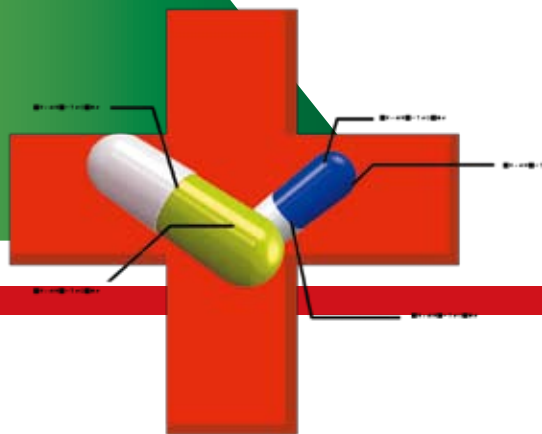


专题译述

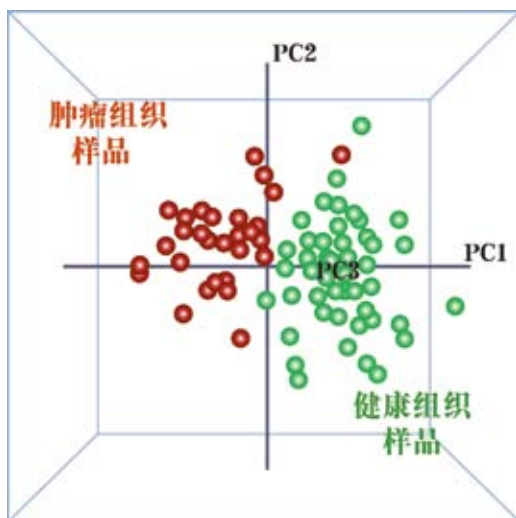


本期导读：

生物标志物——药物研发的催化剂

序言 生物标志物是医药领域一个十分热门的课题。随着对个性化用药的愈发关注与重视，研究人员开始致力于寻找那些在人体内因服用药物而发生相应变化的分子。随着基因组学、蛋白质组学、代谢组学及生物信息学等新兴学科的兴起，生物标志物的研究工作也得到了长足的发展。

1 生物标志物基础知识概述



每年，全球有多达700万人因癌症而死亡，我国也有100万人因此而失去生命。一直以来，罹患常见恶性肿瘤的患者的生存率都

非常低，尤其是那些直到晚期才被诊断出患有癌症的患者，其生存率更加低。例如，只有不到10%的结肠癌转移患者和5%左右的胰腺癌患者可以生存5年或5年以上。实际上，目前肿瘤诊治仍然采取“无差别对待”的方式，即对所有病人都采取同一种方法——按照肿瘤类型和分期进行诊断，在完全不考虑患病个体生物学特性的情况下对病人给予相同的治疗。在研究人员的不断努力、探求更好更合适的方法来治疗患者的过程中，他们发现，肿瘤生物标志物将有可能改变肿瘤治疗中这种不尽如人意的现状。

1.1 生物标志物概念

生物标志物（Biomarker）这一概念首次出现于国家研究委员会（NRC）在1983年出版的红皮书《联邦政府风险评估》中。它是指可以标志系统、器官、组织、细胞及亚细胞结构或功能的改变或可能发生的改变的生化指标，具有非常广泛的用途。例如，在医学领域，生物标志物可用于疾病诊断（例如前列腺特异性抗原PSA可用于前列腺癌诊断）、判断疾病分期（例如恶性肿瘤的分期）或者用来评价新药或新疗法在目标人群中的安全性及有效性。



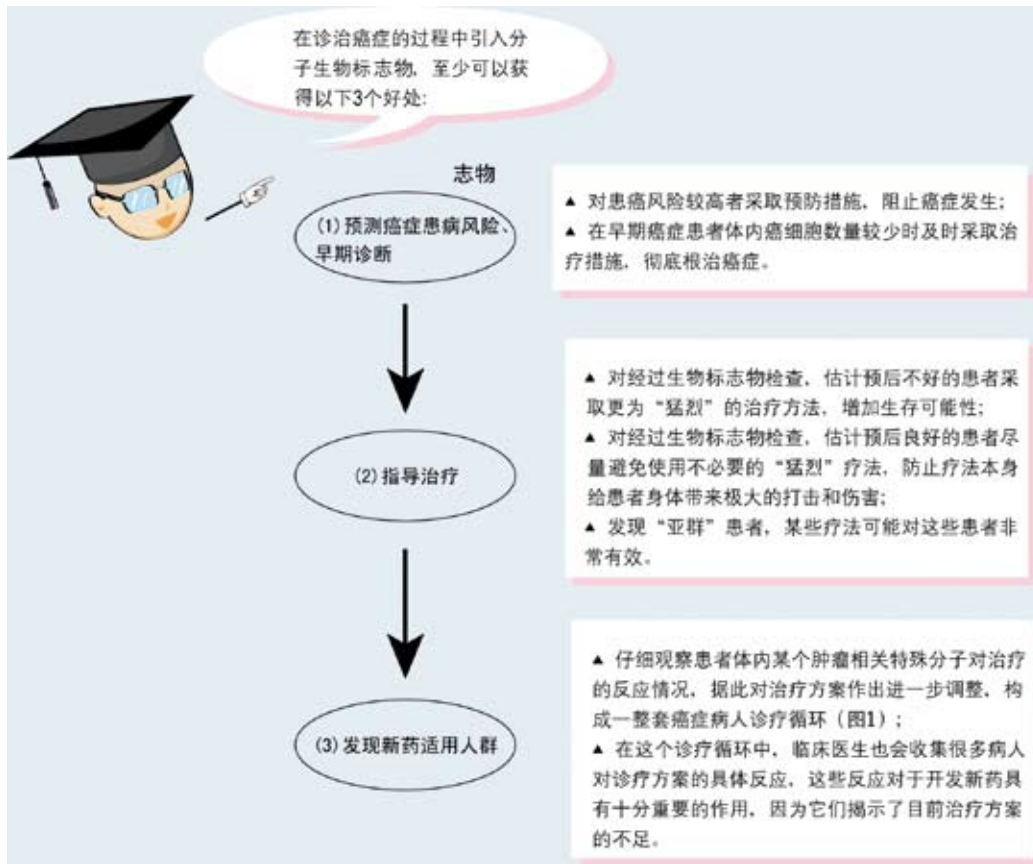
此外，生物标志物还可以用于流行病学或毒理学研究，以判断生物是否暴露于某些环境因素中（如有毒化学品、微生物等）。

人体内存在多种不同的生物标志物，包括基本身体状况、影像资料（例如乳房X线照片）、特定的分子（例如前列腺特异性抗原PSA）、基因突变（例如BRCA突变）、基因或蛋白表达谱（例如电泳检测血清蛋白诊断单克隆丙种球蛋白病）、细胞标志物（例如循环肿瘤细胞）以及其它生物标志物。

随着分子生物学技术的不断进展，肿瘤生物标志物包括的种类也越来越多，例如SNP、基因组、转录组和蛋白质组等等都被列入肿瘤生物标志物的行列，而人们也将生物标志物分为0型、1型以及2型。新型分子生物学技术彻底改变了人们以前那种只考虑肿瘤发病机理中某个单一因素的诊断方式。现在，人们可以从全局的视角去看待整个生物学系统对疾病发生过程中所产生的影响。

目前，生物标志物研究的主要目的是为了改善总体人群的健康状况，并降低罹患以下疾病的风险：如身体机能及消化系统疾病、免疫系统及神经系统功能障碍、动脉粥样硬化、肥胖症、糖尿病、癌症以及骨质疏松症。



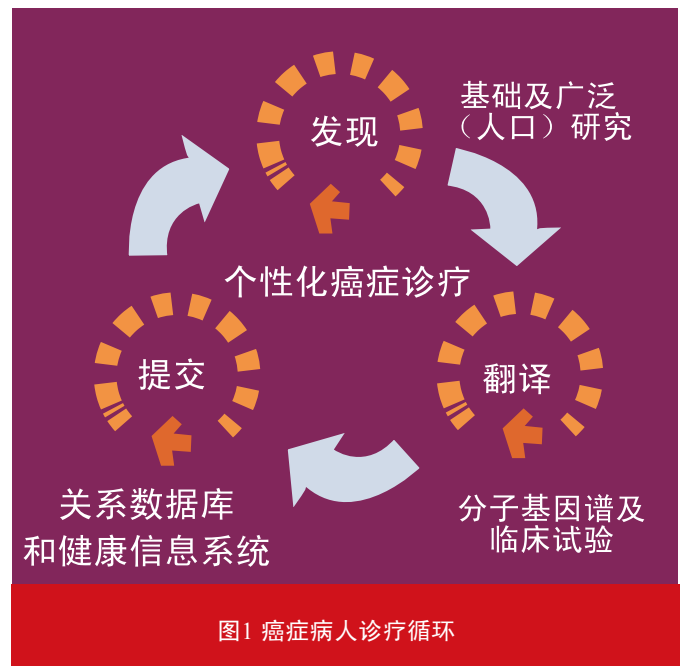


该循环的启动环节是鉴定出发生改变了的肿瘤特异性生物小分子，然后将这些发生改变了的分子与参与临床药物实验的患者的某些预后情况联系起来，以寻找患者的个性化癌症治疗方案。以此为基础所建立的关联数据库和健康信息系统，便可以为其他患者提供诊疗参考，也可为新药开发指引方向。

► 生物标志物是未来的疾病预测因子

现在，研究人员已经开始在貌似健康的、无症状的人群中，使用生物标志物来预测疾病的发生。到目前为止，已经有许多可以独立作为疾病预报因子的炎症标志物被鉴定出来，比如冠心病、II型糖尿病、胰岛素抵抗（IR，见文后小词典1）、高血压、血脂异常及肥胖等。

这些炎症生物标志物包括C反应蛋白、唾液酸、纤维蛋白原等，也包括一大类别的炎症因子前体，例如白介素6、肿瘤坏死因子α等。



目前的研究主要侧重于同时评价不同的炎症生物标志物，更好地了解它们与慢性疾病的关系，从而改进预测的准确性及灵敏性。

为了充分开发生物标志物的潜力，便需在大规模的人群中进行相应研究。因此采用一种操作更简单、费用更低廉的检测方法极为重要。只有经过科学、严谨的试验，才能得到令人信服的科学结果，并确保这些成分的有效性在今后可以为人们所接受。

1.2 生物标志物研究的新方向 ——突破从发现到确认的瓶颈



生物标志物研究包括为了解复杂的生物系统，对蛋白质混合物的定量研究，从而发现大量蛋白样品中真正起作用的蛋白分子之间的功能联系。研究生物标志物的目的就是为发现并证实那些能在临床上用来诊断病人、疗效监测或帮助药物开发的生物标志物。

一般而言，生物标志物的研究是一个连续的过程：从发现到确认，直至最终应用于临床。发现生物标志物的首要一步是筛选候选分子，对这些候选分子进行定量研究，根据它们在某些环境改变、给予药物治疗后或发生疾病的情况下是否发生相应的改变来判断这些候选分子是否为真正的生物标志物。上述整个过程都需要在大范围的患者或人群中进行这一筛选工作。为了避免因为大量样本可能给整个过程带来的瓶颈限制，一个能够预先初步检索候选生物标志物的验证步骤就显得尤为重要。这样，便可以保证只有那些最有可能“出线”的“候选分子”进入下一步的研究阶段——即昂贵的确认阶段。可见这个初筛的验证步骤对于节约成本来说具有非常重要的意义。一个高通量、高特异性及高灵敏度，同时所需的样品准备时间又较短的检测设备是用于初筛标志物的较为理想的设备，同时该设备最好还能够进行初步确认。此外，验证阶段的测量结果要能够用于确认阶段，也十分重要。

► 鉴定生物标志物工作流程图



生物标志物最终进入临床前要经历3个阶段

► 更快、更有效率地发现候选生物标志物

蛋白质生物标志物的发现

低丰度的样品要求、广覆盖的样品检测

在发现生物标志物的过程中，最困难的事情莫过于在成分复杂的生物样品中发现中等丰度或低丰度的蛋白质了。比如人类的血浆中含有超过 1×10^6 种不同的蛋白质分子，它们的含量差别可达到 10^{10} (Anderson, *J. Physiol* 563.1 (2005), 23-60)。在这么多的蛋白质中，含量最丰富的22种蛋白质就占了血浆蛋白质总数的99%，而某一个生物标志物的含量在样品中又是波动的。一种优良的能够发现生物标志物的设备，必须能够在血浆这样的生物样品中准确地发现一系列的生物标志物，并对其进行定量。

► 生物标志物研究系统

美国应用生物系统与其合作伙伴MDS Sciex公司正在为生物标志物研究工作制定新的标准。他们设计、开发了更加高级的基于质谱技术的设备来帮助研究人员发现、验证及确认生物标志物分子，并能提供高质量的数据，这套设备对样品的检测范围非常广。



带QSTAR® Elite LC/MS/MS System系统的
BIOiTRAQ™ Discovery System QS设备。

- 高速
- 高准确
- 高效
- 多功能



带4800 MALDI TOF/TOF™分析设备的BIOiTRAQ™
Discovery System TT设备。

- 最大的检测范围
- 强大的LC/MALDI
- 使用简便的MALDI，高敏感度



因为正常范围的临床差异性或生物差异性的存在，在发现生物标志物候选分子后的确认过程中需要对大量样品进行检测。因此，研究人员需要一种耗时更少的检测手段来检测尽可能多的样品，样品可以多达数百个至数千个。显然，对如此多的样品进行检验将是一个十分费时费力的工作，同时也是一个瓶颈步骤，如果对抗体或人工合成的肽段进行检测则更是如此。

对生物标志物候选分子进行验证——即初筛，有可能突破这一瓶颈。因为经过初筛后，只有那些最有希望的候选分子才能进入到下一步的确认程序中。MIDAS™系统就能同时进行多路（10个至100个分子）验证工作。成分复杂的样品，例如血浆或组织样品需要尽量短的准备时间，并且要能够在数分钟之内分析完毕。经过这个初筛步骤之后，最有可能成为最终的生物标志物的候选分子就可以进入下一步更加严格的确认阶段了。由于有了验证步骤，使人们节约了时间和经费，同时也提高了分析质量。

► 用于验证及确认阶段的设备



有4000 Q TRAP®和Tempo™ nano MDLC系统的MIDAS TRAQ™ 确认系统。

- 极高的敏感性
- 出色的定量表现
- 大范围的定量区间



有4000 Q TRAP®和Tempo™ nano MDLC系统的MIDAS TRAQ™ 确认系统。

- 一次试验定量测定及蛋白质确认
- 经济高效
- 高度的敏感性及出色的定量表现

1.3 新一代基于基因组学和蛋白质组学的生物标志物

尽管出现了多种不同的“组学（omics）”技术，但还没有可以用于临床的生物标志物产品的出现。美国食品与药物监督管理局（FDA）最近批准了几款属于“特定靶向性治疗”的新药，其中包括甲磺酸伊马替尼、吉非替尼以及曲妥单抗。

上述三种药物有一个共同特点，那就是它们各自针对特定的分子，而这些分子都是肿瘤发生所必需的。不过，慢性髓细胞性白血病（CML）则有所不同——只针对其中一个分子进行治疗并不能取得很好的治疗效果，即使只是将病情长期

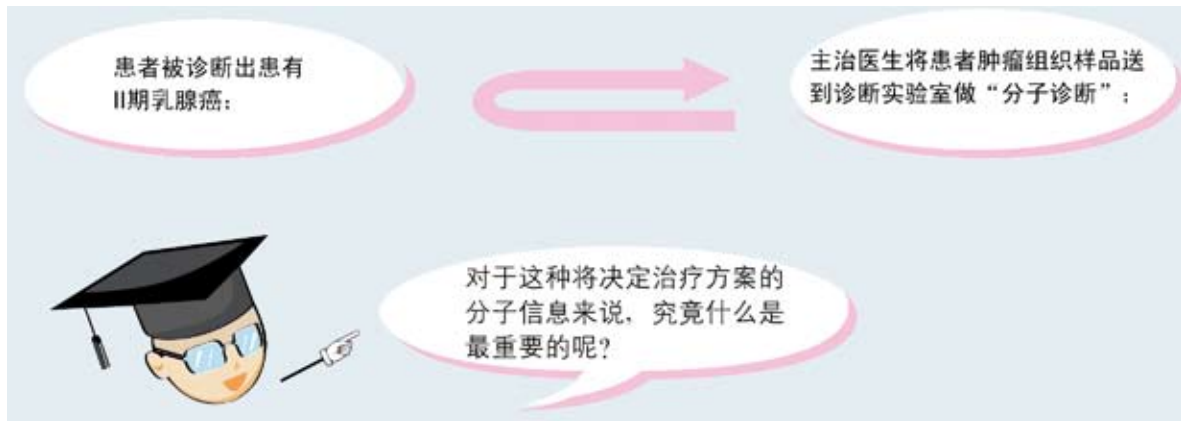
维持在缓解状态也十分困难。

人们越来越清楚，癌症是一个有很多因素参与的病理过程，甚至可能是由于身体的整个生理系统出现问题而导致的。只有针对整个系统进行治疗而不只是针对单个分子，才有可能取得更好的治疗效果。

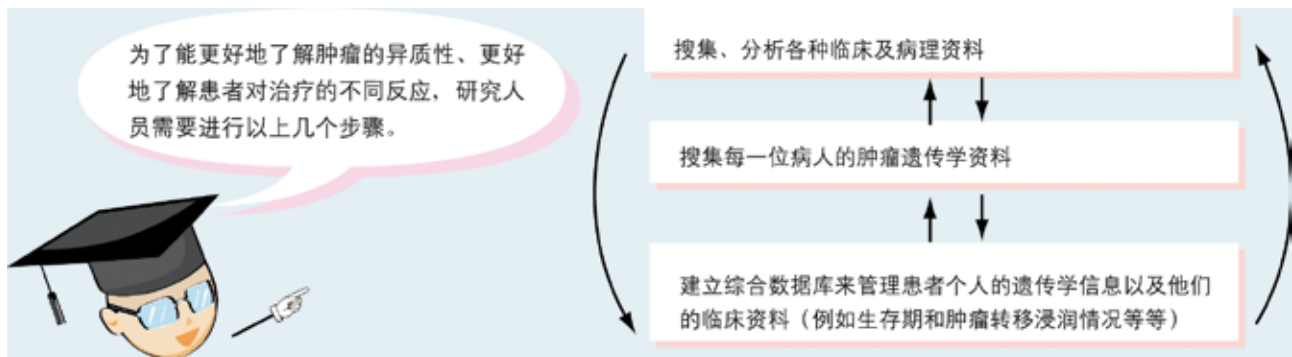
有鉴于此，“omics”技术，因其能发现生物系统不正常的表达方式，所以正成为既能评价癌症病人恶性程度即异质性（heterogeneity），又能在新药研发过程中参与评价生物系统反应的新技术。



既然如此，那还等什么呢？为什么这项技术还没有得到普遍应用呢？这是因为，在这项新技术被用于对临床肿瘤病人进行分析以前，还需要对这一复杂的技术本身进行反复测试，以证明其可靠性和可重复性等，此外还需要其它资料的辅助。



“omics”技术是能最准确分析肿瘤组织样品的技术。



要建立一个对于临床有指导意义并能帮助发现生物标志物的综合数据库，必需要搜集至少上万个患者以及他们所患肿瘤类别的详细资料。建立这样的数据库需要一整套标准的方法和最好的数据分析系统，该数据分析系统应该能将回顾性数据分析与前瞻性临床药物试验设计，以及患者一生的临床资料结合起来进行分析。这种关系数据库有助于判断使用某种疗法来治疗一个被临床诊断为乳腺癌的患者是否有效，因为人们可以从数据库中了解到和他情况相似、生物标志物也相似的其它病例的治疗情

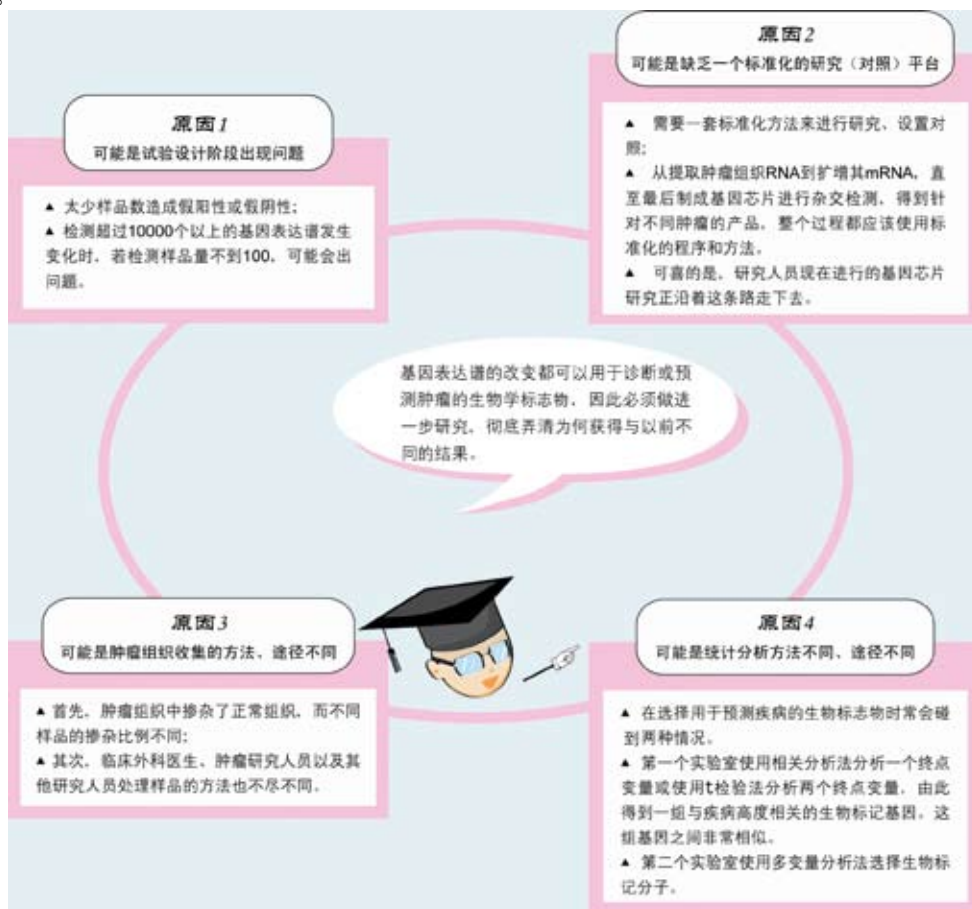
况，以资借鉴。理论上说，使用该数据库可以帮助临床医生为患者选择目前最有效、最合适的治疗方案。

毫无疑问，使用该方法也可以发现目前无法医治的患者，因此，这些患者就可以避免接受那些“摧残性的”而事实上是无效的治疗方法，转而考虑选择其它新型疗法。这项系统应该被纳入公众健康医疗信息系统，以便于社区肿瘤医生查询。不过目前，有关患者隐私的问题和公众对于遗传检查的担忧还在阻碍着这个系统的建立，不过，相信研究人员一定能找出解决的办法去应对。

► 基因组生物标志物足够强大吗？

最初，使用基因组学技术对肿瘤生物学标志物方面所进行的研究，都关注于发掘那些具有不同侵袭性的肿瘤组织在基因表达方面的差异情况。目前，研究人员已发现，这些标志疾病进展的生物学标志物在诸如乳腺癌、淋巴瘤乃至其它类型的肿瘤中并非单一分子，而是一系列基因的表达水平都发生上调或者下调的改变，这种改变与疾病的进展过程相关。

早期以酵母为模式生物所进行的工作，使研究人员掌握了如何利用多基因表达谱——或者叫做基因表达标志物以及形态搭配技术（**pattern-matching technology**）来寻求新的生物学发现。在过去5年内，许多研究小组陆续发现了一些可以预测侵袭性疾病的表达谱改变。出乎人们预料的是，里面的许多基因表达改变与以往的研究结果相比，仅有很少的部分是一致的，而接下来的研究工作又进一步证实了在早期研究中发现的路径机制的正确性。



多变量分析法从最有可能的基因（可能就是第一个实验室得到的基因）入手开始研究，然后再加入其它与该基因不同的基因进行研究。在最特殊的情况下，使用同一组数据，这两种方法得到的基因可能只有一种是共同的。这就会让临床工作者感到困惑，为什么会得到这样的结果呢？最终，尽管临床上的问题可能只有几个，比如哪一个生物标志物分子能用于预测侵袭性疾病，但实际上，在研究过程中人们选择的方法、实验数据不同，得到的结果（生物标志物分子）就会完全不同。

有一个很好的例子可以说明这个问题。这是在一家医院对乳腺癌患者进行的一个配对试验研究，该实验是为了研究与乳腺癌病情进展情况相关的生物标志物。研究人员分别选取了该院55岁以下的患者以及所有合适的患者作为两组试验对象。这两组试验对象得到了完全不同的两组试验结果，这明显和在选择实验对象时加入了年龄因素有关。

当所有的实验室、研究人员们都使用同样的样品、同样的技术平台以及用同样的标准来选择病人，那么得到的数据就要可靠得多了。虽然在这种情况下还是会有变量，但这些变量是能够反应实际生物学变化的变量的。在处于生物学标志物分子研究的起步阶段时，很重要的一点就是，上面提到的那些影响因素会妨碍研究人员建立一个能使患者利益最大化的标准。这甚至比研究预测治疗反应性的生物标志物分子本身还要重要。实际上，在研究能用于发现对药物敏感的亚人群的生物标志物分子方面，研究人员已经获得了一定的进展。

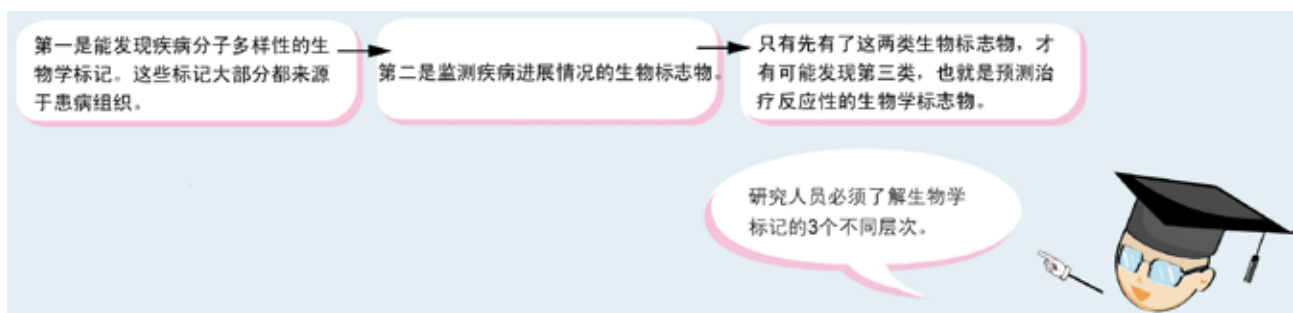
这些进展中的一项就是发现了EGFR基因突变是导致患者对EGFR抑制剂敏感或耐受的原因。其它用表达谱分析法发现的成果包括：发现了对紫杉醇、氟尿嘧啶和其它细胞毒性药物敏感的生物标志物。用同样的方法研究与疾病进展有关的生物标志物分子时也会发现与治疗相关的生物标志物分子。



他们的这种希望说明，他们完全忘记了在过去30年来研究肿瘤的过程中所经历的种种失败与教训——肿瘤并不像想象的那么简单，相反，它们非常复杂。不过现在，研究人员有能力面对这种复杂情况了：人们要全面了解每一个生物学标志物所代表的含义。如果要进一步改进，更好地为某一类患者开发出更适合的药物，那么就需要有一个更大的更富关联性的数据库，该数据库应该包含患者多年的临床资料，同时，还需要基于DNA、RNA和蛋白质的技术平台，只有这样才能取得成功。

如果单纯地希望能发现一个生物标志物分子就可以预测治疗反应性，这是非常幼稚的。

由此便引出了一个基本问题——需要对各种疾病状态进行检测吗？如果肿瘤有几百个甚至数千个不同的生存环境，那将是多么庞大的一个数据量啊，不过即便如此，也未必完全没有可能。



► 将生物标志物分子真正应用到临床中，我们准备好了吗？

这将给我们带来许多紧迫的问题和政治上的麻烦。我们已经在各个方面展开了全方位的紧密合作，我们共享技术平台，研究同一个问题。目前学术界、制药企业以及政府组织都准备投入到图1所示的那种循环模式中。

学术界一直都在为获得那些“看不见的”原创性知识产权而激烈竞争，同时也在为获得“看得见的”地位和资金而竞争。制药企业和生物技术企业也在为他们开发产品所担负的风险，以及是否能获得丰厚回报而竞争。与此相比，在过去的几年间，政府对于他们推动生物标志物的合作研究显得过于乐观了。

► 我们该何去何从？

我们真的应该好好学习互联网界——依靠学术界、企业界和政府组织间的大协作，才获得了如此成就。

互联网界的成功依赖于他们靠团队合作解决项目难题，同时业界有统一的标准，所有人都在这一标准下进行工作，由这一标准将大家整合起来，同时，该标准并不是由哪个政府或哪家企业单独制定的。这种大协作对互联网界的重要性从他们制定操作系统标准这件事就可以看出来。而在我们肿瘤生物标志物界，还有太多的参与者是研究领域之外的资助者、调整者和患者。当我们要把他们通通整合起来时，我们是不可能建立一个简单、快速的问题解决模式的，因此我们只能建立一个长期的新型的合作模式来研究肿瘤生物标志物。

如果不仅仅只是将重心放在设计临床实验以便证明治疗对某个人群具有效益，而是进行相应试验，对已知对治疗有反应的人群在接受治疗后所获得的收益进行统计学的评价，又会给我们带来怎样新的视角与疾病治疗方式呢？比较可能会出现的情况是，临床医师和（或者）癌症研究中心或治疗中心的研究人员不仅需要对接受治疗者

进行安全性、疗效方面的试验，还需要与肿瘤药物公司的科学家们合作，进行相应试验，并对患者接受治疗前的各项数据进行贮存，这种合作关系即使在药物已经获得FDA的批准之后还将延续。上述合作可以帮助我们建立一个科学、有事实依据为基础的系统，从而利于为今后接受治疗的患者提供最佳的治疗，因为在该系统的数据库中已经囊括了所有之前接受过治疗的患者的各项信息。

建立这样一个癌症病人治疗模式可能会使目前学术界、医疗界、制药界和政府间的职能界限变得模糊。通过目前的健康保险便利及责任法案（HIPPA）对患者的隐私进行保护，以及学术界已有的保护知识产权的方案，我们是建立这样一个治疗模式的。

这要求生物公司能找出还处于竞争前期的项目参与合作。在癌症治疗中心和政府组织间展开合作是最有可能，同时也是最有效率的合作方式。当然，要实现该领域的有效合作，我们还需要解决许多问题；但同时我们也应当看到，所有与此相关的研究者们都十分期待合作的开始。食品及药物管理局（FDA）最近所采取的举措表明，展开全方位合作的时机业已成熟。



1 胰岛素抵抗 (Insulin Resistance, IR) :

是指胰岛素作用的靶器官对胰岛素作用的敏感性下降，即正常剂量的胰岛素产生低于正常生物学效应的一种状态。目前认为，IR不仅是 II型糖尿病的发病基础，更是贯穿多种代谢相关疾病的主线，是连结它们的纽带，它可以作为这些疾病共同的病理生理基础。

胰岛素抵抗的根本原因是人体内胰岛素的接收器出现了问题。所以如果不彻底修复胰岛素接收器，而单纯的刺激胰岛素分泌，则无法从根本上治疗胰岛素抵抗。

1 代谢组学VS系统生物学：系统生物学的中心任务是要针对生物系统整体（无论它是生物细胞、多细胞组织、器官还是生物整体），建立定量、普适、整体和可预测 (QUIP) 的认知。具体而言，系统生物学研究就是要将给定生物系统的基因、蛋白质和代谢水平所发生的事件、相关性及其对所涉及生物过程的意义进行整体性认识。从而出现了许多的“组”和“组学”的新概念。但是现已提出的一百多个“组”和“组学”，可以大体归纳为“基因组”、“基因组学”、“转录组”、“转录组学”、“蛋白质组”、“蛋白质组学”、“代谢组”、“代谢组学”四个方面。显而易见，DNA、mRNA以及蛋白质的存在为生物过程的发生提供了物质基础（但这个过程有可能不发生），而代谢物质所反映的是已经发生了的生物学事件。因此代谢组学是对一个生物系统进行全面认识的不可缺少的一部分，是全局系统生物学 (global systems biology) 的重要基础，也是系统生物学的一个重要组成部分。

在现有的英文表述中，代谢组学同时存在两个不同的词汇和概念，即 **metabonomics** 和 **metabolomics**。尽管前者多用在动物系统而后者多用于植物和微生物系统，但这些概念的本质从他们的定义中能够得到较为细致的了解。**Metabonomics** 的最初定义是生物系统对生理和病理刺激以及基因改变的代谢应答的定量测定。这个定义现在可以更广泛地表述为：代谢组学

是关于定量描述生物内源性代谢物质的整体及其对内因和外因变化应答规律的科学，其中心任务包括（1）对内源性代谢物质的整体及其动态变化规律进行检测、量化和编录；（2）确定此变化规律和生物过程的有机联系。**Metabolomics** 存在多个定义，但其精髓是：对一个生物系统的细胞在给定时间和给定条件下所有小分子代谢物质的定量分析。因此，**metabolomics** 可以译作“代谢物组学”。不难看出，前者是对生物系统进行的整体和动态的认识，而后者强调分析而且是个静态的认识概念。

代谢组学属于全局系统生物学 (Global systems biology) 研究方法，便于对复杂体系的整体进行认识。譬如，一个正常工作的人体包括“人体”本身和与之共同进化而来且共生的消化道微生物群体（或称菌群），孤立地研究“人体”本身的基因、转录子以及蛋白质虽然可以为人们认识人体生物学提供重要信息，但却无法提供使人体正常工作不可缺少的菌群的信息。人体血液和尿液的代谢组却携带着包括菌群在内的每一个细胞的信息，因此代谢组学方法对研究如人体这样复杂的进化杂合体十分有效。正因如此，代谢组学已经广泛地应用到了包括药物研发、分子生理学、分子病理学、基因功能组学、营养学、环境科学等重要领域。可以预见，这门新兴学科将应用到更为广泛的领域。

2 人类的健康和疾病的防治始终是医学的目标，而其中疾病的早期诊断又是达到这一目标的重要环节。一些蛋白质或者降解后的片段即肽段，可以作为

肿瘤等疾病的生物标志物而用于早期诊断。然而，迄今能在常规临床使用的新型生物指标还非常少。在蛋白质组问世以后，人们将疾病早期诊断的生物指标寄望于蛋白质组的技术和由血清蛋白质组得到的数据。无数的研究小组致力于追踪那些有可能成为疾病标志物的多肽或者代谢产物，使得寻找生物标志物的工作格外繁荣。目前的常用技术中，以双向电泳和质谱最为突出，两种方法都可以用于比较正常和疾病状态下蛋白质组的变化情况。

4同位素标志物相对和绝对定量(iTRAQ)技术是近年来最新开发的一种新的蛋白质组学定量研究技术，具有较好的定量效果、较高的重复性，并可对多达四种不同样本同时进行定量分析。研究人员采用iTRAQ蛋白质定量技术加速蛋白质的定量研究，当然这门新兴的技术仍然需要进一步的完善。iTRAQ的操作程序一般如下。将蛋白质裂解为肽段，然后用iTRAQ试剂进行差异标志物。再将标志物的样本相混合，这样就可以对其进行比较。与样本结合后，通常用MudPIT多维蛋白质鉴定技术进行下一步的操作，用2D液相色谱串联质谱进行分析。在质谱分析鉴定特殊肽离子片断结构的基础上，采用美国应用生物系统公司的软件包MASCOT和Protein Pilot对每一个肽段进行鉴定。

原文检索：

www.sciencemag.org

<http://zh.wikipedia.org/wiki/%E7%94%9F%E7%89%A9%E6%A0%87%E8%AE%B0>

<http://baike.baidu.com/view/63384.htm>

http://www.wellnesswest.ca/dmdocuments/TW_7_low%20res.pdf

http://www3.appliedbiosystems.com/cms/groups/psm_marketing/documents/generaldocuments/cms_040631.pdf



2 生物标志物的开发与应用

2.1 癌症生物标志物的开发与应用

十年来，研究人员一直在寻觅一种有效且对机体无损伤的肿瘤检测方法。终于，在2003年，Cui等人宣称找到了一种理想的肿瘤检测方法^[1]。他们开发了一种基于DNA的血液检测法来评估病人罹患结肠癌(CRC)的风险。Cui的这项研究成果受到肿瘤生物学界密切关注，因为该成果是肿瘤检测生物标志物应用及开发领域所取得的重大进展。

Cui等人对172例结肠癌病人的组织切片及血液样本进行了研究，分析了这些样本中胰岛素样生长因子II(IGF2)基因的印记缺失(LOI, 见文后小词典)

情况。

基因甲基化也可发生于生命晚期。Cui等人证明IGF2基因的印记缺失不仅与CRC家族病史相关，同时也证明与个人结肠腺癌和CRC病史相关^[1]。LOI与家族性CRC之间的关联尤为引人注目，因为多达30%—50%的散发性CRC都具有家族病史，但这其中的遗传机制目前还不清楚。然而人们目前已经发现了症状明显但是十分罕见的遗传性CRC综合症—家族性结肠腺瘤性息肉病(familial adenomatous polyposis coli)及遗传性非息肉性结肠癌的分子发病机制。