

(19)中华人民共和国国家知识产权局



(12)发明专利申请

(10)申请公布号 CN 108319818 A

(43)申请公布日 2018.07.24

(21)申请号 201810122529.9

(22)申请日 2018.02.07

(71)申请人 中国科学院生物物理研究所
地址 100101 北京市朝阳区大屯路15号

(72)发明人 陈小伟 范珍 陈润生

(74)专利代理机构 北京纪凯知识产权代理有限公司 11245

代理人 关畅

(51)Int.Cl.

G06F 19/22(2011.01)

G06F 19/20(2011.01)

权利要求书3页 说明书15页

(54)发明名称

一种预测影响长非编码RNA生物学功能的
SNP位点的方法

(57)摘要

本发明公开了一种预测影响长非编码RNA生物学功能的SNP位点的方法。本发明首先建立了长非编码RNA所包含的RNA结合蛋白能够识别的特异的motif序列的识别及显著性评价方法,然后基于此方法对位于这些特定序列中的SNP位点所产生的影响进行评价。利用此方法可以发现对长非编码RNA生物学功能产生重要影响的SNP位点,为实验人员提供可靠的功能研究线索,也为长非编码RNA的临床应用提供指导。

1. 一种预测影响与RNA结合蛋白结合的长非编码RNA的功能的SNP位点的方法,包括如下步骤:

(1) 收集整理长非编码RNA数据,构建长非编码RNA数据集;

(2) 收集人类基因组SNP位点的注释数据,通过比较SNP位点和长非编码基因的基因组定位,识别位于长非编码基因区的SNP位点;将所述长非编码RNA数据集中的长非编码RNA序列上的所述SNP位点对应的碱基替换为突变后的碱基,构建得到SNP位点碱基突变后的长非编码RNA数据集;

(3) 收集RNA结合蛋白的注释数据,构建RNA结合蛋白的motif数据集;

所述motif数据集中包括motif序列,所述motif序列为RNA结合蛋白能够特异识别的RNA序列;

(4) 基于极值分布对所述motif序列与所述长非编码RNA数据集中的每个长非编码RNA上的目标序列的相似程度进行评价,得到每个长非编码RNA打分最高的目标序列与所述motif序列相似程度的显著性水平 p ,并选择显著性水平 p 低于阈值的长非编码RNA作为SNP位点碱基突变前的所述RNA结合蛋白的靶标长非编码RNA;

所述目标序列是指位于所述长非编码RNA序列上且与所述motif序列长度相等的RNA序列;

假设某长非编码RNA序列长度为 N ;所述motif序列长度为 L ,那么在该长非编码RNA序列上就会得出 $(N-L+1)$ 个目标序列;

(5) 基于极值分布对所述motif序列与所述SNP位点碱基突变后的长非编码RNA数据集中的长非编码RNA上的目标序列的相似程度进行评价,得到每个长非编码RNA打分最高的目标序列与所述motif序列相似程度的显著性水平 p ,并选择相似程度的显著性水平低于阈值的长非编码RNA作为SNP位点碱基突变后的所述RNA结合蛋白的靶标长非编码RNA;

(6) 比较所述SNP位点碱基突变前的所述RNA结合蛋白的靶标长非编码RNA和所述SNP位点碱基突变后的所述RNA结合蛋白的靶标长非编码RNA,得到所述影响与RNA结合蛋白结合的长非编码RNA的功能的SNP位点。

2. 根据权利要求1所述的方法,其特征在于:以所述长非编码RNA数据集中的任意一个长非编码RNA为例,将其记作长非编码RNA甲,对所述motif序列与所述长非编码RNA甲上的目标序列的相似程度进行评价,得到所述长非编码RNA甲打分最高的目标序列与所述motif序列相似程度的显著性水平 p 的方法包括如下步骤:

1) 针对长非编码RNA甲,构建10000条与所述长非编码RNA甲序列长度相同的随机序列;所述长非编码RNA甲序列长度为 N ;所述motif序列长度为 L ;所述 N 大于等于 L ;

2) 针对每一条随机序列,分别计算 $(N-L+1)$ 个目标序列与所述motif序列的相似度打分,分别得到 $(N-L+1)$ 个目标序列与所述motif序列的相似度打分,将最高的相似度打分记作 S_{max} ;

所述目标序列是指位于所述随机序列上且与所述motif序列长度相等的RNA序列;

所述目标序列与所述motif序列的相似度打分 S 的计算公式如下:

$$s = \sum_{i=1}^L w_i$$

其中, w_{ij} 是目标序列第 j 个位置上的碱基与 motif 序列的相似度打分;

所述目标序列第 j 个位置上的碱基与 motif 序列的相似度打分的计算方法如下: 对于序列长度为 L 的目标序列, 位于第 j 个位置的碱基 i 与 motif 序列的相似度打分 $w_{i,j}$ 的计算公式如下:

$$w_{i,j} = \log\left(\frac{f_{i,j}}{p_i}\right);$$

其中, i 为 A、G、C、U; j 为 1、2、3、……, L ; $f_{i,j}$ 是 motif 序列第 j 个位置的碱基 i 出现的频率, p_i 是碱基 i 在人类基因组中出现的频率;

3) 同理, 按照步骤 2) 中的方法, 计算其他随机序列的最高的相似度打分, 共得到 10000 个最高的相似度打分, 并计算所述 10000 个最高的相似度打分的平均值, 将所述平均值记作 \bar{S}_{\max} ;

4) 按照如下公式估计极值分布的参数 μ 和 β :

$$\mu = \bar{S}_{\max} - \gamma\beta;$$

$$\beta = \frac{\sigma_{S_{\max}} \sqrt{6}}{\pi};$$

其中, $\sigma_{S_{\max}}$ 为 S_{\max} 的标准差; γ 为欧拉-马歇罗尼常数;

5) 基于极值分布, 按照如下公式计算长非编码 RNA 甲打分最高的目标序列与所述 motif 序列相似程度的显著性水平 p :

$$p = 1 - e^{-e^{-\frac{S}{\beta}}}$$

其中, S 为长非编码 RNA 甲所对应的所有目标序列中与 motif 序列的相似度打分最高者的打分;

同理, 得到所述长非编码 RNA 数据集中的其他长非编码 RNA 打分最高的目标序列与 motif 序列相似程度的显著性水平 p 。

3. 根据权利要求 1 或 2 所述的方法, 其特征在于: 所述步骤 (6) 中, “比较所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 和所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA, 得到所述影响与 RNA 结合蛋白结合的长非编码 RNA 的功能的 SNP 位点”的方法如下: 属于所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 但不属于所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA 的长非编码 RNA 上的 SNP 位点为对所述 RNA 结合蛋白结合位点有显著影响的 SNP 位点。

4. 根据权利要求 1 或 2 所述的方法, 其特征在于: 所述步骤 (6) 中, “比较所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 和所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA, 得到所述影响与 RNA 结合蛋白结合的长非编码 RNA 的功能的 SNP 位点”的方法如下: 不属于所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 但属于所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA 的长非编码 RNA 上的 SNP 位点即为所述 RNA 结合蛋白能够产生新的结合位点的 SNP 位点。

5. 根据权利要求 1-4 中任一所述的方法, 其特征在于: 所述步骤 (1) 中, 所述收集整理长非编码 RNA 数据, 构建长非编码 RNA 数据集的方法包括如下步骤:

A、收集长非编码 RNA 的注释数据, 得到长非编码 RNA 数据集;

B、过滤所述长非编码RNA数据集中没有实验证据支持、注释数据不一致的长非编码RNA,得到过滤后的长非编码RNA数据集;

C、去除所述过滤后的长非编码RNA数据集中的冗余,得到所述长非编码RNA数据集。

6. 根据权利要求5所述的方法,其特征在于:所述步骤A中,所述注释数据包括名称、基因组定位和序列信息。

7. 根据权利要求5所述的方法,其特征在于:所述步骤C中,利用数据集间的交叉注释、序列相似性比对和基因组定位的方法去除所述过滤后的长非编码RNA数据集中的冗余,使序列相似度大于95%且基因组位置重叠度大于95%。

8. 根据权利要求1-7中任一所述的方法,其特征在于:所述步骤(2)中,所述注释数据包括基因组定位和基因型信息。

9. 根据权利要求1-8中任一所述的方法,其特征在于:所述步骤(3)中,所述注释数据包括名称、motif序列和motif序列的位置权重矩阵。

10. 根据权利要求1-9中任一所述的方法,其特征在于:所述阈值为0.05。

一种预测影响长非编码RNA生物学功能的SNP位点的方法

技术领域

[0001] 本发明涉及分子生物学、功能基因组学和生物信息学领域,更具体涉及一种预测影响长非编码RNA生物学功能的SNP位点的方法。

背景技术

[0002] 2001年人类基因组草图的发布是人类基因组计划成功的里程碑事件,开启了功能基因组研究的序幕。基因组中蕴含的大量基因和功能元件有待于进一步的识别和研究。2012年,DNA元件百科全书计划(ENCODE)公布了一批研究数据,数据显示人类基因上约75%的区域能够发生转录,而能够编码蛋白质的编码区只占人类基因组的不到3%,预示着人类基因组的转录产物中含有大量的未知的非编码RNA。早在1989年,文献中第一次报道了一种全新的非编码RNA——长非编码RNA。长非编码RNA(long noncoding RNA, lncRNA)是一类长度大于200个核苷酸并且不具有编码蛋白质能力的核糖核酸。由于受到当时技术条件的限制,长非编码RNA的发现和进展非常缓慢。随着高通量检测技术(例如基因芯片技术和高通量测序技术)的发展和进步,2007年文献首次报道长非编码RNA系统发现的工作,由此开始了长非编码RNA系统发现和研究的序幕。据GENCODE计划最新公布的数据,人类长非编码RNA的数目已达27,908条。根据长非编码RNA相对于附近蛋白编码基因所在的基因组位置关系,大致可以将其分为以下五类:外显子型lncRNA、内含子型lncRNA、反义型lncRNA、反向型lncRNA和基因间型lncRNA。

[0003] lncRNA的数量巨大超出了绝大多数科学家的预期,其生物学功能的研究逐渐成为生命科学研究的热点,经过近十年的研究积累,生物学家已经对lncRNA发挥生物学功能的作用机制有了初步的认识。首先,lncRNA可以作为诱饵通过与其他分子相结合,阻断其他分子与其靶标的结合,进而影响其他分子的原有功能。例如,lncRNA与mRNA竞争性结合miRNA,阻断了miRNA与mRNA的结合,进而间接影响了mRNA的表达。linc-RoR是一个长非编码RNA,能够跟胚胎干细胞分化相关核心转录因子Oct4、Sox2和Nanog竞争性结合miR-145,阻止miR-145对Oct4等基因的调控,linc-RoR与转录因子和miRNA共同构成了一个调控环路调控胚胎干细胞的状态;其次,lncRNA可以作为脚手架,将生物大分子关联起来形成复合物,例如长非编码RNA HOTAIR能够跟PRC2结合,同时又能跟LSD1/CoREST/REST复合物结合,PRC2具有组蛋白甲基转移酶活性,能够催化H3K27发生三甲基化,从而沉默基因的转录,而LSD1具有去甲基化酶的活性,能够催化H3K4去甲基化。HOTAIR作为脚手架分子将两种不同的染色质修饰复合物关联起来共同沉默基因的表达;第三,lncRNA作为向导,带领蛋白质复合物到特定的细胞位置行使功能。例如,受p53调控的长非编码RNA lincRNA-p21.lincRNA-p21能够抑制p53依赖的转录应答.lincRNA-p21能够跟hnRNP-K相互作用,进而把hnRNP-K引导到特定的基因组位置去抑制基因的表达。

[0004] 人们对lncRNA的作用机制有了初步认识的同时,也在研究lncRNA与疾病的关系。到目前为止,lncRNA与肿瘤、代谢疾病、神经退行性疾病、精神疾病、心血管疾病和自身免疫疾病的发生发展都有明确的报道,HOTAIR是从HOX基因位点转录出来的一个lncRNA,它与很

多种肿瘤密切相关。2010年,Howard Y.Chang实验室发现HOTAIR在乳腺癌的原发灶和转移灶中表达显著上升,因此,HOTAIR在肿瘤组织中的表达水平可以用来预测肿瘤的转移。在上皮肿瘤细胞中过表达HOTAIR,改变了PRC2的靶向目标,影响了H3K27的甲基化以及相应基因的表达,增强了肿瘤细胞侵袭和迁移能力。2011年,研究人员发现HOTAIR在结肠癌组织中的表达水平要高于癌旁组织,而且HOTAIR的高表达与结肠癌的肝转移显著相关。结合患者的临床资料,还发现HOTAIR表达水平高的患者预后较差。前列腺癌是另一个备受关注的疾病,在前列腺癌组织中发现了很多组织特异性的长非编码RNA,例如PCA3/DD3、PCGEM1、PCAT-1、PRNCR1等。此外,在神经母细胞瘤、乳腺癌、胶质瘤、结直肠癌、神经退行性疾病中也发现了相关的长非编码RNA,例如aHIF、ANRIL、Oct4-pg、PTENP1和BC200等。这些研究表明lncRNA在疾病的发生发展过程中起到了至关重要的作用。

[0005] 2001年,人类基因组草图发布,发布的基因组序列是单倍体序列,没有遗传变异的注释。为了发现不同个体基因组之间的差异和研究表型多样化产生的原因,人们展开了大量工作来发现个体之间遗传变异。2007年,人类第一个个体的完整基因组图谱发布,即克莱格·文特尔本人的基因组,随后詹姆斯·沃森的基因组图谱也发布出来。后来,一个中国汉族人和一个非洲的尼日利亚人的基因组序列也陆续发表出来。随着二代测序技术的出现并不断成熟,人类基因组测序所需的时间和成本都显著减少。2012年,千人基因组计划完成了1,092个个体的基因组测序工作,其中包含97个来自中国北京的汉族人。到目前为止,通过比较不同个体的基因组序列,在人类基因组上发现的SNP位点多达2500万个。SNP是一种广泛存在的基因组变异方式。SNP是指某个人群的正常个体中,在基因组DNA的单个碱基对位置上存在不同的碱基的情况。通常认为SNP的出现是由基因组DNA的突变引起的。人的基因组DNA有30亿个碱基对,在同一个位置发生两次甚至三次突变的可能性极低,因此,SNP位点通常都是二态(有两个等位)。人类基因组DNA中平均每67个碱基就有1个SNP位点。但是这些SNP位点在基因组上并不是均匀分布,编码蛋白质的外显子区的SNP位点密度比其他区域低。不在蛋白质编码区的SNP虽然不会改变蛋白质的序列,但是可能通过其他的方式对基因产生影响。

[0006] 基因组上的长非编码基因区存在大量的SNP,这些SNP可能通过改变长非编码RNA的序列、结构和表达水平进而影响长非编码RNA的生物学功能。很多长非编码基因的转录受到转录因子的调控,位于启动子区的SNP位点可能影响转录因子的结合,从而影响长非编码基因的转录。例如,SNP位点rs944289是甲状腺癌的一个易感位点,它位于一个长非编码基因PTCSC3的启动子区的CEBPA和CEBPB的转录因子结合位点处。该位点碱基发生突变后影响了转录因子的结合,进而影响了PTCSC3的转录。染色体空间相互作用研究技术的不断进步,使得研究人员能够获取更加精细的染色体空间相互作用图谱。基因组上的增强子在空间上与基因的启动子区有相互作用,预示着该增强子对基因可能存在远程调控。远程调控元件中的SNP位点可能通过影响空间相互作用进而对基因的表达调控产生影响。RNA分子上存在的修饰目前已经超过100种,这些修饰可能影响RNA分子的活性、细胞定位和稳定性,进而可能与疾病相关联。m6A是RNA分子上分布最广泛的修饰。以往的研究发现,RNA分子上发生m6A修饰的区域含有一段保守序列RRACH。近期发表的m6A-seq数据也证实了这一发现。m6A修饰的生物学功能有待于进一步的研究。已有研究报道发现有些蛋白质能够选择性的结合含有m6A修饰的RNA分子,进而影响RNA的稳定性和生命周期。

[0007] 基因组上绝大部分区域是非编码区,非编码区包含大量的SNP位点,这些位点可能对非编码基因的表达和非编码RNA的序列及结构产生影响。SNP位点只涉及到一个碱基的改变,目前还无法衡量一个碱基改变后所产生的影响。

发明内容

[0008] 本发明的目的在于提供一种预测对长非编码RNA产生重要影响的SNP位点的方法,为长非编码RNA的生物学功能研究和临床应用提供指导。

[0009] 为了解决上述技术问题,本发明提供了一种预测影响与RNA结合蛋白结合的长非编码RNA的功能的SNP位点的方法。

[0010] 本发明提供的预测影响与RNA结合蛋白结合的长非编码RNA的功能的SNP位点的方法包括如下步骤:

[0011] (1) 收集整理长非编码RNA数据,构建长非编码RNA数据集;

[0012] (2) 收集人类基因组SNP位点的注释数据,通过比较SNP位点和长非编码基因的基因组定位,识别位于长非编码基因区的SNP位点;将所述长非编码RNA数据集中的长非编码RNA序列上的所述SNP位点对应的碱基替换为突变后的碱基,构建得到SNP位点碱基突变后的长非编码RNA数据集;

[0013] (3) 收集RNA结合蛋白的注释数据,构建RNA结合蛋白的motif数据集;

[0014] 所述motif数据集中包括motif序列,所述motif序列为RNA结合蛋白能够特异识别的RNA序列;

[0015] (4) 基于极值分布对所述motif序列与所述长非编码RNA数据集中的每个长非编码RNA上的目标序列的相似程度进行评价,得到每个长非编码RNA打分最高的目标序列与所述motif序列相似程度的显著性水平 p ,并选择显著性水平 p 低于阈值的长非编码RNA作为SNP位点碱基突变前的所述RNA结合蛋白的靶标长非编码RNA;

[0016] 所述目标序列是指位于所述长非编码RNA序列上且与所述motif序列长度相等的RNA序列;

[0017] 假设某长非编码RNA序列长度为 N ;所述motif序列长度为 L ,那么在该长非编码RNA序列上就会得出 $(N-L+1)$ 个目标序列;

[0018] (5) 基于极值分布对所述motif序列与所述SNP位点碱基突变后的长非编码RNA数据集中的长非编码RNA上的目标序列的相似程度进行评价,得到每个长非编码RNA打分最高的目标序列与所述motif序列相似程度的显著性水平 p ,并选择相似程度的显著性水平低于阈值的长非编码RNA作为SNP位点碱基突变后的所述RNA结合蛋白的靶标长非编码RNA;

[0019] (6) 比较所述SNP位点碱基突变前的所述RNA结合蛋白的靶标长非编码RNA和所述SNP位点碱基突变后的所述RNA结合蛋白的靶标长非编码RNA,得到所述影响与RNA结合蛋白结合的长非编码RNA的功能的SNP位点。

[0020] 上述方法中,所述步骤(1)中,所述收集整理长非编码RNA数据,构建长非编码RNA数据集的方法包括如下步骤:

[0021] A、收集长非编码RNA的注释数据,得到长非编码RNA数据集;

[0022] B、过滤所述长非编码RNA数据集中没有实验证据支持、注释数据不一致的长非编码RNA,得到过滤后的长非编码RNA数据集;

[0023] C、去除所述过滤后的长非编码RNA数据集中的冗余,得到所述长非编码RNA数据集。

[0024] 所述步骤A中,从公开发表的文献“Cabili等,Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.2011.Genes Dev 25:1915-27”和GENCODE公共数据库(公开GENCODE公共数据库的文献如下:GENCODE.Harrow等,GENCODE:the reference human genome annotation for The ENCODE Project.Genome Research.2012.22:1760-74.,GENCODE公共数据库的查询网址如下:<http://www.encodegenes.org/>)中收集长非编码RNA的注释数据;所述注释数据包括名称、基因组定位和序列信息。

[0025] 所述步骤C中,利用数据集间的交叉注释、序列相似性比对和基因组定位的方法去除所述过滤后的长非编码RNA数据集中的冗余,使序列相似度大于95%且基因组位置重叠度大于95%。

[0026] 上述方法中,所述步骤(2)中,从dbSNP公共数据库(<https://www.ncbi.nlm.nih.gov/projects/SNP/>)中收集人类基因组SNP位点注释数据,通过比较SNP位点和长非编码基因的基因组定位,识别位于长非编码基因区的SNP位点;所述注释数据包括基因组定位和基因型信息。

[0027] 上述方法中,所述步骤(3)中,从公开发表的文献“Ray等,A compendium of RNA-binding motifs for decoding gene regulation.Nature.499:172-177.”中获取人类RNA结合蛋白的注释数据;所述注释数据包括名称、motif序列和motif序列的位置权重矩阵。

[0028] 上述方法中,通常选取0.05作为阈值。阈值选取的越小(即越严格),预测的结果越少。本发明中阈值具体设定为0.05。

[0029] 上述方法中,所述步骤(4)和步骤(5)中,以所述长非编码RNA数据集中的任意一个长非编码RNA为例,将其记作长非编码RNA甲,对所述motif序列与所述长非编码RNA甲上的目标序列的相似程度进行评价,得到所述长非编码RNA甲打分最高的目标序列与所述motif序列相似程度的显著性水平p的方法包括如下步骤:

[0030] 1) 针对长非编码RNA甲,构建10000条与所述长非编码RNA甲序列长度相同的随机序列;所述长非编码RNA甲序列长度为N;所述motif序列长度为L;所述N大于等于L;

[0031] 2) 针对每一条随机序列,分别计算(N-L+1)个目标序列与所述motif序列的相似度打分,分别得到(N-L+1)个目标序列与所述motif序列的相似度打分,将最高的相似度打分记作 S_{max} ;

[0032] 所述目标序列是指位于所述随机序列上且与所述motif序列长度相等的RNA序列;

[0033] 所述目标序列与所述motif序列的相似度打分S的计算公式如下:

$$[0034] \quad S = \sum_{j=1}^L w_j^*$$

[0035] 其中, w_j^* 是目标序列第j个位置上的碱基与motif序列的相似度打分;

[0036] 所述目标序列第j个位置上的碱基与motif序列的相似度打分的计算方法如下:对于序列长度为L的目标序列,位于第j个位置的碱基i与motif序列的相似度打分 $w_{i,j}$ 的计算公式如下:

$$[0037] \quad w_{i,j} = \log\left(\frac{f_{i,j}}{p_i}\right),$$

[0038] 其中, i 为 A、G、C、U; j 为 1、2、3、……, L ; $f_{i,j}$ 是 motif 序列第 j 个位置的碱基 i 出现的频率, p_i 是碱基 i 在人类基因组中出现的频率;

[0039] 3) 同理, 按照步骤 2) 中的方法, 计算其他随机序列的最高的相似度打分, 共得到 10000 个最高的相似度打分, 并计算所述 10000 个最高的相似度打分的平均值, 将所述平均值记作 \bar{S}_{\max} ;

[0040] 4) 按照如下公式估计极值分布的参数 μ 和 β :

$$[0041] \quad \mu = \bar{S}_{\max} - \gamma\beta,$$

$$[0042] \quad \beta = \frac{\sigma_{\bar{S}_{\max}} \sqrt{6}}{\pi},$$

[0043] 其中, $\sigma_{\bar{S}_{\max}}$ 为 \bar{S}_{\max} 的标准差; γ 为欧拉-马歇罗尼常数;

[0044] 5) 基于极值分布, 按照如下公式计算长非编码 RNA 甲打分最高的目标序列与所述 motif 序列相似程度的显著性水平 p :

$$[0045] \quad p = 1 - e^{-e^{-\frac{S-\mu}{\beta}}}$$

[0046] 其中, S 为长非编码 RNA 甲所对应的所有目标序列中与 motif 序列的相似度打分最高者的打分;

[0047] 同理, 得到所述长非编码 RNA 数据集中的其他长非编码 RNA 打分最高的目标序列与 motif 序列相似程度的显著性水平 p 。

[0048] 上述方法中, 所述步骤 (6) 中, “比较所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 和所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA, 得到所述影响 RNA 结合蛋白结合位点的 SNP 位点”的方法如下:

[0049] 属于所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 但不属于所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA 的长非编码 RNA 上的 SNP 位点即为对所述 RNA 结合蛋白结合位点有显著影响的 SNP 位点;

[0050] 不属于所述 SNP 位点碱基突变前的所述 RNA 结合蛋白的靶标长非编码 RNA 但属于所述 SNP 位点碱基突变后的所述 RNA 结合蛋白的靶标长非编码 RNA 的长非编码 RNA 上的 SNP 位点即为所述 RNA 结合蛋白能够产生新的结合位点的 SNP 位点。

[0051] 在本发明的具体实施例中, 所述 RNA 结合蛋白为人类 RNA 结合蛋白, 具体为 TARDBP, 按照上述方法预测到的 SNP 位点碱基突变前的 RNA 结合蛋白 TARDBP 的靶标长非编码 RNA 具体如表 1 所示; 预测到的 SNP 位点碱基突变后的 RNA 结合蛋白 TARDBP 的靶标长非编码 RNA 具体如表 2 所示。对 RNA 结合蛋白 TARDBP 的结合位点有显著影响的 SNP 位点具体如表 4 所示, RNA 结合蛋白 TARDBP 能够产生新的结合位点的 SNP 位点具体如表 3 所示。这些 SNP 位点对长非编码 RNA 的生物学功能可能产生重要的影响。

[0052] 长非编码 RNA 发挥生物学功能通常需要借助特定的碱基序列, 例如长非编码 RNA 所包含的 RNA 结合蛋白能够识别的特异的 motif 序列。本发明提供了预测影响 RNA 结合蛋白结合位点的 SNP 位点的方法, 这些 SNP 位点更有可能对长非编码 RNA 生物学功能产生重大影响。本发明提供的方法首先建立了长非编码基因相关的特定序列的识别及显著性评价方法, 然

后基于此方法对SNP位点所产生的影响进行评价。利用此方法可以发现对长非编码RNA生物学功能产生重要影响的SNP位点,为实验人员提供可靠的功能研究线索。

具体实施方式

[0053] 为了使发明的技术方案和优点更加清楚明白,以下结合实施例对本发明进行进一步说明。此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0054] 下述实施例中的Motif序列是指序列(氨基酸序列、DNA序列或RNA序列)中局部的保守区域,或者是一组序列中共有的一小段序列模式。一般情况下是指有可能具有分子功能、结构性质或家族成员相关的任何序列模式。

[0055] 实施例1、对长非编码RNA产生重要影响的SNP位点的获得方法

[0056] 一、构建长非编码RNA完备数据集

[0057] 1、收集整理长非编码RNA数据

[0058] 从公开发表的文献“Cabili等, Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. 2011. Genes Dev 25:1915-27”和GENCODE公共数据库(公开GENCODE公共数据库的文献如下: GENCODE. Harrow等, GENCODE: the reference human genome annotation for The ENCODE Project. Genome Research. 2012. 22:1760-74., GENCODE公共数据库的查询网址如下: <http://www.encodegenes.org/>)中收集长非编码RNA的注释数据,包括名称、基因组定位、序列等信息,得到长非编码RNA数据集。

[0059] 2、对长非编码RNA数据集进行过滤

[0060] 从长非编码RNA数据集中过滤掉没有实验证据支持、注释数据不一致的长非编码RNA,得到过滤后的长非编码RNA数据集。

[0061] 3、去除长非编码RNA数据集之间的冗余序列

[0062] 利用数据集间的交叉注释、序列相似性比对和基因组定位的方法去除数据集之间的冗余,使序列相似度大于95%且基因组位置重叠度大于95%,得到去除冗余后的数据。

[0063] 4、将去除冗余后的数据汇总,构建得到完备的长非编码RNA数据集。该数据集包括明确的长非编码RNA名称、基因组定位和序列信息。

[0064] 二、构建SNP位点碱基突变后的长非编码RNA数据集

[0065] 1、从dbSNP公共数据库(<https://www.ncbi.nlm.nih.gov/projects/SNP/>)获取人类基因组SNP位点注释数据,包括基因组定位、基因型等信息。

[0066] 2、通过比较SNP位点和长非编码基因的基因组定位,识别位于长非编码基因区的SNP位点。

[0067] 3、在长非编码RNA序列上,将SNP位点处的碱基替换成突变后的碱基,构建得到SNP位点碱基突变后的长非编码RNA数据集。

[0068] 三、构建RNA结合蛋白的motif数据集

[0069] 从公开发表的文献“Ray等, A compendium of RNA-binding motifs for decoding gene regulation. Nature. 499:172-177.”中,获取人类RNA结合蛋白的注释数据,包括名称、motif序列、motif序列的位置权重矩阵,构建得到RNA结合蛋白的motif数据集。例如,对于RNA结合蛋白TARDBP共有1个motif序列。

[0070] 四、构建motif序列与目标序列相似度打分系统

[0071] 1、定义碱基相似度打分公式

[0072] 如果motif序列的长度为L,那么对于长度为L的目标序列(目标序列的长度要大于等于motif序列的长度),位于第j个位置的碱基i与motif序列的相似度打分 $w_{i,j}$ 的计算公式为:

$$[0073] \quad w_{i,j} = \log\left(\frac{f_{i,j}}{p_i}\right),$$

[0074] 其中,i为A、G、C、U;j为1、2、3、……,L; $f_{i,j}$ 是motif序列第j个位置的碱基i出现的频率,来自motif序列的位置权重矩阵,motif序列的位置权重矩阵包含了每个位置上所有四种碱基的频率,可从步骤三中公开发表的文献中获取; p_i 是碱基i在人类基因组中出现的频率,具体计算方法如下:首先数一下人类基因组总共有多少个碱基,然后再数一下四种碱基分别有多少个,用四种碱基的个数分别除以人类基因组碱基总数,即分别得到四种碱基在人类基因组中出现的频率。

[0075] 2、定义motif序列与目标序列的相似度打分公式

[0076] Motif序列与目标序列的相似度打分LOD scoreS的计算公式为:

$$[0077] \quad S = \sum_{j=1}^L w_j^*$$

[0078] 其中, w_j^* 是目标序列第j个位置上的碱基与motif的相似度打分。

[0079] 五、基于极值分布构建RNA结合蛋白的motif序列与长非编码RNA上的目标序列相似程度评价体系

[0080] 1、对于motif序列长度为L的RNA结合蛋白,可以在序列长度为N的长非编码RNA上计算(N-L+1)个目标序列与motif序列的相似度打分,其中最高打分的分布近似于Gumbel分布。例如,序列长度为5的长非编码RNA(如ATCGC)上,可计算3个目标序列为:ATC、TCG和CGC与长度为3的motif序列的相似度打分。

[0081] 2、针对步骤一中的完备的长非编码RNA数据集中每条长非编码RNA,分别构建10000条与之序列长度相同的随机序列。以长非编码RNA数据集中的任一条长非编码RNA为例,将其记作长非编码RNA甲,构建10000条与长非编码RNA甲序列长度相同的随机序列。对于每一条长度为N随机序列,按照步骤1中的方法可以得到N-L+1个目标序列与motif序列的相似度打分, S_{max} 为最高打分,可以视为极值分布的一次抽样。

[0082] 3、对所有的随机序列按照步骤2的方法可以得到10000个最高打分,并计算10000个最高打分的平均值,将其记作 \bar{S}_{max} 。

[0083] 4、估计极值分布的参数 μ 和 β :

$$[0084] \quad \mu = \bar{S}_{max} - \gamma\beta$$

$$[0085] \quad \beta = \frac{\sigma_{S_{max}} \sqrt{6}}{\pi}$$

[0086] 其中, $\sigma_{S_{max}}$ 为 S_{max} 的标准差, γ 为欧拉-马歇罗尼常数。

[0087] 5、基于极值分布,按照如下公式计算长非编码RNA甲打分最高的目标序列与motif序列相似程度的显著性水平p:

[0088]
$$p = 1 - e^{-\frac{S}{M}}$$

[0089] 其中,S为长非编码RNA甲所有目标序列中与motif序列相似度打分最高者的打分。

[0090] 6、按照上述步骤1-5中的方法,计算长非编码RNA数据集中的其他长非编码RNA打分最高的目标序列与motif序列相似程度的显著性水平p。

[0091] 7、设定阈值,筛选相似程度的显著性水平低于阈值的长非编码RNA作为RNA结合蛋白的靶标。通常选取0.05作为阈值,阈值选取的越小(即越严格),预测的结果越少。本发明中阈值具体设定为0.05。

[0092] 以RNA结合蛋白TARDBP为例,基于上述步骤1-7中的方法,在长非编码RNA数据集(SNP位点碱基突变前的长非编码RNA)中筛选得到RNA结合蛋白TARDBP的靶标长非编码RNA,具体预测结果如表1所示。

[0093] 表1、SNP位点碱基突变前TARDBP靶标长非编码RNA预测结果

TCONS_00009382	TCONS_00025321	TCONS_00026047	TCONS_00000707	TCONS_00008663
TCONS_00013656	TCONS_00005989	TCONS_00012990	TCONS_00001764	TCONS_00025436
TCONS_00013107	TCONS_00005990	TCONS_00014754	TCONS_00029230	TCONS_00000838
TCONS_00020447	TCONS_00001577	TCONS_00015298	TCONS_00018161	TCONS_00015945
TCONS_00008252	TCONS_00003580	TCONS_00003262	TCONS_00010335	TCONS_00015167

[0094]

[0095]

TCONS 00021085	TCONS 00004702	TCONS 00024773	TCONS 00003931	TCONS 00014225
TCONS 00022101	TCONS 00009742	TCONS 00025088	TCONS 00001806	TCONS 00004694
TCONS 00008606	TCONS 00007987	TCONS 00024280	TCONS 00002579	TCONS 00001081
TCONS 00025013	TCONS 00007988	TCONS 00008498	TCONS 00022118	TCONS 00018669
TCONS 00017954	TCONS 00008755	TCONS 00010392	TCONS 00024453	TCONS 00018670
TCONS 00009640	TCONS 00006643	TCONS 00023815	TCONS 00021437	TCONS 00018671
TCONS 00020679	TCONS 00002017	TCONS 00019083	TCONS 00021439	TCONS 00017773
TCONS 00012603	TCONS 00008433	TCONS 00020870	TCONS 00021440	TCONS 00020374
TCONS 00012604	TCONS 00029433	TCONS 00008217	TCONS 00021438	TCONS 00028321
TCONS 00021569	TCONS 00010493	TCONS 00020885	TCONS 00021442	TCONS 00001045
TCONS 00008469	TCONS 00001117	TCONS 00018619	TCONS 00020976	TCONS 00012478
TCONS 00002206	TCONS 00020574	TCONS 00022244	TCONS 00020977	TCONS 00012480
TCONS 00024213	TCONS 00029566	TCONS 00022247	TCONS 00028985	TCONS 00012479
TCONS 00023781	TCONS 00029784	TCONS 00005729	TCONS 00016265	TCONS 00011295
TCONS 00023782	TCONS 00007374	TCONS 00000480	TCONS 00012946	TCONS 00006776
TCONS 00015975	TCONS 00008672	TCONS 00006373	TCONS 00012947	TCONS 00021897
TCONS 00019165	TCONS 00018430	TCONS 00006455	TCONS 00005850	TCONS 00021594
TCONS 00004275	TCONS 00017911	TCONS 00007129	TCONS 00003837	TCONS 00017629
TCONS 00028963	TCONS 00016664	TCONS 00026337	TCONS 00021224	TCONS 00005674
TCONS 00000740	TCONS 00026504	TCONS 00013157	TCONS 00017301	TCONS 00003907
TCONS 00024758	TCONS 00009345	TCONS 00005191	TCONS 00009403	TCONS 00027712
TCONS 00003973	TCONS 00005806	TCONS 00004595	TCONS 00015389	TCONS 00017649
TCONS 00018335	TCONS 00005807	TCONS 00003869	TCONS 00015390	TCONS 00009474
TCONS 00004738	TCONS 00007316	TCONS 00022772	TCONS 00014901	TCONS 00014353
TCONS 00002691	TCONS 00012352	TCONS 00025195	TCONS 00017870	TCONS 00013888
TCONS 00003621	TCONS 00005650	TCONS 00019423	TCONS 00021641	TCONS 00002410
TCONS 00013502	TCONS 00005651	TCONS 00012548	TCONS 00009802	TCONS 00001741
TCONS 00009394	TCONS 00017446	TCONS 00011893	TCONS 00015362	TCONS 00004332
TCONS 00029564	TCONS 00014957	TCONS 00011894	TCONS 00015361	TCONS 00005605
TCONS 00023757	TCONS 00028890	TCONS 00023723	TCONS 00022867	TCONS 00023821
TCONS 00022971	TCONS 00025573	TCONS 00025541	TCONS 00026571	TCONS 00013975
TCONS 00028304	TCONS 00000198	TCONS 00020399	TCONS 00025556	TCONS 00010090
TCONS 00022069	TCONS 00003410	TCONS 00021114	TCONS 00003063	TCONS 00019764
TCONS 00020708	TCONS 00006096	TCONS 00022636	TCONS 00003960	TCONS 00028640
TCONS 00000574	TCONS 00011408	TCONS 00009501	TCONS 00003518	TCONS 00002888
TCONS 00017765	TCONS 00009520	TCONS 00014790	TCONS 00007832	TCONS 00002889
TCONS 00011363	TCONS 00012236	TCONS 00009611	TCONS 00002951	TCONS 00016916
TCONS 00019142	TCONS 00012764	TCONS 00020694	TCONS 00006134	TCONS 00016915
TCONS 00009499	TCONS 00006447	TCONS 00016859	TCONS 00001093	TCONS 00016957
TCONS 00016928	TCONS 00016076	TCONS 00009532	TCONS 00015773	TCONS 00001880
TCONS 00017550	TCONS 00008301	TCONS 00007471	TCONS 00023787	TCONS 00028695
TCONS 00022567	TCONS 00003965	TCONS 00028097	TCONS 00023786	TCONS 00024839
TCONS 00001322	TCONS 00024739	TCONS 00028098	TCONS 00011323	TCONS 00024387
TCONS 00018842	TCONS 00017346	TCONS 00000742	TCONS 00008588	TCONS 00006527
TCONS 00000694	TCONS 00018527	TCONS 00026906	TCONS 00003759	TCONS 00013752
TCONS 00005628	TCONS 00009069	TCONS 00003929	TCONS 00006568	TCONS 00013634
TCONS 00005861	TCONS 00008501	TCONS 00006934	TCONS 00003235	TCONS 00013635
TCONS 00005862	TCONS 00015478	TCONS 00006936	TCONS 00017566	TCONS 00016827
TCONS 00001891	TCONS 00030008	TCONS 00006935	TCONS 00017565	TCONS 00019277
TCONS 00020490	TCONS 00029403	TCONS 00006684	TCONS 00017599	TCONS 00017923
TCONS 00025311	TCONS 00026420	TCONS 00017043	TCONS 00003152	TCONS 00004042
TCONS 00025312	TCONS 00020691	TCONS 00016317	TCONS 00015796	TCONS 00019871
TCONS 00000952	TCONS 00028860	TCONS 00011543	TCONS 00017107	TCONS 00019259
TCONS 00003680	TCONS 00018124	TCONS 00016140	TCONS 00016200	TCONS 00004843
TCONS 00026626	TCONS 00020433	TCONS 00026788	TCONS 00000435	TCONS 00016862
TCONS 00019544	TCONS 00021137	TCONS 00026396	TCONS 00013119	TCONS 00013695
TCONS 00000746	TCONS 00021138	TCONS 00015377	TCONS 00028805	TCONS 00017171
TCONS 00003582	TCONS 00020355	TCONS 00014937	TCONS 00028764	TCONS 00003348
TCONS 00013154	TCONS 00019616	TCONS 00015355	TCONS 00019642	TCONS 00028848
TCONS 00008088	TCONS 00023891	TCONS 00015356	TCONS 00020721	TCONS 00018162
TCONS 00008437	TCONS 00007514	TCONS 00015358	TCONS 00023507	TCONS 00000348
TCONS 00017140	TCONS 00009176	TCONS 00005765	TCONS 00020541	TCONS 00025588
TCONS 00016965	TCONS 00015205	TCONS 00015757	TCONS 00018458	TCONS 00028435
TCONS 00003592	TCONS 00024428	TCONS 00009829	TCONS 00013606	TCONS 00001897
TCONS 00014799	TCONS 00006576	TCONS 00019441	TCONS 00003553	TCONS 00008508
TCONS 00018094	TCONS 00003683	TCONS 00020811	TCONS 00001874	TCONS 00022239

TCONS_00026497	TCONS_00017294	TCONS_00020209	TCONS_00019591	TCONS_00024693
TCONS_00008729	TCONS_00012531	TCONS_00020812	TCONS_00021013	TCONS_00025610
TCONS_00021551	TCONS_00018648	TCONS_00019890	TCONS_00021014	TCONS_00022500
TCONS_00017449	TCONS_00020984	TCONS_00000856	TCONS_00021461	TCONS_00013564
TCONS_00014341	TCONS_00029834	TCONS_00011686	TCONS_00019159	TCONS_00008412
TCONS_00021599	TCONS_00006528	TCONS_00026373	TCONS_00011399	TCONS_00020867
TCONS_00021600	TCONS_00008447	TCONS_00004898	TCONS_00001622	TCONS_00012313
TCONS_00015427	TCONS_00029371	TCONS_00011850	TCONS_00029795	TCONS_00012314
TCONS_00021584	TCONS_00016139	TCONS_00019652	TCONS_00016160	TCONS_00012315
TCONS_00028615	TCONS_00016138	TCONS_00020434	TCONS_00005117	TCONS_00008984
TCONS_00005508	TCONS_00015233	TCONS_00011935	TCONS_00009331	TCONS_00012778
TCONS_00011831	TCONS_00014628	TCONS_00020763	TCONS_00009851	TCONS_00012777
TCONS_00017369	TCONS_00013376	TCONS_00018492	TCONS_00011715	TCONS_00028118
TCONS_00006904	TCONS_00000216	TCONS_00026416	TCONS_00018424	TCONS_00023963
TCONS_00006903	TCONS_00007918	TCONS_00021674	TCONS_00007517	TCONS_00021344
TCONS_00006902	TCONS_00009622	TCONS_00010401	TCONS_00002918	TCONS_00008306
TCONS_00021854	TCONS_00010283	TCONS_00010757	TCONS_00002919	TCONS_00016663
TCONS_00007855	TCONS_00014781	TCONS_00004038	TCONS_00012993	TCONS_00016114
TCONS_00005519	TCONS_00003082	TCONS_00013677	TCONS_00011938	TCONS_00020823
TCONS_00003424	TCONS_00007980	TCONS_00012898	TCONS_00007627	TCONS_00023972
TCONS_00019123	TCONS_00011615	TCONS_00021409	TCONS_00026485	TCONS_00028469
TCONS_00029084	TCONS_00007042	TCONS_00006097	TCONS_00028560	TCONS_00004032
TCONS_00009795	TCONS_00024119	TCONS_00012436	TCONS_00001459	TCONS_00028763
TCONS_00022623	TCONS_00000524	TCONS_00011734	TCONS_00026919	TCONS_00029315
TCONS_00001557	TCONS_00006016	TCONS_00013521	TCONS_00027491	TCONS_00008048
TCONS_00001319	TCONS_00013270	TCONS_00013522	TCONS_00015792	TCONS_00024863
TCONS_00008289	TCONS_00029197	TCONS_00013523	TCONS_00004809	TCONS_00004745
TCONS_00017415	TCONS_00019680	TCONS_00014171	TCONS_00005563	TCONS_00012088
TCONS_00016207	TCONS_00020590	TCONS_00011700	TCONS_00023520	TCONS_00020455
TCONS_00012739	TCONS_00028340	TCONS_00008341	TCONS_00001791	TCONS_00021089
TCONS_00003339	TCONS_00011354	TCONS_00017290	TCONS_00005551	TCONS_00000347
TCONS_00003212	TCONS_00002077	TCONS_00017291	TCONS_00028213	TCONS_00025984
TCONS_00007931	TCONS_00002078	TCONS_00004223	TCONS_00018898	TCONS_00010055
TCONS_00011306	TCONS_00002079	TCONS_00003664	TCONS_00028086	TCONS_00005952
TCONS_00015070	TCONS_00006223	TCONS_00026124	TCONS_00012208	TCONS_00015204
TCONS_00018398	TCONS_00015879	TCONS_00026125	TCONS_00013329	TCONS_00009877
TCONS_00007731	TCONS_00023293	TCONS_00001454	TCONS_00020864	TCONS_00027178
TCONS_00007733	TCONS_00011379	TCONS_00017632	TCONS_00020865	TCONS_00003414
TCONS_00008453	TCONS_00022349	TCONS_00020388	TCONS_00013294	TCONS_00014643
TCONS_00029515	TCONS_00007657	TCONS_00020390	TCONS_00006085	TCONS_00015248
TCONS_00000674	TCONS_00005755	TCONS_00020389	TCONS_00012022	TCONS_00023451
TCONS_00006112	TCONS_00019479	TCONS_00020392	TCONS_00013203	TCONS_00010247
TCONS_00006113	TCONS_00019412	TCONS_00020394	TCONS_00012855	TCONS_00005851
TCONS_00006900	TCONS_00004557	TCONS_00009830	TCONS_00012917	TCONS_00007791
TCONS_00006707	TCONS_00029558	TCONS_00010598	TCONS_00010983	TCONS_00011421
TCONS_00014916	TCONS_00001508	TCONS_00010599	TCONS_00016264	TCONS_00020566
TCONS_00014917	TCONS_00018031	TCONS_00008015	TCONS_00029004	TCONS_00020028
TCONS_00004665	TCONS_00025533	TCONS_00005098	TCONS_00003831	TCONS_00019549
TCONS_00004666	TCONS_00019634	TCONS_00005099	TCONS_00003832	TCONS_00021206
TCONS_00004664	TCONS_00013602	TCONS_00004166	TCONS_00008286	TCONS_00021207
TCONS_00019741	TCONS_00021915	TCONS_00005103	TCONS_00007654	TCONS_00028115
TCONS_00014733	TCONS_00012988	TCONS_00005102	TCONS_00009060	TCONS_00013142
TCONS_00008093	TCONS_00013406	TCONS_00004167	TCONS_00015943	TCONS_00024755
TCONS_00029835	TCONS_00013189	TCONS_00022458	TCONS_00024147	TCONS_00004196
TCONS_00018764	TCONS_00006753	TCONS_00024153	TCONS_00023714	TCONS_00006330
TCONS_00017854	TCONS_00023689	TCONS_00000192	TCONS_00021259	TCONS_00010468
TCONS_00008550	TCONS_00025520	TCONS_00007242	TCONS_00020384	TCONS_00014835
TCONS_00005931	TCONS_00028873	TCONS_00016209	TCONS_00013699	TCONS_00017357
TCONS_00023231	TCONS_00026907	TCONS_00016210	TCONS_00003179	TCONS_00011887
TCONS_00009123	TCONS_00017313	TCONS_00016211	TCONS_00003180	TCONS_00001269
TCONS_00006482	TCONS_00024432	TCONS_00002202	TCONS_00003181	TCONS_00019939
TCONS_00006483	TCONS_00010053	TCONS_00007047	TCONS_00013402	TCONS_00005222
TCONS_00014838	TCONS_00004943	TCONS_00002580	TCONS_00003998	TCONS_00001536
TCONS_00001535	TCONS_00024667	TCONS_00025511	TCONS_00004985	TCONS_00019639
TCONS_00017074	TCONS_00024668	TCONS_00005654	TCONS_00015007	TCONS_00003897
TCONS_00009070	TCONS_00005553			

[0096]

[0097] 六、基于极值分布构建RNA结合蛋白的motif序列与SNP位点碱基突变后的长非编

码RNA上的目标序列相似程度评价体系

[0098] 以RNA结合蛋白TARDBP为例,按照步骤四和步骤五建立的方法,在SNP位点碱基突变后的长非编码RNA数据集中筛选得到RNA结合蛋白TARDBP的靶标长非编码RNA,具体预测结果如表2所示。

[0099] 表2、SNP位点碱基突变后TARDBP靶标长非编码RNA预测结果

TCONS_00009382	TCONS_00006483	TCONS_00009340	TCONS_00006544	TCONS_00016265
TCONS_00013656	TCONS_00014838	TCONS_00020590	TCONS_00008276	TCONS_00012946
TCONS_00013107	TCONS_00014726	TCONS_00028340	TCONS_00016317	TCONS_00012947
TCONS_00020447	TCONS_00001535	TCONS_00011354	TCONS_00011543	TCONS_00005850
TCONS_00008252	TCONS_00008663	TCONS_00002077	TCONS_00016140	TCONS_00021224
TCONS_00021085	TCONS_00025436	TCONS_00002078	TCONS_00026788	TCONS_00017301
TCONS_00005684	TCONS_00023960	TCONS_00002079	TCONS_00026396	TCONS_00009403
TCONS_00001326	TCONS_00014225	TCONS_00024338	TCONS_00015355	TCONS_00015389
TCONS_00024429	TCONS_00001081	TCONS_00015879	TCONS_00015356	TCONS_00015390
TCONS_00017954	TCONS_00018669	TCONS_00023293	TCONS_00015358	TCONS_00014901
TCONS_00020679	TCONS_00018670	TCONS_00007657	TCONS_00005765	TCONS_00017870
TCONS_00012603	TCONS_00018671	TCONS_00005755	TCONS_00015757	TCONS_00021641
TCONS_00012604	TCONS_00017773	TCONS_00019479	TCONS_00009829	TCONS_00009802
TCONS_00021569	TCONS_00005009	TCONS_00019412	TCONS_00019441	TCONS_00015362
TCONS_00008469	TCONS_00020374	TCONS_00006207	TCONS_00020811	TCONS_00015361
TCONS_00002206	TCONS_00028321	TCONS_00004557	TCONS_00020209	TCONS_00006187
TCONS_00020621	TCONS_00012478	TCONS_00029552	TCONS_00020812	TCONS_00022867
TCONS_00020622	TCONS_00012480	TCONS_00001508	TCONS_00025344	TCONS_00026571
TCONS_00023781	TCONS_00012479	TCONS_00025533	TCONS_00025163	TCONS_00004195
TCONS_00023782	TCONS_00011295	TCONS_00019634	TCONS_00000856	TCONS_00003063
TCONS_00009719	TCONS_00021897	TCONS_00013602	TCONS_00026553	TCONS_00003960
TCONS_00015975	TCONS_00021594	TCONS_00021915	TCONS_00026729	TCONS_00003518
TCONS_00019165	TCONS_00017629	TCONS_00012988	TCONS_00026554	TCONS_00007832
TCONS_00004279	TCONS_00027712	TCONS_00013406	TCONS_00022453	TCONS_00002951
TCONS_00025789	TCONS_00011812	TCONS_00006753	TCONS_00022454	TCONS_00005118
TCONS_00013779	TCONS_00017649	TCONS_00023689	TCONS_00022944	TCONS_00021450
TCONS_00023365	TCONS_00009474	TCONS_00015789	TCONS_00011686	TCONS_00006134
TCONS_00023368	TCONS_00013888	TCONS_00028873	TCONS_00026373	TCONS_00001093
TCONS_00028963	TCONS_00002410	TCONS_00026907	TCONS_00004898	TCONS_00015773
TCONS_00000740	TCONS_00022728	TCONS_00017313	TCONS_00011850	TCONS_00023787
TCONS_00012607	TCONS_00023100	TCONS_00024432	TCONS_00020434	TCONS_00023786
TCONS_00011381	TCONS_00001741	TCONS_00010053	TCONS_00011935	TCONS_00011323
TCONS_00024758	TCONS_00004332	TCONS_00004943	TCONS_00014999	TCONS_00008588
TCONS_00018147	TCONS_00005605	TCONS_00024667	TCONS_00020763	TCONS_00003759
TCONS_00003973	TCONS_00025321	TCONS_00024668	TCONS_00018492	TCONS_00006568
TCONS_00018335	TCONS_00005989	TCONS_00002265	TCONS_00026416	TCONS_00005115
TCONS_00004738	TCONS_00005990	TCONS_00002266	TCONS_00010211	TCONS_00003235
TCONS_00002691	TCONS_00007988	TCONS_00002264	TCONS_00010401	TCONS_00016015
TCONS_00003621	TCONS_00008755	TCONS_00002263	TCONS_00004038	TCONS_00017566
TCONS_00013502	TCONS_00006643	TCONS_00005553	TCONS_00013677	TCONS_00017565
TCONS_00009394	TCONS_00002017	TCONS_00023821	TCONS_00012898	TCONS_00017599
TCONS_00029564	TCONS_00008433	TCONS_00013975	TCONS_00005612	TCONS_00013652
TCONS_00023757	TCONS_00010493	TCONS_00010090	TCONS_00006097	TCONS_00003152
TCONS_00022971	TCONS_00020574	TCONS_00019764	TCONS_00012436	TCONS_00009761
TCONS_00028304	TCONS_00019783	TCONS_00002840	TCONS_00011734	TCONS_00017107
TCONS_00017898	TCONS_00007374	TCONS_00002888	TCONS_00013521	TCONS_00000435
TCONS_00022069	TCONS_00008672	TCONS_00002889	TCONS_00013522	TCONS_00013119
TCONS_00000574	TCONS_00018430	TCONS_00016916	TCONS_00013523	TCONS_00019642
TCONS_00017765	TCONS_00016664	TCONS_00016915	TCONS_00014171	TCONS_00023507
TCONS_00028034	TCONS_00026504	TCONS_00016957	TCONS_00011700	TCONS_00020541
TCONS_00009499	TCONS_00023915	TCONS_00001880	TCONS_00008341	TCONS_00018458
TCONS_00026638	TCONS_00007316	TCONS_00028695	TCONS_00017290	TCONS_00013606
TCONS_00016928	TCONS_00012352	TCONS_00024839	TCONS_00017291	TCONS_00027380
TCONS_00017550	TCONS_00005650	TCONS_00024387	TCONS_00003664	TCONS_00003553

[0100]

[0101]

TCONS_00001322	TCONS_00005651	TCONS_00006527	TCONS_00026124	TCONS_00001874
TCONS_00000694	TCONS_00017446	TCONS_00013752	TCONS_00026125	TCONS_00019591
TCONS_00005628	TCONS_00002826	TCONS_00005763	TCONS_00001454	TCONS_00021013
TCONS_00005861	TCONS_00014957	TCONS_00004073	TCONS_00017632	TCONS_00021014
TCONS_00005862	TCONS_00004420	TCONS_00013634	TCONS_00020388	TCONS_00021461
TCONS_00001891	TCONS_00004422	TCONS_00016827	TCONS_00020390	TCONS_00019159
TCONS_00020490	TCONS_00028890	TCONS_00017923	TCONS_00020389	TCONS_00011399
TCONS_00000952	TCONS_00025573	TCONS_00023286	TCONS_00020392	TCONS_00001622
TCONS_00003680	TCONS_00015043	TCONS_00004042	TCONS_00020394	TCONS_00029795
TCONS_00026626	TCONS_00000198	TCONS_00019871	TCONS_00009830	TCONS_00016160
TCONS_00000746	TCONS_00018449	TCONS_00019259	TCONS_00010598	TCONS_00019753
TCONS_00003582	TCONS_00003410	TCONS_00002412	TCONS_00010599	TCONS_00005117
TCONS_00008088	TCONS_00006096	TCONS_00016862	TCONS_00005099	TCONS_00009331
TCONS_00008437	TCONS_00023810	TCONS_00013695	TCONS_00024153	TCONS_00009851
TCONS_00017140	TCONS_00008531	TCONS_00017171	TCONS_00014813	TCONS_00011715
TCONS_00016965	TCONS_00011408	TCONS_00003348	TCONS_00000192	TCONS_00020424
TCONS_00003592	TCONS_00010100	TCONS_00028848	TCONS_00007242	TCONS_00007517
TCONS_00002651	TCONS_00009520	TCONS_00018162	TCONS_00023115	TCONS_00012993
TCONS_00026497	TCONS_00012236	TCONS_00000348	TCONS_00016210	TCONS_00011938
TCONS_00008729	TCONS_00012764	TCONS_00025588	TCONS_00016211	TCONS_00007627
TCONS_00003617	TCONS_00006447	TCONS_00026047	TCONS_00002202	TCONS_00026485
TCONS_00010828	TCONS_00016076	TCONS_00012990	TCONS_00002580	TCONS_00028560
TCONS_00020135	TCONS_00008301	TCONS_00014754	TCONS_00019615	TCONS_00001459
TCONS_00017449	TCONS_00003965	TCONS_00015298	TCONS_00025511	TCONS_00027927
TCONS_00014341	TCONS_00017346	TCONS_00003262	TCONS_00003897	TCONS_00004098
TCONS_00013867	TCONS_00009069	TCONS_00024773	TCONS_00008508	TCONS_00026919
TCONS_00021599	TCONS_00008501	TCONS_00025088	TCONS_00022239	TCONS_00027491
TCONS_00021600	TCONS_00015478	TCONS_00024280	TCONS_00024693	TCONS_00005563
TCONS_00021584	TCONS_00030008	TCONS_00024692	TCONS_00025610	TCONS_00009468
TCONS_00028615	TCONS_00029403	TCONS_00008498	TCONS_00022500	TCONS_00001791
TCONS_00005508	TCONS_00026420	TCONS_00010392	TCONS_00001129	TCONS_00006106
TCONS_00011831	TCONS_00020691	TCONS_00023815	TCONS_00013564	TCONS_00028213
TCONS_00017369	TCONS_00028860	TCONS_00019083	TCONS_00008412	TCONS_00018898
TCONS_00006904	TCONS_00009016	TCONS_00020870	TCONS_00008984	TCONS_00012208
TCONS_00006903	TCONS_00018124	TCONS_00008217	TCONS_00012778	TCONS_00013329
TCONS_00006902	TCONS_00020355	TCONS_00020885	TCONS_00012777	TCONS_00020864
TCONS_00021854	TCONS_00026651	TCONS_00018619	TCONS_00028118	TCONS_00020865
TCONS_00007855	TCONS_00019616	TCONS_00022244	TCONS_00023963	TCONS_00013294
TCONS_00005519	TCONS_00023891	TCONS_00022247	TCONS_00021344	TCONS_00012022
TCONS_00003424	TCONS_00007514	TCONS_00005729	TCONS_00008306	TCONS_00013203
TCONS_00019123	TCONS_00015205	TCONS_00000480	TCONS_00016663	TCONS_00012855
TCONS_00029084	TCONS_00024428	TCONS_00000984	TCONS_00016114	TCONS_00012917
TCONS_00009795	TCONS_00006576	TCONS_00006455	TCONS_00028317	TCONS_00007670
TCONS_00001557	TCONS_00003683	TCONS_00007129	TCONS_00028620	TCONS_00029004
TCONS_00001319	TCONS_00010372	TCONS_00020997	TCONS_00028617	TCONS_00003831
TCONS_00008289	TCONS_00012531	TCONS_00026337	TCONS_00020823	TCONS_00003832
TCONS_00005769	TCONS_00004415	TCONS_00013825	TCONS_00023972	TCONS_00008286
TCONS_00006500	TCONS_00018648	TCONS_00013181	TCONS_00028469	TCONS_00007654
TCONS_00016207	TCONS_00024393	TCONS_00013182	TCONS_00004032	TCONS_00009060
TCONS_00012739	TCONS_00020984	TCONS_00012221	TCONS_00028763	TCONS_00024147
TCONS_00003339	TCONS_00029834	TCONS_00004595	TCONS_00029315	TCONS_00023714
TCONS_00003212	TCONS_00006528	TCONS_00003869	TCONS_00008048	TCONS_00020384
TCONS_00007931	TCONS_00000471	TCONS_00022772	TCONS_00024863	TCONS_00013699
TCONS_00017639	TCONS_00029371	TCONS_00025195	TCONS_00011455	TCONS_00003179
TCONS_00002109	TCONS_00016139	TCONS_00012548	TCONS_00012085	TCONS_00003180
TCONS_00011306	TCONS_00016138	TCONS_00011893	TCONS_00004745	TCONS_00003181
TCONS_00018398	TCONS_00015233	TCONS_00011894	TCONS_00012088	TCONS_00013402
TCONS_00007731	TCONS_00014628	TCONS_00023723	TCONS_00021089	TCONS_00003998
TCONS_00007733	TCONS_00013376	TCONS_00025541	TCONS_00021090	TCONS_00004985
TCONS_00008453	TCONS_00000216	TCONS_00020399	TCONS_00028435	TCONS_00015007
TCONS_00029515	TCONS_00007918	TCONS_00021114	TCONS_00000707	TCONS_00019639
TCONS_00000674	TCONS_00010257	TCONS_00014790	TCONS_00001764	TCONS_00001215
TCONS_00006707	TCONS_00018294	TCONS_00009611	TCONS_00029230	TCONS_00000347
TCONS_00004665	TCONS_00009622	TCONS_00020694	TCONS_00027229	TCONS_00025984
TCONS_00004666	TCONS_00020471	TCONS_00022841	TCONS_00010335	TCONS_00010055
TCONS_00004664	TCONS_00010283	TCONS_00009532	TCONS_00003931	TCONS_00005952

[0102]	TCONS_00019741	TCONS_00014781	TCONS_00017158	TCONS_00001806	TCONS_00027154
	TCONS_00014733	TCONS_00003082	TCONS_00025203	TCONS_00002579	TCONS_00009877
	TCONS_00029835	TCONS_00017756	TCONS_00007471	TCONS_00022118	TCONS_00012218
	TCONS_00018764	TCONS_00007980	TCONS_00028097	TCONS_00024453	TCONS_00027178
	TCONS_00003385	TCONS_00011615	TCONS_00028098	TCONS_00021437	TCONS_00003414
	TCONS_00018032	TCONS_00015292	TCONS_00000742	TCONS_00021439	TCONS_00015246
	TCONS_00017854	TCONS_00007042	TCONS_00026906	TCONS_00021440	TCONS_00015247
	TCONS_00008309	TCONS_00024119	TCONS_00003929	TCONS_00021438	TCONS_00014643
	TCONS_00028384	TCONS_00023673	TCONS_00006934	TCONS_00021442	TCONS_00015248
	TCONS_00008550	TCONS_00000524	TCONS_00006936	TCONS_00020976	TCONS_00023451
	TCONS_00005931	TCONS_00001282	TCONS_00006935	TCONS_00020177	TCONS_00010247
	TCONS_00023231	TCONS_00006016	TCONS_00006684	TCONS_00020981	TCONS_00007791
	TCONS_00009123	TCONS_00013270	TCONS_00014914	TCONS_00017230	TCONS_00011420
	TCONS_00006482	TCONS_00029197	TCONS_00017043	TCONS_00028985	

[0103] 七、对长非编码RNA产生重要影响的SNP位点的获得

[0104] 比较步骤五和步骤六中的SNP位点碱基突变前后的RNA结合蛋白TARDBP靶标长非编码RNA的预测结果,发现对RNA结合蛋白TARDBP的结合位点有显著影响的SNP位点(表4, SNP位点碱基突变前,存在于表1中TARDBP的靶标lncRNA,而SNP位点突变后,不再是表2中的TARDBP的靶标lncRNA上的SNP位点,由于SNP位点,使TARDBP的结合位点消失)以及能够产生新的结合位点的SNP位点(表3,SNP位点碱基突变前,不存在于表1中TARDBP的靶标lncRNA,而SNP位点突变后,存在于表2中的TARDBP的靶标lncRNA上的SNP位点,由于SNP突变,使TARDBP出现了新的结合位点)。这些SNP位点对长非编码RNA的生物学功能可能产生重要的影响。

[0105] 表3、SNP位点碱基突变后增加的TARDBP结合位点的SNP位点

	lncRNA	SNP	lncRNA	SNP
[0106]	TCONS_00002651	rs1004923932	TCONS_00010372	rs559635461
	TCONS_00023960	rs1010410259	TCONS_00024338	rs560663585
	TCONS_00005009	rs1020842811	TCONS_00022944	rs563008629
	TCONS_00018032	rs1021936797	TCONS_00011812	rs563353079
	TCONS_00023286	rs1044763179	TCONS_00013161	rs563536165
	TCONS_00013652	rs1050407492	TCONS_00022728	rs566729079
	TCONS_00022701	rs1052706334	TCONS_00023100	rs566729079
	TCONS_00022702	rs1052706334	TCONS_00017756	rs568529287
	TCONS_00014999	rs1053805743	TCONS_00016074	rs569245353
	TCONS_00009340	rs112515658	TCONS_00004415	rs571611512
	TCONS_00004420	rs112889063	TCONS_00022841	rs571954447
	TCONS_00004422	rs112889063	TCONS_00020424	rs57341229
	TCONS_00004098	rs115423500	TCONS_00010100	rs574225773
	TCONS_00004195	rs116406052	TCONS_00025203	rs575146417
	TCONS_00015246	rs118190295	TCONS_00009761	rs575240666
	TCONS_00015247	rs118190295	TCONS_00025344	rs578031053
	TCONS_00027643	rs145351429	TCONS_00025163	rs578031053
	TCONS_00024692	rs145646538	TCONS_00012221	rs578117182
	TCONS_00005684	rs150548820	TCONS_00017230	rs58657573
	TCONS_00024393	rs16944841	TCONS_00001129	rs587614516
	TCONS_00010828	rs180998598	TCONS_00028384	rs6025609
	TCONS_00000984	rs184515929	TCONS_00009719	rs6414946
	TCONS_00023115	rs186627870	TCONS_00023365	rs72731414
	TCONS_00004073	rs188382713	TCONS_00023368	rs72731414

	TCONS_00009016	rs188613583	TCONS_00002109	rs74106083
	TCONS_00011455	rs188935251	TCONS_00001326	rs750275215
	TCONS_00012085	rs188935251	TCONS_00005769	rs751545502
	TCONS_00003385	rs189915095	TCONS_00006500	rs751545502
	TCONS_00007670	rs191572469	TCONS_00013779	rs752530767
	TCONS_00009468	rs368057114	TCONS_00008309	rs75694467
	TCONS_00005118	rs368933643	TCONS_00014914	rs76586127
	TCONS_00020997	rs374570691	TCONS_00010257	rs767747049
	TCONS_00021090	rs374607480	TCONS_00017639	rs770210661
	TCONS_00024429	rs376758127	TCONS_00029431	rs770677856
	TCONS_00008276	rs529823172	TCONS_00001215	rs77648262
	TCONS_00003617	rs531971849	TCONS_00028317	rs778796923
	TCONS_00020621	rs535298044	TCONS_00028620	rs778796923
	TCONS_00020622	rs535298044	TCONS_00028617	rs778796923
	TCONS_00006106	rs535570306	TCONS_00014813	rs7829904
	TCONS_00026638	rs537629445	TCONS_00006544	rs796801790
	TCONS_00002412	rs538759157	TCONS_00019753	rs886707451
[0107]	TCONS_00019783	rs538763715	TCONS_00027927	rs891461579
	TCONS_00022453	rs540370422	TCONS_00020177	rs895280176
	TCONS_00022454	rs540370422	TCONS_00020981	rs895280176
	TCONS_00002265	rs541895358	TCONS_00015043	rs903848162
	TCONS_00002266	rs541895358	TCONS_00017158	rs912939329
	TCONS_00002264	rs541895358	TCONS_00026651	rs914544317
	TCONS_00002263	rs541895358	TCONS_00027154	rs916231756
	TCONS_00018449	rs543807927	TCONS_00012607	rs936225082
	TCONS_00028034	rs544785979	TCONS_00011381	rs936225082
	TCONS_00027229	rs544879340	TCONS_00011420	rs938324559
	TCONS_00005612	rs550567197	TCONS_00012218	rs942470532
	TCONS_00016015	rs551028831	TCONS_00027380	rs951046479
	TCONS_00020471	rs551616147	TCONS_00015292	rs978756196
	TCONS_00018294	rs555969125	TCONS_00029552	rs982963981
	TCONS_00013825	rs556805291	TCONS_00019615	rs986488936
	TCONS_00013181	rs556805291	TCONS_00006442	rs993898324
	TCONS_00013182	rs556805291	TCONS_00026553	rs995627872
	TCONS_00021450	rs558489415	TCONS_00026729	rs995627872
	TCONS_00017898	rs558812600	TCONS_00026554	rs995627872

[0108] 表4、SNP位点碱基突变后TARDBP结合位点缺失的SNP位点

	lncRNA	SNP	lncRNA	SNP
	TCONS_00025311	rs1000153552	TCONS_00015943	rs549632898
	TCONS_00025312	rs1000153552	TCONS_00015167	rs550317976
	TCONS_00007047	rs1005803568	TCONS_00029566	rs552255363
	TCONS_00009176	rs1007571842	TCONS_00029784	rs552255363
	TCONS_00002918	rs1008299409	TCONS_00020721	rs553157457
	TCONS_00002919	rs1008299409	TCONS_00023520	rs553375502
[0109]	TCONS_00003907	rs1014484393	TCONS_00004196	rs553818437
	TCONS_00015204	rs1020906904	TCONS_00011363	rs554813088
	TCONS_00018161	rs1023104806	TCONS_00004223	rs557819351
	TCONS_00005103	rs1030407636	TCONS_00021409	rs558734951
	TCONS_00005102	rs1030407636	TCONS_00022349	rs561386618
	TCONS_00004167	rs1030407636	TCONS_00016200	rs563918135
	TCONS_00016209	rs1043162038	TCONS_00017911	rs565333471
	TCONS_00010757	rs1052099041	TCONS_00022623	rs565360130

[0110]

TCONS_00028086	rs1060490	TCONS_00021551	rs566365462
TCONS_00008015	rs116240306	TCONS_00021674	rs567041770
TCONS_00009742	rs116362930	TCONS_00005098	rs570642421
TCONS_00019939	rs141753053	TCONS_00004166	rs570642421
TCONS_00014916	rs142285150	TCONS_00028805	rs570691714
TCONS_00014917	rs142285150	TCONS_00028764	rs570691714
TCONS_00009345	rs142543515	TCONS_00024213	rs570754511
TCONS_00005222	rs145246679	TCONS_00011379	rs571638589
TCONS_00019142	rs145647232	TCONS_00020433	rs573089081
TCONS_00022636	rs145871875	TCONS_00021137	rs573089081
TCONS_00007987	rs146004143	TCONS_00021138	rs573089081
TCONS_00022458	rs147284543	TCONS_00019680	rs574374432
TCONS_00004275	rs148577523	TCONS_00022101	rs575377781
TCONS_00005551	rs148969881	TCONS_00019277	rs577385258
TCONS_00015792	rs151264164	TCONS_00015377	rs587652931
TCONS_00009501	rs181120153	TCONS_00001045	rs61772227
TCONS_00006223	rs183343562	TCONS_00001897	rs61833251
TCONS_00017415	rs186377124	TCONS_00025013	rs71393060
TCONS_00006112	rs187529298	TCONS_00017074	rs73486562
TCONS_00006113	rs187529298	TCONS_00019423	rs73520934
TCONS_00006900	rs187529298	TCONS_00001577	rs74099414
TCONS_00001117	rs187559458	TCONS_00008093	rs763569805
TCONS_00006373	rs188858758	TCONS_00005674	rs866901722
TCONS_00029558	rs189283209	TCONS_00005191	rs898201078
TCONS_00009070	rs190081551	TCONS_00006085	rs903795892
TCONS_00021259	rs199500836	TCONS_00015070	rs909718263
TCONS_00008606	rs375859490	TCONS_00029433	rs911594961
TCONS_00019890	rs3763827	TCONS_00013635	rs919291682
TCONS_00018842	rs527896060	TCONS_00015945	rs922906644
TCONS_00005806	rs528862967	TCONS_00020867	rs928620169
TCONS_00005807	rs528862967	TCONS_00016264	rs932133361
TCONS_00010468	rs530968274	TCONS_00014353	rs948852892
TCONS_00009640	rs531716762	TCONS_00018527	rs949076788
TCONS_00016859	rs531862109	TCONS_00018424	rs953778834
TCONS_00024755	rs532021143	TCONS_00003580	rs954277373
TCONS_00019652	rs532370116	TCONS_00004702	rs954277373
TCONS_00005851	rs535903422	TCONS_00017294	rs959272000
TCONS_00025520	rs536885753	TCONS_00014799	rs966748986
TCONS_00013189	rs537036015	TCONS_00003837	rs968771596
TCONS_00014937	rs538607975	TCONS_00019544	rs975948869
TCONS_00015796	rs539005305	TCONS_00020708	rs983336877
TCONS_00020028	rs540516083	TCONS_00025556	rs987115712
TCONS_00019549	rs540516083	TCONS_00012313	rs988902691
TCONS_00024739	rs543648228	TCONS_00012314	rs988902691
TCONS_00004843	rs543684736	TCONS_00012315	rs988902691
TCONS_00010983	rs547633820	TCONS_00008447	rs990794764
TCONS_00018094	rs548290829	TCONS_00004809	rs991514376