

新一代测序很快就应用到了和医学相关的领域，如癌症基因组学。这些技术也许有利于分析突变、改进基因表达的量化研究和发现/研究调节RNA分子（也就是非编码RNA），其许多功能都有助于了解癌症的特点。然而，尽管有这些巨大的潜力，下一代测序技术的应用现在仍然被一些因素阻碍，如每次运行上涨的成本（至少8000美元）和每次实验都会产生巨大的数据，而生物信息学处理这些数据仍面临着一定的难度。

## 5. 质谱技术



质谱分析技术是一个强大的分析技术。它可以用来识别未知化合物和量化已知化合物，并阐明一个分子的化学性质。质谱（MS）是几乎所有蛋白质组学实验的核心。基本上，占主导地位MS工作流程开始于一个特定位点的酶消化，使蛋白质变成肽。接下来，肽被处理成挥发性物质，质谱仪产生每个样品的光谱图。最后，把光谱图与数据库里的肽序列进行比较，推断蛋白质序列。

MS实验的第一步，酶消化是由1个蛋白酶（如胰蛋白酶）完成的。为了使肽变成挥发性物质，人们通常采用两种方法——基质辅助激光解吸/电离（MALDI）和电喷雾（ESI）。MALDI用于挥发含有少量肽的混合物；电喷雾用于挥发含有大量的肽的肽混合物。最后是生物信息学方案，即光谱（观察到的峰）的处理和与可以消化的蛋白质序列数据库的肽序列比较。由于蛋白质鉴定依赖于与序列数据库匹配，目前蛋白质组学主要限制那些综合序列数据库可用的物种。

MS可直接应用的是检测蛋白质或肽的峰。它们的质量和所带电荷不同，从而可以将癌症患者的样品与正常个体的样品进行比较。例如，采用这一方法，Nakagawa等人确定了两个乳腺癌相关的多肽；Hao等人确定了一些有关胃癌的多肽；而Sun等人则确定了116个可以用来区分肝癌和正常肝细胞的蛋白质。

即使MS的应用对癌症研究有巨大的潜力，但由于个体间的基因变异和血浆蛋白质组动态变化是多种因素（如性别、年龄和健康状况）综合的结果，所以这种类型的分析仍然存在一些大的障碍。



## 五、各种“组学”信息的整合

正如以前指出的那样，在基因组学、转录组学和蛋白质组学方面的进展为生物学家提供了大量要处理的数据。然而，这些数据大多来自不同的平台或存储在不同的数据库中，数据整合是非常不容易的，甚至是不可行的。鉴于此，人们希望形成一个生物信息学的特定领域，以便解决这些很重要的问题，这就是“整合基因组学”。

在整合基因组学方面，人们已经研发出了算法和工具，这使得基因组学、转录组学和蛋白质组学外部和内部的数据能够被整合。从理论上讲，整合基因组学的方法可分为两种亚型：主要数据的高水平和低水平整合。

第一亚型涉及来自原始实验数据的推定事实和结论整合；第二亚型涉及低水平数据的整合，包括平台、数据的组织和数据挖掘技术之间的处理。两种亚型使用的方法可以在以下网站看到：加州大学圣克鲁斯分校（UCSC）（<http://genome.ucsc.edu>）、美国国立生物技术信息中心（NCBI）（<http://www.ncbi.nih.gov>）以及欧洲生物信息学中心（EBI）（<http://www.ebi.ac.uk>）。

数据整合中一个有趣的例子是基因的命名和基因产物的分类。通常，一个基因会被不同的群体发现，因此名字会超过1个。例如，基因TP53有5个不同的名字，基因产物也存在这种情况。为了规范基因的名称，人们创建了HUGO（<http://www.hugointernational.org>）。HUGO给每个人类基因分配一个独特的符号，每个新发现的人类基因都应提交到HUGO，以获得一个正式的名字。目前，有几项举措试图分类基因产物，其中最流行的是Gene Ontology (GO) consortium。它根据生物学处理过程、细胞组成和分子功能的不同，使用一些容易控制的词汇来给每个基因产物分类。

即使存在上述几点困难，数据整合仍然非常重要。这有多方面的原因，包括技术上的需要（如需要加强信号和需要去除一些平台“噪声”）以及从多种资料中提炼有效信息的需要（如发展癌症的分子药理学等）。尽管许多方法和协议已经整合到Integromics公司，但还有仍多“组学”的数据有待整合。



## 六、临床生物信息学

临床生物信息学的目的是应用生物信息学知识和技术来帮助诊断、治疗、预防和控制疾病，以及发展化学的、结构的和生化的方法来应用于临床研究。癌症研究中，在癌症发生的不同阶段，如起始、持续和发展时期，生物信息学工具被用于检测几种癌症的生物标记。根据NCI的解释：生物标记的定义是细胞的、生化的、分子的（遗传的和表观遗传的）改变。有了生物标记，一个正常的、异常的或简单的生物学过程就可以被识别或监测。生物标记可以通过生物媒介，如组织、细胞或流体来衡量，也可用于评估癌症的早期诊断、风险、癌症分类和预断癌症病情。

1999年，NCI生物标记发展实验室（<http://edrn.nci.nih.gov/about-edrn/scicomponents/bdl>）建成。它的目的是寻找癌症分子的、遗传的和生物的早期预警信号。生物标记发展实验室连同生物标记参考实验室（前身为著名生物标志物验证实验室）、临床流行病学和验证中心（曾被称为临床与流行病学中心）以及数据管理和协调中心（DMCC）形成了NCI的早期检测研究网络（EDRN—<http://edrn.nci.nih.gov/about-edrn>）。为了癌症早期鉴定和风险评估，它有责任开发、评估和审定生物标记。除了分类和评估风险，肿瘤生物标记还有可能被转化成疫苗发挥效用的靶标。