

文章编号: 1000-6281(2007)05-0488-06

基于纳米操纵的基因组测序技术与策略

吕 鸣¹, 李雪玲^{1*}, 吕军鸿², 邵志峰³, 胡 钧^{1, 2}

(1. 上海交通大学生命科学技术学院纳米生物学实验室, 上海 200240;

2. 中国科学院上海应用物理研究所, 上海 201800;

3. Dept of Mol Physiol & Biol Phys, University of Virginia Medical School, Charlottesville, VA 22908-0736, USA)

摘 要: 人类基因组计划所遗留下来的未能解决的序列(如人类基因组缺口, gaps)以及越来越多的测序需求, 都迫切需要发展新一代的测序技术与策略。我们实验室长期从事基于原子力显微镜(AFM)对单个DNA分子进行纳米操纵方面的工作, 在此基础上, 最近提出并原理演示完成了一种单分子有序化测序策略。这种有序化测序策略与现有测序技术以及新兴高通量单分子测序技术的结合, 有望解决基因组测序中的许多困难。

关键词: 有序化测序; 单分子测序; 测序策略; 纳米操纵; 原子力显微镜

中图分类号: Q336; Q523; TG115.21+5.7 文献标识码: A

人类基因组计划的巨大成功, 极大改变了当前生物学及医学研究, 而它的成功与过去几十年测序技术的快速发展是分不开的。Sanger^[1]发明的双脱氧链终止 DNA 测序方法是整个测序计划的基础, 而此后该方法的并行化与自动化^[2]则在加快了测序的速度同时也大大降低了测序的成本, 开始了 DNA 测序的大规模应用。直到鸟枪法测序策略^[3]提出, 大规模的基因组测序技术才得以成熟。目前, 人类基因组计划虽然已经完成^[4], 但却留下了一些现有技术很难甚至无法完成的基因组序列^[4, 5]。这部分 DNA 在全基因组中所占的比例并不高, 但是却有可能包含具有重要生物学意义的信息。另一方面, 人类基因组计划完成后, 基因组测序的需求不仅没有减少反而是增加了, 更多的基因组需要测序, 甚至今后的医学研究与疾病治疗要求针对个人进行个人基因组测序^[6], 这就需要发展新一代的更廉价、更快速、更灵敏的测序技术^[5]。传统的测序技术已很成熟, 成本的下降和速度的提高也很有限, 虽然相对于最初成本已下降很多, 速度也提高很多, 但仍无法满足今后测序的要求, 唯有发展全新的技术与策略才有可能带来根本性的改变。

正是在这种背景下, 最近几年先后涌现出了一批全新的测序技术与策略。总的来说, 这些技术与策略可分为三大类。一类是基于 DNA 合成的策略^[7-9], 通过 DNA 聚合酶添加带有荧光的核苷酸,

检测荧光信号来获得序列的信息。这类方法能够以大量 DNA 分子为模板, 并行的获得大量 DNA 分子的序列, 然后通过类似于鸟枪法的策略组装序列。由于高度的并行性和简单的操纵过程, 测序成本和速度都得到了非常可观的改善。但这类方法获得的序列片段长度却很有限, 甚至还不如传统的 Sanger 法^[10]。更短的片段长度造成了序列组装存在的问题就更大, 而难于组装却是传统方法难以解决部分序列测序的一个主要原因。

另一类策略是通过能够区分碱基信号差异的原理直接顺序读取序列信息。比如纳米孔测序^[11-13], 是通过检测单个 DNA 分子经过纳米孔时电流的变化来获得序列信息。这类方法的优点是原则上可以读取任意长度的 DNA 序列。但是, 这类策略当前还存在很多技术瓶颈需要解决, 还无法达到对单个碱基的准确辨别。

第三类策略是有序化测序, 就是按照次序将 DNA 分子分离出来, 然后测序。这类策略的最大优点在于: 1) 可以确定各个片段的位置, 因此序列的组装就变得非常简单, 这不仅解决了复杂序列, 尤其是重复序列难以组装的难题; 2) 可以只对部分目标序列进行测序, 从而加快了速度, 这在单核苷酸多态性(SNP)的筛选和疾病基因的定位克隆方面具有潜在应用前景; 3) 可以与当前的测序方法或新兴的测序方法相结合, 具有很好的集成性与整合性。我们

收稿日期: 2007-06-07; 修订日期: 2007-06-21

基金项目: 国家自然科学基金资助项目(No. 10335070, No. 10604061, No. 10674147)。

Foundation item: National Natural Science Foundation of China(No. 10335070, No. 10604061, No. 10674147)。

作者简介: 吕鸣(1981-), 男(汉族), 江西人, 博士研究生。

* 通讯作者: 李雪玲(1972-), 女(汉族), 山东人, 讲师。

实验室多年来面向这种测序策略开展工作, 在先后解决了 DNA 纳米操纵一系列技术问题的基础上, 最近完成了一种基于纳米操纵的基因组测序技术与策略。

1 DNA 纳米操纵技术

单分子 DNA 纳米操纵是指利用原子力显微镜 (AFM) 探针与衬底上的 DNA 分子间的力相互作用, 对单个 DNA 分子进行切割、推移甚至拾取等物理操纵^[14-17]。DNA 纳米操纵技术, 可以切割并分离吸附

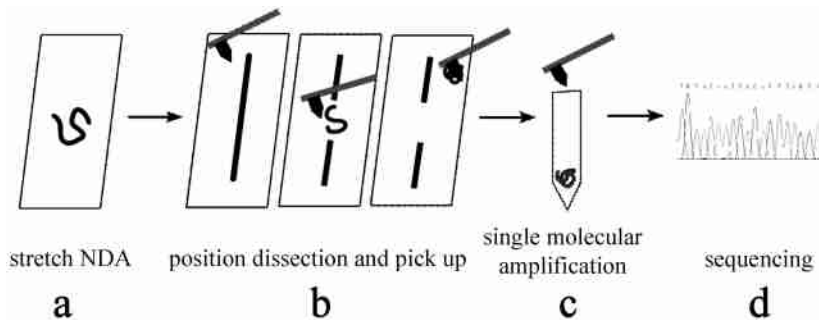


图 1 基于纳米操纵的有序化单分子测序。

Fig. 1 Ordered single molecule sequencing based on nanomanipulation.

为了对单个 DNA 分子进行定位切割和分离, 先决条件是要制备拉直在衬底上的 DNA 样品。人们先后发展了多种 DNA 拉直的方法, 譬如较为简单和快速的分子梳^[22,23]方法。与单纯的用于观察的不同, 纳米操纵的 DNA 样品要求 DNA 分子在衬底上的吸附强度必须恰当。吸附力太弱则 DNA 分子不容易吸附在衬底上, 太强的吸附则导致后续的操作十分困难。通过对云母表面的硅烷化修饰及其进一步钝化, 可以恰当地控制 DNA 分子与衬底的结合强度, 制备出适合纳米操纵的样品^[15,17,24]。

样品制备好后, 就可利用原子力显微镜探针 DNA 分子进行纳米操纵。然而, 由于 DNA 柔软的特性, 还要避免外来的污染, 只能选择在空气中的 Tapping 模式。由于探针工作于振动状态, 控制它与样品之间的作用就变得不那么容易, 而精确的操作需要对探针的精确控制。为此, 我们发展了一种基于抬升模式的纳米操纵方法, 很好地解决了用同一个针尖很难同时完成成像与操纵的问题, 从而能够以一种十分精确的方式对 DNA 分子进行切割、推移、折叠等一系列的纳米操纵^[14]。然而如何拾取切割后的 DNA 分子片段仍是一个难题, 主要因为探针与生物大分子相互作用的复杂性。我们尝试了多种方法以后, 发展了一种基于经验的“纳米分离”技

在衬底上的 DNA 分子的特定片段, 并通过后续高灵敏的单分子 DNA 扩增技术, 放大该 DNA 片段的拷贝数量, 就可以进一步的进行序列的测定与分析^[18,19] (图 1)。由于我们获得的片段在原 DNA 上的位置是确定的, 这样就实现测序的有序化。我们称之为有序化单分子纳米测序 (Ordered single molecule Sequencing based on Nanomanipulation, OsmSN)^[20,21]。DNA 样品制备技术、单个 DNA 分子的分离技术以及单分子 DNA 扩增技术是该方法的三个关键技术。

术^[18] (图 2)。

对拾取出的单个 DNA 分子进行序列分析是一个极具挑战的问题。虽然生物学家已经发展了 DNA 分子的扩增技术——聚合酶链式反应 (PCR), 能够指数级地增加 DNA 分子的拷贝数, 但常规的 PCR 都是以一定数量的 DNA 分子为模板, 对于单个 DNA 分子, 其扩增灵敏度与特异性都还存在不足。我们后来发展了一种采用纳米金颗粒作为辅助成分纳米粒子 PCR^[25] 技术, 能够很好地提高 PCR 反应的性能, 从而解决了单个分子 DNA 扩增这一后续分析的瓶颈问题。

2 单分子有序化测序策略

2.1 连续拾取

如前所述, 为了获得长片段的全部序列, 最简单的思想是将长片段依顺序分成一次测序可以完成的足够短的片段, 分别测序。而序列的组装则非常简单, 只需按顺序连接起来即可。这种策略在原理上比较简单, 不过要真正实现有序化测序仍有一些问题需要解决。首先是切割精度的问题, 因为目前测序反应一次能够测得的平均长度大约在 500 bp 左右, 因此切割精度必须控制在 150 bp (50 nm) 以下。目前我们的切割精度可以控制在 30 nm 左右, 保证

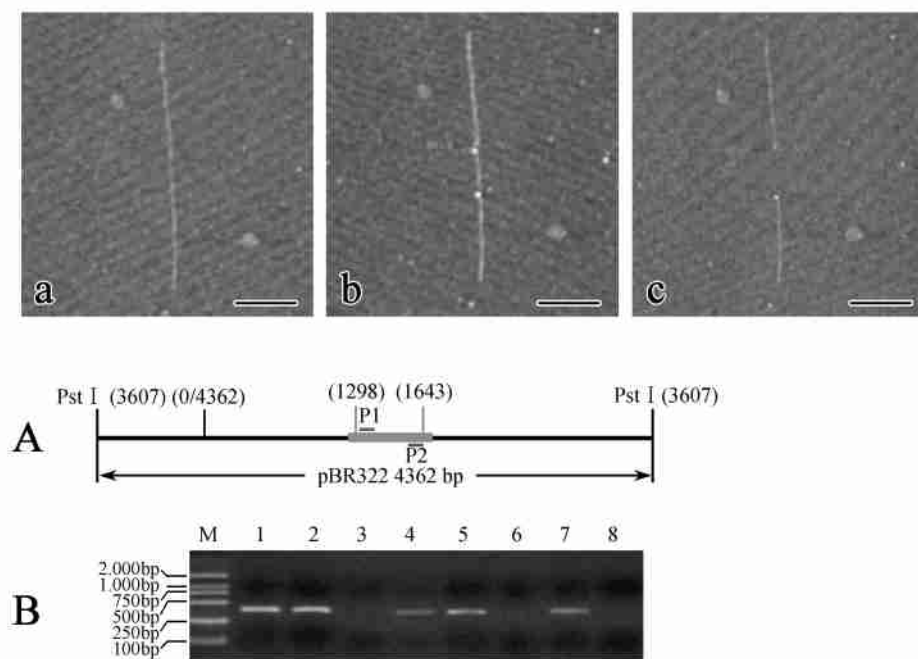


图 2 对单个 DNA 分子的定位分离以及后续的单分子扩增^[18]。a: 操纵前 DNA 的 AFM 图像; b: 切割后 DNA 的 AFM 图像; c: 片段被拾取后 DNA 的 AFM 的图像; A: 目的片段在 pBR322 质粒上位置示意图; B: 拾取片段单分子扩增的电泳结果。Bar= 500 nm

Fig. 2 Positioning isolation of single DNA molecule and subsequent single molecule amplification^[18]. a: The AFM image of DNA before manipulation. b: The AFM image of the cut DNA. c: The AFM image of the DNA after the segment was picked up. A: The schema of the segment's position on pBR322 plasmid. B: The electrophoresis of the single molecule amplification products of the picked DNA segment. Bar= 500 nm

了切割的高精度。此外, AFM 针尖总是有一定大小, 切割部位的 DNA 势必被破坏, 从而导致无法获得这部分的序列。为了获得完整的序列, 可以分别从不同 DNA 分子上获取不同位置的片段, 使它们各自之间有一定的重叠。这样做同时带来其它优点, 一是避免了重新定位难的问题, 不用每次都去寻找同一个分子; 二是一定的重叠确保最终能获得完整的序列。但这样同时带来另一个需要解决的问题, 即 AFM 下对 DNA 方向的判定, 因为 DNA 是均一的线性分子, AFM 下无法区分头尾。我们利用一种 DNA 分子的标记技术来解决这个问题。通过在 DNA 分子的某一端修饰上生物素基团, 然后进行抗生物素蛋白(Avidin)与生物素基团的特异性结合, 这样在 AFM 下就可以观察到 DNA 一端结合的蛋白, 从而确定 DNA 在 AFM 下的方向。最近, 我们以 pBR322 DNA 为模式分子, 通过对相互重叠的两个片段的分别拾取和扩增、测序^[26], 最终实现了这种有序化测序的完整过程(图 3, 图 4)。

2.2 与鸟枪法策略的结合

为了克服纳米操纵技术面前自动化程度低所导

致的测序效率不高这一问题, 我们提出了与现行鸟枪法测序方法相结合的综合策略^[21]。鸟枪法策略是目前对长片段及基因组 DNA 测序普遍采用的一种策略, 其基本原理是将长片段或基因组 DNA 随机打断为短的片段, 然后对短片段分别测序, 最后根据短片段序列之间的重叠, 组装成最终的序列。该策略的优点在于各个片段的测序可以同时进行, 并行程度非常高。对于不含重复片段的基因组来说, 鸟枪法策略无疑是非常成功的。但由于片段的随机性, 重复序列的存在是鸟枪法测序组装过程中面临的重大问题。而基于纳米操纵的有序化测序策略与鸟枪法的结合刚好可以克服这一问题。为了验证这一策略的可行性, 我们选取了一个人工构建的重复片段为例子, 进行了计算机模拟(图 5)。由于其序列的重复性, 单靠鸟枪法无法正确确定其序列。采用 OsmSN 方法, 事先获得几个指定位置的序列, 且这几个位置的前后顺序和它们之间的间隔都是已知的, 然后以这几个序列作为分子标记, 对全部片段进行拼接。结果表明: 通过结合 OsmSN 策略的确可以很好的解决含重复序列的片段的拼接问题。

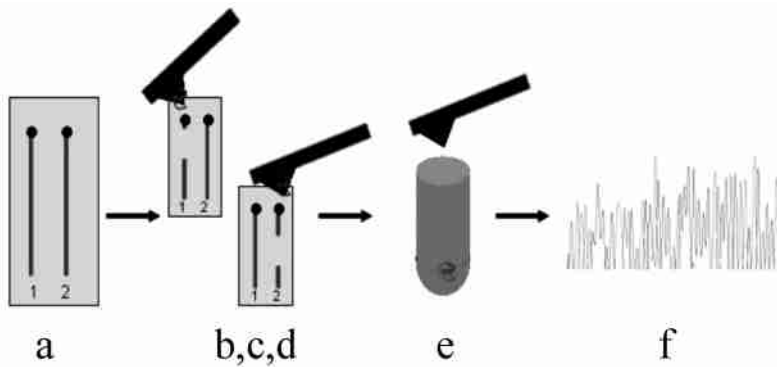


图 3 基于 OsmSN 两段重叠 DNA 片段测序示意图。a: 相同序列的两个 DNA 分子; b, c, d: 对两个 DNA 分子不同位置但序列重叠的片段分别进行拾取; e: 拾取片段的单分子扩增; f: 测序

Fig. 3 Schematics of sequencing two overlapping DNA fragments by OsmSN. a: Two DNA molecules with the same sequence; b, c, d: Pick up segments of different DNA molecules overlapped; e: Single molecule amplification of the picked DNA segments; f: DNA sequencing.

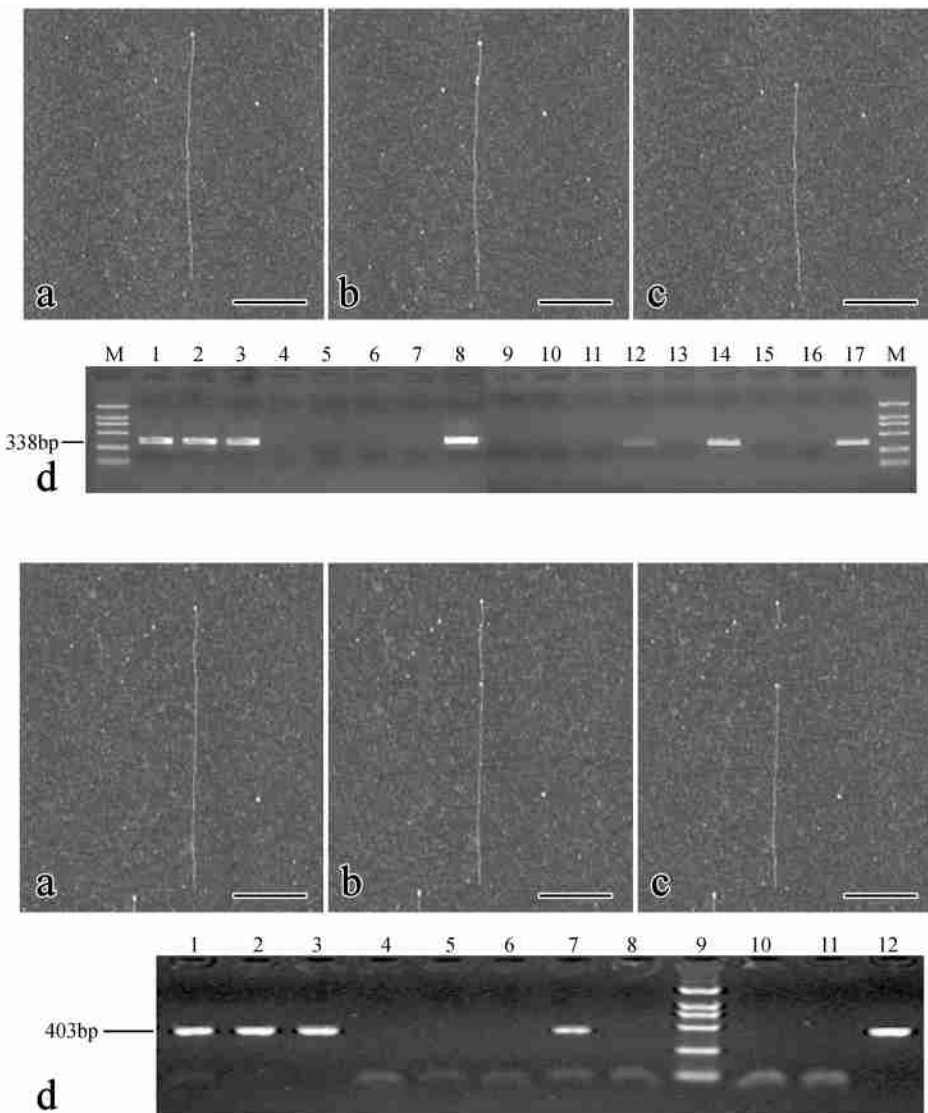


图 4 连续拾取及扩增两段 DNA^[26]。上: 拾取第一个 DNA 片段前后的 AFM 图像及拾取片段单分子扩增结果。下: 拾取第二个 DNA 片段前后的 AFM 图像及拾取片段单分子扩增结果。Bar= 500 nm

Fig. 4 Continuously pick up and amplify two segments of DNA^[26]. Top: The AFM image of the DNA before and after the 1st segment was picked up and the electrophoresis of the single molecule amplification product of the picked segment. Bottom: The AFM image of the DNA before and after the 2nd segment was picked up and the electrophoresis of the single molecule amplification product of the picked segment. Bar= 500 nm

2.3 解决目前基因组测序中的难题

到目前为止,人类基因组中仍还存在大大小小的缺口(gaps) 280 个(数据来自 NCBI),此外还有占全基因组约 6% 的异染色质部分未进行测序。对这些缺口周边序列的研究发现其绝大部分都与重复序列有紧密联系。同样的情况也存在于其它基因组的测序中。由于这些缺口部分的 DNA 通常是难克隆的区域^[27],为了填补这些缺口,通常的做法是通过 PCR 扩增的方法获得这部分 DNA 并进行后续测序。对于常规的基因组 DNA,这是非常有效的方法,但对于这些缺口部分的 DNA,有如下两个原因使得它难以成功进行。首先是这部分 DNA 的扩增通常都存在特异性差的问题,往往很难得到特异性扩增产物,或即使得到扩增产物,但由于产物混杂而无法进行测序反应。二是这些区域由于未知原因,往往难以克隆。针对上述情况,我们最近提出一种将纳米粒子 PCR 扩增技术^[25]、单分子 DNA 纳米操纵技术与新兴的高通量测序方法^[28]相结合的策略。基本思路是纳米粒子 PCR 能够提高扩增反应的特异性,提高扩增产物的质量,使得产物更适合进行后面的分析。通过采用单分子 DNA 纳米操纵:一方面可以拾取单个 DNA 分子并进行单分子扩增,从而实现一种不依赖于细胞的克隆方法,解决有些 DNA 细胞克隆难的问题;另一方面更为重要的是,有序的定位拾取能够对序列的正确组装起到关键的作用。而结合

a sequence with repeats

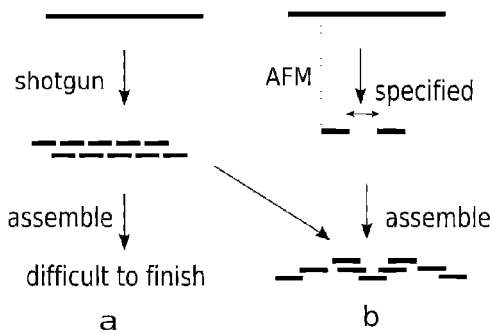


图5 鸟枪法与有序化测序相结合。a: 鸟枪法测序,含有重复片段的序列拼接很困难; b: OsmSN 策略与鸟枪法结合,能够获得已知位置的序列信息,有助于指导拼接过程。

Fig. 5 Assembling sequences with OsmSN and shotgun. a: Shotgun sequencing, assembly of sequence with repeats will be very difficult; b: OsmSN strategy, combined with shotgun method, getting some sequence information with known position guides the assembling process.

新的高通量的测序方法,则能大大加快整个测序进程,并解决一些片段用 Sanger 方法测序难的问题。目前,相关工作正在进行中。

3 讨论与展望

无论传统的测序技术还是新兴的测序方法,由于基本上都是采用了随机化的测序策略,因此在基因组测序中存在很多无法逾越的障碍。基于纳米操纵的测序技术与策略通过定位切割拾取,将原来随机的测序变得有序化了,从而有望解决目前测序过程中的一些困难。同时,将它与新的高通量测序方法相结合,利用各自的优点,将更加拓宽它的应用范围。但是同其它新技术一样,基于纳米操纵的测序技术还处在发展之中,要将它应用到实际的大规模测序,还有一些技术问题需要解决,这也是我们目前正在研究的内容。首先,如果拾取的片段序列是完全未知的,那么未知序列片段的单分子扩增在目前仍存在困难。通过 DNA 连接反应可以在 DNA 分子两端加上特定的引物接头,但目前的单分子 DNA 连接效率仍很低。一种可行的方法就是拾取处在不同分子同一位置的多个片段,增加分子的个数,以弥补目前 DNA 连接反应效率不足的问题。另外,目前的纳米操纵无论从速度上还是效率上都还不高,尚不能满足大规模应用的要求。纳米操纵与拾取样品后处理自动化的实现将大大改善这一状况。

参考文献:

- [1] Sanger F, Nicklen S, Coulson A R. DNA sequencing with chain-terminating inhibitors[J]. Proc Natl Acad Sci, 1977, 74(12): 5463- 5467.
- [2] Smith L M, Sanders J Z, Kaiser R J, et al. Fluorescence detection in automated DNA sequence analysis[J]. Nature, 1986, 321 (6071): 674- 679.
- [3] Venter J C, Smith H O & Hood L. A new strategy for genome sequencing[J]. Nature, 1996, 381 (6581): 364 - 366.
- [4] The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome [J]. Nature, 2004, 431 (7011): 931- 945.
- [5] Eichler E E, Clark R A & She X. An assessment of the sequence gaps: unfinished business in a finished human genome[J]. Nat Rev Genet, 2004, 5 (5): 345- 354.
- [6] Shendure J, Mitra R D, Varma C, et al. Advanced sequencing technologies: methods and goals[J]. Nat Rev Genet, 2004, 5 (5): 335- 344.
- [7] Magulies M, Egholm M, Altman W E, et al. Genome sequencing in microfabricated high density picolitre reactors

- [J]. *Nature*, 2005, 437 (7057): 376–380.
- [8] Mitra R D, Shendure J, Olejnik J, et al. Fluorescent in situ sequencing on polymerase colonies[J]. *Anal Biochem*, 2003, 320 (1): 55–65.
- [9] Shendure J, Porreca G J, Reppas N B, et al. Accurate multiplex polony sequencing of an evolved bacterial genome [J]. *Science*, 2005, 309 (5741): 1728–1732.
- [10] Whiteford N, Haslam N, Weber G, et al. An analysis of the feasibility of short read sequencing[J]. *Nucleic Acids Res*, 2005, 33 (19): e171.
- [11] Deamer D W, Akeson M. Nanopore and nucleic acids: prospects for ultrarapid sequencing[J]. *Trends Biotechnol*, 2000, 18: 147–151.
- [12] Meller A, Nivon L, Brandin E, et al. Rapid nanopore discrimination between single polynucleotide molecules[J]. *Proc Natl Acad Sci*, 2000, 97 (3): 1079–1084.
- [13] Ashkenasy N, Sanchez-Quesada J, Bayley H, et al. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores[J]. *Angew Chem Int Ed Engl*, 2005, 44 (9): 1401–1404.
- [14] Hu J, Zhang Y, Gao H, et al. Artificial DNA patterns by mechanical nanomanipulation[J]. *Nano Letters*, 2002, 2 (1): 55–57.
- [15] 吕军鸿, 吴世英, 王国华, 等. DNA 单分子的纳米定位切割与拾取研究[J]. *电子显微学报*, 2003, 22 (3): 185–188.
- [16] Hu J, Zhang Y, Li B, et al. Nanomanipulation of single DNA molecules and its applications [J]. *Surf Interface Anal*, 2004, (36): 124–126.
- [17] L J H, Wu S Y, Zhang Y, et al. Single DNA molecular manipulation with atomic force microscopy [J]. *Nuclear science and techniques*, 2004, 15(5): 257–260.
- [18] L J H, Li H K, An H J, et al. Positioning isolation and biochemical analysis of single DNA molecules based on nanomanipulation and single-molecule PCR [J]. *J Am Chem Soc*, 2004, 126 (36): 11136–11137.
- [19] L J H, An H J, Li H K, et al. Nanodissection, isolation and PCR amplification of single DNA molecules [J]. *Surf Interface Anal*, 2006, 38: 1010–1013.
- [20] Zhang Y, L J H, Li M Q, Hu J. A strategy for ordered single molecule sequencing based on nanomanipulation (OsmSN) [J]. *International Journal of Nanotechnology*, 2007, 4 (1/2): 163–170.
- [21] 吕鸣, 石宝晨, 李雪玲, 等. 一种基于单分子纳米操纵的有序化测序策略[J]. *生物化学与生物物理进展*, 2006, 33 (7): 660–664.
- [22] Bensimon A, Simon A, Chiffaudel A, et al. Alignment and sensitive detection of DNA by a moving interface [J]. *Science*, 1994, 265 (5181): 2096–2098.
- [23] Michalet X, Ekong R, Fougereousse F, et al. Dynamic molecular combing: stretching the whole human genome for high-resolution studies [J]. *Science*, 1997, 277 (5331): 1518–1523.
- [24] An H J, Guo Y C, Zhang X D, et al. Nanodissection of single- and double-stranded DNA by atomic force microscopy [J]. *J Nanosci Nanotechnol*, 2005, 5(10): 1656–1659.
- [25] Li H K, Huang J H, L J H, et al. Nanoparticle PCR: nanogold-assisted PCR with enhanced specificity[J]. *Angew Chem Int Ed Engl*, 2005, 44 (32): 5100–5103.
- [26] An H J, Huang J H, L M, et al. Single-base resolution and long-coverage sequencing based on single-molecule nanomanipulation [J]. *Nanotechnology*, 2007, 18, 225101 (5pp).
- [27] Razin S V, Ioudinkova E S, Trifonov E N, et al. Non-clonability correlates with genomic instability: a case study of a unique DNA region[J]. *Mol Biol*, 2001, 307 (2): 481–486.
- [28] Kling J. The search for a sequencing thoroughbred [J]. *Nat Biotechnol*, 2005, 23 (11): 1333–1335.

A sequencing strategy based on nanomanipulation

LÜ Ming¹, LI Xue-ling^{1*}, LÜ Jun-hong², SHAO Zh-feng³, HU Jun^{1, 2}

(1. Nanobiology Laboratory, College of Life Science & Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China;

2. Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China;

3. Dept of Mol Physiol & Biol Phys, University of Virginia Medical School, Charlottesville, VA 22908-0736, USA)

Abstract: The increasing requirement for sequencing large amount of genomes calls for cheaper, faster, and more sensitive sequencing technology to overcome the limitations in current sequencing methods. We have devoted in developing new methods to manipulate single DNA molecules with atomic force microscope (AFM) at high spatial resolution. Based on AFM nanomanipulation, we have proposed a novel strategy for ordered single molecule sequencing recently. In this paper, we reviewed our latest progresses and discussed the possibility of combination between our sequencing strategy and others, including conventional sequencing methods and the emerging ultra-high throughput single-molecule sequencing techniques.

Keywords: ordered sequencing; single molecule sequencing; sequencing strategy; nanomanipulation; atomic force microscope

* Corresponding author