.5,引物二聚体(Cross Dimer)ΔG 值为-5.0。各值均符合引物设计一般规则。针对 FOXL2(C134W)基因,得到 4 对引物。4 对引物各值均基本符合引物设计的一般规则,引物 1 相对于其他 3 对引物,无发卡结构及错配修复,引物二聚体及上下游引物之间的二聚体 ΔG 值较小(表 1),故选取 1 号引物进行订购及验证。

表 1 FOXL2 4 对引物各值比较

特征	引物 1	引物 2	引物 3	引物 4
引物长度(3'端/5'端)	20 bp/19 bp	23 bp/23 bp	20 bp/21 bp	19 bp/20 bp
Tm 值(3'端/5'端)	59.28/59.02	59.19/59.96	59.97/59.12	58.87/60.53
GC(3'端/5'端)	55.00/57.80	50.00/55.00	55.00/52.38	52.63/55.00
Hairpin ΔG	—	-0.4	0.0	—
Dimer ΔG	-9.4	-4.5	-4.2	-10.7
Cross Dimer ΔG	-6.1	-5.6	-6.2	-9.4
False Priming	—	—	-14.7	-16.8

注:— 为此项无数据

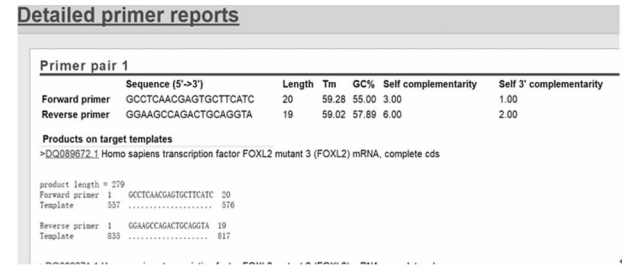
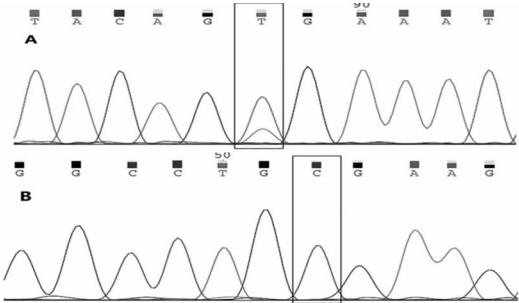


图 3 FOXL2 基因使用 BLAST 网页后设计出的引物

2.2 引物验证 核查各引物序列无误后,向引物合成厂商(上海生工生物公司)订购 BRAF (V600E)、FOXL2(C134W)基因引物各 1 对。用石蜡包埋甲醛固定的标本对引物进行验证。将被测模板扩增、纯化后,用毛细电泳分析仪(ABI3500Dx, ThermoFisher Scientific, America)得到被检测的基因序列。将数据用 Chromas (Technelysium, Australia) 打开,测序结果无明显干扰数据,没有杂峰、重叠峰,测序片段长度

大于 400 bp。经验证, FOXL2 (C134W)、BRAF (V600E)引物可用,分别检测到 BRAF(V600E)测序序列为:CTAGCTACAGTGAAA (BRAF 突变标本,突变碱基为下划线所标位置);FOXL2(C134W)测序序列为 CGGCCTGCGAAG(FOXL2 阴性标本),见图 3、4。



注:A 为 BRAF(V600E)基因引物设计完成后,使用已知突变结果的 FFPE 标本检测出的测序图;B 为 FOXL2(C134W)基因引物设计完成后,使用已知突变结果的 FFPE 标本检测出的测序图

图 4 治疗序列图

3 讨 论

对于 Sanger 测序引物设计而言,查找目的基因突变片段所在序列的 mRNA 序列、氨基酸序列、基因组序列是设计目的基因突变点引物非常重要的一步。目前临床报告基因位点突变还是习惯于使用氨基酸位置,但引物设计又必须使用核酸位置。因此,目的基因的基因组序列(NC/NG 号)及 mRNA 序列(NM 号)相匹配,mRNA 序列(NM 号)及氨基酸序列(NP 号)相匹配,对引物设计和临床报告均非常重要。

笔者选择的用于引物设计的目的片段长度为 400 bp,一般情况下,Primer3 都可生成相应的引物,但也会有不能生成引物的情况出现。此时,可以尝试增加或减少碱基片段,找到合适引物。

本研究发现,在 LRG 网页上会出现部分基因收录的氨基酸、mRNA 序列与临床报告中所取的氨基酸位点偏差情况。因此,基因各序列的准确核查很重要。必要时检索相关的原始报道,以明确突变的准确位点。如 MYD88(LRG_157),MYD88 突变点位 L265P,LRG 提供的氨基酸序列中的 265 位并非“L”(亮氨酸)相应碱基代码,造成笔者也无法验证 mRNA 序列(NM 号)是否正确,也无法以此核查基因组序列(NG 号)中的特异序列。

又如 H3F3B 基因的转录起点后有多多个 ATG,如何选择它的翻译起始位点就必须准确地核对各个版本号,经核查,H3F3B 的 NM 后 ATG 筛选是从第 4 个 ATG 而非常规的第 1 个 ATG 开始。因此,在引物设计过程中,要准确核实基因转录本号的匹配度,以减少因版本号的不匹配,对引物设计的不利影响。

NCBI 和 ENSEMBL 两大数据库都是基因查找常用的数据库,NCBI 数据库中会清晰显示出氨基酸、蛋白等版本号;而 ENSEMBL 数据库中会显示基因相对应的 LRG 号和基因突变位置信息,两个数据库各有特色和重点,但是本文推荐使用 NCBI 数据库,方便对于基因氨基酸和蛋白的版本追踪、溯源和参考,并制作 Sanger 测序流程单(其中包含基因名称、mRNA 序列号、蛋白序列号、基因组序列号,引物长度、引物序列、实验验证结果及验证人等信息)为日后实验复核和临床报告发布提供便利。

本文中使用的 Lasergene(DNASTAR, USA)为付费软件。免费款的 UGENE(Unipro GENE, Russia)也可以作为替代软件使用。

综上所述,可以通过便捷的规范化方式设计 Sanger 测序引物。但是由于 Sanger 测序引物设计流程中有一定不确定性(如各基因不同版本号位点不同;部分基因的特殊设计规则等),目前很难统一用全自

动化的流程或软件设计。期待今后可以出现更为规范化和自动化的引物设计方式出现。

参考文献

- [1] 李婕,张鹏,李小宁.高通量测序在分子诊断中的应用[J].锦州医科大学学报,2018,39(2):106-108.
- [2] 邵向阳,徐伟文.下一代测序(NGS)技术的发展及在肿瘤研究的应用[J].分子诊断与治疗杂志,2016,8(5):289-296.
- [3] LIN M T, MOSIER S L, THIESS M, et al. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing[J]. Am J Clin Pathol, 2014, 141(6):856-866.
- [4] STROM SP, LEE H, DAS K, et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory[J]. Genet Med, 2014, 16(7):510-515.
- [5] BECK T F, MULLIKIN J C, NISC Comparative Sequencing Program, et al. Systematic evaluation of sanger validation of next-generation sequencing variants[J]. Clin Chem, 2016, 62(4):647-654.
- [6] MU W, LU H M, CHEN J, et al. Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing[J]. J Mol Diagn, 2016, 18(6):923-932.
- [7] 胡京美. BPES 综合征伴不孕症家系 FOXL2 基因突变检测及功能研究[D]. 济南:山东大学, 2017.
- [8] 梁莹. 肿瘤基因组结构变异的检测方法及应用研究[D]. 长沙:湖南大学, 2017.
- [9] FU W, O'CONNOR T D, JUN G, et al. Analysis of 6 515 exomes reveals the recent origin of most human protein-coding variants[J]. Nature, 2013, 493(7431):216-220.
- [10] 王静,范慧君,梁兴国.一种用于优化 PCR 引物设计的短链 DNA 熔点分析方法[J].中国生物化学与分子生物学报, 2016, 32(10):1168-1176.
- [11] 林刻智,楼迪栋.错配引物诱导酶切技术检测 GC 多态性[J].中国法医学杂志, 2009, 24(5):323-325.
- [12] 谭学贤.四种错配修复基因蛋白在结直肠癌中的表达及临床意义[D].广州:广州医科大学, 2017.
- [13] 翟中会,陈希南,王娟.利用 Primer Premier 5.0 进行引物设计[J].西北医学教育, 2008, 16(4):695-698.
- [14] 高梅,刘树人,李强,张凌峰.高 GC 含量 DNA 序列 PCR 扩增的引物设计[J].华南国防医学杂志, 2014, 28(7):633-638.
- [15] 张见影,伦志军,李正红. NCBI 基因序列数据库使用和检索方法[J].现代情报, 2003, 24(12):224-225.

(收稿日期:2018-09-11 修回日期:2018-12-18)