

全基因组测序在结核分枝杆菌研究中的应用

张洁 任怡宣 潘丽萍 张宗德

【摘要】 全基因组测序技术已经广泛应用于结核分枝杆菌的研究中,包括结核分枝杆菌谱系或亚种鉴定、区分患者的发病模式,以及识别传染源、超级传播者和确定传播方向,以及用于结核分枝杆菌的耐药性预测和微进化研究等,在临床和公共卫生领域发挥了很重要的作用。即便如此,目前还存在很多问题需要解决。作者对全基因组测序技术在结核分枝杆菌研究中的相关内容、应用成果及进展进行综述。

【关键词】 分枝杆菌; 结核; 全基因组关联研究; 高通量核苷酸测序; 分子诊断技术; 综述

Application of whole genome sequencing in research of *Mycobacterium tuberculosis* ZHANG Jie*, REN Yi-xuan, PAN Li-ping, ZHANG Zong-de. * Beijing Key Laboratory of Drug-resistant Tuberculosis Research, Institute of Tuberculosis and Thoracic Tumor, Beijing Chest Hospital, Capital Medical University, Beijing 101149, China (* Doctoral student); * Central Laboratory of Beijing Research Institute for Tuberculosis Control, Beijing 100035, China (* Work unit)

Corresponding author: ZHANG Zong-de, E-mail: zzd417@163.com

【Abstract】 Whole genome sequencing has been widely used in the research of *Mycobacterium tuberculosis* (*M. tuberculosis*), including classification of *M. tuberculosis* lineages or subspecies, differentiation of disease patterns, identification of putative source cases and super-spreaders, determination of the transmission chains and their directions, as well as prediction of drug resistance of *M. tuberculosis* and detection of microevolution of *M. tuberculosis*. Although the whole genome sequencing has played an important role in clinical and public health applications, many problems are still existed and should be resolved in future. Herein, the current research directions and achievements of whole genome sequencing are reviewed, and the future applications are prospected.

【Key words】 *Mycobacterium tuberculosis*; Genome-wide association study; High-throughput nucleotide sequencing; Molecular diagnostic techniques; Review

在全球范围内,结核病是导致死亡的十大原因之一,据估计每年新增的结核病患者超过 1000 万例^[1]。越来越多的新发及复发结核病患者对一线抗结核药品利福平和异烟肼产生了耐药性(即耐多药结核病)^[1]。在过去的几十年中,分子生物学方法已经被广泛应用于结核病的诊断和结核分枝杆菌(*Mycobacterium tuberculosis*, MTB)的耐药性分析。基因分型技术为许多低发病率国家的结核病分子流行病学研究提供了重要的帮助,有助于发现疫情^[2]。随着生物信息的发

展,测序技术的日益成熟,且成本逐渐降低。以基因组 DNA 为基础、计算机技术为辅助的全基因组测序技术已经广泛应用于对 MTB 的研究中,并迅速从实验室研究转向临床和公共卫生应用,为解决 MTB 的传播和耐药问题提供了很大的帮助。

一、全基因组测序在结核病流行病学研究中的优势

高收入且结核病低流行国家的结核病监测系统中,重要的一个环节是接触者追踪,针对的对象是接触传染性结核病的人群,以便确定应接受预防性治疗的接触者,并减少进一步的传播^[3]。但是流行病学调查耗费时间长,劳动强度大,患者之间通过短期或临时接触而产生的流行病学联系不容易被发现。而分子流行病学整合了现有的流行病学数据和分子分型结果,在结核病传播链的研究中起到了极其重要的作用。

在过去的 20 年里,分子分型方法极大地提高了大家对结核病流行病的认识。目前有 3 种主要的基因分型方法:即 IS6110 限制性片段长度多态性 (restriction fragment length polymorphism, RFLP) 分型法、间隔区寡核苷酸分型 (Spoligotyping) 和多位点数目可变串联重复序列 (variable number of tandem repeats, VNTR) 分型。这些常规的基因分型方法在评估同源性方面价值有限。全基因组测序



开放科学(资源服务)标识码(OSID)的开放科学计划以二维码为入口,提供丰富的线上扩展功能,包括作者对论文背景的语音介绍、该研究的附加说明、与读者的交互问答、拓展学术圈等。读者“扫一扫”此二维码即可获得上述增值服务。

doi:10.3969/j.issn.1000-6621.2020.07.017

基金项目:“十三五”国家科技重大专项(2017ZX10201301004);北京市医院管理局“登峰”人才培养计划(DFL20181601)

作者单位:101149 首都医科大学附属北京胸科医院 北京市结核病胸部肿瘤研究所 耐药结核病研究北京市重点实验室[张洁(博士研究生)、潘丽萍、张宗德];北京结核病控制研究所中心实验室(张洁、任怡宣)

通信作者:张宗德,Email: zzd417@163.com

(whole-genome sequencing, WGS)方法是利用 DNA 测序平台重新构建生物体基因组的完整 DNA 序列。瑞士 Stucki 等^[4]在一项全国性的研究中,对 520 例结核病患者分离出来的标本进行了标准分子分型(Spoligotyping 和 VNTR 分型),鉴定出来的 35 个传播簇中的 18 个被 WGS 证明大于 12 个单核苷酸多态性(single nucleotide polymorphisms, SNPs),这表明常规的基因分型方法可能会高估 MTB 在瑞士这种发病率较低国家的传播^[4]。而也有研究认为当使用常规基因分型工具时,外源性再感染而引起的结核病由于与第一次疾病发作间期出现的菌株基因型相似,可能会被错误地归类为复发。这些错误分类低估了 MTB 在高发病率环境中的传播^[5]。由于不同的 MTB 菌株 SNPs 发生相同突变的可能性很小,因此,基于鉴定 SNPs 的 WGS 基因分型较其他分子分型方法具有更高的分辨率。

二、WGS 在 MTB 研究中的应用

(一)WGS 在 MTB 谱系或亚种鉴定(菌株分型)中的应用

有研究认为,MTB 最有可能起源于非洲,而全球的传播主要受人类活动的推动,即人类的探索、迁徙、贸易和战争等活动均可推动 MTB 在全球的传播^[6-7]。人感染的 MTB 具有高度的克隆性,根据 SNP 的差异和缺失可以将 MTB 复合群分为 7 个主要的适应人类的系统发育谱系,即第 1 谱系至第 7 谱系。而东亚地区占主导地位的是第 2 谱系,其中“北京家族”菌株是在全球传播和流行最为广泛的菌株之一^[8]。WGS 证实了东格陵兰及加拿大魁北克省努纳维克分离的 MTB 菌株大多来自第 4 谱系(欧美系)^[9-10],越南的研究表明,长期潜伏性感染的 MTB 第 1 谱系以及最近引进传播的“北京家族”和第 4 谱系菌株,导致了该地区结核病传播的复杂性^[11]。上海的结果研究表明,虽然我国结核病患者数多,但 MTB 的遗传多样性却不大,我国 MTB 99.4% 属于第 2 谱系和第 4 谱系的 3 个子谱系^[12]。

(二)WGS 在结核病发病模式及疫情控制中的应用

结核病患者发病主要有两种途径,一种是既往感染过,之后经过较长时间的潜伏而发病,称为内源性复燃^[13];另一种为近期传播导致的结核病。MTB 产生耐药可能是两种机制造成的:一是治疗不足引发的获得性耐药;二是感染耐药 MTB(原发耐药)。WGS 可以通过比较不同患者分离出来的菌株基因型是否相同以及同一患者前后两次发病时分离出来的菌株基因型是否相同从而判断结核病发病原因及耐药原因。而区分患者的发病模式对结核病控制策略的制定具有非常重要的意义,如果一个地区由于近期传播导致的结核病患者比例较高,说明结核病正在传播,应该将主动迅速发现传染源、切断传播途径作为防治工作的重点;而如果大部分患者是由于内源性复燃引起的,证明结核病的传播发生在过去,应该将对患者进行规范化治疗和管理作为防治工作的重心,可以对既往感染者中的高危人群进行筛查和采取预防性干预措施^[13]。

结核病高负担国家的结核病暴发调查往往仅限于调查

家庭成员和其他密切接触者,而在结核病发病率较低的国家,结核病控制的目的是制止结核病的传播,因此,除了结核病的早期诊断和治疗外,调查结核病暴发情况也是非常重要的^[14]。越来越多的研究表明 WGS 在 MTB 疫情控制方面发挥着重要的作用^[15]。在一项对伦敦广泛耐药结核病(XDR-TB)暴发的调查中,采用 WGS 后证实了患者之间的关联性,并指导了对早期患者的治疗^[16]。使用 WGS 进行全国疫情调查被越来越多的国家使用,WGS 也被应用于国际跨境疫情的调查^[17]。

在缺乏广泛的流行病学数据的情况下,基于 WGS 的系统发育树可以用来识别可能的传染源、超级传播者和确定传播方向。上海的一项研究表明,大约 1/3 的耐多药结核病患者归因于最近的传播,与传统的基因分型方法相比,WGS 目前在区分临床 MTB 分离株基因型方面具有许多优势,并且可以有效地研究 MTB 分离株的传播动力学。WGS 可提供更明确的证据,证明耐药突变是沿传输链出现和固定的^[18]。

(三)WGS 在 MTB 耐药性及其预测中的应用

据 WHO 报道,2017 年全世界有 55.8 万例结核病患者对一线抗结核药品利福平耐药,其中 82% 的患者为耐多药结核病患者;而这些耐多药结核病患者中,大约有 8.5% 为 XDR-TB^[1]。基于培养的传统技术方法仍然是耐药结核病诊断和药物敏感性试验(简称“药敏试验”)的参考标准,但是这些方法耗时长,并且对实验室的生物安全要求较高。近几年,在全球以 GeneXpert MTB/RIF 技术为代表的快速分子学方法检测 MTB 及利福平耐药已经普及。WGS 技术已经鉴定出与 MTB 耐药相关的许多突变,其中一些突变与适应性相关,而另一些突变与额外的代偿性突变有关,这些突变可能增加适应性并能进行传播^[19]。

WGS 完整地描述了 MTB 耐药株的基因组特征,填补了我们对经典抗结核药品和新的抗结核药品是如何起作用的认识上的空白,并确定了使 MTB 能够逃避这些药物影响的特定突变。一些耐药基因的存在是众所周知的,例如, *rpoB* 和 *rpoC* 基因突变导致利福平耐药, *KatG* 和 *inhA* 基因的突变导致异烟肼耐药, *dfrA-thyA* 双缺失与对氨基水杨酸耐药性相关^[20-21]。WGS 确认外排泵导致的药物外流是 MTB 对氯法齐明和贝达喹啉的耐药机制,这一方法仍然是鉴定正在开发的抗结核药物的作用机制的主要途径^[22]。WGS 还揭示了药物联合化疗如何能够有效地防止结核病治疗期间出现耐药。Trauner 等^[23]使用 WGS 证明了多种活性药物的联合作用下,能阻止暂时性突变体在菌株种群内的固定而避免成为耐药优势菌;使用的药物越少,耐药性就越有可能产生并固定下来。

一些将基因组数据和其他功能数据结合起来的 SNP 数据库已经建立,MTB 耐药株相关基因突变的一些开拓性数据库也在不断完善^[24],基于软件和网络的工具,如 MTB 复合群综合分析服务器(the comprehensive analysis server for the *Mycobacterium tuberculosis* complex, CASTB)^[25],能够从 WGS 数据网络中推断菌株的耐药性。有研究比较了这些工

具在预测耐药性方面的敏感度和特异度,发现上述工具对一线抗结核药品有很好的效果;但对于二线药物则不太理想^[26-27]。数据集和数据库的创建需要实施“大数据”分析方法(例如机器学习方法)^[28],非参数机器学习方法,如分类树(CTs)和梯度提升树(GBTs)以全基因组关联(GWAS)为基础,从复杂数据集中提取信息,通过精密的计算和分析,在预测耐药性和发现新的基因突变方面将发挥重要作用。

(四) WGS 在 MTB 微进化研究中的应用

临床上从同一患者体内分离出来的 MTB 样本中同时存在敏感菌株和耐药菌株的情况越来越引起重视,这种由于敏感和耐药菌株共感染或敏感菌株部分亚群通过微进化而发生的耐药突变称之为异质性耐药^[29]。结核感染中的微进化现象比较复杂,宿主体内多个 MTB 亚克隆同时存在和发生微进化导致的异质性耐药,必将影响患者的治疗方案和监测,并使得推断传播链的工作复杂化^[30-31]。WGS 不仅可以区分患者是否为混合感染,还可以用来掌握 MTB 的耐药性形成过程,从而对 MTB 的变异程度进行更详尽的分析,还能够在传播链中观察到变异性,从而提高我们对 MTB 感染动力学的理解^[32]。及时发现异质性耐药现象,有利于及时调整治疗方案并进行有效的临床治疗。

三、WGS 技术面临的挑战

虽然 WGS 技术在 MTB 的传播检测中敏感度较高,但是用 WGS 构建传播网络的细节是很困难的。传播网映射高度依赖于采样范围、对疫情的全面抽样,以及宿主内环境的多样性。不同患者的菌株基因成簇,但却没有其他任何直接的流行病学联系,说明这种传播很大程度上是在社区环境中的偶然接触引起的^[33]。传播网络的系统发育重建特别具有挑战性,因为菌株之间关系非常密切,但是 MTB 的进化速度很慢^[24]。

WGS 的操作流程还需统一化、标准化。工作流程的所有步骤,从 DNA 提取到测序、数据分析和报告,都应该标准化,并有良好的文档记录,还应该建立一个外部质量评估方案。2016 年,美国食品和药物管理局发布了基于测序的传染病诊断指南草案,WHO 和欧洲疾病预防控制中心等机构正在采取措施,实现 MTB WGS 的国际标准化^[34-35]。

WGS 用于 MTB 的检测结果及数据共享目前仍没有国际标准。报告应具有足够的灵活性,以解决最终用户对基因组数据解释的差异,并允许针对特定地区的治疗准则和格式要求进行定制。报告单的设计要避免使用缩写,重要信息使用加底色、粗体和其他方式标记,必要时纳入简要说明解释关键结果^[36-37]。

WGS 起源于一项基础科研技术,现在慢慢成为现代临床微生物学实验室的一部分,有望快速、更好地防控和诊断结核病,并对结核病暴发疫情进行详细、实时的流行病学研究。

参 考 文 献

[1] Goosby E, Jamison D, Swaminathan S, et al. The Lancet

Commission on tuberculosis: building a tuberculosis-free world. *Lancet*, 2018, 391(10126): 1132-1133. doi: 10.1016/S0140-6736(18)30666-4.

[2] Lalor MK, Anderson LF, Hamblion EL, et al. Recent household transmission of tuberculosis in England, 2010—2012: retrospective national cohort study combining epidemiological and molecular strain typing data. *BMC Med*, 2017, 15(1): 105. doi: 10.1186/s12916-017-0864-y.

[3] Zakhm F, Laurent S, Esteves Carreira AL, et al. Whole-genome sequencing for rapid, reliable and routine investigation of *Mycobacterium tuberculosis* transmission in local communities. *New Microbes New Infect*, 2019, 31: 100582. doi: 10.1016/j.nmni.2019.100582.

[4] Stucki D, Ballif M, Egger M, et al. Standard genotyping overestimates transmission of *Mycobacterium tuberculosis* among immigrants in a low-incidence country. *J Clin Microbiol*, 2016, 54(7): 1862-1870. doi: 10.1128/JCM.00126-16.

[5] Dippenaar A, De Vos M, Marx FM, et al. Whole genome sequencing provides additional insights into recurrent tuberculosis classified as endogenous reactivation by IS6110 DNA fingerprinting. *Infect Genet Evol*, 2019, 75: 103948. doi: 10.1016/j.meegid.2019.103948.

[6] Pepperell CS, Granka JM, Alexander DC, et al. Dispersal of *Mycobacterium tuberculosis* via the Canadian fur trade. *Proc Natl Acad Sci USA*, 2011, 108(16): 6526-6531. doi: 10.1073/pnas.1016708108.

[7] O'Neill MB, Shockey A, Zarley A, et al. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *Mol Ecol*, 2019, 28(13): 3241-3256. doi: 10.1111/mec.15120.

[8] Luo T, Comas I, Luo D, et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proc Natl Acad Sci USA*, 2015, 112(26): 8136-8141. doi: 10.1073/pnas.1424063112.

[9] Bjorn-Mortensen K, Soborg B, Koch A, et al. Tracing *Mycobacterium tuberculosis* transmission by whole genome sequencing in a high incidence setting: a retrospective population-based study in East Greenland. *Sci Rep*, 2016, 6: 33180. doi: 10.1038/srep33180.

[10] Lee RS, Radomski N, Proulx JF, et al. Population genomics of *Mycobacterium tuberculosis* in the Inuit. *Proc Natl Acad Sci USA*, 2015, 112(44): 13609-13614. doi: 10.1073/pnas.1507071112.

[11] Holt KE, McAdam P, Thai PVK, et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat Genet*, 2018, 50(6): 849-856. doi: 10.1038/s41588-018-0117-9.

[12] Liu Q, Ma A, Wei L, et al. China's tuberculosis epidemic stems from historical expansion of four strains of *Mycobacterium tuberculosis*. *Nat Ecol Evol*, 2018, 2(12): 1982-1992. doi: 10.1038/s41559-018-0680-6.

[13] 高谦,杨崇广.我国结核病近期传播与控制策略. *结核病与肺部健康杂志*, 2017, 6(3): 193-198. doi: 10.3969/j.issn.2095-3755.2017.03.001

[14] World Health Organization. Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries. Geneva: World Health Organization, 2012.

[15] van der Werf MJ, Ködmön C. Whole-Genome sequencing as tool for investigating international tuberculosis outbreaks: a systematic review. *Front Public Health*, 2019, 7: 87. doi: 10.3389/fpubh.2019.00087.

[16] Arnold A, Witney AA, Vergnano S, et al. XDR-TB transmission in London: Case management and contact tracing investi-

- gation assisted by early whole genome sequencing. *J Infect*, 2016, 73(3): 210—218. doi: 10.1016/j.jinf.2016.04.037.
- [17] Walker TM, Merker M, Knoblauch AM, et al. A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *Lancet Infect Dis*, 2018, 18(4): 431-440. doi: 10.1016/S1473-3099(18)30004-5.
- [18] Yang C, Luo T, Shen X, et al. Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect Dis*, 2017, 17(3): 275-284. doi: 10.1016/S1473-3099(16)30418-2.
- [19] Liu Q, Zuo T, Xu P, et al. Have compensatory mutations facilitated the current epidemic of multidrug-resistant tuberculosis? *Emerg Microbes Infect*, 2018, 7(1): 98. doi: 10.1038/s41426-018-0101-6.
- [20] Coll F, McNerney R, Preston MD, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med*, 2015, 7(1): 51. doi: 10.1186/s13073-015-0164-0.
- [21] Moradigaravand D, Grandjean L, Martinez E, et al. *dfpA thyA* Double Deletion in para-Aminosalicylic Acid-Resistant *Mycobacterium tuberculosis* Beijing Strains. *Antimicrob Agents Chemother*, 2016, 60(6): 3864-3867. doi: 10.1128/AAC.00253-16.
- [22] Almeida D, Ioerger T, Tyagi S, et al. Mutations in *pepQ* Confer Low-Level Resistance to Bedaquiline and Clofazimine in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 2016, 60(8): 4590-4599. doi: 10.1128/AAC.00753-16.
- [23] Trauner A, Liu Q, Via LE, et al. The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome Biol*, 2017, 18(1): 71. doi: 10.1186/s13059-017-1196-0.
- [24] Cohen KA, Manson AL, Desjardins CA, et al. Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: progress, promise, and challenges. *Genome Med*, 2019, 11(1): 45. doi: 10.1186/s13073-019-0660-8.
- [25] Iwai H, Kato-Miyazawa M, Kirikae T, et al. CASTB (the comprehensive analysis server for the *Mycobacterium tuberculosis* complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. *Tuberculosis (Edinb)*, 2015, 95(6): 843-844. doi: 10.1016/j.tube.2015.09.002.
- [26] Schleusener V, Koser CU, Beckert P, et al. *Mycobacterium tuberculosis* resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. *Sci Rep*, 2017, 7: 46327. doi: 10.1038/srep46327.
- [27] van Beek J, Haanpera M, Smit PW, et al. Evaluation of whole genome sequencing and software tools for drug susceptibility testing of *Mycobacterium tuberculosis*. *Clin Microbiol Infect*, 2019, 25(1): 82-86. doi: 10.1016/j.cmi.2018.03.041.
- [28] Deelder W, Christakoudi S, Phelan J, et al. Machine Learning Predicts Accurately *Mycobacterium tuberculosis* Drug Resistance From Whole Genome Sequencing Data. *Front Genet*, 2019, 10: 922. doi: 10.3389/fgene.2019.00922.
- [29] 高旭, 李静, 柳清云, 等. 异质性耐药对结核分枝杆菌表型和基因型耐药检测结果的影响. *中华结核和呼吸杂志*, 2014, 37(4): 260-265. doi: 10.3760/cma.j.issn.1001-0939.2014.04.007
- [30] Navarro Y, Perez-Lago L, Herranz M, et al. In-Depth Characterization and Functional Analysis of Clonal Variants in a *Mycobacterium tuberculosis* Strain Prone to Microevolution. *Front Microbiol*, 2017, 8: 694. doi: 10.3389/fmicb.2017.00694.
- [31] Ley SD, de Vos M, Van Rie A, et al. Deciphering Within-Host Microevolution of *Mycobacterium tuberculosis* through Whole-Genome Sequencing: the Phenotypic Impact and Way Forward. *Microbiol Mol Biol Rev*, 2019, 83(2): e00062-18. doi: 10.1128/MMBR.00062-18.
- [32] Bespyatykh J, Shitikov E, Bespyatykh D, et al. Metabolic Changes of *Mycobacterium tuberculosis* during the Anti-Tuberculosis Therapy. *Pathogens*, 2020, 9(2): E131. doi: 10.3390/pathogens9020131.
- [33] Auld SC, Shah NS, Mathema B, et al. Extensively drug-resistant tuberculosis in South Africa: genomic evidence supporting transmission in communities. *Eur Respir J*, 2018, 52(4): 1800246. doi: 10.1183/13993003.00246-2018.
- [34] Tagliani E, Cirillo DM, Ködmön C, et al. EUSeqMyTB to set standards and build capacity for whole genome sequencing for tuberculosis in the EU. *Lancet Infect Dis*, 2018, 18(4): 377. doi: 10.1016/S1473-3099(18)30132-4.
- [35] Meehan CJ, Goig GA, Kohl TA, et al. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat Rev Microbiol*, 2019, 17(9): 533-545. doi: 10.1038/s41579-019-0214-5.
- [36] Crisan A, McKee G, Munzner T, et al. Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. *PeerJ*, 2018, 6: e4218. doi: 10.7717/peerj.4218.
- [37] Tornheim JA, Starks AM, Rodwell TC, et al. Building the framework for standardized clinical laboratory reporting of next generation sequencing data for resistance-associated mutations in *Mycobacterium tuberculosis* complex. *Clin Infect Dis*, 2019, 69(9): 1631-1633. doi: 10.1093/cid/ciz219.

(收稿日期:2020-03-10)

(本文编辑:范永德)