
第 1 部分

乳腺癌组学方法

第1章

乳腺癌研究的组学技术

Mariana Panal Cusati, Maria Herrera
de la Muela, Ignacio Zapardiel

摘要

组学技术是大范围定量检测细胞内组分的技术。用高通量的基因表达谱方法研究基因组已经在分子水平加深了对各种疾病特别是肿瘤的发病机制的了解。组学技术从一诞生起就用于乳腺癌的研究,它揭示了肿瘤的异质性、基因组复杂性以及驱动肿瘤发生的分子事件。组学技术产生的整套遗传信息使我们了解到乳腺癌是一种异质性的疾病,包括多变的形态、不同的分子特征以及不同的临床表现。因此,组学技术目前可用于鉴别基因标签,以对肿瘤进行更精准的诊断、预后和治疗。多个研究团队从乳腺癌研究中获得大量宝贵的遗传信息并保存在公共数据库中,以供科学家进行分析、整合,以便获得对乳腺癌更深入、更彻底的了解,且最终改善乳腺癌患者的临床治疗效果。

关键词

乳腺癌 组学技术 高通量 基因表达谱 分子谱学技术 基因组学 表观基因组学 蛋白质组学 转录组学

引言

虽然都用相同的名字,但现在则认为乳腺癌指的不是单纯的一种疾病,它可分为不同的亚型,有不同的组织学、生物学及分子特征。这些变异性导致癌组织对治疗会产生不同的反应,因此就存在不同的预后^[1]。目前在临床上,肿瘤大小、组织分级、淋巴结状态、激素受体表达情况被用于制订治疗方案和临床预后评估。但由于乳腺癌的很多遗传和分子机制还没有充分研究清楚,这些预后指标缺乏精确性^[2]。由于害怕肿瘤复发,一些患者被过度治

疗。因此,发现乳腺癌新的标志物来对它进行检测并进行更细致的亚型分类以及预测患者对治疗的反应有着重要的意义^[3,4]。

组学时代的诞生

1920年,植物学家 Hans Winkler 使用了一个新词“genome(基因组)”来描述动物或植物的遗传物质,genome是由gene(基因)和chromosome(染色体)合并而来的。一些科学家解释后缀“ome”是单元的集合,例如基因是一个单元,基因组就是基因的集合。随后“ome”这个后缀用来生成新的词汇。第一个

新词汇是“proteome”，它描述的是从基因组产生的整套蛋白。在 1990 年末，单词“genomics（基因组学）”开始用于描述研究和应用基因组上的信息，这样就产生了“omics”这个单词后缀。这个新的遗传语言开启了 omics（组学）时代，现在组学是指对生物系统的综合分析^[5]。

组学是高通量技术，用于大范围定量分析细胞内的组分，例如基因组和蛋白质组。这种技术上的突破已产生了肿瘤基因组学，或者说是通过各种层次的方法学，例如 DNA 拷贝数、DNA 甲基化、转录水平和基因组测序对肿瘤细胞产生的谱学数据。肿瘤基因组学使得我们认识了肿瘤的遗传通路，让我们更深入了解了肿瘤生物学，并且由此发现了新的诊断、预后判断和治疗方法。组学技术揭示了基因组的复杂性、肿瘤的异质性和驱动肿瘤发生的分子事件，通过识别这些现象，可以提高肿瘤治疗的特异性^[6]。

总之，我们可利用组学技术来深入研究肿瘤细胞，包括使用 DNA 或 RNA 拷贝数、转录水平、DNA 甲基化多种水平的遗传信息，直至代谢信息，这些来自组学技术所产生的数据^[7]。

我们将描述与分子生物学中心法则“基因-DNA-RNA-蛋白”相对应的组学研究“基因组学-表观基因组学-转录组学-蛋白质组学”策略，展示它们是如何应用在乳腺癌研究中的。图 1.1 解释了组成分子生物学中心法则的基因、DNA、RNA 和蛋白质，以及获得中心法则中组分信息的组学技术之间的关系。

基因组学

基因组学是指对基因、DNA 结构和功能进行综合分析或者寻找整个基因组信息的学科。在组学技术诞生之前，科学家是对一个一个的基因进行 DNA 测序；但在基因组学方法中，改进的信息分析方法和生物技术整合到一起后，能对多种生物的完整基因组 DNA 进行测序，并对其进行分类后存储在基因组数据库中。基因或 DNA 序列的变异表现在基因扩增、基因缺失或基因重排中^[8]。

拷贝数变异（copy number aberrations, CNA; 或 copy number variations, CNV）是指一个细胞内的 DNA 量或结构发生改变。这些遗传性的改变包括染色体缺失、重复、倒位、易

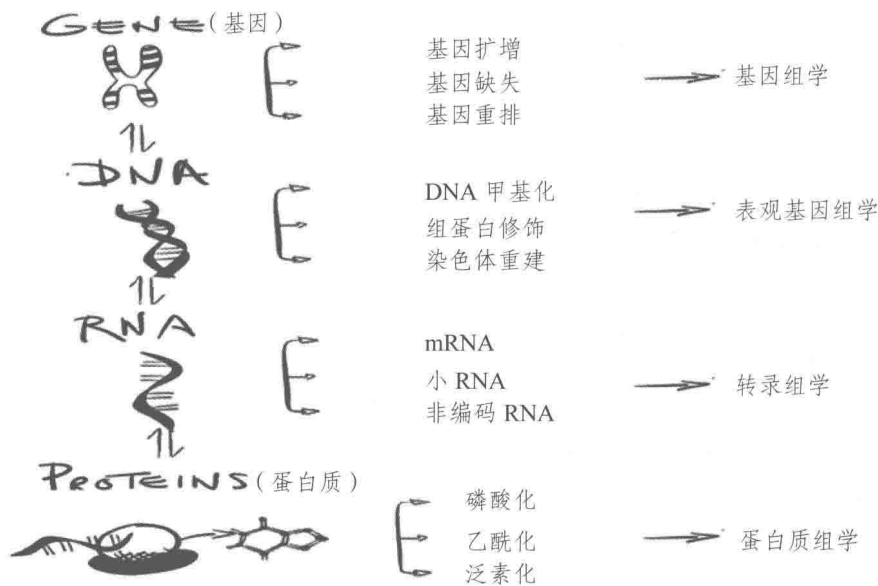


图 1.1 分子生物学中心法则中的各种组学技术包括的基因、DNA、RNA 和蛋白之间的关系。

位,而且这些变异是可遗传的。拷贝数变异能改变基因转录的活性,增强或减弱基因的表达水平^[9]。

基因扩增是导致基因过表达的原因之一,也是癌基因被激活的原因之一,与疾病的发展和不良的预后相关。在肿瘤发生过程中,基因扩增常在染色体上包含癌基因的特定区域发生,导致癌基因被激活^[10]。HER2 基因在乳腺癌中是研究得最多的癌基因,它位于染色体 17q21.1 区段。HER2 基因编码一个跨膜蛋白,与表皮生长因子 HER1 类似。一个正常的细胞包含 2 个拷贝的 HER2 基因,表达 5 万个拷贝的 mRNA。而在乳腺癌细胞中,有多个 HER2 基因拷贝出现,编码超过 100 万个 mRNA 拷贝^[11]。

高通量技术可用来发现、监测、定量这些 DNA 改变。用得最广的组学技术是微阵列芯片技术,它最初用来监测成千上万个基因表达水平。微阵列芯片技术发展很快,现在可以在芯片表面固定组织或一组 DNA/RNA 分子、蛋白、抗体,这样的芯片可同时用于研究一个样本中的所有 DNA、RNA、蛋白和抗体。基因组微阵列芯片也称为比较基因组杂交芯片(array-CGH),可在整个基因组水平用于定量研究染色体数目异常、微缺失、微重复以及拷贝数变异。之前的比较基因组染色体杂交(chromosomal-CGH)只能大致定位 DNA 缺失或重复的基因位置,而 array-CGH 则可精细定位^[1,12]。

chromosomal-CGH 技术发现 16q 缺失是在浸润性小叶癌(ILC)中最常见的染色体变异。Etzell 等在研究原位小叶癌(LCIS)基因组变异时,他们用 array-CGH 发现 LCIS 患者中 88% 都出现了 16q 缺失^[13]。Mastracci 等在研究非典型性小叶增生(ALH)和 LCIS 时,用 array-CGH 发现在 ALH 和 LCIS 患者中都存在 16q21-q23.1 的缺失^[14]。共同的遗传缺失区段表明 ILC 和 LCIS 之间存在联系,ILC 和 LCIS 可能是同一种疾病的两种状态。LCIS 可能是 ILC 的前体形式,16q 缺失也就

成为预测乳腺癌风险的一种标志物。这就是一个利用基因组学研究成果应用于发现乳腺癌的发生、进展和转移标志物的例子^[13,14]。

乳腺癌组织基因表达谱可检测肿瘤细胞中所有基因的表达信息,包括基因之间互作、基因和环境之间互作引起的基因表达改变。具有相同表达特征或和某一临床症状相关的一组基因称为基因谱或基因标签。基因表达谱已用于乳腺癌的分子分型。临床上不同特点的乳腺癌同其基因表达谱的多样性是相吻合的,这种基因表达谱信息多是通过 DNA 微阵列芯片技术得到的^[1]。Perou 等利用 DNA 微阵列芯片技术比较乳腺癌组织和正常乳腺组织发现了一组基因,他们称之为“内在基因”,因为这些基因固定在某一个患者的样本中表达,而不在其他患者的样本中表达。Perou 等通过这组基因表达的信息,把乳腺癌分成了 4 种分子亚型,包括管腔样(ER+/luminal-like)、基底样(basal-like)、HER2 阳性型(HER2 enriched)以及类正常型(normal breast-like)^[15]。分子分型后,每种分型可适用不同的治疗方法^[16,17]。表 1.1 总结了几个研究乳腺癌分子分型的新发现以及所使用的技术。

基于基因表达谱对乳腺癌进行的分型

管腔样

这种亚型的乳腺癌患者行内分泌治疗是有效的,化疗敏感性差,但其在所有亚型中预后是最好的。Sorlie 等发现管腔样还可分为 A 型和 B 型, A 型的预后要好于 B 型^[18]。在管腔样 B 型中,HER2 的表达要高于 A 型, B 型对化疗的反应略好于 A 型,但 A 型复发的风险较低^[18,19]。

基底样

基底样乳腺癌有细胞角蛋白的表达,但没

表 1.1 几个乳腺癌内在亚型的研究

研究者	所用的主要技术	其他技术	新发现
Sørlie 等,2001 ^[18]	cDNA 微阵列芯片	等阶聚类	发现管腔样还可分为 A 型和 B 型
Sørlie 等,2003 ^[19]	cDNA 微阵列芯片	等阶聚类	BRCA1 基因和基底样的关系
Sørlie 等,2003 ^[19]	cDNA 微阵列芯片	等阶聚类	ER 状态和内在亚型之间的关系
Abd El-Rehim 等,2004 ^[21]	组织微阵列芯片	免疫组织化学	细胞因子之间的关系
Carey 等,2006 ^[22]	微阵列芯片	免疫组织化学	各亚型在乳腺癌患者中的比例
Hu 等,2006 ^[65]	微阵列芯片	等阶聚类	预测生存率

有雌激素受体 (ER) 和 HER2 的表达。无论淋巴结的大小和状态是什么情况,基底样乳腺癌的预后都是很差的。当患者出现 BRCA1 基因突变时,该亚型的乳腺癌是最常见的^[16, 19, 20]。

HER2 阳性型

这类乳腺癌 HER2 基因高表达 (指 IHC+3 或 FISH 阳性),并且在管腔样中特异表达的基因在该类型的乳腺癌中不表达。并不是所有经过免疫组织化学检出 HER2 表达的样本在分子分型上都属于 HER2 阳性型。基底样和 HER2 阳性型乳腺癌患者预后最差,管腔样 A 型预后最好,管腔样 B 型预后中等^[18, 19]。

目前乳腺癌通过激素受体的表达情况分成生物学性质不同的表型,这已成为临床预后评估和治疗的基础。基因组学通过研究乳腺癌的内在亚型和临床特征的关系,把表型和基因型联系起来。乳腺癌的分子分型和预后的关系通过基因表达谱得到了深入研究。通过表达谱研究,发现了激素受体表达情况、HER2 基因表达情况和其他一些基因的表达特性在乳腺癌组织中是不同的,从而奠定了这些差异特征在临床进行应用的基础。在乳腺癌各亚型中,管腔样 A 型和 B 型激素受体高表达,而 HER2 阳性型和基底样的激素受体低表达。在后续的研究中,有发现乳腺癌中还存在紧密连接蛋白低表达型 (low claudin)、干扰素高表达型 (interferon rich)、雄激素受体型 (androgen receptor)、类正常型 (normal-like) 等新亚

型^[19]。还有很多研究发现传统的预后因素和基因表达标签之间存在联系^[21-23]。

基因组测试

乳腺癌的分子分型已揭示了乳腺癌具有不同的肿瘤特征,包括侵袭能力和对化疗反应的不同。乳腺癌癌细胞根据侵袭不同组织的倾向性而表现出不同的转移性。Smid 等利用微阵列芯片和芯片统计学显著性分析软件 (SAM 软件) 研究了乳腺癌不同分型具有器官偏向性的转移能力,转移处的癌细胞和原发灶的癌细胞具有相同的遗传特性^[24]。

由于采用了基因表达分析技术,我们发现乳腺癌分成不同的亚型。并且新的实验数据表明,乳腺癌的亚型可能会更多,但这些新亚型目前还未进入到临床应用,除非针对这些新的亚型乳腺癌能找到针对性的治疗方案^[11, 15]。

基于乳腺癌相关的基因表达信息和以此做出的乳腺癌内在分型,目前已经有几种分子诊断技术在实验室得到应用,用于患者的个体化诊断,以评估预后和治疗方案。这些来自基因组学的分子诊断方法计算出临床的风险性,例如若患者不用化疗或只用激素治疗,患者复发的风险有多大? 或者患者用任意的方法治疗,复发的风险有多大? 这些分子诊断方法尽管选择的基因不同,但在对乳腺癌患者预后进行评估时,结果是类似的。MammaPrint 和 Oncotype DX 目前是两种用得最多的检测方

法^[1,3]。

MammaPrint 方法通过检测 70 个基因的表达情况来确定患者是属于低风险的亚型,从而可以避免化疗。它也能鉴别出哪些表面上临床指标较好的患者,但预示预后不良的基因表达,这意味着这些患者的预后并不好^[1,3,4]。发现 MammaPrint 方法的 Van de Vijver 是使用了喷墨技术合成的寡聚核苷酸微阵列芯片来发现与乳腺癌患者预后相关的 70 个基因的,这 70 个基因的表达可预测患者在 5 年内远处转移的可能性^[25]。

Oncotype DX 是用 21 个基因的表达情况对雌激素和孕激素受体阳性以及淋巴结阴性的乳腺癌患者进行个体化诊断,以避免过度化疗^[1,3]。Paik 等利用反转录 PCR 发现这 21 个基因的表达情况与接受他莫昔芬治疗、淋巴结阴性患者的癌细胞远处复发转移有关,并且可以给每个受检的患者注明复发的风险系数^[26,27]。

尽管这些分子诊断方法开始应用于临床的日常工作,但由于这些检测的效果还没有经过长时间的考验,因此对这些检测的可靠性还没有形成统一的认识,也还没有形成统一的临床指南。但业内人士都认为,这些新型的检测方法需要同传统方法结合起来应用^[16,17]。

区分哪些患者需要接受化疗、哪些患者不需要是个体化医学要做的事情,此时需要找到一组基因,这组基因的表达情况和患者的临床结局是相关的^[4,11]。为了获得更多的科学数据,前瞻性的临床试验——针对 MammaPrint 检测的临床试验 (MINDACT) 和针对 Oncotype DX (TAILORx),都在进行当中,以期待改善这些检测的准确性^[28-30]。

把基因组学得到的数据应用在临床上最大的不足是这些数据往往基于较少的临床样本数目。因此,最新的研究倾向于采用更多的活检组织样本,并检测其中更多的遗传信息,从而产生了乳腺癌的基因组数据图谱^[6]。利用大样本来整合研究乳腺癌的基因组、转录组,并寻找肿瘤事件的驱动基因,得到数据的

准确性肯定要好于小样本^[6,31]。现在很多医院开始建立生物样本库,保存了不同患者、不同肿瘤、不同治疗方案的样本。生物样本库中的样本,可按照是否属于同一个家系、治疗前和治疗后以及种族特性等不同的标准来提供样本,使得基因组学获得的数据库范围得以延伸。但在使用生物样品库中的样本时,针对不同的研究目的,必须要考虑伦理问题^[32]。

一个存储新产生的癌基因组公共数据库的例子就是癌症基因组图谱 (the Cancer Genome Atlas, TCGA)。这个数据库是多方协助努力的结果,力求通过应用基因组分析技术,包括大规模基因组测序技术,来促进对癌症分子基础的理解。TCGA 数据库最初是由美国国家癌症研究所 (NCI) 和美国人类基因组研究所 (NHGRI) 发起创建的,应用了多种组学技术来分析可以同癌症关联的 DNA、RNA、蛋白和其他生物成分。目前这些数据已经有多个癌症组织的样本,产生了一系列的基因突变数据。

现在针对乳腺癌已产生了海量的遗传学数据,但还缺乏同患者临床症状的最终关联^[6,11]。Curtis 等从 2000 多个乳腺癌组织样本中分析了基因组和转录组水平的信息,包括肿瘤体细胞 DNA 拷贝数变异、单核苷酸多态性。他们发现了乳腺癌新的分子分型^[33]。TCGA 协作组在 2012 年通过该数据库中用 DNA 拷贝数芯片、DNA 甲基化、外显子测序、信使 RNA 芯片和 microRNA 测序以及反相蛋白芯片等技术产生的数据,发现通过几个基因的突变和一组蛋白的表达信息,可以把乳腺癌分成 4 个主要的类别。他们还发现基底样乳腺癌和高级别严重卵巢癌之间存在分子水平的相似性^[34]。

表观基因组学

表观遗传修饰是不改变 DNA 一级序列结构变化的,例如 DNA 甲基化、组蛋白修饰、染色质或核小体重建。这些改变是可遗传的,

也是可逆的。表观基因组学就是研究以上全部表观遗传修饰现象的学科^[7,35]。

DNA 甲基化、组蛋白修饰、染色质或核小体重建相互作用来调节基因表达。这些遗传修饰有时会导致细胞的正常基因表达失调。表观遗传学修饰过程中包括染色质形成疏松的状态,从而使上面的基因转录变得活跃,此时的染色质称为常染色质;若形成聚缩的状态,基因转录静止态,此时的染色质称为异染色质。这些过程就可调控基因的表达或关闭。

癌症似乎是这些表观遗传改变所驱动的,这些改变使得与细胞增殖、生存、分化相关的基因表达过程失调。因此,在乳腺癌细胞中,DNA 甲基化、组蛋白修饰、染色质或核小体重建可引起肿瘤的发生、发展和转移,以及治疗手段的失败。为了找到表观基因组学在临床上的应用,了解表观遗传改变如何触发和维持癌细胞中抑癌基因沉默和癌基因激活,是十分必要的^[36-38]。

DNA 甲基化

DNA 甲基化是指 DNA 甲基化转移酶把 -CH₃ 甲基基团加到 DNA 的某个区域上。甲基化现象可导致基因组 DNA 不稳定及重排、癌基因被激活、抑癌基因被抑制,最后导致细胞的癌变^[36]。在一个细胞内可能发生原有的甲基化丢失,出现去甲基化(hypomethylation)或原不发生甲基化的位置出现过度甲基化(hypermethylation)的现象。目前我们已经可以检出乳腺癌组织中发生的甲基化事件。通过甲基化图谱检出技术,我们能够筛选出一组基因的甲基化谱来区分乳腺癌的亚型,诊断乳腺癌患者的临床分期和预后情况。

DNA 甲基化分析方法可分为基因组学甲基化分析和单基因特异性甲基化分析。前者是用于测试细胞内甲基化的胞嘧啶总体水平,后者是通过甲基化特异性酶切方式确定到具体甲基化的基因。单基因特异性甲基化分析中需要用到 PCR 技术,该技术在组学中经常用到。PCR 是通过酶学反应,可将目标区段扩

增上百万倍。而甲基化 PCR 技术(MSP-PCR)则可检出基因组中发生甲基化的区段^[39,40]。

Dejeux 等在进展期乳腺癌患者接受阿霉素新辅助化疗之前,用焦磷酸测序法分析了癌组织中数个基因的甲基化情况,发现 14 个基因中有 9 个基因出现了甲基化程度的改变,其中 3 个基因的甲基化改变可能用于判断患者的预后以及预测患者对新辅助化疗的反应^[41]。

Hsu 等通过 MSP-PCR 方法研究了 BRCA1 基因启动子区域的甲基化情况。他们发现, BRCA1 基因的甲基化情况和三阴性乳腺癌患者的生存期存在明显的相关性^[42]。

组蛋白

组蛋白和 DNA 一起组装成核小体,形成了染色质的基本单位。核小体由 DNA 分子缠绕 4 个组蛋白分子组成。组蛋白修饰和染色质重建是通过各种酶的作用来完成的。当正常的 DNA 排列模式出现异常后,会启动组蛋白的修饰过程,引起染色质上基因转录的激活或抑制,包括抑癌基因和癌基因的激活或抑制^[37]。

研究组蛋白修饰较 DNA 甲基化难度大,需要高通量的蛋白质组学技术,这些技术在后续的蛋白质组学这一章中会进行专门的描述。其中有一项技术称为基于质谱方法,在培养细胞中对氨基酸进行稳定同位素标记(stable isotope labeling with amino acids in cell culture, SILAC)^[43]。Cuomo 等利用 SILAC 技术比较了乳腺癌细胞系和正常乳腺组织的组蛋白修饰差异,他们发现,在乳腺癌细胞系上的组蛋白修饰发生了很大的改变,这些改变可作为生物标志物,他们称之为“乳腺癌特异的表观遗传标签”^[39]。

免疫组织化学(IHC)

另外一项配合组学研究的技术是 IHC, IHC 可在组织微阵列芯片上,通过荧光物质标记的抗体检出组织中特异的抗原^[44]。Elsheikh 等使用该技术来检测乳腺癌组织中特异的组

蛋白标志物,他们发现当组蛋白标志物含量不同时,乳腺癌患者的临床和病理特性也相差很大,组蛋白修饰也与乳腺癌类型相关,还与乳腺癌生物标志物和表型,例如雌激素受体、孕激素受体、管腔角蛋白表达和其他都有联系^[45]。

转录组学

转录是指遗传信息从DNA复制到RNA的过程。RNA可作为模板翻译出多肽蛋白组。RNA还可参与基因调控和酶活性调节^[46]。转录组是指细胞内所有的RNA集合,而转录组学是指研究转录组的技术,对一个细胞内的所有RNA进行检测和分类。转录组包括不同类型的转录本,例如mRNA、非编码RNA和小RNA。基因的结构决定了转录的起始位点以及在不同条件下转录本的表达丰度^[47]。转录组学技术描述所有类型的转录本,发现基因转录的起始点,评估转录后修饰,解码所有的RNA序列,定量后绘制出RNA表达图谱^[46]。

基因表达

基因表达是对基因活性的研究,需要定量测量DNA转录到mRNA以及从mRNA翻译出蛋白^[12]。了解癌基因的表达后就可清楚驱动肿瘤发生的遗传机制。各种高通量的组学技术,包括cDNA芯片、Oligo芯片和新一代测序可用于检测转录组^[46]。

互补DNA(cDNA)

cDNA是指一段和mRNA序列互补的DNA序列。在实验室中,可通过先提取mRNA,然后以mRNA为模板来形成cDNA。cDNA芯片可以通过把cDNA片段固定在玻璃片表面,然后把mRNA通过化学或荧光方式进行标记,再后通过杂交来检测mRNA的表达丰度。

寡核苷酸(oligonucleotide)是一段短的核苷酸序列。Oligo芯片是把人工合成的Oligo探针放在芯片上。由于Oligo探针是选择性地设计合成,因此杂交的特异性会更好^[1,46]。

Yao等用cDNA芯片分析了导管原位癌、浸润性乳腺癌和淋巴结转移乳腺癌样本的基因组拷贝数差异,同时用SAGE方法分析这些样本的表达变化,找到了两个过表达的基因H2AFJ和EPS8是由于DNA拷贝数扩增导致的。他们还发现,在浸润性的肿瘤中,DNA拷贝数变化频率要高于导管原位癌。这项研究表明,cDNA芯片可作为检测基因表达变化的工具,而这些表达改变的基因可能作为乳腺癌治疗的靶标^[12]。

把新一代测序技术用于RNA测序是一种新型检测RNA丰度的方法。该方法中,先把RNA通过已知的序列转化成cDNA文库,然后对cDNA文库中的cDNA片段进行单分子测序,这些片段化的序列再和mRNA的参考序列进行比对拼接就能鉴别出具体是哪个基因表达的RNA^[47]。Huber-Keener等用RNA测序比较了药物他莫昔芬抗性和敏感的乳腺癌细胞系中转录组基因表达的差异情况。他们发现一些mRNA和小RNA的表达改变和他莫昔芬抗性产生的机制相关,这些基因涉及雌激素受体功能、细胞调节、转录调控、线粒体功能等。因此,通过检测基因表达的改变可用于预测是否会产生他莫昔芬抗性^[48]。

蛋白质组学

蛋白质组学是通过系统性研究蛋白来综合了解它们的结构、功能以及对生物系统的调节作用^[49, 50]。蛋白质是细胞功能的最终体现形式,它们能催化酶学反应、参与细胞构成、传递细胞信号,以及行使其他细胞内和细胞间的功能。蛋白质组的复杂性和可变性超过了基因组^[7, 50]。

蛋白可发生多种形式的修饰,例如磷酸化、乙酰化、泛素化等,这些修饰都称为翻译后修饰(PTM)。多项蛋白质组学技术被用于分离、鉴定、定量、分类乳腺癌组织中低丰度的蛋白。这些技术包括电泳、质谱以及表面增强激光解离质谱(SELDI-MS)、二维差异凝胶电泳(DIGE)和多维蛋白鉴定技术(Mud Pit)等多

技术联合使用新技术。

蛋白质组学也采用微阵列芯片技术。蛋白芯片是在芯片表面固定组织或者一组蛋白、抗体,然后可用来检测样品中的蛋白或抗体。利用蛋白芯片来发现乳腺癌组织中特异的蛋白,可用于未来特异性药物治疗靶标的鉴别,以及发现诊断乳腺癌的新标志物^[40,44,51]。

双向电泳(2DE)是研究蛋白谱的一项技术,它是根据不同蛋白的一些特性不同,例如等电点、分子量不同,通过凝胶电泳把蛋白进行分离,然后再通过质谱法(MS)对分离的蛋白进行鉴定。质谱法是蛋白质组学研究的一项基础技术,它能检出蛋白中的不同多肽,使其中的化学组分离子化产生带电荷的分子,然后测量这些分子的质量和电荷的比值。质谱法能直接鉴定通过凝胶分离到的蛋白^[44]。结合双向电泳和质谱法,Rowell等研究了大豆中提取的主要异黄酮成分(木黄酮)对大鼠乳腺组织的影响。他们发现经过异黄酮处理的乳腺组织,细胞中的蛋白质组发生了变化,这些变化促进了细胞增殖、细胞分化以及腺体成熟,从而增加了患乳腺癌的风险^[52]。这个例子说明蛋白质组学研究可以让我们加深理解乳腺癌是如何启动的以及发现乳腺癌新的分子标志物和治疗靶标。

表面增强激光解离质谱是一项用于蛋白质组学分析的高通量技术,它能根据蛋白或多肽的表面色谱性质差异选择性进行分析^[44]。Hu等用SELDI-MS技术研究乳腺癌患者的蛋白质组学,在比较了乳腺癌患者、良性乳腺疾病患者和正常女性对照后发现了4个新的标志物,并能用它创造了一种具有显著灵敏性和特异性的诊断模式^[53]。

二维差异凝胶电泳是指在同一块胶上从不同的样本中分离蛋白,因此,可在同一张二维凝胶上定量分析2个或3个样本之间的蛋白差异点,使得蛋白质组中相对意义上的大部分蛋白可以同时可视化^[54,55]。Davalieva等用DIGE分析浸润性导管癌患者的癌组织和癌旁的正常组织的蛋白质组学,他们发现了在癌组

织中过表达的蛋白,这些蛋白在以前没有发现它们与乳腺癌有关,但确实在肿瘤发生的通路中^[56]。

多维蛋白鉴定技术把二维液相色谱和质谱法结合在一起来鉴别多肽,它能改善不同样本中的蛋白分类准确性。Sandhu等利用该技术来研究乳腺癌细胞的蛋白谱,他们发现与癌细胞恶化程度相关的蛋白发生改变,包括细胞周期、信号转导、细胞凋亡、转录调控、细胞代谢等关键的调控蛋白水平发生了改变^[57]。

其他的组学技术

代谢组学

代谢组学是对所检测有机液体中所有小分子物质的研究,这些液体包含细胞生命过程中的终产物^[7]。目前实验室新技术能检出更多种类的细胞代谢产物。全代谢组分析和全套分子检测在有机液体或组织中产生的细胞代谢产物的应用,包括疾病诊断、药物鉴定或化学暴露。利用代谢组学技术可知道一个个体的代谢组谱和预测某些药物的毒性,主要依据患者的代谢能力判别。随着能力检测、高准确性和代谢组学水平这些新方法的应用,我们将发现生物治疗的分子反应和肿瘤细胞的不同代谢特征^[58]。

药物基因组学

药物基因组学是研究患者的基因型对不同药物反应的一种组学技术,它的目的是根据人的基因型来获得对疾病的最好治疗效果,同时避免药物副作用。药物基因组学能找出患者的基因型和药物反应之间的关系,是个体化诊疗的基础。药物基因组学的内容包括药物靶点、药物代谢、药物分子运输、疾病易感性和药物安全性^[59]。

相互作用组学

一个细胞中的分子和蛋白所发生的所有

相互作用称为相互作用组学,因此它是研究相互作用过程以及相互作用所产生结果的技术^[60]。蛋白-蛋白、蛋白-分子之间的完整网络使得信号通路、代谢通路、细胞过程等细胞生存必要的活动协调一致。对相互作用组学的深入研究能加深我们对包括肿瘤在内的疾病的了解^[61,62]。

总结与展望

个体化肿瘤用药是指通过分子检测技术,针对恰当的人、在恰当的时间、找出恰当的治疗策略,以及先要在人群水平上找出患病的易感位点,然后给出预防的策略^[63]。

癌细胞的遗传物质发生改变,如果对癌细胞是如何演化来的每一步都有所了解,那么就会发现更好的治疗策略。一个肿瘤会包含许多遗传性质的改变,但只有少数的几个因素是可以干预后让肿瘤消失的,这些就是我们需要寻找的治疗肿瘤的靶点^[7,64]。每个肿瘤患者的遗传改变不同,因此需要根据每个人具体的遗传改变来确定治疗方案,有的肿瘤患者还会接受针对不同靶点的组合治疗方案^[64]。

曲妥珠单抗是一种单克隆抗体,它可以和HER2受体结合,而HER2受体是可调节基因开关的蛋白,从而刺激细胞的增殖。作为一个经典的肿瘤个体化用药案例,若在不知乳腺癌患者HER2的情况下针对每个患者都用曲妥珠单抗(商品名为赫赛汀)和只针对HER2阳性型的患者才使用曲妥珠单抗,后者的疗效肯定要好,因为曲妥珠单抗只针对HER2阳性型的乳腺癌患者才有效^[7,64]。

利用之前研究出来的针对最合适的人群来进行治疗的手段将通过降低医疗费用、增加药物疗效、减少药物毒性来改善医疗健康服务效果。组学研究的目的是尽可能少用侵入性的方法,而能鉴别出高风险的患者,从而应用针对性的肿瘤筛查和预防方法^[7,63,64]。

高通量分子谱学方法加快了我们对肿瘤生物学的认识,研究者有望把组学方法得出的

海量数据用于癌症患者管理的新领域。组学技术的应用正在日益得到拓展,目前组学技术所获得的成果尽管只有一小部分在临床上得到应用,但不妨碍它被认为是标准的研究方法。组学技术创新性地应用在乳腺癌研究中,欲解开乳腺癌的发病机制,并利用更多的分子特征来对乳腺癌进行分子分型,从而做到精准预测疾病预后,并针对疾病发生的靶标分子找到最有疗效的药物。

(张亮 译)

参考文献

1. Gevaert O, De Moor B: Prediction of cancer outcome using DNA microarray technology: past, present and future. *Expert Opin Med Diagn* 2009, 3(2):157-165.
2. Dowsett M, Dunbier AK: Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer. *Clin Cancer Res* 2008, 14(24):8019-8026.
3. Daidone MG, Zaffaroni N, Cappelletti V: Strategies to translate preclinical information to breast cancer patient benefit. *J Natl Cancer Inst Monogr* 2011, 2011(43):55-59.
4. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT et al: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, 415(6871):530-536.
5. Lederberg J MA: 'Ome Sweet 'Omics-- A Genealogical Treasury of Words. *Scientist* 2001, 15(7):8.
6. Ellis MJ, Perou CM: The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discov* 2013, 3(1):27-34.
7. Damia G, Brogini M, Marsoni S, Venturini S, Generali D: New omics information for clinical trial utility in the primary setting. *J Natl Cancer Inst Monogr* 2011, 2011(43):128-133.
8. Institute NHGR: A brief guide to genomics. 2010. In: *Genomegov* Retrieved 2011.
9. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al: Global variation in copy number in the human genome. *Nature* 2006, 444(7118):444-454.
10. Savelyeva L, Schwab M: Amplification of oncogenes revisited: from expression profiling to clinical application. *Cancer Lett* 2001, 167(2):115-123.
11. Vucic EA, Thu KL, Robison K, Rybaczyk LA, Chari R, Alvarez CE, Lam WL: Translating cancer 'omics' to improved outcomes. *Genome Res* 2012, 22(2):188-195.

12. Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R, Brennan C, Polyak K: Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res* 2006, 66(8):4065-4078.
13. Etzell JE, Devries S, Chew K, Florendo C, Molinaro A, Ljung BM, Waldman FM: Loss of chromosome 16q in lobular carcinoma in situ. *Hum Pathol* 2001, 32(3):292-296.
14. Mastracci TL, Shadco A, Colby SM, Tuck AB, O'Malley FP, Bull SB, Lam WL, Andrulis IL: Genomic alterations in lobular neoplasia: a microarray comparative genomic hybridization signature for early neoplastic proliferation in the breast. *Genes Chromosomes Cancer* 2006, 45(11):1007-1017.
15. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA et al: Molecular portraits of human breast tumours. *Nature* 2000, 406(6797):747-752.
16. Strehl JD, Wachter DL, Fasching PA, Beckmann MW, Hartmann A: Invasive Breast Cancer: Recognition of Molecular Subtypes. *Breast Care (Basel)* 2011, 6(4):258-264.
17. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF et al: Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* 2009, 4(7):e6146.
18. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001, 98(19):10869-10874.
19. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003, 100(14):8418-8423.
20. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L et al: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004, 10(16):5367-5374.
21. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Praderwand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schutzh F et al: Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 2008, 10(4):R65.
22. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S et al: Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006, 295(21):2492-2502.
23. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey RW, Robertson JF, Nicholson RI, Ellis IO: Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* 2004, 203(2):661-671.
24. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG, Foekens JA, Martens JW: Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 2008, 68(9):3108-3114.
25. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ et al: A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002, 347(25):1999-2009.
26. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J et al: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006, 24(23):3726-3734.
27. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004, 351(27):2817-2826.
28. Perou CM, Borresen-Dale AL: Systems biology and genomics of breast cancer. *Cold Spring Harb Perspect Biol* 2011, 3(2):1.
29. Lyng MB, Laenkholm AV, Tan Q, Vach W, Gravaard KH, Knoop A, Ditzel HJ: Gene expression signatures that predict outcome of tamoxifen-treated estrogen receptor-positive, high-risk, primary breast cancer patients: a DBCG study. *PLoS One* 2013, 8(1):e54078.
30. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ: Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* 2008, 26(5):729-735.
31. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR et al: The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012, 486(7403):400-404.
32. Cambon-Thomsen A, Ducournau P, Gourraud PA, Pontille D: Biobanks for genomics and genomics for biobanks. *Comp Funct Genomics* 2003, 4(6):628-634.
33. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y et al: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486(7403):346-352.
34. Cancer-Genome-Atlas-Network: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490(7418):61-70.
35. Bernstein BE, Meissner A, Lander ES: The mammalian epigenome. *Cell* 2007, 128(4):669-681.
36. Widschwendter M, Jones PA: DNA methylation and breast carcinogenesis. *Oncogene* 2002, 21(35):5462-5482.
37. Esteller M: Cancer epigenomics: DNA methylomes

- and histone-modification maps. *Nat Rev Genet* 2007, 8(4):286-298.
38. Baylin SB, Ohm JE: Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* 2006, 6(2):107-116.
 39. Cuomo A, Moretti S, Minucci S, Bonaldi T: SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells. *Amino Acids* 2011, 41(2):387-399.
 40. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB: Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 1996, 93(18):9821-9826.
 41. Dejeux E, Ronneberg JA, Solvang H, Bukholm I, Geisler S, Aas T, Gut IG, Borresen-Dale AL, Lønning PE, Kristensen VN et al: DNA methylation profiling in doxorubicin treated primary locally advanced breast tumours identifies novel genes associated with survival and treatment response. *Mol Cancer* 2010, 9:68.
 42. Hsu NC, Huang YF, Yokoyama KK, Chu PY, Chen FM, Hou MF: Methylation of BRCA1 promoter region is associated with unfavorable prognosis in women with early-stage breast cancer. *PLoS One* 2013, 8(2):e56256.
 43. Lo PK, Sukumar S: Epigenomics and breast cancer. *Pharmacogenomics* 2008, 9(12):1879-1902.
 44. Lau TY, O'Connor DP, Brennan DJ, Duffy MJ, Pennington SR, Gallagher WM: Breast cancer proteomics: clinical perspectives. *Expert Opin Biol Ther* 2007, 7(2):209-219.
 45. Elsheikh SE, Green AR, Rakha EA, Powe DG, Ahmed RA, Collins HM, Soria D, Garibaldi JM, Paish CE, Ammar AA et al: Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res* 2009, 69(9):3802-3809.
 46. Culhane AC, Howlin J: Molecular profiling of breast cancer: transcriptomic studies and beyond. *Cell Mol Life Sci* 2007, 64(24):3185-3200.
 47. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009, 10(1):57-63.
 48. Huber-Keener KJ, Liu X, Wang Z, Wang Y, Freeman W, Wu S, Planas-Silva MD, Ren X, Cheng Y, Zhang Y et al: Differential gene expression in tamoxifen-resistant breast cancer cells revealed by a new analytical model of RNA-Seq data. *PLoS One* 2012, 7(7):e41333.
 49. Anderson NL, Anderson NG: Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 1998, 19(11):1853-1861.
 50. Goncalves A, Bertucci F: Clinical application of proteomics in breast cancer: state of the art and perspectives. *Med Princ Pract* 2011, 20(1):4-18.
 51. Stein RC, Zvelebil MJ: The application of 2D gel-based proteomics methods to the study of breast cancer. *J Mammary Gland Biol Neoplasia* 2002, 7(4):385-393.
 52. Rowell C, Carpenter DM, Lamartiniere CA: Chemoprevention of breast cancer, proteomic discovery of genistein action in the rat mammary gland. *J Nutr* 2005, 135(12 Suppl):2953S-2959S.
 53. Hu Y, Zhang S, Yu J, Liu J, Zheng S: SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast* 2005, 14(4):250-255.
 54. Gharbi S, Gaffney P, Yang A, Zvelebil MJ, Cramer R, Waterfield MD, Timms JF: Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Proteomics* 2002, 1(2):91-98.
 55. Marouga R, David S, Hawkins E: The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal Bioanal Chem* 2005, 382(3):669-678.
 56. Davalieva K, Kiprijanovska S, Broussard C, Petrussevska G, Efremov GD: Proteomic analysis of infiltrating ductal carcinoma tissues by coupled 2-D DIGE/MS/MS analysis. *Mol Biol (Mosk)* 2012, 46(3):469-480.
 57. Sandhu C, Connor M, Kislinger T, Slingerland J, Emili A: Global protein shotgun expression profiling of proliferating mcf-7 breast cancer cells. *J Proteome Res* 2005, 4(3):674-689.
 58. Yang C, Richardson AD, Smith JW, Osterman A: Comparative metabolomics of breast cancer. *Pac Symp Biocomput* 2007:181-192.
 59. Ma Q, Lu AY: Pharmacogenetics, pharmacogenomics, and individualized medicine. *Pharmacol Rev* 2011, 63(2):437-459.
 60. Kiemer L, Cesareni G: Comparative interactomics: comparing apples and pears? *Trends Biotechnol* 2007, 25(10):448-454.
 61. Rak J: Extracellular vesicles - biomarkers and effectors of the cellular interactome in cancer. *Front Pharmacol* 2013, 4:21.
 62. Lo SH: Reverse interactomics: from peptides to proteins and to functions. *ACS Chem Biol* 2007, 2(2):93-95.
 63. Gonzalez-Angulo AM, Hennessy BT, Mills GB: Future of personalized medicine in oncology: a systems biology approach. *J Clin Oncol* 2010, 28(16):2777-2783.
 64. Ocana A, Pandiella A: Personalized therapies in the cancer "omics" era. *Mol Cancer* 2010, 9:202.
 65. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*. 2006;7:96.

第2章

遗传性乳腺癌的组学分析

Catherine A. Moroski-Erkul, Burak Yilmaz, Esra Gunduz, Mehmet Gunduz

摘要

乳腺癌是世界范围内女性癌症死亡的首要原因。尽管在过去的10年内,我们对于这种疾病的理解取得了进展,但有的乳腺癌有效治疗手段仍然不足,对于某些种类的难治性乳腺癌尤其如此。遗传性或家族性的乳腺癌治疗是一个严峻的挑战,因为目前我们只确定了少量高外显率的易感基因,即BRCA1和BRCA2。现在猜测大多数的遗传性或家族性乳腺癌是由几个中等和(或)低外显率基因的不同组合协同作用导致的。生物学在研究方法和概念框架上的最新发展已经彻底改变了癌症的研究方式。这种系统研究方法强调生物系统的全面理解,被统称为“组学”。经过10年来的组学研究探索,已有许多新的治疗靶标和生物标志物被鉴定,使得我们能够更精确、早期地诊断和治疗一系列乳腺癌。在此,我们对数个组学领域在遗传性和家族性乳腺癌研究中取得的成就做一个综述,包括基因组学、转录组学、蛋白质组学和代谢组学。

关键词

遗传性乳腺癌 基因组学 转录组学 蛋白质组学 代谢组学

引言

随着人类基因组测序的完成,生物学的研究系统经历了重大转变。许多研究者将他们的工作向更全局化的方向靠拢。由高通量大数据组成的对新领域的研究方法已经发展起来,其目的是了解我们周围及生活在其中的错综复杂的生物系统。自动化高通量技术的发展能够执行复杂的实验程序,获取详细的成像或其他类型的数据,与传统的劳动密集型的方法相比,这种新技术消耗更少的时间,引领我们进入了一个数据繁多难于掌握的时代(表

2.1示出了可公开获取的微阵列芯片数据的表单)。伴随着这些实验室新技术的发展,计算机科学也取得了长足的进步,使没有或缺少软件工程背景的研究者们能够从如山的数据中筛选得到有意义的信息^[1]。这些全新而又令人兴奋的研究领域逐渐被统称为“组学”^[2]。

各种组学领域发挥它们巨大威力的基础在于现在可以从一个生物样品相对快速和容易地提取到大量而详细的信息。如图2.1所示,数个组学领域(如基因组学、蛋白质组学和代谢组学)的数据整合可为待研究的系统提供更丰富、更全面的视野。这将引领我们在疾病预防(基因组测序)和检测(生物标志物)领域

表 2.1 网络上公开可获得的微阵列芯片数据库

数据库	管理者	参考文献
基因表达数据库 (Gene Expression Omnibus ,GEO)	美国国家生物技术信息中心 (National Center for Biotechnology Information,NCBI)	[3]
Riken 表达阵列数据库 (Riken Expression Array Database,READ)	Riken	[4]
ArrayTrack	美国食品与药物管理局 (Food and Drug Administration,FDA)	[5]
ArrayExpress	欧洲分子生物学实验室-欧洲生物信息研究所 (European Molecular Biology Laboratory-European Bioinformatics Institute,EMBL-EBI)	[6]
BioGPS	诺华制药基因组学研究所 (Genomics Institute of the Novartis Research Institute)	[7]
Microarray Retriever(MaRe)	莱顿大学医学中心 (Leiden University Medical Center,LUMC)	[8]

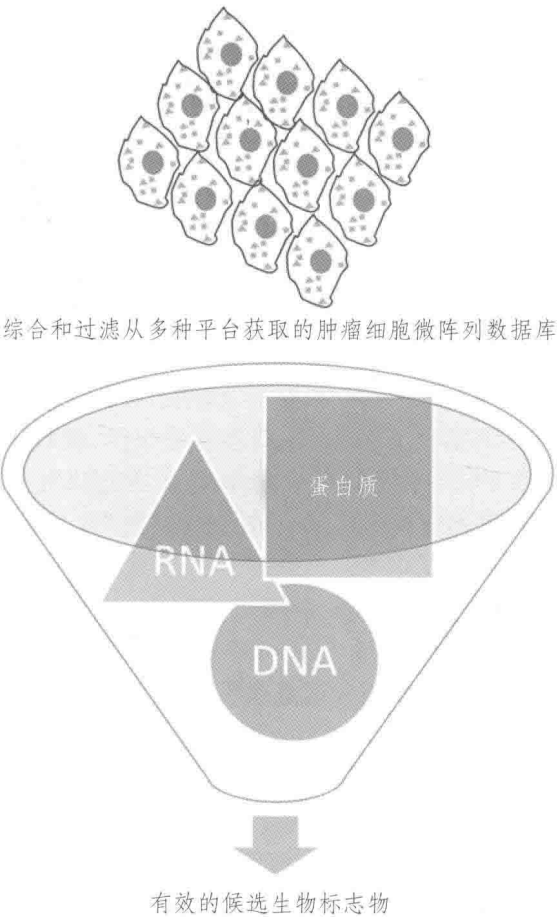


图 2.1 可视化显示用于识别更可靠生物标志物的高通量数据的整合。

取得进步,为患者提供更好、更个体化的治疗(药物基因组学分析)以及对疾病的预后进行更全面和准确的分析(生物分子谱)。

尽管在 21 世纪的第一个 10 年内,新研究方法的爆炸式发展使我们对生物系统的探索更加详细和全面,但组学的检测和分析方法仍然需要改进 [1, 9, 10]。最终,在医药应用方面,组学革命的最终目的是为了更好地了解病理进程,对疾病进行预防、诊断、干预和治疗。本章旨在介绍目前在癌症研究领域应用的主要组学方法以及这些方法如何提升了我们当前对遗传性乳腺癌的认识。

一个日益增长的问题

癌症为卫生保健系统带来了沉重的经济负担,对于癌症的早期诊断和更有效的治疗不仅是医疗上的当务之急,同时也是一个经济问题 [11-13]。目前的治疗手段虽然不断进步并且更具有靶向性,但仍然有害于健康的细胞和组织。对健康细胞的这种伤害引发令人不快的副作用,并有可能导致继发性癌症的产生 [14-16]。

癌症在高收入国家是导致死亡的第一原因,而在低收入和中等收入国家的致死原因中也排名第二。对于女性来说,乳腺癌是致死率最高的癌症。2008年,乳腺癌占到了所有女性新发癌症的23%,同时导致了14%的癌症相关死亡。50%的乳腺癌发生在发展中国家,而60%的乳腺癌相关死亡也发生在这些国家,这意味着我们更迫切地需要更好的早期诊断方式以及更有针对性的、更经济划算的治疗方法^[17, 18]。乳腺癌对患者及其家庭造成了身体上和精神上的双重毁灭性打击,在某些缺乏完善医疗保障体系的国家,例如美国,乳腺癌还会使患者的家庭陷入沉重的经济负担。尽管在早期诊断和新方法认定上我们已经迈出了一大步,但仍然有大量的工作要做,需要我们集中精力提高对影响乳腺癌发生发展的基因、生化和环境因素的认识。

异质性

认识任何癌症的主要挑战之一是癌症本身固有的异质性^[18]。事实上,当“癌症”这个词被用作一般的描述时,就已经给非专业人士带来了大量的困惑和不解,他们难于理解不同类型的癌症之间与某种癌症之中都存在着极大的异质性。我们能够更好地领会癌症的这种特性,这归功于组学领域认识的突破。通常乳腺癌是说明异质性这个术语的一个特别好的例子,但同时乳腺癌又不能完全代表肿瘤异质性^[19]。乳腺癌常常被分为遗传性或家族性乳腺癌和散发性乳腺癌,基于各种病理、遗传和生物分子特性也被分为各种亚型。过去10年来的研究,越来越清楚每个类型的肿瘤既与其他肿瘤有许多相似的特性,却也有自己独一无二的特性。因此,当我们在寻找新的靶标时,既要集中不同亚型的相似特性,同时还必须意识到每个肿瘤的独自特性,这可能导致其对多种有效治疗方法产生抵抗。组学革命的目标就是让常规的、廉价的个体肿瘤分子生物学分析引领真正的个体化治疗模式。

致癌性转化

致癌性转化是一个复杂的、多步骤的过程,在不同癌症之间,甚至相同癌症的不同亚型之间有着广泛的差异。然而,每个癌症病例可能是不同的,而共同的是原癌基因激活和抑癌基因突变的特性,以及涉及大量不同信号通路的其他基因,累积产生一个癌细胞的表型^[20]。使用转录组学、代谢组学和蛋白质组学等方法对高风险患者的生物样本(如血液、血清或尿液)进行长时间监测,可能有助于我们了解癌症转化过程中出现的早期变化。最近的一些研究已经利用乳腺癌细胞系和(或)患者来源的样本来检测癌症转移转化过程的全局变化,以期寻找更加特异和敏感的分子标志物。人们希望的是早期诊断和治疗可阻止一个癌症的转移^[21]。也许有一天,随着检测技术的进步,我们对于癌症的认识甚至能使我们检测到致癌性转化被阻断的一个阶段。

在本章的内容中,我们将描述4个组学科以及它们在研究遗传性乳腺癌过程的贡献:基因组学、转录组学、蛋白质组学和代谢组学。它们的排列顺序代表了分子信息产生的流程:首先是基因组学,生命中相对固定的分子密码;然后是转录组学,遗传密码翻译成“有用”成分的第一步;接下来是蛋白质组学,细胞活性的中心;最后以代谢组学结束,经过上述复杂的细胞进程后形成下游“最终产物”。

癌症基因组学

基因组学现今已是众所周知的一门学科,始于20世纪70年代的DNA克隆的发明及随后的人类基因组测序^[22]。“经典”基因组学主要关注基因组测序,包含特定基因组内的所有基因的鉴定以及已知基因结构和基因与环境之间复杂的相互作用。现如今在这个领域已有许多亚学科出现,比如结构和功能基因组学、表观基因组学、病理和毒理基因组学。而

这些都是为了更好地了解基因序列和生物学进程或结果之间的关系。

我们目前正处于所谓的“后基因组”时代。现已检测出会增加人们患上各种癌症风险的基因突变。在高风险乳腺癌家系中,实施基因筛查以便开展预防措施是可行的,比如改变生活方式、尽早开始乳腺X线检查或接受预防性乳腺切除术^[23-26]。

遗传性乳腺癌基因组学

我们已经在将乳腺癌分成有意义的亚型方面进行了许多尝试,希望在确定乳腺癌诊断、最佳诊疗方式及预后上提供帮助。随着我们对乳腺癌异质性认识的不断提升,对乳腺癌分类系统也在不断发展。根据临床和治疗特征乳腺癌可被分成4个主要类型。第一类是管腔样乳腺癌,往往被细分为管腔样A型和管腔样B型,是最复杂的乳腺癌亚型,目前有数个基因组检测能够预测内分泌治疗的疗效。第二类是HER2阳性型(或HER2过表达型),该类型对于靶向HER2单克隆抗体反应良好。第三类被称为类正常型。最后一类被称为三阴性(或基底样),因为该类型乳腺癌缺少雌激素受体(ER)、孕激素受体(PR)和HER2表达^[27]。它们在BRCA1突变的种系或非裔人群中高发,发病率占到所有乳腺癌的15%^[28]。

2009年,Parker等报道了一种通过定量RT-PCR和基因芯片技术来预测乳腺癌亚型的方法,后来被称为微阵列50预测分析(PAM50),常用来预测最佳的个体化治疗方案^[29]。2011年,Ebbert等报道称,PAM50系统通常是准确的,该测定用来分类的多变量分析(MVA)很少出现错误。然而,在肿瘤与现有参数符合度不太好的情况下,该系统可能会导致不准确的结论^[30]。2012年,IMPAKT工作组比较了PAM50与ER、HER2和Ki67三联免疫组织化学两种方法的有效性,结果发现用前者来决定全身性治疗方案还是“不够健全”,他们建议通过联合ER和HER2的IHC

应用来替代。

除了PAM50,还有针对BRCA1、BRCA2和CYP2D6的生殖细胞基因检测和乳腺癌指数(BCI)评估。Oncotype DX和MammaPrint检测已在美国和欧洲应用于临床决策^[31]。最近通过大样本基因组(以及其他组学)谱分析已经使得更广泛和有意义的亚型分类成为可能^[18, 32]。这种分型对于识别并应用合理的治疗组合至关重要。

首个与遗传性乳腺癌相关基因也可能最广为人知的是BRCA1和BRCA2;这两个乳腺癌易感基因,具有常染色体显性遗传的特性并具有较高的外显率^[33, 34]。这两个基因引发了约30%的家族性乳腺癌病例^[35]。这些基因中生殖细胞基因突变会导致所谓的遗传性乳腺癌和卵巢癌(HBOC)综合征,使得携带者终身有50%~80%患乳腺癌和30%~50%患卵巢癌的风险^[36]。有趣的是,尽管BRCA基因主要与乳腺癌关联,但它们与卵巢癌关联度更高,在诊断为卵巢癌的女性中约有12%的总突变率^[37]。这些基因的确定让人兴奋,人们希望能够发现更多高外显率基因。然而,事实却并非如此,这也是为何寄希望于通过组学方法来进一步了解乳腺癌的一个原因。

虽然两个BRCA基因作用于相同的DNA修复通路,即同源重组(HR)修复(图2.2),但由BRCA1和BRCA2基因突变所造成的肿瘤却有着显著的差异。BRCA2肿瘤有类似散发病例的特点,另一方面,BRCA1肿瘤更具有侵袭性,难以治疗,并且通常是ER阴性^[36, 38]。在最近一篇综述中,Roy等提出几种理论来解释,为什么参与相同DNA修复通路的两个基因所产生的肿瘤在遗传性和临床特性上显著不同。这可能是由于其他基因突变或多态性是BRCA1连锁遗传造成的,虽然他们指出目前没有可用来支持这一理论的证据;他们提出另一种可能性是BRCA1的转录共激活或共抑制能够修饰ER的表达,而BRCA2不具有这种功能。为了证明这种理论,需要对ER阴性BRCA1和ER阴性散发性肿瘤的表达谱进