

乳腺癌中 SRA1 基因表达调控的生物信息学分析

王青¹,傅泉男¹,洪祝平¹,张玉涵¹,郑春红¹,王语嫣²,丁先锋¹

(1. 浙江理工大学生命科学学院生物工程研究所,杭州 310018;2. 浙江经贸职业技术学院应用工程系,杭州 310018)

摘要: 分析类固醇受体 RNA 激活物 (steroid receptor RNA activator, SRA1) 以长链非编码 RNA 形式 (lncRNA SRA1) 和蛋白质形式 (SRA1 蛋白) 在乳腺癌中的分子调控网络。运用 RegRNA、Targetscan、MicroCosm Targets、PicTar 软件和 HMDD、DAVID 在线数据库,预测 lncRNA SRA1 与 microRNA 及下游靶基因间的多种调控途径,对靶基因进行功能富集分析,绘制 lncRNA SRA1 的核心调控网络图。同时运用 TCGA 数据库、string 软件预测分析与 SRA1 蛋白相互作用的蛋白质。发现 lncRNA SRA1 上存在 hsa-let-7b、hsa-let-7c、hsa-let-7g、hsa-let-7i、has-miR-92、has-miR-103、has-miR-194、has-miR-874、has-miR-141、has-miR-146a、has-miR-661 这 11 个与乳腺癌相关 microRNA 的可能结合位点,从而调节下游 230 个靶基因,构建了 lncRNA SRA1-miRNAs-mRNAs 调控网络,提示 lncRNA SRA1 可能参与基因转录调控、生物合成、MAPK、癌症相关通路、紧密连接等信号通路。并且发现 SRA1 蛋白和 DDX17、ESR1、NCOA2、NCOA1、NR1P1 等多种蛋白存在相互作用。对 SRA1 基因调控网络的生物信息学分析有助于理解其在乳腺癌发生发展过程中的分子机制。

关键词: 乳腺癌;类固醇受体激活物;lncRNA SRA1-miRNAs-mRNAs;生物信息学

中图分类号: Q786 **文献标志码:** A **文章编号:** 1673-3851 (2016) 04-0283-07 **引用页码:** 030703

0 引言

乳腺癌是危害妇女健康的最常见恶性肿瘤之一,据统计,2014 年美国的癌症统计研究报告中,大约有 23.27 万女性罹患乳腺癌,占全部女性所患癌症总人数的 29%,位列妇女恶性肿瘤的首位^[1]。乳腺癌的科学有效治疗成为临床面临的巨大挑战。

类固醇受体 RNA 激活物 (steroid receptor RNA activator, SRA1) 是一种能以长链非编码 RNA (lncRNA SRA1) 形式和蛋白质形式 (SRA1 蛋白) 调节甾体激素受体活性的分子^[2]。目前,多项研究证实,SRA1 基因在卵巢癌^[3-4]、乳腺癌^[4-6]、子宫肿瘤^[4]等癌症中异常表达,与肿瘤的发生、发展、预后密切相关。Liu 等^[3]报道 lncRNA SRA1 在多囊卵巢综合征患者中比正常显著高表达。Yan 等^[5]研

究发现在雌激素受体阳性 (ER+) 乳腺癌患者中,低表达 SRA1 蛋白患者的生存率较高表达的高,他们推测 SRA1 蛋白可作为 ER+ 乳腺癌的预后指标。Lanz 等^[4]通过 Northern 和荧光定量发现 SRA1 基因在卵巢肿瘤,子宫肿瘤和乳腺癌组织中较周围组织高表达,尤其在 29 例乳腺癌组织样本中 SRA1 基因比对应癌旁样本表达量上调 6 倍。

然而,这些研究只停留在 SRA1 基因的表达情况,并没有深入研究其作用机制。SRA1 基因作为 lncRNA 时发挥竞争性内源性 RNA 作用调控下游 miRNA 及其靶基因与 SRA1 蛋白和其他蛋白相互作用的机制并不清楚。运用生物信息学方法预测并分析 lncRNA SRA1 与下游 microRNA 及 microRNA 靶基因间的相互关系,并预测 SRA1 蛋白和其他蛋白的相互作用,以期为进一步研究

收稿日期: 2015-07-30

基金项目: 国家自然科学基金项目 (11105121); 浙江省自然科学基金项目 (LY15C050002); 浙江省创新团队项目 (2014R406078); 浙江经贸职业技术学院大学生创新项目 (20140421)

作者简介: 王青 (1991-), 女, 江苏盐城人, 硕士研究生, 主要从事肿瘤非编码 RNA 方面的研究。

通信作者: 丁先锋, E-mail: xfding@zstu.edu.cn

SRA1 基因的靶基因验证、蛋白质间相互作用以及其生物学功能等的研究提供数据支持和理论指导,为更深层次研究其在乳腺癌发生发展过程中的分子机制提供可靠线索,获得更好的研究乳腺癌的治疗方法。

1 方法

1.1 与 lncRNA SRA1 相互作用的 microRNA 预测

运用 RegRNA 生物学软件预测 lncRNA SRA1 序列上可能存在的 microRNA 结合位点 (<http://regRNA.mbc.nctu.edu.tw/html/prediction.html>)。在预测 lncRNA SRA1 与 microRNA 结合位点时,选取最小折叠自由能 (minimum folding free energy, MFE) 小于或者等于 -25 , 并且根据 lncRNA SRA1 与其 microRNA 配对时的得分,选取 score 大于或等于 150, 得分越高,表示 lncRNA SRA1-microRNA 的结合能力越强。同时,以人类 microRNA 疾病数据库 (HMDD) (<http://www.cuilab.cn/hmdd>) 检索分析与乳腺癌有关的 microRNAs,取其交集分析在乳腺癌中可能与 lncRNA SRA1 相互作用的 microRNA。

1.2 microRNA 靶基因的预测

利用 Targetscan (http://www.targetscan.org/vert_50/seedmatch.html)、PicTar (<http://pictar.mdc-berlin.de/>) 和软件 MicroCosm Targets (<http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/>) 预测 microRNAs 和靶基因上可能的结合位点。为了避免过多的假阳性结果,取至少两个软件预测结果交集,作为进一步结果分析其调控网络。

1.3 构建 lncRNA SRA1-miRNAs-mRNAs 相互作用网络

选取了 lncRNA SRA1 潜在调控的 microRNAs,分别预测每个 microRNA 调控的靶基因,并利用 Cytoscape 整合 lncRNA SRA1-microRNAs 和 microRNAs-mRNAs 网络,构建成 lncRNA SRA1-miRNAs-mRNAs 网络调控关系。

1.4 microRNA 靶基因的功能分析

利用 DAVID (<http://david.abcc.ncifcrf.gov/>) 数据库分析靶基因的生物学功能。选择 Gene Ontology 中的生物过程 (biological process, BP)、分子功能 (molecular function) 与细胞组分 (cell component) 条目和 Pathways 中的 KEGG 通路条目,进行分析。

1.5 SRA1 蛋白 string 分析

为了进一步分析 SRA1 蛋白与之相互作用蛋白之间的关系,将 SRA1 蛋白上传至 string 数据库 (<http://string-db.org/>),string 数据库可以根据之前的实验数据和预测数据给蛋白之间相互作用的可能性打分,分数越高,表明这两个蛋白之间关联度越高,发生相互作用的可能性就越高。设置可信度 (confidence) 为高可信度, score 大于或等于 0.7, 对与 SRA1 蛋白相关蛋白及其之间相互作用进行分析。

1.6 运用公共数据库分析 SRA1 蛋白在乳腺癌组织中的表达

通过在线工具 cBioPortal for Cancer Genomics (<http://www.cbioportal.org/>) 以肿瘤基因组图谱 (the cancer genome atlas, TCGA) 提供的大量样本,分析 SRA1 蛋白在乳腺癌组织中的表达情况及其他蛋白随之变化情况。

2 结果

2.1 与 lncRNA SRA1 相互作用的 microRNAs

运用 RegRNA 生物学软件预测能够与 lncRNA SRA1 结合的 microRNAs 共 71 个。通过 HMDD 数据库检索,结果显示与乳腺癌有关的有 11 个 microRNAs (hsa-let-7b, hsa-let-7c, hsa-let-7g, hsa-let-7i, has-miR-92, has-miR-103, has-miR-194, has-miR-874, has-miR-141, has-miR-146a, has-miR-661)。在这 11 个 microRNAs 中, lncRNA SRA1 与 miR-661 配对的得分最高,提示 lncRNA SRA1 与 miR-661 结合能力可能最强。因此预测了 lncRNA SRA1 与 miR-661 可能形成的配对结构,见图 1。从图 1 中可见该结构由多个发夹结构组成,说明能够形成稳定的结构。并且计算该结构的最小折叠自由能,结果显示 MFE 为 -30.70 kcal/mol,提示该结构具有较低的折叠自由能, lncRNA SRA1 与 miR-661 结合的可能性非常大。



图1 lncRNA SRA1 与 miR-661 可能形成的配对结构

2.2 microRNAs 的靶基因

运用 targetsScan, picTar 和 MicroCosm Targets 软件共同分析这 11 个 microRNAs 的靶基因,预测有 230 个靶基因可能受这 11 个 microRNAs 调控。各 microRNA 靶基因的统计结果见图 2,其中发现 let-7b 的靶基因有 5 个,let-7c 的靶基因有 5 个,let-7g 的靶基因有 12 个,let-7i 的靶基因有 9 个,miR-92 的靶基因有 5 个,miR-103 的靶基因有 6 个,miR-194 的靶基因有 70 个,miR-874 的靶基因有 36 个,miR-141 的靶基因有 34 个,miR-146a 的靶基因有 27 个,miR-661 的靶基因有 25 个。其中 miR-194、miR-874 和 miR-141 的靶基因较多,说明 miR-194、miR-874 和 miR-141 参与更为复杂的调控通路。并且发现 miR-146a 和 miR-661

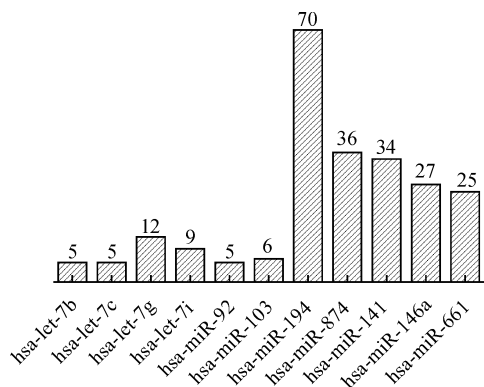


图 2 microRNAs 调控靶基因统计图

共同调控白介素 1 受体相关激酶 1 (IRAK1) 基因, let-7b 和 miR-874 共同调控 L 型钙通道 $\alpha 1$ 亚单位基因 (CACNA1D), miR-141 和 miR-194 共同调控转录因子分析叉头框蛋白 A1 基因 (FOXA1) 和色素域解旋酶 DNA 结合蛋白 (CHD9) 基因。

2.3 lncRNA SRA1-miRNAs-mRNAs 相互作用网络

在建立 lncRNA SRA1-miRNAs 调控网络的基础上,增加了 miRNAs-mRNAs 网络,最终建立了 lncRNA SRA1 在乳腺癌中的 lncRNA SRA1-miRNAs-mRNAs 调控网络,如图 3 所示。从图中可以看出该网络总共由 242 个节点和 245 条边组成,242 个节点代表 lncRNA SRA1,11 个 miRNAs 和 230 个 mRNAs,其中菱形代表 lncRNA SRA1,三角形代表 microRNA,圆形代表 microRNA 靶基因,245 条边表示它们之间存在 245 种相互作用关系。lncRNA SRA1 在该网络的中心,调节与之结合的 hsa-let-7b、hsa-let-7c、hsa-let-7g、hsa-let-7i、has-miR-92、has-miR-103、has-miR-194、has-miR-874、has-miR-141、has-miR-146a、has-miR-661 这 11 个 miRNAs,进而调控下游 230 个靶基因。所有相关基因构成一个复杂的网络调控,相互协调,提示 lncRNA SRA1 可能从不同的层面上参与乳腺癌细胞的发生发展过程。

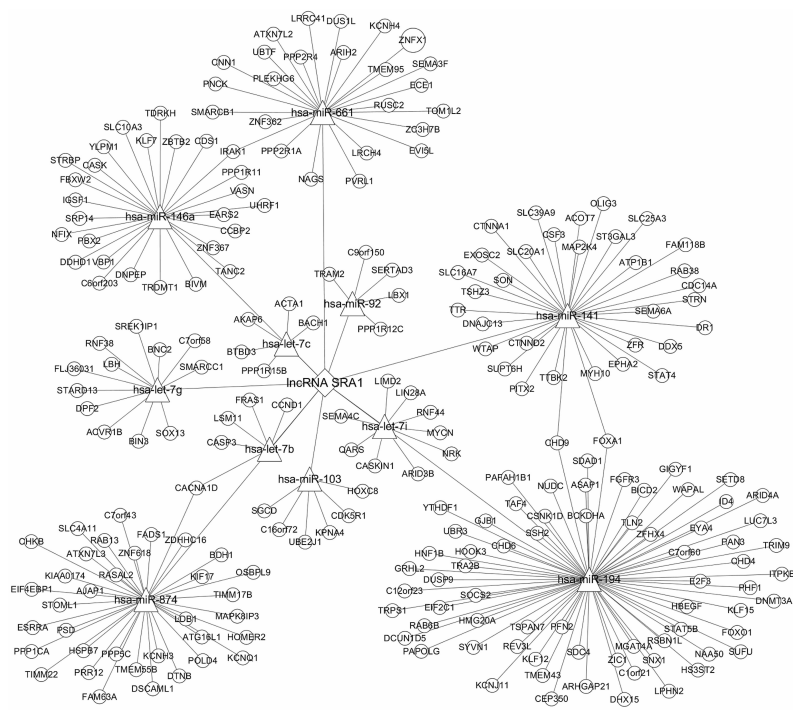


图 3 lncRNA SRA1-miRNA-mRNA 调控网络

注:菱形代表 lncRNA SRA1,三角形代表 microRNA,圆形代表 microRNA 靶基因。

2.4 microRNAs 靶基因的 GO 和 pathway 功能富集分析

为了进一步从总体上了解乳腺癌中 lncRNA SRA1 的生物学功能,将参与 lncRNA SRA1-miRNAs-mRNAs 的 230 个 microRNA 靶基因映射到 DAVID 数据库,进行 GO 和 KEGG pathway 功能聚类分析,推测 lncRNA SRA1 可能参与的生物

过程以及所在的信号通路。在 microRNA 的靶基因 GO 分析中,结果显示:在生物学过程中, microRNA 靶基因高度富集到转录调节、RNA 代谢、磷酸化、DNA 结合等过程,见图 4。在 microRNA 的靶基因 KEGG pathway 分析中, microRNA 靶基因高度富集到 MAPK 信号通路,癌症相关通路,紧密连接和轴突导向信号转导通路。

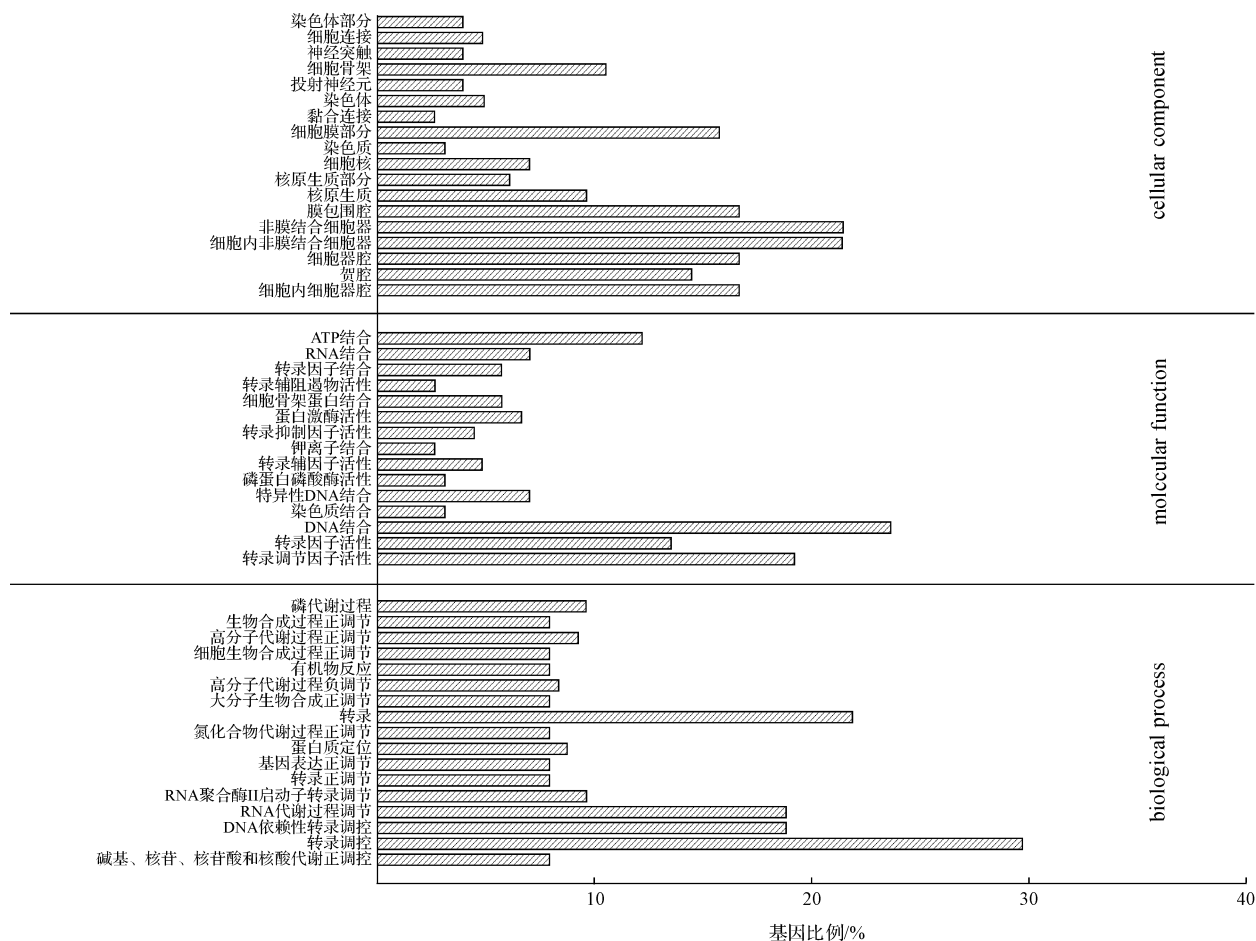


图4 microRNA 靶基因的 GO 聚类分析

2.5 SRA1 蛋白相互作用调控网络

采用 string 在线软件,分析提示 SRA1 蛋白和 DDX17、ESR1、NCOA2、NCOA1、NRIP1、CREBBP、PPARA、EP300、MED1、ACOX1、NR1H3、APOA1、ME1、HSD17B4、APOA2、CITED2、LPL、FABP1、CPT1A 和 PPARGC1A 等多种蛋白存在相互作用。由图 5 可知,SRA1 蛋白处于网络的中心,多种蛋白之间存在密切复杂的相互作用关系,为深入研究 SRA1 蛋白提供了一定的线索。

同时利用 cBioPortal 软件通过 TCGA 大量样本分析乳腺癌相关基因表达情况,结果显示 SRA1 基因在 9.1%($n=507$)乳腺癌患者中变异。进一步

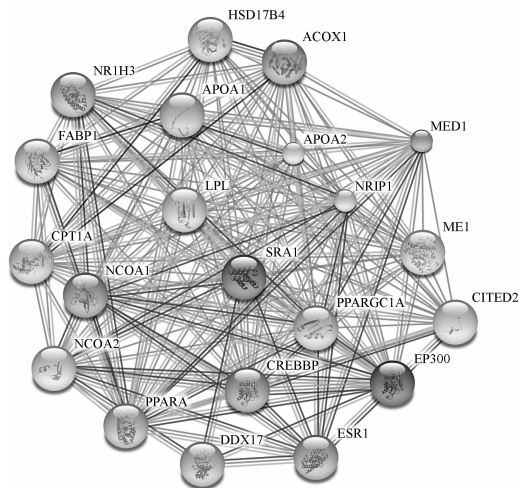


图5 以 SRA1 为中心的蛋白网络

分析蛋白表达情况,发现 SRA1 蛋白的突变在乳腺癌中与 ESR1 蛋白的表达量下调显著相关($P=0.0009$)(图6)。说明 SRA1 蛋白可能通过和 ESR1 相互作用,调控下游基因的表达,参与乳腺癌发生发展过程。从蛋白质相互作用网络出发,在蛋白质水平宏观分析乳腺癌发病的整体联系,能更客观的发现乳腺癌发生发展的关键靶点。

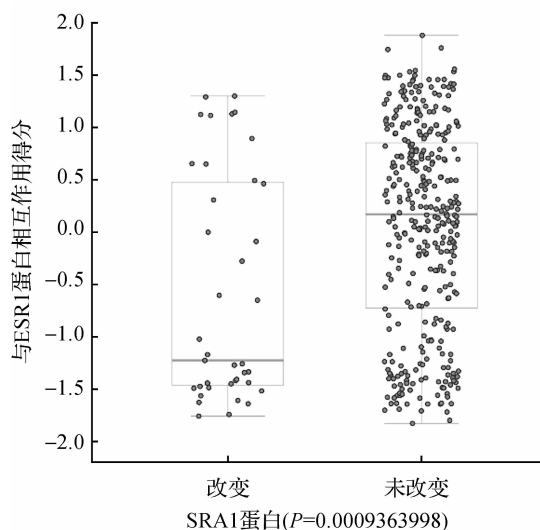


图6 ESR1 蛋白随 SRA1 蛋白在乳腺癌中变化
注:改变代表 SRA1 蛋白改变的样本,未改变代表 SRA1 蛋白没有改变的样本。

3 讨论

近年来,随着对肿瘤的深入研究,人们逐步认识到许多分子和蛋白与肿瘤的发生发展密切相关。SRA1 是首次发现的可以从 RNA 和蛋白质两个水平上发挥功能的分子。目前对 lncRNA SRA1 和 SRA1 蛋白的功能研究较少,尤其在乳腺癌中 lncRNA 调控 miRNA 方面的报道更为少见。

lncRNA SRA1 参与复杂的调节过程。lncRNA 通过多种机制在染色质重塑、转录调控、转录后调控和翻译调控等生命活动中发挥功能。2011 年, Pandolfi 教授领导的课题组提出了竞争性内源 RNA (competing endogenous RNA, ceRNA) 假说^[7]。该假说认为 mRNA, lncRNA 等转录物可以竞争性结合一些相同的微小 RNA, 从而相互影响彼此的表达水平。Wang 等^[8]的研究表明 lncRNA HULC 具有 miR-372 的结合位点, 可以作为一种内源性“海绵”, HULC 的高表达可以使 miR-372 的表达水平降低, 导致 miR-372 靶基因 Prkacb 的表达上调, 最终使得 CREB 磷酸化, 促进 PKA 信号通路,

促进肝癌的发生。Maxim 等^[9]研究了新型 lncRNA BARD1 9' L 在多种癌症发生发展中的作用, 发现 BARD1 9' L mRNA 的 3'-UTR 能竞争性结合 miR-203 和 miR-101, 从而调控肿瘤抑制基因 BARD1 的表达水平, 促进癌细胞的增殖和迁移。虽然类似这样的例子在乳腺癌中也有描述^[10], 但多角度关于 lncRNA SRA1 的调控, 尤其是可能存在 lncRNA SRA1-miRNA 相互作用, 这需要进一步研究。利用 RegRNA 软件预测与 lncRNA SRA1 结合的靶 microRNA, 再结合 HMDD 数据库分析, 发现在乳腺癌中有 11 个 microRNAs 可能受 lncRNA SRA1 调控, 其中 miR-141、miR-146a 和 miR-661 研究相对广泛, 同时运用生物学软件预测这 11 个 microRNAs 的靶基因, 发现有 230 个靶基因可能受这 11 个 microRNAs 调控, 最终建立了 lncRNA SRA1 在乳腺癌中的 lncRNA-miRNA-mRNA 调控网络。如果其中任何一个基因表达失调也许会使得多个信号通路发生改变, 从而促使乳腺癌的发生。此外, 根据 ceRNA 假说对 lncRNA SRA1 的功能进行了预测, 发现 lncRNA SRA1 可能影响 RNA 的转录、神经分化、磷代谢过程、MAPK 信号通路等多种生物学功能, 这有利于生物学家更有针对性地锁定感兴趣的靶基因并对其进行研究, 为阐明 SRA1 基因的生物学功能提供了数据支持和理论指导。

长期以来, 类固醇受体激活物 SRA1 基因被认为作为非编码转录产物发挥活性^[11], 但 Hube 研究发现如果 SRA1 前体选择另一种剪切方式, 将会产生 mRNA, 翻译出 SRA1 蛋白^[12]。一系列研究发现 SRA1 的 lncRNA 形式主要与类固醇激素受体辅激活物 SRC-1 结合发挥转录共激活作用^[11]。SRA1 的蛋白质形式也具有类似的功能, 但是不同于 RNA 形式, SRA1 蛋白可结合到特定基因的启动子区域, 并起到抑制物的作用。Kawashima 改变 SRA1 蛋白的开放阅读框序列来阻遏 SRA1 蛋白的表达, 促使调节功能丧失, 同时发现 SRA1 蛋白在体外能直接结合雄激素受体 AR 的激活功能区 AF-2^[13]。SRA1 蛋白的异常表达参与肿瘤细胞的发生、发展, 并且影响患者的生存率。因此, 运用生物学软件预测 SRA1 蛋白与其他蛋白的相互关系, 发现 SRA1 蛋白和许多蛋白之间都存在直接或间接的相互作用, 这也提示了生物调控网络具有多层次和复杂性。此外, 通过 TCGA 数据库得到 507 例乳腺癌患者测序信息, 发现 SRA1 蛋白的突变在乳腺癌中与

ESR1 蛋白的表达量下调显著相关($P=0.0009$)。虽然目前在乳腺癌研究上已取得一定成果,但其中很多复杂的蛋白之间相互作用仍未被发现,这需要更深入的研究。

SRA1 基因能以长链非编码 RNA 和蛋白质两个形式发挥功能,但在乳腺癌发生、发展中什么阶段是在长链非编码 RNA 水平上发挥作用,什么阶段在蛋白质水平上起作用,两者之间又是如何相互协调的,这都是需要进一步去解决的问题。虽然构建了 lncRNA SRA1-miRNAs-mRNAs 调控网络和分析了 SRA1 蛋白与其他蛋白的相互作用,但也存在一定的局限性。一方面,由于各种预测软件具有一定的局限性,预测结果存在假阳性。另一方面,以上结果仅仅是从生物信息学角度分析而来,还需要进一步的实验验证,以获得更可信的结果。

4 结 论

利用生物信息学方法,研究发现 lncRNA SRA1 上存在 11 个与乳腺癌相关 microRNA 的可能结合位点和 230 个 microRNA 靶基因,构建了 lncRNA SRA1 紧密复杂的基因表达调控网络,并根据竞争性内源假说预测 lncRNA SRA1 的功能,结果显示 lncRNA SRA1 可能参与基因转录调控、生物合成、MAPK 等信号通路。同时构成以 SRA1 蛋白为中心的蛋白质网络,并且 SRA1 蛋白可能通过和 ESR1 相互作用,调控下游基因的表达,为研究其在乳腺癌发生发展中的分子机制及后续实验提供新思路。

参考文献:

- [1] SIEGEL R, MA J, ZOU Z, et al. Cancer statistics, 2014[J]. CA-cancer J Clin, 2014, 64(1):9-29.
- [2] CHOONIEDASS-KOTHARI S, EMBERLEY E, HAMEDNI M, et al. The steroid receptor RNA activator is the first functional RNA encoding a protein [J]. FEBS Lett, 2004, 566(1):43-47.
- [3] LIU Z, HAO C, HUANG X, et al. Peripheral blood leukocyte expression level of lncRNA steroid receptor RNA activator (SRA) and its association with polycystic ovary syndrome: a case control study [J]. Gynecol Endocrinol, 2015, 31(5):1-6.
- [4] LANZ R B, CHUA S S, BARRON N, et al. Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo [J]. Mol Cell Biol, 2003, 23(20): 7163-7176.
- [5] YAN Y, PENNER CC, SKLIRIS GP, et al. Steroid receptor RNA activator protein (SRAP) expression as a prognostic factor in ER+ human breast tumors [J]. J Cancer Res Clin Oncol, 2013, 139(10):1637-1647.
- [6] FOULDS C E, TSIMELZON A, LONG W, et al. Research resource: expression profiling reveals unexpected targets and functions of the human steroid receptor RNA activator (SRA) gene [J]. Molecular Endocrinology, 2010, 24(5): 1090-1105.
- [7] SALMENA L, POLISENO L, TAY Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? [J] Cell, 2011, 146(3):353-358.
- [8] WANG J, LIU X, WU H, et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer [J]. Nucleic acids Res, 2010, 38(16):5366-5383.
- [9] PILYUGIN M, IRMINGER-FINGER I. Long non-coding RNA and microRNAs might act in regulating the expression of BARD1 mRNAs [J]. Int J Biochem Cell B, 2014, 54(8):356-367.
- [10] AUGOFF K, MCCUE B, PLOW E F, et al. miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer [J]. Mol cancer, 2012, 11(1):5.
- [11] LANZ RB, MCKENNA NJ, ONATE SA, et al. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex [J]. Cell, 1999, 97(1):17-27.
- [12] HUBE F, GUO J, CHOONIEDASS-KOTHARI S, et al. Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines [J]. Dna Cell Biol, 2006, 25(7):418-428.
- [13] KAWASHIMA H, TAKANO H, SUGITA S, et al. A novel steroid receptor co-activator protein (SRAP) as an alternative form of steroid receptor RNA-activator gene: expression in prostate cancer cells and enhancement of androgen receptor activity [J]. Biochem J, 2003, 369:163-171.

Bioinformatic Analysis of SRA1 Gene in Breast Cancer

WANG Qing¹, FU Xiaonan¹, HONG Zhuping¹, ZHANG Yuhan¹, ZHENG Chunhong¹, WANG Yuyan², DING Xianfeng¹

(1. Institute of Bioengineering, College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. Department of Application Engineering, Zhejiang Economic & Trade Polytechnic, Hangzhou 310018, China)

Abstract: To predict the regulatory network of SRA1 both at RNA and protein level in breast cancer by the methods of bioinformatics. The RegRNA, Targetscan, MicroCosm Targets, PicTar bioinformatics applications and HMDD, DAVID databases were employed to predict the interaction between lncRNA SRA1 and microRNAs. At the same time, TCGA database and string application were utilized to develop the SRA1-centric proteins network. The potential binding sites of 11 miRNAs (hsa-let-7b, hsa-let-7c, hsa-let-7g, hsa-let-7i, has-miR-92, has-miR-103, has-miR-194, has-miR-874, has-miR-141, has-miR-146a, has-miR-661) on lncRNA SRA1 sequence, and 230 target genes of the 11 miRNAs were found. The results indicated that the lncRNA SRA1 may play some parts in transcription, biosynthetic process, MAPK signaling pathway, cancer related pathways and tight junction. In addition, it also showed SRA1 protein interacted with multitude proteins. It was demonstrated that the regulatory network of SRA1 by bioinformatics provided reliable clues for exploring the molecular mechanisms of SRA1 gene in breast cancer.

Key words: breast cancer; steroid receptor RNA activator; lncRNA SRA1-miRNAs-mRNAs; bioinformatics

(责任编辑: 许惠儿)