

下一个改变世界的技术——基因测序

叶碧然 21301050250 临床医学（八年制）

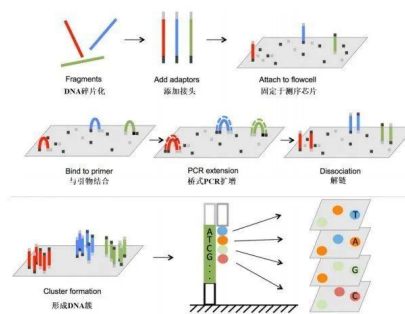
1 基因测序的原理

基因测序是一种新型基因检测技术，能够从血液或唾液中分析测定基因全序列，预测罹患多种疾病的可能性，个体的行为特征及行为合理。基因测序技术能锁定个人病变基因，提前预防和治疗。基因测序相关产品和技術已由实验室研究演变到临床使用，可以说基因测序技术是下一个改变世界的技术。

基因测序技术的发展历史，是效率、通量和成本的变革历史，促进了基因测序的普及，对生命科学和医学研究起到重大推动作用，也使得大规模商业化的应用变为可能。从 Sanger 测序法发明以来，基因测序的发展大致历经三个发展阶段：

1.1 第一代测序：

1977 年 Sanger 发明了双脱氧核糖核酸链末端终止法，自上世纪 90 年代起，大量基因测序均采用半自动化毛细管电泳 Sanger 测序法。该测序方法的主要步骤和原理如下：



DNA 片段扩增：将碎片化的 DNA 连接到质粒载体中，跟随质粒进行自我复制；

循环测序：加入一定比例带有荧光标记的、可以用来中断 DNA 合成反应的反应物，DNA 复制过程中在任一位置都有可能终止，经历大量循环合成后将形成长短不同的、带末端标记的 DNA 链；

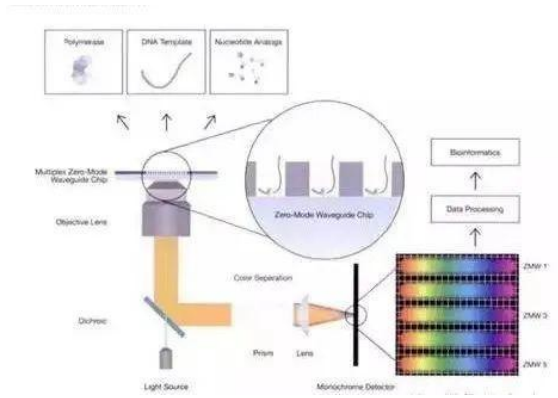
凝胶电泳：通过凝胶电泳，DNA 链将根据长短次序依次排开，利用放射自显影各链末端的荧光标记，实现 DNA 序列读取。

该技术的突出优势是长度长及高准确性，一次读取 DNA 片段长度可 1,000bp，准确率可到 99.99%；但测序通量低，耗时长，成本高，因此应用范围有限。目前该技术主要用于单基因病多外显子的测序或者少量基因多位点的检测，以及验证高通量测序中出现的阳性结果。虽然其他测序技术有很大的通量，但基于 Sanger 原理的毛细管电泳测序仍是超高精度测序的金标准，目前其他新发展

的测序技术结果都必须应用 Sanger 测序技术对其结果进行认证。因此，该测序技术尚未完全被取代。

1.2 第二代测序：

第二发展阶段的代表性测序技术主要是高通量测序技术，又称下一代测序技术（即 Next Generation Sequencing, NGS），是目前也是未来较长时间内的主流基因测序技术。高通量测序技术的核心思想是边合成边测序，即通过捕捉新合成的末端的标记来确定 DNA 的序列。该测序方法的主要步骤和原理如下：



建库： DNA 碎片化，并添加接头到 DNA 片段两端；

桥式扩增并形成 DNA 簇： 携带引物的 DNA 两端固定在测序芯片上，通过桥式 PCR 反应实现 DNA 扩增，以实现信号放大，DNA 在各自的位置解链后形成 DNA 簇，该 DNA 簇内的 DNA 完全相同；

边合成边测序： 反应体系内的 4 种碱基用不同的荧光标记分别标记，DNA 簇中的各 DNA 链同步复制，每延伸一个碱基读取一次信号，实现边合成边测序，且各 DNA 簇同时读取，实现高通量。

高通量测序的不同 DNA 片段固定在同一个基因芯片中同时边合成边测序，因此通量和速度大幅提高；但由于在高通量测序中，单个 DNA 必须扩增成由相同 DNA 组成的 DNA 簇，来增强荧光信号强度从而读出 DNA 序列，随着 DNA 片段长度增长，在扩增过程中，碱基添加出现错误的概率也随之增加、且基因簇复制的协同性降低，将导致测序质量下降，这严格限制了高通量测序的读长（不超过 500bp）。因此，高通量测序具有通量高、读长短的特点，需要在建库阶段打断 DNA 成为小片段，测序完毕后经由生物信息技术作拼接，因此对实验技术和生物信息技术有较高的要求。

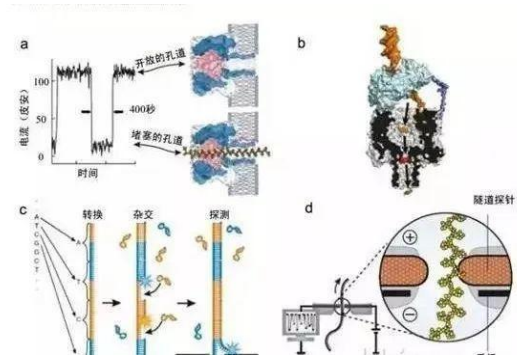
高通量测序大幅降低了测序成本、缩短了测序时间，同时保持了较高的准确性，目前是基因测序主流技术，也是基因测序技术商业化应用普及的主要推动力。

以人类基因组测序为例，高通量测序将一人份基因组测序的时间从数年缩短至一周，测序成本由上亿美金降低到数千美金甚至更低。

1.3 第三代测序：

不经过扩增的单分子测序和长读长为标志的DNA测序技术被称为第三代测序技术，因其测序时DNA分子无需PCR扩增，实现了对每一条DNA

分子的单独测序，也称为单分子测序技术。以PacBio公司的SMRT和Oxford Nanopore Technologies纳米孔单分子测序技术为代表。



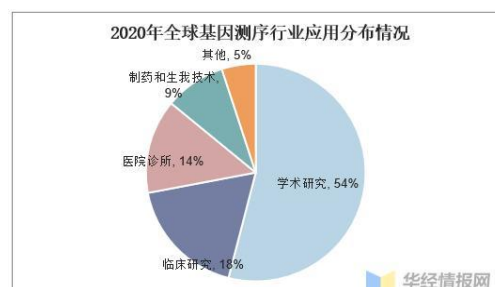
PacBio SMRT 技术应用了边合成边测序的思想，并以 SMRT 芯片为测序载体，芯片上有很多小孔，每个孔中均有 DNA 聚合酶。测序基本原理是：DNA 聚合酶和模板结合，4 色荧光标记 4 种碱基（即是 dNTP），在碱基配对阶段，不同碱基的加入，会发出不同光，根据光的波长与峰值可判断进入的碱基类型。DNA 聚合酶是实现超长读长的关键之一，读长主要跟酶的活性保持有关，它主要受激光对其造成的损伤所影响。另外，可以通过检测相邻两个碱基之间的测序时间，来检测一些碱基修饰情况，既如果碱基存在修饰，则通过聚合酶时的速度会减慢，相邻两峰之间的距离增大，可以通过这个来检测甲基化等信息。SMRT 技术的测序速度很快，每秒约数个 dNTP。但是，同时其测序错误率比较高（这几乎是目前单分子测序技术的通病），达到 15%，但好在它的出错是随机的，并不会像第二代测序技术那样存在测序错误的偏向，因而可以通过多次测序来进行有效的纠错（代价是重复测序，也就是成本会增加）。

Oxford Nanopore Technologies 所开发的纳米单分子测序技术与以往的测序技术皆不同，它是基于电信号而不是光信号的测序技术。该技术的关键之一是，设计了一种特殊的纳米孔（只能容纳单分子通过），孔内共价结合有分子接头。当 DNA 碱基通过纳米孔时，它们使电荷发生变化，从而短暂地影响流过纳米孔的电流强度（每种碱基所影响的电流变化幅度是不同的），灵敏的电子设备检测到这些变化从而鉴定所通过的碱基。

2 基因测序的应用

2.1 应用的领域

学术研究一直是基因测序行业最大的用途，根据数据显示，2020年学术研究在基因测序行业的市场份额已经达到



54%。可以预见，随着基因测序技术的进一步提升，在科学研究领域的市场份额将进一步增大。其次，在临床研究中，基因测序市场占总市场的约18%；医院诊所的应用市场占比约为14%，位列第三。

2.2 学术研究

从前瞻经济学人发布的全球基因测序数据来看，2020年学术研究占比超过50%，预计2021年到2026年以稳定的复合年进一步增长，而这种增长可归因于NGS解决方案在研究项目中的应用在大学和研究中心进行。

去年，中国科学家向全球第一时间分享的首个新冠病毒基因组序列就是在因美纳的技术平台上完成测序的，为后续的疫情防控，疫苗研发奠定了基础。而在今年第四届进博会上，因美纳更是展出了在全球最多国家和地区通过临床注册的中通量新一代测序平台（NGS），以及获得2021年红点设计大奖的的基因测序平台——NextSeq™ 2000。NextSeq™ 2000也是因美纳首个将文库扩增、测序以及快速基因组分析整合在一个平台上的测序仪，其数据分析效率在同类产品中出类拔萃。

随着国家政策的开放，我国基因测序行业得到了极大的发展，助推了各大巨头争相抢占主要市场，比如，为加速本土基因测序的发展，因美纳带来了通量最高的新一代测序平台，NovaSeq™ 6000Dx-CN技术原型机，目前已在中国完成了试生产。据悉，目前，位于上海浦东的因美纳中国生产基地正在积极筹备，预计将于明年第三季度投入生产。

2.3 临床应用和临床研究

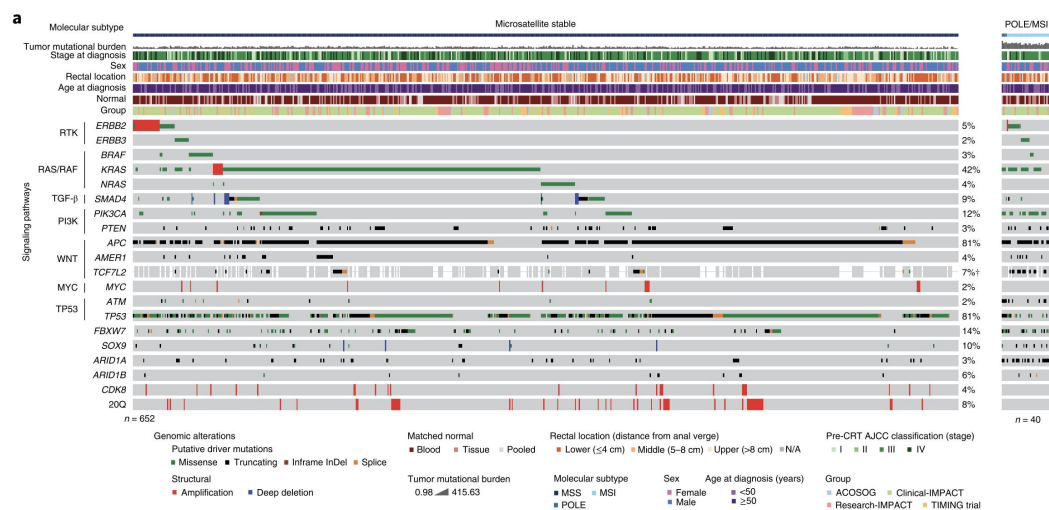
在临床上，孕前基因筛查是人们目前最广泛地能接触到这项高精尖技术的渠道。NGS可以进行胚胎移植前的基因学检查、基于cfDNA的无创产前诊断、

DNA 甲基化研究、新生儿基因筛查等。优生方面，包括产前筛查，包括叶酸代谢能力、无创产前基因检测、染色体异常筛查；儿童的优生包括遗传疾病的筛查、代谢性营养元素的基因检测、儿童安全用药、遗传代谢疾病等。该方法具有快速、快速、高灵敏度和高特异性等优点。除了常规的染色体检查外，还可以进行单基因检测，可以有效的预防胎儿的出生缺陷。传统的检测技术存在着一定的危险性，需要花费大量的时间和精力。NGS 还可以从孕妇外周血液中提取 DNA 碎片，进行无创产前检查。

同时，基因测序也是目前临床研究中最重要的手段。例如肿瘤的精准医疗，是针对肿瘤异质性强的特点，为了创造更好的治疗和达到更好的预后，提高病人的生存质量而提出的新概念，而其中比较重要的研究就是寻找新的生物标志物对于病情进展以及预后的相关性，这其中，基因测序则是最重要的工具手段。

精准医疗是完善的人体基因检测。目前临床上将肺癌划分为 NSCLC 和 NSCLC；NSCLC 分为鳞癌、腺癌、未分化癌、大细胞癌等。肿瘤的分子层次不同，可以分为 EGFR 阳性和 EGFR 阴性。然后再根据不同的基因表现，制定相应的治疗方案，形成个性化治疗。准确的肿瘤诊断应从以往的影像水平、细胞层面的病理学检查逐渐发展为基因层面的分子诊断。

在肿瘤的前沿研究中，最不可或缺的一项就是通过对于病人关键靶点的测序来展示队列的基因景观。刚刚发表在 Nature Medicine 杂志的研究论文 Genomic and transcriptomic determinants of response to neoadjuvant therapy in rectal cancer 中,就可以看见作者首先在 FIG1 中就以热图的形式展示了研究队列的基因景观。科研成果的展现离不开基因测序这项重要的生物技术。



3 基因测序的局限和发展方向

第一代 Sanger 测序因为它的原理和通量限制，并不适用于检测级别快速全面的获得海量基因组信息。目前市场上，二代测序由于上、中、下游技术成熟、优化完善、成本不断下滑，无疑是领军者，但它始终无法解决两大约束：一是需要分子扩增，二是使用光学系统检测。这两个缺点是造成二代测序难以进入基层医院和诊所的主要原因。

技术迭代要做的事情就是解决二代测序存在的问题和缺憾。其中 Ion Torrent（2.5 代）和 Pacific Biosciences（三代）的技术是在 Illumina 上迭代的，但是仍未能撼动其市场地位。2.5 代和三代测序也各有局限。

2.5 代的代表是 Ion Torrent，其技术特点是使用离子感应场效应晶体管取代光学传感器，简单来说就是将复杂昂贵大型的超敏光学信号采集转换成检测化学反应 pH 信号，这一变革大大降低了仪器成本，一举占领了中国 50% 的市场。然而，其生物信号的产生依然依赖于边合成边测序，并且没有解决分子扩增步骤带来的读长短等问题。

第三代的突破在于去掉了分子扩增步骤，即所谓的单分子测序。单分子测序在生物学及临床诊断领域具有划时代的意义。单分子测序解决了分子扩增步骤引入的人为误差，样品制备步骤的大大简化则满足了临床诊断的需求。然而，不幸的是其配合三代测序的检测部件是比二代测序仪更为高精尖的光学仪器，因而导致成本直线上升。以三代技术的杰出代表 Pacific Biosciences（PacBio）为例，其平均测序成本是 Illumina 的十倍以上。这在价格敏感度高的市场中竞争力不高。因此 PacBio 虽然在生物学上做出了重大突破却只占有全球 3% 的市场份额。第四代的出现是则是颠覆性的：不仅能做到单分子长读长测序，还实现了检测部件芯片化从而大大降低成本——第四代技术也是单分子测序，核心在于长读长、小型芯片化、适合临床应用市场。这正是测序发展趋势，适用于测序和诊断的融合。

三代测序和四代测序都是基于纳米孔技术的单分子测序，而未来最有可能取代二代测序的技术可能就是操作和成本都更具亲和力的四代测序。技术的更迭是必然趋势，但四代测序尚处于青少年发展时期，想要真正撼动二代测序的地位，仍需时日。

基因检测市场前景一片大好，各方企业都想借此机会分一杯羹，激烈的竞争中不乏“半吊”商人“鱼龙混杂”，给整个市场带来了负面影响，造成了基因检测行业的多种乱象。这与基因检测技术在国内尚处于起步阶段,基因检测公司的技术标准、市场准入标准等都是空白有关。这项技术的普遍推广应用仍面临许多挑战。

相信每一位基因检测人的初心都不希望这项技术被“滥用”！这是一项非常伟大的科学技术，是精准医学概念提出及未来发展的基础之一。黄尚志教授也在他的发言里强调了认知滞后、技术滥用、无序竞争等是整个行业发展目前面临的主要问题，希望未来在制度和行业标准更加完善的条件下，基因检测可以在各方努力下更健康的发展，辅助推动精准医学。