

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。

临床生物信息学是一个癌症个性化治疗的基本组成部分。人们需要寻找遗传改变和临床治疗结果之间的联系，这些相关的研究很大程度上依赖于生物信息学的方法，而这些方法也对生物标记与临床知识的整合具有重要作用。

## 1. 药物和疫苗的基因靶标或蛋白质靶标鉴定

癌症是一种遗传失调，因此，它的特点是遗传，这是在癌症初期和发展期逐步积累起来的。如前所述，由基因表达改变、体细胞变异和基因组失稳造成的基因和蛋白质标记物，可以让从正常组织中区分癌症组织成为可能，并且可以运用这些知识诊断、预后、开发新型药物和制定相应疗法等。在过去几年中，生物信息学技术的发展加速了癌症标记物的鉴定工作的发展，也促进了人们更好地理解致瘤机制。目前检测的方法涉及高通量基因和蛋白质的筛选及量化它们的表达谱和变异。癌症研究的生物信息学方法需要鉴定很多标记，这些标记在可能应用于临床实践之前，将需要实验和临床验证。

### 药物发现过程的五个阶段

目标识别	旨在展示基因或蛋白质的表达谱变化与特定癌症之间的具体关系；
靶点验证	目的是确立靶点的临床意义； 癌症药物研发进程中的主要挑战就在这个阶段；
高通量的化合物筛选	目的是识别化合物； 这些化合物在体外可以抑制已经确定的蛋白质靶标； 在这个阶段，一个特定的化合物可能能够在体外阻止靶标，但它并没有表现出基本化学和物理制剂药物的特点；
高通量的化合物优化	目的是提高选定目标药物的效率，摒弃非特异性药物； 不幸的是，常见的增强效果的副作用可能会限制药物的临床应用；
前临床试验	根据预先设计好的方案，进行生物医学和健康相关的研究，实验效果根据临床实验来衡量和评价。

上述每个阶段中，某些因素，如样品采集、处理和储存、病人分类、生物信息学工具和数据库分析不完整，都会影响最终结果。

尽管存在一定困难，组学分析正加快鉴定多种生物标记物，这些标记有望转化成潜在的治疗药物的靶标。



这些靶标可以分为三大类：

第一类 致癌信号中的小分子抑制剂；

第二类 表面元件和细胞间信号分子的抗体；

第三类 分子疫苗。

在第一类中，最明显的靶标是激酶，因为它们与肿瘤发展的联系是显而易见的。这种药物研发战略在临床使用中已取得积极效果的例子就是伊马替尼甲磺酸（格列卫）。它目前已用于治疗某些成人和小儿慢性粒细胞性白血病（CML），也用于治疗一种罕见癌症，如胃肠道间质瘤（<http://www.cancer.gov/clinicaltrials/digestpage/gleevec>）。

第二和第三类中，很容易被治疗性抗体和疫苗鉴别的靶标就是那些限于一个或多个非必需组织和细胞中的表达靶标，不论细胞疾病的状态如何都是这样。通过这种方式，器官中免疫干预治疗造成的损伤，如乳腺癌和前列腺器官都可以经临床治疗去除，而不会造成严重伤害。

其它已成功应用于临床的例子还有：单克隆抗体Rituxan——用于治疗非霍奇金淋巴瘤；阿瓦斯丁——用于治疗结直肠癌和非小细胞肺癌；赫赛汀——用于治疗乳腺癌。虽然涉及靶向激酶和免疫治疗系统的方法正在发展过程中，并已取得了一些成就，但是两个方法都被同样一个事实所阻碍——潜在的靶标仅存在于某一肿瘤类型的亚群中。这种分子异质性使得可用的普遍治疗方式非常复杂并且困难重重。

## 2. 基于分子和遗传差异的个性治疗

人类基因组序列测序的完成，开启了功能基因组分析时代，也带来了新的疾病疗法的发现。不过，尽管我们对人类基因组有了一些了解，也对癌症的遗传改变有了更好的认识，但要发展普遍的癌症疗法仍有很长的路要走。

传统上，癌症的治疗取决于肿瘤的大小和位置，以及癌细胞扩散的程度和人体的基本健康状况。虽然所有的这些事实都很重要，但肿瘤定位已被确认为临床肿瘤学的基本原则，肿瘤的组织起源决定了治疗癌症方法。典型的治疗原发性肝癌的方法通常是手术切除原发肿瘤、根据手术的类型和疾病程度局部照射，以及根据肿瘤生物学辅助全身治疗（<http://www.cancer.net/portal/site/patient>）。然而，近年来基因组学的研究资料表明，每种癌症，甚至具有同一起源的癌症都有自身的特点，都由不同的基因变异导致，因此可能需要有针对性的药物。据Lengauer等人说：“很可能1癌症药物针对1个癌症的惯例，将最终被取代为由多个靶向药物用于多个不同的遗传缺陷。”

事实上，癌症个性化治疗的想法并不新鲜，数十年前就已经由Scott提出，但它的实现则依赖于科学的进一步发展。个性化的癌症化疗（ICC）是一个程序，目的是最大限度地提高化疗效果，尽量减少不利影响。基于对一个特定肿瘤发病情况及对药物反应的知识，临床医生可以规定药物治疗的最佳组合。

这个策略基于以下信息：

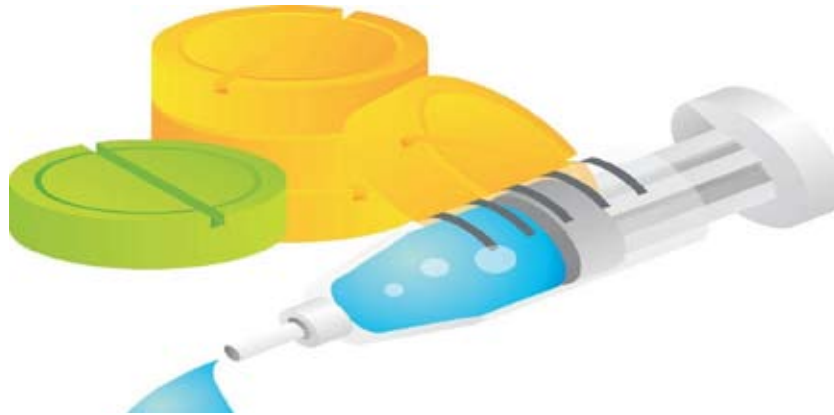
- (1) 使用该药物的敏感试验，与手术切除比较，分析潜在候选药物的抗癌功能，找到反应效果最佳的药物，并将其应用于往后的治疗中；
- (2) 肿瘤生成的标记作为药物拮抗作用和破坏的特殊靶标；
- (3) 整合涉及原发肿瘤位置和扩散模式的形态学信息，和一个合理的药物设计组合。



其它方法用于癌症个体化治疗的目的是：确定预处理的分子标记，预测化学疗法对头部和颈部的毒性；排除成本最高和化疗毒性最大的1个样本；找出可靠的可验证标记，这些标记可以辅助患者，例如乳腺癌患者选择特定的治疗方法。

虽然了解癌症的基因表达和蛋白质谱对全面了解特定癌症的特点是一项有吸引力的策略，但该策略要想发展成为能够成功预测癌症的工具，并成功用于癌症的临床治疗已被证明是非常困难的。

癌症治疗的复杂性会给癌症治疗增加巨大的困难，解决办法就是要加强基础研究和应用癌症研究领域的联系，包括生殖细胞基因突变分析、遗传和基因表达谱分析、靶向治疗分析、癌症干细胞分析、肿瘤细胞循环和单核苷酸多态性分析等。靶向治疗依赖于科学发展的水平，并能够促进药物发现和研发过程，这被认为是未来在个体基础上完善预防和干预的策略。



### 3. 生物信息学与抗癌药物设计

一种新的方法可以使研究人员能够将癌症细胞基因组和蛋白质组信息与药物的结构信息联系起来，该方法将成为一种功能强大的新药设计和研发工具。新的生物信息学方法，是美国国家癌症研究所（NCI）和美国俄亥俄州哥伦布市的LeadScope公司这两个组织的科学家共同努力的结果。新方法将基因表达的方式和肿瘤细胞中27,000多个基础化合物联系起来。它改进了以往的方法——将分子资料和化合物的抗癌活性联系，而不和化合物的基础结构联系。新系统虽然由基因表达资料研发而得，但可以很容易用于蛋白质表达研究。

新方法非常重要，因为现在研究人员已开始使用最新的技术创新来获得癌细胞中有关基因和蛋白质的信息。这种与癌症相关的分子途径的知识有助于找到潜在的靶标，预计能够大大加快药物研发的步伐。

药物结构数据的综合运用极大地提高了生物信息学的系统潜力。根据该研究报告的主要作者所言，“谁如果试图设计出完美和理想的抗癌药物，就要将涉及的基因或蛋白质的表达直接和药物结构联系起来”。新系统正是如此，这使研究人员能够了解那些化学药物的结构特征和如何制约它在不同细胞类型中的行为。研究人员希望这种新技术可以帮助化学家从大量收集到的化合物中进一步筛选最有价值的药物。

下面以错配修复基因MLH1为例。首先，通过连锁分析和小鼠错配修复蛋白mMr同源性分析，得知这一基因与家族性非息肉型大肠癌的发病有关，进而拿它作靶点进行药物设计。通过翻译软件把它的核酸序列翻译为蛋白质序列，然后进行序列比对，找到模式生物中的同源序列，利用模式生物蛋白质的可靠结构信息推测人类这一分子的三维结构，设计出可以和这个分子相互作用的配体，并将其合成。最后，进行实际的生物化学试验，验证其生物学活性。