# 遗传研究中的分型与测序之争

何云刚

国际斯坦福研究所分子遗传项目组，美国 加州 门罗公园 94025

**摘要：**DNA 序列测定技术的飞速进步，使利用测序方法进行基因型检测的成本和时间消耗大大降低。由此有人预测 DNA 分型很快将被序列测定所完全取代。本文从多个角度探讨了分型技术相对于序列测定技术仍然具有优势的原因。说明分型技术在以后一个相当长的时期内仍然不可能被序列测定完全取代。

**关键词：**DNA 测序；基因分型；单倍域；Hapmap 计划

## Debate on Roles of Genotyping and Sequencing in Genetic Research

HE Yungang

Molecular genetic program, Stanford Research Institute International, Menlo Park, CA94025

ABSTRACT: Advances in DNA sequencing technology greatly reduced time and financial cost of sequencing-based genotyping. Some geneticists predicted that large-scale genotyping platforms would be vanished with arising of whole-genome sequencing strategy. This paper pointed out that whole-genome genotyping is still with advantages and will continue his role in future.

Key words: DNA sequencing; genotyping; haplotype block; HapMap project

耗时四年的 HAPMAP 计划终于在今年五月画上一个句号。然而这个项目，对于当前和将来的人类遗传研究并没有带来像当初人类基因组计划那样巨大的冲击。相反，一些人始终持一种观点：HAPMAP 计划的产品，人类染色体单倍体图，只是廉价的大规模序列测定技术发展成熟前，研究设计中使用的一个阶段性替代品[1]。究竟该如何来看待这一测序与分型之争？对此，我们在这里进行一些剖析。

由于 DNA 序列测定技术的发展与成熟都较早，随着技术的进步，当前 DNA 序列测定的成本已经降到很低的水平。自从双脱氧核甘酸在测序中得到应用以来，测序技术在原理上已经基本成熟。直到目前，都没有重大变化。而后期的发展和主要进步，体现在以信号分离和拾取为中心的检测方法和仪器构造的改进上。标记物经历了从放射性同位素标记到多色荧光标记的转变；而信号分离采集方法从平板变性凝胶电泳基础上的放射自显影，过度到应用毛细管电泳和 CCD 荧光成像。所以，科学工作者对序列检测技术的开发和应用由来已久，并在人类基因组计划完成时达到一个顶峰。

但是，遗传多态性研究通常需要通过对大量个体的遗传物质多态性进行测定。建立在统计学基础上的遗传学数据分析方法是导致样本量成为研究中的重要因素的直接原因。以毛细管电泳为手段的序列测定的检测成本和检测速度仍然远远不能满足对大样本进行全基因组规模的精细尺度研究的需要。在最近几年，高通量基因分型技术得到飞速发展，特别是以 Illumina 和 Affymetrix 的寡核甘酸阵列技术为代表的高通量分型技术发展成熟。这些技术均达到了一次检测百万个单核甘酸多态性遗传标记的目标[2,3]。近年来，这些技术已经被大量用于进行遗传疾病全基因组关联研究。此现象表明这些技术从技术和成本两方面都已经完全地商用化，而且实用化了。

正是在这些崭新的分型技术和其他相关科学发现的推动下，诞生了雄心勃勃的 HAPMAP 计划[4]。科研工作者观察到，当研究对象是一个大小有限（大约几到几十 kb）的基因组片段时，在这样一个片段内观察到的遗传标记的组合类型，通常大大小于将他们随机组合时候的理论预期。这一现象，体现了这些在物理上临近遗传标记之间非随机的相互联系。根据单体型，通常用少数几个遗传标记就可以准确描述大部分人所携带的这样一个小片段内其他多态性位点的情况。利用这一发现，可以在不损失大部分遗传信息的前提下，达成大幅度减少遗传多态检测的工作量之目的。从而减少全基因组规模的研究工作的开销。HAPMAP 计划的理论基础，就扎根于此。该计划进行四年来，在人类遗传学领域已经产生一系列的重要研究成果。更重要的是，通过直接或间接利用该计划提供的信息，对复杂人类性状展开的全基因组关联分析已经取得了一些鼓舞

**收稿日期：**2007 年 5 月 25 日 **联系人：**何云刚 HeYunGang@gmail.com

 

人心的进展[5]。

实际上，从单体型计划开始提出的时候就一直存在反对的声音。 部分研究人员认为，基因分型成本的下降和测序技术的飞速进步会让这个计划在完成的时候就已经过时。因为，这些技术进步将导致实验成本已经不是限制研究人员进行研究设计的一个障碍。相对而言，研究人员会更有兴趣获取更加全面的数据，而不再满足于在考虑成本时候采取的抓大放小的策略。该观点虽然比较激进，但是反映了科研工作者对研究技术手段的进步的信心。

技术进步在推动以测序为重要手段的遗传学研究上的巨大力量，已经通过 Celera 在人类基因序列测定上的完美成功得到完全的体现。而 454 测序平台的诞生，是继 DNA 序列测定技术在从平板凝胶电泳过渡到毛细管电泳后的又一次巨大进步[6]。通过将鸟枪法测序策略与先进的仪器制造技术结合，一个大的序列测定任务被分解成为数十万个乃至更多的独立的小任务并得以在一台仪器上同时运行。最后这些小任务产生的短的片段序列被组装起来以获取所希望得到的大片段的基因组序列。目前，以 454 平台和 Illumina 的序列测定技术为代表的技术进步使在几个月，甚至数个周内完成高等生物的基因组序列测定成为可能 [6,7]。

但是，回顾过去数年来的研究工作，虽然大规模甚至全基因组规模分型研究已经得到蓬勃发展，建立在对有限位点（几到几十个遗传标记）的基因分型基础上的大样本量研究却仍然在遗传关联研究中占据重要地位。其归因于表型多样性自身的复杂性和以统计学方法为基础的遗传数据分析手段。因为遗传因素的贡献通常只是表型差异中的一小部分，所以在样本大小有限的情况下，大部分表型相关的遗传多态位点，由于对表型贡献不大，并不能在统计分析中被可靠的检测出来。当前的全基因组关联分析，进行的项目很多而产出的论文却相对很少，其原因就在于此。所以对进行数据分析的遗传学研究者来说，虽然他们乐于见到更多的高质量的数据，但是从有限样本得到的更多的基因组多态性数据（多于当前全基因组分型平台所能提供的信息）显然并不能显著增加他们的研究产出。

从另外一方面来看，目前的和将来即将进入商业化的大规模序列测定平台的测定速度和成本还无法和当前成熟的大规模分型平台相抗衡。由于这些测序平台和先进的分型平台共享大量的技术，例如同时依靠高密度的寡核甘酸阵列等，测序总是消耗比较多的试剂和实验时间。虽然消耗上的差距已经在逐渐缩小，但是就目前情况来看，消耗数周时间和数万至数十万美金用于测定一个个体的全基因组序列，对于需要处理很大样本量的那些研究项目来说还是过于昂贵。在迎来下一次序列测定平台的技术革命前，测序平台还不可能和分型平台在研究成本上相抗衡。

况且，从遗传研究历史上看，大多数的古老技术会与新兴技术共存，至少是共存相当长的一段时间。所以，从短期内来看，序列测定平台完全取代分型平台在遗传研究中的位置的可能性，是几乎不存在的。

**参考文献**

1. Check E (2007) Time runs short for HapMap. *Nature* 447:242-243.
2. Illumina Inc. dna analysis solutions: snp genotyping. http://www.illumina.com/pages.ilmn?ID=39
3. Affymetrix Inc. Whole Genome Analysis. http://www.affymetrix.com/products/application/whole_genome.affx
4. The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426: 789-796.
5. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-678.
6. 454 Life Sciences. Measuring Life One Genome at a Time. http://www.454.com/enabling-technology/the-system.asp
7. Illumina Inc. DNA Sequencing with Solexa® Technology. http://www.illumina.com/downloads/SS_DNAsequencing.pdf