



二代测序技术在法医学的研究进展与应用^{*}

杨子豪

1. 华东政法大学, 上海;
2. 司法鉴定科学研究院, 上海

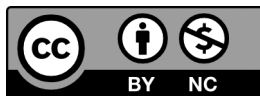
摘要 | 随着近些年来二代测序 (Second Generation Sequencing, SGS) 技术的兴起, 法医学者将其引入并应用于法医学领域, 从而充分发挥了其测序通量大、测序速度快且成本低的优势, 同时大大促进了法医学的发展以及在案件侦破中的应用。本文从测序技术在法医学领域内的应用历程展开, 回顾了二代测序技术在法医学遗传标记检测中的重要性。以现有法医学研究应用相关的 Illumina 和 Thermo Fisher Scientific 测序技术公司推出的二代测序技术平台为例, 探讨其在法医学领域中的应用现状及潜能。就目前研究而言, 这些测序平台可以完成 STR 遗传标记、SNP 遗传标记、mtDNA 及 mRNA 等遗传标记的测序。然而, 测序试剂盒中部分遗传标记的优化、推出及验证针对中国人群测序需求的测序试剂盒、制定一套具有广泛公信力的数据采集标准、测序数据分析软件的优化、与现有法医遗传学数据库的对接、测序数据使用中的伦理问题等均是决定二代测序技术能否代替 (或补充) 成熟 PCR 毛细管电泳技术以及普遍应用于刑事案件检测的关键。

关键词 | 法医遗传学; 二代测序; 遗传标记

Copyright © 2021 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



一、引言

为了解决一代测序技术成本高和通量低等缺陷, 响应科研人员对更为先进、更低成本、更高通量的测序技术的诉求, 基于大规模平行测序 (Massively Parallel Sequencing, MPS) 思想的二代测序技术 (Second Generation Sequencing, SGS) 亦称新一代测序技术 (Next Generation Sequencing, NGS) 应运而生。依靠二代测序技术研发的 SGS 测序平台主要包括基于焦磷酸测序原理的 Roche 公

司推出的 454 基因组测序仪^[1]、Illumina 公司推出的基于边合成边测序原理的 Illumina 测序平台^[2]以

^{*} 基金项目: 十三五国家重点研发计划 (项目编号: 2018YFC0830400)。

[1] Rothberg J M, Leamon J H. The development and impact of 454 sequencing [J]. Nat Biotechnol, 2008 (26): 1117-1124.

[2] Bentley D S, Balasubramanian S, Swerdlow H, et al. Accurate whole human genome sequencing using reversible terminator chemistry [J]. Nature, 2008 (456): 53-59.

及 Thermo Fisher Scientific 公司推出的基于离子流半导体芯片 DNA 测序技术的 Ion Torrent 测序平台^[1]。与第一代测序技术 (Sanger Sequencing) 相比, 第二代测序技术克服了 Sanger 测序操作繁琐且成本高的缺点, 同时提高了测序通量和速度^[2]; 与第三代测序技术 (Single Molecule Sequencing, SMS) 相比, 第二代测序技术降低了成本的同时测序错误率较低^[3]。目前, 由于市场运营等问题, Roche 公司宣布其 454 测序仪退出市场, Illumina 平台和 Ion Torrent 平台成为当下全球测序市场上主要的 SGS 测序平台, 在法医学领域引起了学者的广泛关注和研究热潮。本文就应用 SGS 测序技术检测法医遗传标记的研究进展及常见的应用到法医学领域的 SGS 测序平台进行介绍和展望。

二、应用 SGS 技术在法医遗传标记检测中的研究现状

个体的单位遗传性状作为标志用于法医

物证分析时, 这种遗传性状就称为遗传标记 (Genetic Marker, GM)。具有个体特异性的遗传标记的检测与分析是法医学进行个体识别的重要依据。目前在法医学中推广应用的遗传标记主要有以下三类: (1) 短串联重复序列 (Short Tandem Repeat, STR), 是一类存在于人类基因组 DNA 中的由 2 ~ 6 个碱基的串联重复单位组成的 DNA 序列^[4]。STR 基因座根据分布的染色体位置可分为常染色体 STR 基因座 (A-STR)^[5]、X 染色体 STR 基因座 (X-STR)^[6] 和 Y 染色体 STR 基因座 (Y-STR)^[7]; (2) 单核苷酸多态性 (Single nucleotide Polymorphism, SNP), 是人类基因组内特定位点上单个碱基序列的变异^[8]。在法医学领域 SNP 遗传标记可划分为个人识别 SNP (Individual Identifying SNP, IISNP)^[9]、祖先信息 SNP (Ancestry Informative SNP, AISNP)^[10]、表型信息 SNP (Penotype Informative SNP, PISNP)^[11]、Y-SNP^[12] 等;

[1] Rothberg J M, Hinz W, Rearick T M, et al. An integrated semiconductor device enabling non-optical genome sequencing [J]. Nature, 2011 (475): 348-352.

[2] Sanger F, Nicklen S, Coulson A R. DNA sequencing with chain-terminating inhibitors. 1977 [J]. Biotechnology, 1992 (24): 104-108.

[3] Harris T D, Buzby P R, Babcock H, et al. Single-molecule DNA sequencing of a viral genome [J]. Science, 2008 (320): 106-109.

[4] Butler J M. Chapter 5 - Short Tandem Repeat (STR) Loci and Kits [M] // J M Butler, Advanced Topics in Forensic DNA Typing: Methodology, 2012: 99-139.

[5] Gettings K B, Aponte R A, Vallone P M, et al. STR allele sequence variation: Current knowledge and future issues [J]. Forensic Sci Int Genet, 2015 (18): 118-130.

[6] Yang Z, Chen C, Zhang J, et al. Genetic polymorphisms in 16 X-STR loci analyzed in the She population from Zhejiang Province, China [J]. Leg Med (Tokyo), 2019 (39): 25-28.

[7] Diegoli T M. Forensic typing of short tandem repeat markers on the X and Y chromosomes [J]. Forensic Sci Int Genet, 2015 (18): 140-151.

[8] Butler J M. Chapter 12 - Single Nucleotide Polymorphisms and Applications [M] // J M Butler. Advanced Topics in Forensic DNA Typing: Methodology, 2012: 347-369.

[9] Pakstis A J, Speed W C, Fang R, et al. SNPs for a universal individual identification panel [J]. Hum Genet, 2010 (127): 315-324.

[10] Kidd K K, Speed W C, Pakstis A J, et al. Progress toward an efficient panel of SNPs for ancestry inference [J]. Forensic Sci Int Genet, 2014 (10): 23-32.

[11] Weise N A, Roberts K A, Hardy W R. Reliability of phenotype estimation and extended classification of ancestry using decedent samples [J]. Int J Legal Med, 2021 (135): 2221-2233.

[12] Karafet T M, Mendez F L, Meilerman M B, et al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree [J]. Genome Res, 2008 (18): 830-838.

(3) 其它, 如能够对组织特异性进行检测的 mtDNA^[1] 以及 mRNA^[2] 等。

(一) STR 遗传标记

从 20 世纪 80 年代至今, 人类基因组中广泛存在的微卫星 DNA 区域, 也被称为简单序列重复 (Simple Sequence Repeats, SSR) 或短串联重复 (Short Tandem Repeats, STR), 是由核心序列 2 ~ 6 个碱基组成的串联重复 DNA 序列。STR 序列在人类基因组中占 5% 左右, 平均每 6 ~ 10kb 就出现一个 STR 基因座, 其中大约一半具有遗传多态性。另外, STR 重复单位小从而杂合子个体的两个等位基因在长度上差异小且两等位基因间不存在差异扩增的问题, 因而易于 PCR 扩增成为法医学中最为常用的一类 DNA 遗传标记。目前国内外与法医学相关的 DNA 数据库主要是围绕 STR 基因座建立的。因此, SGS 技术要想在法医学领域推广应用就必须能对这一类遗传标记进行测序检测。然而要将 SGS 技术应用于 STR 基因座检测仍有以下问题需要解决: 一是现有平台测序读长的限制。STR 基因座扩增子长度跨度大, 大多数 SGS 测序平台的读长不能完全满足其检测的需求。二是检测和读取比对重复结构较多的复杂核心序列基因座或复合核心序列基因座 (如 D21S11、FGA 等) 的序列信息存在困难。三是要有符合法医使用习惯的且 STR 分型准确的生信分析软件, 在数据处理层面上, 不同重复结构的核心序列会使 STR 基

因座更加复杂, 倘若生信分析软件无法提供准确的 STR 分型, 即使得到正确的序列信息, 也需要大量的时间和人力去检查和清理数据。四是等位基因命名问题^[3], 通过测序揭示了 STR 基因座多态性的本质就是长度多态性和序列多态性, 面对 STR 基因座中具有长度一致但序列结构信息不一致的同等位基因该如何命名是应用 SGS 测序技术检测 STR 基因座迫切需要解决的问题。随着测序平台技术的进步以及法医科研人员和 SGS 测序平台的厂商们的不懈努力下取得了一定的成果, 早在 2012 年, 由 Bornman DM 等人^[4] 针对 13 个 CODIS 核心 STR 基因座的序列多态性通过 Illumina GAI 测序仪进行测序研究, 其结果获取 13 个 CODIS 核心 STR 基因座的全部等位基因序列, 这更加证实了 SGS 测序技术在 STR 基因座检测中的优势。之后, Illumina 公司就针对 Miseq 测序平台先后推出相关测序试剂盒, 如 PowerSeqTM Auto System^[5], ForenSeqTM DNA Signature Prep Kit^[6], 尤其是后者, 它可以同时检测超过 200 个遗传标记, 包含多种遗传标记 (Amelogenin、27 个 A-STR、7 个 X-STR、24 个 Y-STR、94 个 IISNP、22 个 PISNP 和 56 个 AISNP), 显示了 SGS 测序技术在法医学领域的多种应用方向, 同时也反映了采用 SGS 测序平台进行 STR 基因座研究的优势。相信随着 SGS 测序技术不断完善, 利用 SGS 测序技术检测 STR 基因座会更加成熟。

[1] Merheb M, Matar R, Hodeify R, et al. Mitochondrial DNA, a Powerful Tool to Decipher Ancient Human Civilization from Domestication to Music, and to Uncover Historical Murder Cases [J]. *Cells*, 2019, 8 (5).

[2] Zubakov D, Kokmeuer I, Ralf A, et al. Towards simultaneous individual and tissue identification: A proof-of-principle study on parallel sequencing of STRs, amelogenin, and mRNAs with the Ion Torrent PGM [J]. *Forensic Sci Int Genet*, 2015 (17): 122-128.

[3] Zeng X, King J L, Stoljarova M, et al. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing [J]. *Forensic Sci Int Genet*, 2015 (16): 38-47.

[4] Bornman D M, Hester M E, Schuetter J M, et al. Short-read, high-throughput sequencing technology for STR genotyping [J]. *Biotech Rapid Dispatches*, 2012: 1-6.

[5] Zeng X P, King J, Hermanson S, et al. An evaluation of the PowerSeq (TM) Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing [J]. *Forensic Science International-Genetics*, 2015 (19): 172-179.

[6] Guo F, Yu J, Zhang L, et al. Massively parallel sequencing of forensic STRs and SNPs using the Illumina ForenSeq DNA Signature Prep Kit on the MiSeq FGx Forensic Genomics System [J]. *Forensic Sci Int Genet*, 2017 (31): 135-148.

（二）SNP 遗传标记

单核苷酸多态性 (Single nucleotide Polymorphism, SNP) 是人类基因组中分布最广泛的二等位基因遗传标记。具有稳定可靠, PCR 产物较短以及适用于降解检材检测等特点。作为法医学中最常见的一类多态位点, 与 STR 基因座相比, 单个二等位基因 SNP 遗传标记所含的遗传信息含量比单个多等位基因的 STR 遗传标记低, 因而单个二等位基因 SNP 遗传标记的识别能力也低于单个多等位基因的 STR 遗传标记。但是二等位基因 SNP 遗传标记在基因组中分布广泛且数量多, 所以必须通过检测更多的 SNP 遗传标记, 才能满足个人识别和亲缘鉴定的需求。就 SNP 遗传标记的法医学应用来说主要可分为: (1) 个人识别: 一是降解检材个体身份信息识别, 当生物检材高度降解时, STR 基因座检测无法获得完整的分型信息, 而 SNP 遗传标记是发生在单个碱基上的变异, 设计的引物可尽量靠近变异区域因而得到较短的 PCR 扩增产物, 更适用于分析法医降解检材。如个人识别 SNP (Individual Identifying SNP, IISNP) 遗传标记, 在不同群体间具有较高的杂合度、扩增片段短 (通常 <200bp)、群体间分化度低以及不

与常用 STR 基因座发生连锁反应等特点, 在保证不同遗传标记间的中立性同时降低了遗传结构差异对 IISNP 遗传标记造成的影响, 使 IISNP 组合适用于不同地区的群体。早在 2006 年 Sanchez JJ 团队^[1]所发起的“SNPforID”计划就是旨在开发可用于 DNA 分析的 SNP 遗传标记检测方法, 通过 SNaPshot 技术构建的 52 个 SNP 遗传标记检测体系。Kidd KK 团队^[2]发表了用于个人识别 SNP 遗传标记的筛选标准。随后, 众多法医学者在构建更多的 SNP 位点的检测体系同时缩短检测片段方面取得了一定的成果^{[3][4]}。近年来随着测序技术的进步和测序平台的发展, 法医工作者开始关注应用 SGS 测序技术检测 SNP 遗传标记的目标片段以提升降解检材的识别能力^{[5][6][7][8]}, 如在 2017 年由 GUO Fei 等在 MiSeq FGx™ 平台上对 ForenSeq™ DNA Signature Prep 试剂盒 (包含 Amelogenin、27 个 A-STR、7 个 X-STR、24 个 Y-STR、94 个 IISNP、22 个 PISNP 和 56 个 AISNP) 进行验证, 结果显示 DNA 含量低至 200pg, 即可获得全部 SNP 位点的完整的分型结果。二是推断样本的种族来源, 正是由于 SNP 遗传标记具有低突变率, 突变率大约在 $10^{-10} \sim 10^{-8}$, 因而才比 STR 遗传标记更易在人群中稳定遗传。而这种能较准

[1] Sanchez J J, Phillips C, Borsting C, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification [J]. Electrophoresis, 2006 (27): 1713-1724.

[2] Kidd K K, Pakstis A J, Speed W C, et al. Developing a SNP panel for forensic identification of individuals [J]. Forensic Sci Int, 2006 (164): 20-32.

[3] Fordyce S L, Mogensen H S, Børsting C, et al. Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™ [J]. Forensic Science International: Genetics, 2015 (14): 132-140.

[4] Gelardi C, Rockenbauer E, Dalgaard S, et al. Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles [J]. Forensic Sci Int Genet, 2014 (12): 38-41.

[5] Cho S, Yu H J, Han J, et al. Forensic application of SNP-based resequencing array for individual identification [J]. Forensic Science International-Genetics, 2014 (13): 45-52.

[6] Gettings K B, Kiesler K M, Vallone P M. Performance of a next generation sequencing SNP assay on degraded DNA [J]. Forensic Science International-Genetics, 2015 (19): 1-9.

[7] Lee E Y, Lee H Y, Oh S Y, et al. Massively parallel sequencing of the entire control region and targeted coding region SNPs of degraded mtDNA using a simplified library preparation method [J]. Forensic Sci Int Genet, 2016 (22): 37-43.

[8] Jager A C, Alvarez M L, Davis C P, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories [J]. Forensic Sci Int Genet, 2017 (28): 52-70.

确地判断样本始祖来源的 SNP 遗传标记称为祖先信息 SNP (Ancestry Informative SNP, AISNP) 遗传标记。AISNP 遗传标记因为在预测样本的祖先信息方面的优势而应用于刑事案件调查。近年来,许多学者通过筛选构建了有效的 AISNP 遗传标记检测体系。如 2011 年, Kidd JR 团队^[1]构建了一个包含 128 个 AISNP 位点的检测体系, 可用于对全球 119 个人类群体进行验证。2016 年, 李彩霞等^[2]发表一个包含 27 个 SNP 位点的检测体系针对东亚、非洲和欧洲人群进行区分。之后, 2019 年, 张林等^[3]构建了包含 18 个 SNP 位点的检测体系更加针对性地细分亚洲族群。通过不间断地筛选有效的 AISNP 遗传标记并建立相关的检测体系, 为推断样本的种族来源提供了一种有效的检测手段。三是预测 DNA 身源者的体貌特征, 与之相关的是表型信息 SNP (Penotype Informative SNP, PISNP) 遗传标记能通过推断个体的表型信息为案件侦破提供有用信息。2010 年 Walsh S 团队^[4]构建的包含 6 个 SNP 位点的 IrisPlex 系统, 可用于区分 DNA 身源者的眼球虹膜颜色 (蓝色 - 棕色)。但是多基因表型与复杂的外界因素 (如老化和环境因素), 即便经过筛选的 SNP 遗传标记也不能够准确地呈现 DNA 身源者的表型信息。即便如此, 法医学者对这一热点的研究仍将继续, 希望将来可为案件调查提供有用信息。(2) 亲缘关系鉴定: 人类基因组中含有 300 多万个 SNP 位点,

理论上若能检测足够多的 SNP 位点, 就有助于亲缘关系的鉴定^[5]。这一原理主要是利用 SNP 单倍型的形式进行检测, 而 SNP 单倍型是由连锁遗传且不产生重组的常染色体 SNP 遗传标记组合而成。SNP 单倍型可提供足够多的 SNP 位点, 有助于亲缘关系、迁徙模式、家系检索的判断。2010 年 Ge J 等通过在 22 条常染色体上筛选出 253 个 SNP 单倍型区块来验证 SNP 单倍型在鉴定亲缘关系中的作用, 结果表明 SNP 单倍型区块可提供比单个 SNP 遗传标记更多的遗传信息有助于亲缘关系的鉴定。随后, Morimoto C 等^[6]基于 ICS (Index of Chromosome Sharing) 指数利用包含 174254 个 SNP 位点的 Human Core-24 Bead Chip 研究多层级的亲缘关系, 结果显示可对 5 级内的亲缘关系和无关个体进行区分。2018 年, 鉴于大规模 SNP 遗传标记检测技术难度大、数据分析压力大及相关实验成本高的问题背景下, Mo SK 团队^[7]发表了一个包含 472 个 SNP 位点的检测体系可用于亲缘关系的鉴定, 为 SNP 位点应用于鉴定亲缘关系提供有力的检测工具。

(三) 其他遗传标记

除了上述常用的遗传标记外, SGS 测序技术还可以应用于检测其他的法医遗传标记, 例如微单倍型 (MH), 利用 MH 可以单次测序片段内检测 3 个及以上等位基因的位点的特点, 为亲缘

[1] Kidd J R, Friedlaender F R, Speed W C, et al. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples [J]. *Investig Genet*, 2011 (2): 1.

[2] Wei Y L, Wei L, Zhao L, et al. A single-tube 27-plex SNP assay for estimating individual ancestry and admixture from three continents [J]. *Int J Legal Med*, 2016 (130): 27-37.

[3] Qu S Q, Zhu J, Wang Y J, et al. Establishing a second-tier panel of 18 ancestry informative markers to improve ancestry distinctions among Asian populations [J]. *Forensic Science International-Genetics*, 2019 (41): 159-167.

[4] Walsh S, Liu F, Ballantyne K N, et al. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information [J]. *Forensic Sci Int Genet*, 2011 (5): 170-180.

[5] Ge J, Budowle B, Planz J V, et al. Haplotype block: a new type of forensic DNA markers [J]. *Int J Legal Med*, 2010, 124: 353-361.

[6] Morimoto C, Manabe S, Kawaguchi T, et al. Pairwise Kinship Analysis by the Index of Chromosome Sharing Using High-Density Single Nucleotide Polymorphisms [J]. *Plos One*, 2016, 11 (7).

[7] Mo S K, Ren Z L, Yang Y R, et al. A 472-SNP panel for pairwise kinship testing of second-degree relatives [J]. *Forensic Sci Int Genet*, 2018 (34): 178-185.

关系判定、未知个体身份信息识别以及医学诊断等提供重要的研究价值^{[1][2][3][4][5][6][7]}；通过检测 mtDNA 单倍群信息^{[8][9]}、mRNA 表达情况来分析具有组织特异性的案件检材来源；通过对多种遗传标记联合检测，如 STR-SNP 复合遗传标记，为不同的刑事案件类型灵活地定制具有特定位点的遗传标记检测体系。通过利用 SGS 测序技术可以同时检测多种遗传标记，高通量的优势，提升了这些遗传标记在法医学领域中的应用能力。

三、常见的应用到法医学领域的 SGS 测序平台

如今市场上的 SGS 测序平台根据测序原理和测序通量的不同可划分为多种平台类型，以达到适用于各种科学研究的目的。本文仅介绍常见的适用于法医学领域的 SGS 测序平台，探讨其原理以及在法医学领域中的应用前景。

（一）Illumina 测序平台

通量适中、较低成本以及较为操作简单的 MiSeq FGx™ 测序平台是 Illumina 公司专门针对法医学领域发行的 SGS 测序平台，其工作原理采用边合成边测序（SBS）测序技术，其核心是利用可逆终止子的方法来检测测序片段中每个碱基与 DNA 模板链的结合。由于每个测序循环中存在全部四种可逆终止子结合的 dNTP（脱氧核糖核苷三磷酸），标记的荧光基团被激发从而可识别核苷酸种类。与其他测序技术相比，这种天然竞争最大限度地减少了结合偏向，并大大降低了测序错误率，从而更快更准确地识别出待检检材的遗传信息^{[10][11][12]}。即使对于碱基重复序列区域和均聚物，同样可以进行高度精准的逐个碱基测序，并且几乎避免了序列背景特异的错误。更重要的是，MPS 检测的 STR 等位基因与目前的数据库格式完全兼容，提供了 CE 与 MPS 数据的无缝衔接。MPS 数据的质量和准确性对法医基因组学至关重要，特别是在分

[1] Oldoni F, Kidd K K, Podini D. Microhaplotypes in forensic genetics [J]. *Forensic Sci Int Genet*, 2019 (38) : 54-69.

[2] Kidd K K, Pakstis A J, Speed W C, et al. Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics [J]. *Forensic Sci Int Genet*, 2014 (12) : 215-224.

[3] Kidd K K, Speed W C, Pakstis A J, et al. Evaluating 130 microhaplotypes across a global set of 83 populations [J]. *Forensic Sci Int Genet*, 2017 (29) : 29-37.

[4] Turchi C, Melchionda F, Pesaresi M, et al. Evaluation of a microhaplotypes panel for forensic genetics using massive parallel sequencing technology [J]. *Forensic Science International-Genetics*, 2019 (41) : 120-127.

[5] Chen P, Zhu J, Pu Y, et al. Microhaplotype identified and performed in genetic investigation using PCR-SSCP [J]. *Forensic Sci Int Genet*, 2017 (28) : e1-e7.

[6] Chen P, Yin C, Li Z, et al. Evaluation of the Microhaplotypes panel for DNA mixture analyses [J]. *Forensic Sci Int Genet*, 2018 (35) : 149-155.

[7] Kidd K K, Speed W C. Criteria for selecting microhaplotypes: mixture detection and deconvolution [J]. *Investig Genet*, 2015 (6) : 1.

[8] Just R S, Scheible M K, Fast S A, et al. Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U. S. populations [J]. *Forensic Sci Int Genet*, 2015 (14) : 141-155.

[9] Kakuda T, Shoji H, Tanaka M, et al. Multiplex APLP System for High-Resolution Haplogrouping of Extremely Degraded East-Asian Mitochondrial DNAs [J]. *Plos One*, 2016 (11) : e0158463.

[10] Jünemann S, Sedlazeck F, Prior K, et al. Updating benchtop sequencing performance comparison [J]. *Nature biotechnology*, 2013 (31) : 294-296.

[11] Loman N, Misar R, Dallman T, et al. Performance comparison of benchtop high-throughput sequencing platforms [J]. *Nature biotechnology*, 2012 (30) : 434-439.

[12] Quail M, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers [J]. *BMC Genomics*, 2012 (13) : 341.

析混合 DNA 样本、mtDNA 异质性或 STR 和 SNP 数据结果时。目前, Verogen 公司针对 MiSeq FGx™ 测序平台推出了 Forenseq™ DNA Signature Prep 试剂盒以及 ForenSeq™ Universal Analysis Software (UAS) 数据分析软件。2017 年, Guo Fei 团队对该试剂盒基于 MiSeq FGx™ 测序平台进行了法医学验证; 随后 2019 年, Wu J 等^[1]应用该试剂盒针对实际案件中的异常分型现象进行测序检测。相关研究结果均表明 Forenseq™ DNA Signature Prep 试剂盒及 MiSeq FGx™ 测序平台适用于法医学研究。

(二) Ion Torrent PGM 平台

测序速度快、通量适中的 Ion Torrent PGM 平台的核心技术为半导体芯片上离子流测序, 核苷酸依次通过半导体芯片, 当 NTP 和 DNA 模板结合后, 释放出 H⁺ 离子引起 PH 值变化, 通过半导体传感器阵列将该化学信号转化成电压差信号, 最后结合特定的碱基流顺序定序, 无须化学级联酶促反应, 无须荧光和化学发光反应。在对多聚物的检测准确性上不如 Illumina 平台^[2]。目前, 基于 Ion Torrent PGM 平台, Thermo Fisher Scientific 公司推出了适用于法医学中 STR 基因座检测的 Precision ID GlobalFiler™ NGS STR Panel v2 试剂盒、针对个体识别的 Precision ID Identity Panel 试剂盒以及针对线粒体检测的 Precision ID mtDNA Control Region Panel 试剂盒和 Precision ID mtDNA Whole Genome Panel 试剂盒等。对于目前针对法医学常用的遗传标记 (STR、SNP、mtDNA 及 microRNA) 进行的相关研究而言, 均表明 Ion Torrent PGM 平台对法医学研究具有较大的吸引力。

四、展望

二代测序技术的研发和应用使法医遗传学进入了一个全新的发展阶段, 随着 SGS 测序技术及相关科研平台的发展与成熟, 与常用的毛细管电泳技术 (PCR-CE 技术) 相比, 它可以同时容纳数百甚至数千个 STR 和 SNP 等多种遗传标记的同时并行检测分析。一方面 SGS 测序技术可针对样本不依靠荧光标记系统进行 STR 和 SNP 等多种遗传标记识别, 同时可将二代测序文库构建的测序片段设计得更短, 以容纳更多的遗传标记, 从而达到对多个样本的并行检测分析; 另一方面, 多种遗传标记的联合应用, 容纳的新遗传标记增多, 检测系统的效能

大大提高。基于上述优势, 应用二代测序技术检测 STR 和 SNP 等多种遗传标记在法医遗传学领域是十分重要的研究方向。在关注二代测序技术具有众多检测优势的同时, 也要关注其不足之处。其一, 测序平台生产商基于多种测序检测的需求而推出的多种商业化测序试剂盒, 可应用于未知个体识别、复杂亲缘关系和族源推断等鉴定案件。但是, 试剂盒中部分遗传标记在检测结果上表现并不理想, 例如 Illumina 公司的 Forenseq™ DNA Signature Prep 试剂盒含有 58 个 STR 基因座、94 个人识别 SNP (IISNP)、22 个刻画表型特征的 SNP (PISNP) 和 56 个用于祖先推断的 SNP (AISNP)。其中, STR 基因座方面, D22S1045 基因座在杂合子等位基因之间看到的片段计数在不稳定性上可能比其他基因座位点要高, 在判断是否存在 DNA 混合液时, D22S1045 基因座要考虑多位点基因型; SNP 遗传标记方面, 用于刻画表型特征的 PISNP 主要针对眼睛和头发的颜色, 这在中国人人群中应用价值较少。总而言之, 仍需加大力度开发针对中国人需求的二代测序试剂盒。其二, 随着二代测序技术在法医学领域中的应用和推广, 法医学者需要考虑如何科学合理地利用测序所得到的数据, 如何合理合法、公正科学地解释这些测序数据, 如何制定一套具有广泛认可度和公信力的数据采集标准是目前迫切需要解决的问题。另外, 二代测序技术相关的试剂及科研平台的投入成本高, 相关商业化试剂盒的推出与验证评估, 与现有法医遗传学数据库应用分析的衔接等方面是决定该技术能否在法医遗传学领域内广泛应用的关键因素之一。同时, 对案件样本测序可能涉及的伦理问题等均是决定二代测序技术何时能够替代或补充成熟 PCR 毛细管电泳技术成为主流的 DNA 检测技术以及普遍应用于刑事案件检测的关键。

(责任编辑: 王巧丽)

[1] Wu J, Li J L, Wan M L, et al. Evaluation of the MiSeq FGx system for use in forensic casework [J]. International Journal of Legal Medicine, 2019 (133): 689-697.

[2] Borsting C, Morling N. Next generation sequencing and its applications in forensic genetics [J]. Forensic Sci Int Genet, 2015 (18): 78-89.

Research Advances and Applications of Second-generation Sequencing Technology in Forensic Science

Yang Zihao

1. East China University of Political Science and Law, Shanghai;

2. Academy of Forensic Science, Shanghai

Abstract: With the rise of Second Generation Sequencing (SGS) technology in recent years, forensic scientists have introduced and applied it to the field of forensic science, thus taking full advantage of its high throughput, fast sequencing speed and low cost, and greatly contributing to the development of forensic science and its application in case detection. This paper reviews the importance of second-generation sequencing technology in forensic genetic marker detection, starting from the history of the application of sequencing technology in forensic science. The current and potential applications of second-generation sequencing technology in forensic science are explored using the existing platforms from Illumina and Thermo Fisher Scientific Sequencing Technologies, which are relevant to forensic science research. For current research, these sequencing platforms can sequence genetic markers such as STR genetic markers, SNP genetic markers, mtDNA and mRNA. However, the optimisation of some of the genetic markers in sequencing kits, the introduction and validation of sequencing kits for Chinese populations, the development of a widely credible data collection standard, the optimisation of sequencing data analysis software, the interface with existing forensic genetics databases, and the ethical issues in the use of sequencing data are all important factors in determining whether second-generation sequencing technologies can replace (or complement) established PCR capillary tube sequencing technologies. The key to whether second-generation sequencing technology can replace (or complement) established PCR capillary techniques and be universally applied in criminal cases.

Key words: Forensic genetics; Second generation sequencing; Genetic markers