

整合分析多组学数据筛选疾病靶点的精准医学策略

谢兵兵^{1,2}, 杨亚东¹, 丁楠^{1,2}, 方向东¹

1. 中国科学院北京基因组研究所, 中国科学院基因组科学与信息重点实验室, 北京 100101;

2. 中国科学院大学, 北京 100049

摘要: 随着高通量测序技术的不断发展与完善, 对于不同层次和类型的生物组学数据的获取及分析方法也日趋成熟与完善。基于单组学数据的疾病研究已经发现了诸多新的疾病相关因子, 而整合多组学数据研究疾病靶点的工作方兴未艾。生命体是一个复杂的调控系统, 疾病的发生与发展涉及基因变异、表观遗传改变、基因表达异常以及信号通路紊乱等诸多层次的复杂调控机制, 利用单一组学数据分析致病因子的局限性愈发显著。通过对多种层次和来源的高通量组学数据的整合分析, 系统地研究临床发病机理、确定最佳疾病靶点已经成为精准医学研究的重要发展方向, 将为疾病研究提供新的思路, 并对疾病的早期诊断、个体化治疗和指导用药等提供新的理论依据。本文详细介绍了基因组、转录组和表观组等系统组学研究在疾病靶点筛选方面出现的新技术手段和研究进展, 并对它们之间的整合分析新策略和优势进行了讨论。

关键词: 基因组; 转录组; 表观遗传组; 整合分析; 疾病靶点

Identification of disease targets for precision medicine by integrative analysis of multi-omics data

Bingbing Xie^{1,2}, Yadong Yang¹, Nan Ding^{1,2}, Xiangdong Fang¹

1. CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: With the development and improvement of high-throughput sequencing technologies, the acquisition and processing approaches of various biological omics data on different levels are becoming more mature. Despite several new disease-associated factors have been discovered based on single omics data analysis, identification of disease targets by integrative analysis of multi-omics data is still growing. Since life is a complex regulatory system in which the regulation of gene mutations, epigenetic alterations, abnormal gene expression as well as anomalous variations in signal pathway are related with the occurrence and development of diseases, it is obvious that finding therapeutic factors using single omics data analysis has its limitation. Systematical studies of clinical and pathological mechanisms and identification of optimal therapeutic targets through integrative analysis of multi-omics data from different levels and resources have become an important research direction of precision medicine, which would pro-

收稿日期: 2015-02-02; 修回日期: 2015-03-31

基金项目: 中国科学院干细胞与再生医学研究战略性科技先导专项子课题(编号: XDA01040405), 国家自然科学基金面上项目(编号: 31471115)和青年项目(编号: 31401160)资助

作者简介: 谢兵兵, 在读硕士研究生, 专业方向: 基因组学数据挖掘。Tel: 010-84097538; E-mail: xiebb@big.ac.cn

通讯作者: 方向东, 博士, 研究员, 研究方向: 干细胞与重要疾病的组学与转化医学。E-mail: fangxd@big.ac.cn

DOI: 10.16288/j.ycz.15-061

网络出版时间: 2015-5-22 17:16:19

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150522.1716.002.html>

vide innovative perspectives on disease study and new theoretical basis for early diagnosis, personalized treatment and medicine guide. In this review, we introduce new technologies and research progresses in screening therapeutic targets using systematic omics such as genomics, transcriptomics and epigenomics, and also discuss new strategies and advantages of integrative analysis among them.

Keywords: genomic; transcriptomic; epigenetic; integrative analysis; disease targets

十几年来,以第二代测序技术为代表的高通量测序技术取得了长足的发展,个人基因组测序的费用已经从 2001 年的 1 亿美元降低到如今的 1000 美元。得益于此,通过焦磷酸测序、聚合酶测序、连接酶测序等技术手段获得的基因组^[1]、转录组^[2]及表观遗传组^[3]为代表的高通量组学数据应运而生。代表性的国际合作项目有国际人类基因组单体型图计划(HapMap^[4])、DNA 元件百科全书计划(ENCODE^[5,6])、千人基因组计划(1000 Genomes Project^[7])等。生命科学领域的数据呈现出爆炸式增长,人类疾病研究正在由传统的假说导向型向数据驱动型转变,基因组学的研究正在改变人类疾病研究以及临床治疗的进程。2015 年 1 月,美国总统奥巴马向国会提议斥资 2.15 亿美元,以开展美国 100 万人基因组研究的精准医疗(Precision medicine)计划,致力于治愈癌症和糖尿病等疾病,让所有人获得需要保障自己和家人健康的个性化信息。

通过高通量组学数据分析,研究人员已经挖掘到了肿瘤抑制基因之外的新的候选疾病靶点——*SETD2*^[8]、*VGLL4*^[9]等,这些发现都将为疾病的临床检测与治疗提供重要的参考信息。随着利用高通量组学测序技术对疾病研究的不断深入,研究人员还发现了许多新的疾病相关调控因子,如 lncRNA^[10]、m6A 甲基化^[11]和 circRNAs^[12]等。这些新因子的检测技术也迅速发展起来^[13,14],极大地丰富了组学数据资源。

疾病的发生与发展涉及到基因组、转录组、表观组、蛋白组及代谢组等多个不同层次的病理过程。单组学数据的分析往往只能体现出疾病样本其中一个层面的变化,在筛选疾病靶点方面具有很大的局限性。通过对多层次疾病组学数据的综合分析,将有助于人们对疾病形成更加系统全面的认识,为药物研发、临床诊断及个性化治疗提供更多有用的参考信息。但多组学数据的整合分析研究还未成熟,亟待开发出通用的数据整合和分析方法,以充分利用已产生的多组学数据。文章总结了不同层次单一

组学数据、两种及以上组学数据整合策略在疾病靶点筛选中的应用。

1 疾病相关组学研究概况

1.1 基因组

基因组是指生命体内包含的所有 DNA 分子,主要研究内容包括生殖细胞和体细胞的点突变、结构变异、重排突变等。高通量测序技术可以使科研人员更细致地看到这些变化。在群体水平,全基因组关联分析(Genome-wide associated studies, GWAS)通过病例-对照样本的基因组学数据对比分析获得疾病相关的靶点。美国国家人类基因组研究所 2013 年发表的文章^[15]中, GWAS 相关文章已达到 1751 篇,发现与疾病相关的 SNP 达到 11 912 个。在个体水平,已有全基因组测序用于诊断和治疗的报道^[16]。除全基因组测序之外,外显子组测序技术通过特异捕获外显子区域 DNA 序列进行测序,成本明显低于全基因组测序的同时也能有效地检测疾病相关基因^[17]。例如,研究人员通过对 3 例乳头状颅咽管瘤病人外显子组进行测序分析,结果发现 *BRAF* 在该病的临床诊断与治疗方面具有潜在的重要作用^[18]。除了单纯的检测 DNA 序列的变化,研究者还利用 MutSig^[19]和 InVEx^[20]等算法分析突变属于驱动突变(Driver mutation)还是乘客突变(Passenger mutation)。随着高通量测序和分析技术的不断发展,人们将会发现更多的基因组变异,对变异与疾病之间的关系也将会有更全面、深入的理解。

1.2 转录组

转录组是特定组织或细胞在某一发育阶段或功能状态下转录生成的所有 RNA 的集合,包括编码蛋白的 mRNA 和各种非编码 RNA(如 microRNA、lncRNA、circRNAs 等)。早期的芯片技术只能检测已知基因的表达水平,现在以 RNA-seq 为代表的转录组学研究方法已经可以检测新的转录本^[21]、可变

剪接^[22]、基因融合^[23]和 SNP^[24]等。Alarcon 等^[25]通过对 47 例大脑皮质样本 mRNA 转录组检测,发现脂蛋白受体相关蛋白 6 亚型(LRP6-3)在阿尔兹海默症(Alzheimer's disease, AD)患者中转录水平显著升高,并在小鼠实验中得到验证,这些结果表明新的可变剪切形式 LRP6-3 在 AD 中具有重要功能。除了传统的 mRNA 研究外, lncRNA、circRNA 等非编码 RNA 也越发受到研究者的关注。RNA-seq 和 ChIRP-seq^[13]技术已用于 lncRNA 的检测并发现 lncRNA 在前列腺癌^[26]、心脏疾病^[27]、肺癌^[28]等疾病中具有重要作用。最近一篇研究通过对 7256 组 RNA-seq 数据进行分析,发现人类转录产物的 68% (58 648)为 lncRNA,约 7% (3900)的 lncRNA 和疾病相关的 SNP 重合^[29]。circRNA 是一种环形的 RNA,保守地存在于哺乳动物转录组中^[30],与动脉粥样硬化^[31]、阿尔兹海默症^[32]等疾病有关。随着更多的基因组“暗物质”被发现,人们对疾病的基因表达调控也将会更加全面的认识。

1.3 表观组

表观遗传是指 DNA 序列不发生变化,但基因表达却发生了可遗传性的改变,并且这种改变在发育和细胞增殖过程中能稳定地传递。表观遗传组主要包括 DNA 甲基化^[33]、组蛋白修饰^[34]和基因组印记^[35]等。表观基因组学路线图计划^[36] (Roadmap Epigenomics Project)整合分析了 111 个组织/细胞的人类表观基因组图谱,揭示了表观信息在基因调控、细胞分化和人类疾病的中心作用。6-甲基腺嘌呤(N6-mA)是一种可逆的 mRNA 甲基化修饰,常见于高等生物,其作用于 mRNA 外显子编码区和 3'非编码区,受甲基转移酶复合物(METTL3、METTL14 和 WTAP)和去甲基化酶(FTO和ALKBH5)调控^[37~39]。N6-mA 可以调控 mRNA 剪切加工和脂肪细胞分化,为 RNA 甲基化开展肥胖症的研究提供了理论依据^[37]。组蛋白修饰主要包括甲基化、乙酰化、磷酸化、泛素化、ADP 核糖基化等。甲基化和乙酰化修饰较为稳定,更适合作为表观遗传信息。常见的组蛋白甲基化标志物包括 H3K4me2、H3K9me2、H3K9me3 和 H3K27me3 等,对癌症的发生有直接的相关性^[40~42]。组蛋白乙酰化则通过控制基因表达影响多种疾病的发生,像运动神经元退行性疾病^[43]、阿尔兹海默症等^[44]。表观遗传学的发展弥补了许多孟德尔遗传规律

无法解释的现象,完善了经典的遗传学理论。

2 整合多组学数据分析疾病靶点的研究进展

2.1 整合多组学数据分析疾病靶点的优势

单组学数据分析是传统的疾病靶点研究策略,帮助科研及临床研究人员确定了许多候选疾病靶点。随着研究的深入,其缺点也愈发的显著,以转录组基因表达为例:基因的表达和蛋白质的表达,生物活性和代谢组之间的相关性较差^[45]。另外,细胞中的许多调控是发生在基因表达之后,如转录后调控、翻译、翻译后调控以及各种形式的异构体和反馈调控^[46]。例如, Kuile 等^[47]发现糖酵解过程和代谢组、蛋白组和基因组都相关,因此功能基因组学不能仅停止在 mRNA 水平。疾病产生是一个持续的多过程连锁事件,一个层面数据的改变很难解释疾病的整个发生过程。当前全基因组检测技术已可以检测到基因组、转录组和表观组等多个层面的数据。通过它们之间线性或非线性关系等^[48,49]可以将三者联系到一起,进而使多组学数据的整合分析成为可能。关联 3 个层面组学数据的整合分析可以有效去除单个层面的随机事件,并观察到真正的候选疾病候选因子在各个层面的不同变化,从而探究这些候选疾病因子的作用机制,找到最有效的治疗措施。

2.2 整合多组学数据分析疾病靶点的思路

多组学数据整合分析的首要步骤是对不同来源的数据进行标准化处理,然后通过比较建立不同组学数据之间的关联性和差异性,进而根据这种内在联系在不同层次对候选疾病因子进行筛选过滤,最终目标是对疾病的发生过程建立定量模型,通过模拟仿真等手段预测候选疾病因子的作用,通过改变因子的水平预测治疗效果等^[49]。多组学数据分析流程见图 1。

2.3 整合多组学数据分析疾病靶点的策略

2.3.1 基因组与转录组

数量性状位点(QTLs)是基因组的一个基本统计指标,用于标识基因组性状关联区域的基因数^[50]。全基因组的 QTLs 分析能找到很多与疾病相关的 QTLs,但它通常不能明确地找出疾病基因。可能是 QTLs 跨越的染色体片段含有很多不相关的基因,所

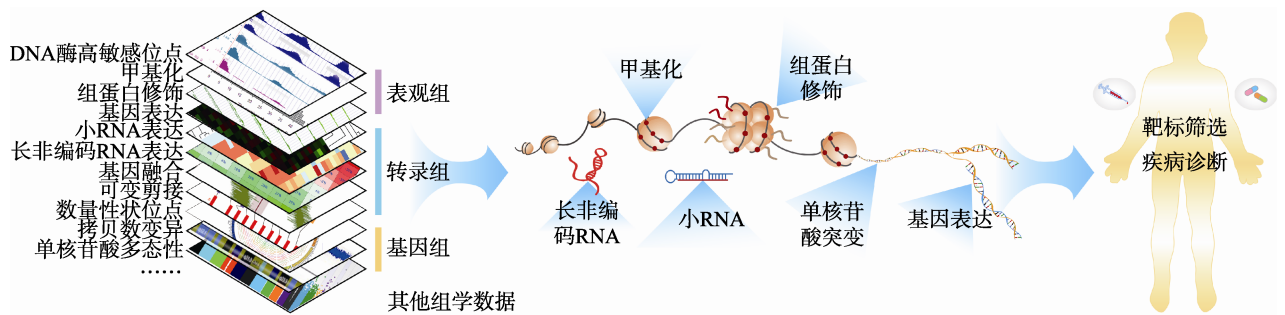


图 1 多组学数据分析流程

现有的组学数据主要包括：基因组(数量性状位点、拷贝数变异和单核苷酸多态性)，转录组(基因表达、小 RNA 表达、长非编码 RNA 表达、基因融合和可变剪接)和表观组(DNA 酶高敏感位点、甲基化和组蛋白修饰)。

以 QTLs 分析面临的主要困难是尽可能地滤除没有参与疾病发病机制的基因^[51]。转录组测序得到的差异基因表达谱广泛用于疾病遗传基础研究。但是单独使用转录组测序会遇到很大困难，因为许多的差异表达基因源于人群中的个体差异，这些基因和疾病并没有直接联系。通过整合基因组数据和转录组数据为复杂疾病研究提供了一个新策略。这种整合策略基于筛选出来的基因是差异表达的，并且映射到了疾病相关的 QTLs。那么这些基因很有可能包含在疾病的病理生理学过程中^[52]。这种整合分析的方法已应用于前列腺癌、高血压、乳腺癌等^[52-54]疾病的研究中。Schadt 等^[55]在基因组和转录组数据整合过程中，进一步探究了突变位点、基因和表型之间的作用方式，提出了 5 种可能的相互调控模型，并采用基于似然方法的因果选择模型检测三者之间最可靠的调控模型，找到了 3 个新的肥胖症的易感基因(*Zfp90*、*C3ar1* 和 *Tgfb2*)。根据乳腺癌转录组的研究，研究人员将乳腺癌分为 5 类：Normal-like, Luminal A, Luminal B, ERBB2+和 Basal-like^[56,57]，并被广为使用。而 Curtis 等^[58]通过整合基因组和转录组数据对乳腺癌重新聚类，得到了 10 个子类，并在将近 1000 个临床样本中得到了证实，解决了之前乳腺癌亚型分类的异质性问题。

2.3.2 基因组与表观组

基因组水平的改变在于 DNA 序列的变化，而表观组水平的改变不涉及 DNA 序列的变化，这两个数据的整合基础在于两者互为因果的关系。如甲基化相关基因序列的变化会影响酶的功能^[59]，而酶功能的改变会进一步改变表观的修饰状态。反之，基因

组稳定性相关基因的表观调控变化会改变基因组序列^[60]。通过基因组和表观组数据的整合，人们可以对两者的作用机制有更深入的理解。Robert 等^[61]通过卵巢癌转录组数据寻找基因组数据和表观组数据的关系：利用 MOMA-ROMA 实验和 TCGA 的数据分别筛选与基因表达改变一致的 CNV 和 DNA 甲基化，综合 3 种数据发现 *CDCA8*、*ATAD2* 和 *BOP1* 基因的三组数据变化均一致；通过整合两组学分析筛选到的 3 个基因都是已知的致癌基因，说明该分析方法的可行性并为三者之间机制验证提供理论基础，也为其他疾病数据的整合分析提供了技术支持。华盛顿大学的研究者通过基因组位置信息分析人类 349 种细胞和组织样本的全基因组 DNase I 图谱和 GWAS 中的 SNPs 数据，建立了一套疾病与生物学性状之间关联性的调控网络模型，发现约 93% 与疾病和性状相关的 SNPs 位于非编码序列内，并且集中在 DNase I 高敏感位点区域(DHSs)^[62]。Yi 等^[63]通过已发表的数据整理出结肠癌中发生突变的基因，进而对这些基因进行了甲基化检测，发现了一系列和结肠癌有关的基因和通路；实验和临床证据表明这些基因/通路与结肠癌的发生及其术后生存期有关。

2.3.3 转录组与表观组

基因表达是一个复杂的多因素调控过程，除了受到自身启动子、增强子和沉默子等顺式调控元件的调控作用，也受到像转录因子和表观遗传修饰等的反式调控作用。但顺式作用元件本身仅能提供作用位点而不能编码蛋白质，真正对基因表达调控起作用的是外部的反式调控元件^[64]。了解表观遗传调控在基因表达过程中的角色，对发育、再生和疾病

发生等生物学过程的完整描述非常关键^[65]。通过转录组测序数据可以鉴定差异表达基因和相应的调控通路^[66]。通过表观遗传组分析可以精确定位转录因子和其它调控蛋白的基因组区域^[67]。基因表达数据和表观组数据的整合有助于人们了解信号通路活性、转录因子和基因组靶标三者之间的作用关系。研究人员通过整合组蛋白的表观谱和全基因组的转录数据,确定了成人肺泡上皮细胞分化过程中分子信号传导的激活和抑制事件,以及与之相关的表观遗传变化,获得了一个新的信号通路——视黄醇 X 受体信号通路^[68]。DNA 甲基化也是重要的表观遗传修饰,甲基化的变化通常与疾病基因的表达异常相关^[69]。MicroRNA 是基因的调控因子,其在大肠癌中呈现出全局下调状态,但是其作用机制还不清楚。由于 CpG 岛的高甲基化能使 microRNA 发生沉默^[70], Balaguer 等^[71]利用转录组和表观组数据探究甲基化、microRNA 和大肠癌之间的关系,结果发现 miR-137CpG 岛的甲基化在大肠癌细胞系样本中是稳定的,而正常样本中没有观察到;并且 miR-137 的表达与其 CpG 岛甲基化呈现明显的负相关关系,后续的 RT-PCR 实验找到了 miR-137 的靶点 LSD1。这种整合分析的方法发现了 miRNA 可能的作用机制,并表明 miR-137 是大肠癌中的一个潜在癌症抑制子。

2.3.4 基因组、转录组与表观组

3 个组学数据涉及到多个层面的生物学过程,其整合还需后续大量的研究。Zhang 等^[72]对基因组、转录组和表观组的 4 个子类数据即基因表达、DNA 甲基化、miRNA 表达和拷贝数改变(Copy number alteration, CNA)进行了整合分析,通过 COX 比例风险模型找到 512 个在卵巢癌样本不同组学数据中起作用的 4526 个风险因子。然后采用贝叶斯信息准则将风险因子降到了 37 维,并用无监督的超级 K 聚类方法找到了卵巢癌的 7 个不同的亚型。在整合分析的过程中还发现主要的分类因子来自于基因组和转录组的数据,而表观组的数据贡献不大,说明表观组数据在卵巢癌分型中的作用不明显,为卵巢癌的针对性治疗和研究提供了很大帮助。

3 结语与展望

随着各种基于高通量测序平台的组学研究技术

的发展与完善,不同层次的调控因子及其在全基因组的分布状态与疾病的发生、发展、预后之间的关联性陆续被发掘和解析。尤其是 lncRNA、m6A 甲基化及基因融合等新的致病机制和疾病靶点研究的逐步深入,也将为疾病的分子机制探索、临床早期诊断、预后判断和用药指导等个性化治疗提供更加全面、精准的崭新思路和策略。同时,结合基于高通量质谱和 iTRAQ 等技术的蛋白质组以及代谢组的分析工作,能够使疾病的靶点研究更加具有说服力^[73,74]。

生命科学研究领域的数据分析是日益严峻的挑战,目前全世界每年的生物数据产生总量已经高达 EB 级,生命科学在某种程度上也成为大数据科学重要的组成部分。同时,高通量测序平台产生的生物大数据和海量的医院病历文档、生化检验、病理学和影像学等数字化信息的有效整合又使得人们面临这一挑战的形势更加急迫和严峻。疾病的复杂性和个体化医疗等要求有生物学意义上更加精准的数据整合方法。首先,需要高效能计算机和云计算等信息领域关键技术的飞跃发展以应对日益显著的数据存储和运算需求;其次,期待数学和物理等学科领域建立更加系统、有效的统计学分析方法和数理模型。值得欣慰的是,以 TCGA(The cancer genome Atlas)为代表的国际大型癌症数据库,已经收集了较为完备的多组学数据供研究人员使用,其接收和存储多组学数据的成功经验为人们提供了有益的参考。全面开展针对生物医学大数据的系统研究与挖掘,成为生物医药科学技术发展的必将趋势。我国正在建立国家生物医学数据中心,希望提高生物医学大数据管理、分析、服务和利用水平,提升我国在生命科学领域数据管理与利用的国际竞争力,促使我国尽快构建实时、便捷、全方位的生物医药领域研究与应用系统,并以此带动我国生物、医疗、健康、医药、公共卫生、能源、环境等重要领域的发展。在此基础之上,整合分析多组学高通量测序数据筛选疾病靶点,在科学理论和技术操作的层面都将会更加系统深入并全面有效地展开。

参考文献

- [1] Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell*, 2013, 155(1):

- 27–38. [DOI]
- [2] Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, Viale A, Wright C, Schweitzer PA, Gao Y, Kim D, Boland J, Hicks B, Kim R, Chhangawala S, Jafari N, Raghavachari N, Gandara J, Garcia-Reyero N, Hendrickson C, Roberson D, Rosenfeld JA, Smith T, Underwood JG, Wang M, Zumbo P, Baldwin DA, Grills GS, Mason CE. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol*, 2014, 32(9): 915–925. [DOI]
- [3] Rivera CM, Ren B. Mapping human epigenomes. *Cell*, 2013, 155(1): 39–55. [DOI]
- [4] The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature*, 2010, 467(7311): 52–58. [DOI]
- [5] 丁楠, 渠鸿竹, 方向东. ENCODE 计划和功能基因组研究. *遗传*, 2014, 36(3): 237–247. [DOI]
- [6] Qu HZ, Fang XD. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics Proteomics Bioinformatics*, 2013, 11(3): 135–141. [DOI]
- [7] The 10000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 2012, 491(7422): 56–65. [DOI]
- [8] Zhu XF, He FH, Zeng HM, Ling SP, Chen AL, Wang YQ, Yan XM, Wei W, Pang YK, Cheng H, Hua CL, Zhang Y, Yang XJ, Lu X, Cao LH, Hao LT, Dong LL, Zou W, Wu J, Li X, Zheng S, Yan J, Zhou J, Zhang LX, Mi SL, Wang XJ, Zhang L, Zou Y, Chen YM, Geng Z, Wang JM, Zhou JF, Liu X, Wang JX, Yuan WP, Huang G, Cheng T, Wang QF. Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat Genet*, 2014, 46(3): 287–293. [DOI]
- [9] Zhang WJ, Gao YJ, Li PX, Shi ZB, Guo T, Li F, Han XK, Feng Y, Zheng C, Wang ZY, Li FM, Chen HQ, Zhou ZC, Zhang L, Ji HB. VGLL4 functions as a new tumor suppressor in lung cancer by negatively regulating the YAP-TEAD transcriptional complex. *Cell Res*, 2014, 24(3): 331–343. [DOI]
- [10] Muers M. RNA: Genome-wide views of long non-coding RNAs. *Nat Rev Genet*, 2011, 12(11): 742. [DOI]
- [11] 李语丽, 于军, 宋述慧. RNA 中 6-甲基腺嘌呤的研究进展. *遗传*, 2013, 35(12): 1340–1351. [DOI]
- [12] Zhang XO, Wang HB, Zhang Y, Lu XH, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*, 2014, 159(1): 134–147. [DOI]
- [13] Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*, 2011, 44(4): 667–678. [DOI]
- [14] Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 2012, 149(7): 1635–1646. [DOI]
- [15] Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucl Acids Res*, 2014, 42(D1): D1001–D1006. [DOI]
- [16] Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I, Reid JG, Fink JK, Morgan MB, Gingras MC, Muzny DM, Hoang LD, Yousaf S, Lupski JR, Gibbs RA. Whole-genome sequencing for optimized patient management. *Sci Transl Med*, 2011, 3(87): 87re3. [DOI]
- [17] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*, 2011, 12(11): 745–755. [DOI]
- [18] Brastianos PK, Taylor-Weiner A, Manley PE, Jones RT, Dias-Santagata D, Thorner AR, Lawrence MS, Rodriguez FJ, Bernardo LA, Schubert L, Sunkavalli A, Shillingford N, Calicchio ML, Lidov HG, Taha H, Martinez-Lage M, Santi M, Storm PB, Lee JY, Palmer JN, Adappa ND, Scott RM, Dunn IF, Laws ER Jr, Stewart C, Ligon KL, Hoang MP, Van Hummelen P, Hahn WC, Louis DN, Resnick AC, Kieran MW, Getz G, Santagata S. Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. *Nat Genet*, 2014, 46(2): 161–165. [DOI]
- [19] Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, 2012, 489(7417): 519–525. [DOI]
- [20] Hodis E, Watson IR, Kryukov GV, Arolt ST, Imielinski M, Theurillat J-P, Nickerson E, Auclair D, Li LR, Place C, Dicara D, Ramos AH, Lawrence MS, Cibulskis K, Sivachenko A, Voet D, Saksena G, Stransky N, Onofrio RC, Winckler W, Ardlie K, Wagle N, Wargo J, Chong K, Morton DL, Stenke-Hale K, Chen G, Noble M, Meyerson M. A landscape of driver mutations in melanoma. *Cell*, 2012, 150(2): 251–263. [DOI]
- [21] Rienzo M, Costa V, Scarpato M, Schiano C, Casamassimi A, Grimaldi V, Ciccodicola A, Napoli C. RNA-Seq for the identification of novel Mediator transcripts in endothelial progenitor cells. *Gene*, 2014, 547(1): 98–105. [DOI]
- [22] Kadara H, Fujimoto J, Yoo SY, Maki Y, Gower AC, Kabout M, Garcia MM, Chow CW, Chu Z, Mendoza G, Shen L, Kalhor N, Hong WK, Moran C, Wang J, Spira A,

- Coombes KR, Wistuba II. Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. *J Natl Cancer Inst*, 2014, 106(3): dju004. [DOI]
- [23] Nord KH, Lilljebjörn H, Vezzi F, Nilsson J, Magnusson L, Tayebwa J, de Jong D, Bovée JVMG, Hogendoorn PCW, Szuhai K. GRM1 is upregulated through gene fusion and promoter swapping in chondromyxoid fibroma. *Nat Genet*, 2014, 46(5): 474–477. [DOI]
- [24] Paritosh K, Gupta V, Yadava SK, Singh P, Pradhan AK, Pental D. RNA-seq based SNPs for mapping in Brassica juncea (AABB): synteny analysis between the two constituent genomes A (from *B. rapa*) and B (from *B. nigra*) shows highly divergent gene block arrangement and unique block fragmentation patterns. *BMC Genomics*, 2014, 15: 396. [DOI]
- [25] Alarcón MA, Medina MA, Hu Q, Avila ME, Bustos BI, Pérez-Palma E, Peralta A, Salazar P, Ugarte GD, Reyes AE, Martin GM, Opazo C, Moon RT, De Ferrari GV. A novel functional low-density lipoprotein receptor-related protein 6 gene alternative splice variant is associated with Alzheimer's disease. *Neurobiol Aging*, 2013, 34(6): 1709.e9–1709.e18. [DOI]
- [26] Malik R, Patel L, Prensner JR, Shi Y, Iyer MK, Subramanian S, Carley A, Niknafs YS, Sahu A, Han SM, Ma T, Liu ML, Asangani IA, Jing XJ, Cao XH, Dhanasekaran SM, Robinson DR, Feng FY, Chinnaiyan AM. The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. *Mol Cancer Res*, 2014, 12(8): 1081–1087. [DOI]
- [27] Kataoka M, Wang DZ. Non-coding RNAs including miRNAs and lncRNAs in cardiovascular biology and disease. *Cells*, 2014, 3(3): 883–898. [DOI]
- [28] White NM, Cabanski CR, Silva-Fisher JM, Dang HX, Govindan R, Maher CA. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome Biol*, 2014, 15(8): 429. [DOI]
- [29] Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*, 2015, 47(3): 199–208. [DOI]
- [30] Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu JZ, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, 2013, 19(2): 141–157. [DOI]
- [31] Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet*, 2010, 6(12): e1001233. [DOI]
- [32] Lukiw WJ. Circular RNA (circRNA) in Alzheimer's disease (AD). *Front Genet*, 2013, 4: 307. [DOI]
- [33] Schübeler D. Function and information content of DNA methylation. *Nature*, 2015, 517(7534): 321–326. [DOI]
- [34] Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, 2011, 12(1): 7–18. [DOI]
- [35] Haig D. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity (Edinb)*, 2014, 113(2): 96–103. [DOI]
- [36] Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang ZZ, Wang JR, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang XC, Claussnitzer M, Liu YP, Coarfa C, Harris RA, Shores N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong CB, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li DF, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature*, 2015, 518(7539): 317–330. [DOI]
- [37] Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, Adhikari S, Shi Y, Lv Y, Chen YS, Zhao X, Li A, Yang Y, Dahal U, Lou XM, Liu X, Huang J, Yuan WP, Zhu XF, Cheng T, Zhao YL, Wang XQ, Danielsen JMR, Liu F, Yang YG. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res*, 2014, 24(2): 177–189. [DOI]
- [38] Jia GF, Fu Y, Zhao X, Dai Q, Zheng GQ, Yang Y, Yi CQ, Lindahl T, Pan T, Yang YG, He C. N⁶-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol*, 2012, 7(12): 885–887. [DOI]
- [39] Zheng GQ, Dahl JA, Niu YM, Fedorcsak P, Huang CM, Li

- CJ, Vågbø CB, Shi Y, Wang WL, Song SH, Lu ZK, Bosmans RPG, Dai Q, Hao YJ, Yang X, Zhao WM, Tong WM, Wang XJ, Bogdan F, Furu K, Fu Y, Jia GF, Zhao X, Liu J, Krokan HE, Klungland A, Yang YG, He C. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell*, 2013, 49(1): 18–29. [DOI]
- [40] Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*, 2012, 13(5): 343–357. [DOI]
- [41] Kaelin WG Jr, McKnight SL. Influence of metabolism on epigenetics and disease. *Cell*, 2013, 153(1): 56–69. [DOI]
- [42] Cencioni C, Spallotta F, Martelli F, Valente S, Mai A, Zeiher AM, Gaetano C. Oxidative stress and epigenetic regulation in ageing and age-related diseases. *Int J Mol Sci*, 2013, 14(9): 17643–17663. [DOI]
- [43] Garbes L, Riessland M, Wirth B. Histone acetylation as a potential therapeutic target in motor neuron degenerative diseases. *Curr Pharm Des*, 2013, 19(28): 5093–5104. [DOI]
- [44] Lu X, Deng Y, S Yu DH, Cao HM, Wang L, Liu L, Yu CJ, Zhang YP, Guo XM, Yu G. Histone acetyltransferase p300 mediates histone acetylation of PS1 and BACE1 in a cellular model of Alzheimer's disease. *PLoS One*, 2014, 9(7): e103067. [DOI]
- [45] Zhang B, Wang J, Wang XJ, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, Davies SR, Wang S, Wang P, Kinsinger CR, Rivers RC, Rodriguez H, Townsend RR, Ellis MJC, Carr SA, Tabb DL, Coffey RJ, Slebos RJ, Liebler DC, NCI CPTAC. Proteogenomic characterization of human colon and rectal cancer. *Nature*, 2014, 513(7518): 382–387. [DOI]
- [46] Zhang WW, Li F, Nie L. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 2010, 156(2): 287–301. [DOI]
- [47] ter Kuile BH, Westerhoff HV. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett*, 2001, 500(3): 169–171. [DOI]
- [48] Nie L, Wu G, Zhang WW. Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: A quantitative analysis. *Genetics*, 2006, 174(4): 2229–2243. [DOI]
- [49] Yoon SH, Han M-J, Jeong H, Lee CH, Xia XX, Lee D-H, Shim JH, Lee SY, Oh TK, Kim JF. Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol*, 2012, 13(5): R37. [DOI]
- [50] Korstanje R, Paigen B. From QTL to gene: the quest begins. *Nat Genet*, 2002, 31(3): 235–236. [DOI]
- [51] Rapp JP, Deng AY. Detection and positional cloning of blood pressure quantitative trait loci: is it possible? Identifying the genes for genetic hypertension. *Hypertension*, 1995, 25(6): 1121–1128. [DOI]
- [52] Yagil C, Hubner N, Monti J, Schulz H, Sapojnikov M, Luft FC, Ganten D, Yagil Y. Identification of hypertension-related genes through an integrated genomic-transcriptomic approach. *Circ Res*, 2005, 96(6): 617–625. [DOI]
- [53] Natrajan R, Weigelt B, Mackay A, Geyer FC, Grigoriadis A, Tan DSP, Jones C, Lord CJ, Vatcheva R, Rodriguez-Pinilla SM, Palacios J, Ashworth A, Reis-Filho JS. An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. *Breast Cancer Res Treat*, 2010, 121(3): 575–589. [DOI]
- [54] Spans L, Helsen C, Clinckemalie L, Van den Broeck T, Prekovic S, Joniau S, Lerut E, Claessens F. Comparative genomic and transcriptomic analyses of LNCaP and C4-2B prostate cancer cell lines. *PLoS One*, 2014, 9(2): e90002. [DOI]
- [55] Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guha-Thakurta D, Sieberts SK, Monks S, Reitman M, Zhang CS, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang LM, Castle J, Zhu HY, Kash SF, Drake TA, Sachs A, Lusis AJ. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet*, 2005, 37(7): 710–717. [DOI]
- [56] Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA*, 2003, 100(18): 10393–10398. [DOI]
- [57] Sørli T, Wang YL, Xiao CL, Johnsen H, Naume B, Samaha RR, Borresen-Dale A-L. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics*, 2006, 7: 127. [DOI]
- [58] Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic

- and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 2012, 486(7403): 346–352. [DOI]
- [59] Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nat Genet*, 2011, 43(4): 309–351. [DOI]
- [60] Lahtz C, Pfeifer GP. Epigenetic changes of DNA repair genes in cancer. *J Mol Cell Biol*, 2011, 3(1): 51–58. [DOI]
- [61] Wrzeszczynski KO, Varadan V, Byrnes J, Lum E, Kama-lakaran S, Levine DA, Dimitrova N, Zhang MQ, Lucito R. Identification of tumor suppressors and oncogenes from genomic and epigenetic features in ovarian cancer. *PLoS One*, 2011, 6(12): e28503. [DOI]
- [62] Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu HZ, Brody J, Shafer A, Neri F, Lee K, Kutayavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 2012, 337(6099): 1190–1195. [DOI]
- [63] Yi JM, Dhir M, Van Neste L, Downing SR, Jeschke J, Glöckner SC, de Freitas Calmon M, Hooker CM, Funes JM, Boshoff C, Smits KM, van Engeland M, Weijenberg MP, Iacobuzio-Donahue CA, Herman JG, Schuebel KE, Baylin SB, Ahuja N. Genomic and epigenomic integration identifies a prognostic signature in colon cancer. *Clin Cancer Res*, 2011, 17(6): 1535–1545. [DOI]
- [64] Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. *Nature*, 2004, 430(6995): 85–88. [DOI]
- [65] Rius M, Lyko F. Epigenetic cancer therapy: rationales, targets and drugs. *Oncogene*, 2012, 31(39): 4257–4265. [DOI]
- [66] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10(1): 57–63. [DOI]
- [67] Cuellar-Partida G, Buske FA, McLeay RC, Whittington T, Noble WS, Bailey TL. Epigenetic priors for identifying active transcription factor binding sites. *Bioinformatics*, 2012, 28(1): 56–62. [DOI]
- [68] Marconett CN, Zhou BY, Rieger ME, Selamat SA, Dubourd M, Fang XH, Lynch SK, Stueve TR, Siegmund KD, Berman BP, Borok Z, Laird-Offringa IA. Integrated transcriptomic and epigenomic analysis of primary human lung epithelial cell differentiation. *PLoS Genet*, 2013, 9(6): e1003513. [DOI]
- [69] Rodríguez E, Baurecht H, Wahn AF, Kretschmer A, Hotze M, Zeilinger S, Klopp N, Illig T, Schramm K, Prokisch H, Kuhnel B, Gieger C, Harder J, Cifuentes L, Novak N, Weidinger S. An integrated epigenetic and transcriptomic analysis reveals distinct tissue-specific patterns of DNA methylation associated with atopic dermatitis. *J Invest Dermatol*, 2014, 134(7): 1873–1883. [DOI]
- [70] Bandres E, Agirre X, Bitarte N, Ramirez N, Zarate R, Roman-Gomez J, Prosper F, Garcia-Foncillas J. Epigenetic regulation of microRNA expression in colorectal cancer. *Int J Cancer*, 2009, 125(11): 2737–2743. [DOI]
- [71] Balaguer F, Link A, Lozano JJ, Cuatrecasas M, Nagasaka T, Boland CR, Goel A. Epigenetic silencing of miR-137 is an early event in colorectal carcinogenesis. *Cancer Res*, 2010, 70(16): 6609–6618. [DOI]
- [72] Zhang W, Liu Y, Sun N, Wang D, Boyd-Kirkup J, Dou XY, Han J-DJ. Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor prognosis in ovarian cancer. *Cell Rep*, 2013, 4(3): 542–553. [DOI]
- [73] Butterfield DA, Dalle-Donne I. Redox proteomics: from protein modifications to cellular dysfunction and disease. *Mass Spectrom Rev*, 2014, 33(1): 1–6. [DOI]
- [74] González-Domínguez R, García-Barrera T, Vitorica J, Gómez-Ariza JL. Metabolomics reveals significant impairments in the immune system of the APP/PS1 transgenic mice of Alzheimer's disease. *Electrophoresis*, 2015, 36(4): 577–587. [DOI]

(责任编辑: 赵方庆)