

DOI: 10.13376/j.cblls/2016077

文章编号: 1004-0374(2016)05-0592-10



杨雪瑞, 清华大学生命科学学院研究员, 博导, 清华-北大生命科学联合中心研究员, 中组部青年千人计划获得者。课题组的研究领域为癌症系统生物学及分子生物学, 从交叉学科的多个角度研究肿瘤等复杂疾病中以 RNA 为核心的转录、转录后及翻译调控通路及网络。实验室有机地结合计算分析与实验生物学的优势, 并开展广泛的合作研究。主要研究方向包括: 非编码 RNA 代谢与编辑对肿瘤基因表达的调控机制; 蛋白质翻译水平的全基因组定量分析及翻译调控在肿瘤中的参与机制; 单细胞转录组尺度的肿瘤异质性、标记物及肿瘤的转移机制; 多维组学大数据的计算分析流程及基于此的肿瘤特异性基因多级调控网络。

异向转录基因对的鉴定、调控及其功能

肖正涛^{1,2,3,4}, 王雨亭^{1,3,4,5}, 杨雪瑞^{1,2,3,4*}

(1 清华大学生物信息学教育部重点实验室, 北京 100084; 2 清华-北大生命科学联合中心, 北京 100084; 3 清华大学合成与系统生物学中心, 北京 100084; 4 清华大学生命科学学院, 北京 100084; 5 清华大学PTN研究生项目, 北京 100084)

摘要: 转录起始位置临近的两个基因分别向两个方向的转录(异向转录)在哺乳动物基因组中广泛存在, 例如许多编码基因与非编码基因形成异向转录基因对。近年来的研究表明异向转录受到多种因素的调控, 如转录终止、基因成环结构、染色体修饰、核小体重构等。异向转录产物可以影响相邻基因的表达, 或者产生增强子 RNA 影响其他基因的表达, 同时也被认为是进化过程中产生新基因的途径之一。现总结近年来围绕异向转录基因对的鉴定、调控及功能的研究进展, 并讨论该研究领域可能的发展方向。

关键词: 异向转录; 非编码 RNA; 转录调控

中图分类号: Q52; Q344; Q786 **文献标志码:** A

Identifications, regulations, and functions of divergent transcription

XIAO Zheng-Tao^{1,2,3,4}, WANG Yu-Ting^{1,3,4,5}, YANG Xue-Rui^{1,2,3,4*}

(1 MOE Key Laboratory of Bioinformatics, Tsinghua University, Beijing 100084, China; 2 Tsinghua-Peking Joint Center for Life Sciences, Beijing 100084, China; 3 Center for Synthetic & Systems Biology, Tsinghua University, Beijing 100084, China; 4 School of Life Sciences, Tsinghua University, Beijing 100084, China; 5 Joint Graduate Program of Peking-Tsinghua-National Institute of Biological Science, Tsinghua University, Beijing 100084, China)

Abstract: Divergent transcription is frequently found in mammalian genomes, producing pairs of transcripts from the two strands of DNA. For example, the transcription of many coding genes was found to be accompanied by transcription of a non-coding gene on the opposite direction. Divergent transcription is tightly regulated via various mechanisms, including transcription termination, gene loop, chromatin modification, and nucleosome remodeling. The transcripts from divergent transcription can affect the transcription of nearby genes, or other genes by

收稿日期: 2016-04-26

基金项目: 国家自然科学基金委员会重大研究计划培育项目(91540109), 清华-北大生命科学联合中心、清华大学生物信息学教育部重点实验室及“青年千人计划”

*通信作者: E-mail: yangxuerui@tsinghua.edu.cn; Tel: 010-62783943

generating enhancer RNAs. In addition, divergent transcription may facilitate the origin of new genes in evolution. Here we provide a review on the recent research in the identifications, regulation machineries, and functions of divergent transcription, together with a brief discussion on the future directions of the field.

Key words: divergent transcription; noncoding RNA; transcriptional regulation

基因的转录起始是一个复杂且精细的过程, 是基因表达过程中最重要的一个环节, 其调控机制一直是生物学中的一个研究热点。转录起始的研究对于解释基因的特异性表达, 细胞的发育、分化、成熟至关重要^[1]。在人、小鼠、酵母细胞中, 多数基因的启动子是双向转录的^[2-5], 即这些基因的启动子上游普遍存在反方向转录的非编码 RNA (ncRNA), 这种现象被称为异向转录, 这对基因也被称为异向转录基因对。研究显示, 大多数异向转录基因对的表达水平具有较强的相关性^[6]。例如在受到脂多糖刺激后, 多个基因的表达产物和它上游的异向转录 RNA 成对地上调或者下调, 并且不同异向转录基因对的调控出现在刺激之后的不同时期, 说明异向转录基因对的共表达及其转录调控可能在细胞响应外界刺激中起着重要的作用。进一步的研究认为, 异向转录的非编码 RNA 受到多种形式的调控^[7], 包括相应的上游基因^[8]、某些蛋白质因子^[9]、组蛋白修饰因子^[10]、染色体重构因子^[11]等。这些非编码 RNA 又可以顺式地调控附近的基因, 或者作为增强子 RNA 调控其他位置的基因。此外, 异向转录在新基因的起源和进化过程中也有重要的贡献^[12]。

人类及多种高等生物的基因组包含着大量的非编码 RNA^[13]。非编码 RNA 作为一种新型的遗传信息传递与调控分子的功能得到了广泛的共识, 并且获得了越来越多深入的研究。其中长非编码 RNA (long non-coding RNA, lncRNA) 近年来被大量发现, 并且在不同的物种和组织中得到了系统的鉴定。虽然 lncRNA 数量众多 (GENCODE 数据库统计的人类基因组中约有 15 941 个^[13]), 但是到目前为止, 对于 lncRNA 形成机制、调控模式以及生物学功能的认识还处于起步阶段, 尚有大量未知因素。一般认为, 大量的非编码 RNA 与编码基因以异向转录基因对的形式存在^[6], 因此, 对于异向转录这一类具有明显特征、极具潜力参与基因表达调控的非编码 RNA 的深入研究具有重要的生物学意义。本文就异向转录基因对的模型、系统鉴定和筛选、转录调控机制和生物功能三个方面作一综述。

1 异向转录模型

在真核生物的基因组中, 基因的分布并不是随机的, 例如管家基因通常聚在一起^[14]。根据基因在染色体上的相对位置关系, 我们可以把相邻的基因分为三类^[3]如图 1 所示: (1) 对向转录基因对 (或 tail-to-tail), 两个基因来自不同的链, 转录方向相反, 3' 端相互邻近; (2) 异向转录基因对 (或 head-to-head), 两个基因的来自不同的链, 转录方向相反, 5' 端相互邻近; (3) 同方向转录基因对 (或 tail-to-head), 两个基因来自同一条链, 转录方向一致。研究发现, 在人的基因组中, 大部分基因的上游存在异向转录的现象^[2-3,15-24], 并且相比其他两种类型, 异向转录基因通常呈现共表达的现象^[6]。2008 年, 有三篇文章报道称, 在小鼠的胚胎干细胞、人的肺成纤维细胞、HeLa 细胞系中, 大部分注释基因的启动子上游存在异向转录的现象^[15,22-23]。这些非编码 RNA 也被称为启动子上游 RNA (upstream antisense RNA, uaRNA)、启动子相关的 RNA (promoter-associated ncRNA, PROMPTs) 或起始转录相关的 RNA (TSS associated RNA, TSSa-RNAs)^[15,22-23]。按照长度可以将这些 RNA 分为: 短非编码 RNA (<100 nt)^[23]; 长非编码 RNA (lncRNA, >100~200 nt)^[25]。按照稳

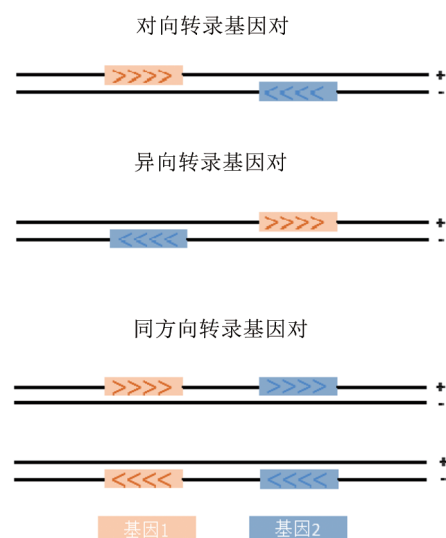


图1 相邻基因的转录

定性可以分为两种类型：不稳定的转录产物 (cryptic unstable transcripts, CUTs) 和稳定转录产物 (stable unannotated transcripts, SUTs)^[7]。大部分的异向转录产物比较短，丰度低，有明显的活跃转录的组蛋白修饰信号：组蛋白乙酰化 (H3K9ac)、组蛋白甲基化 (H3K4me3)^[22-23]，但是缺少与转录延伸相关的组蛋白修饰信号 (H3K36me3 和 H3K79me2)^[26-27]。在酵母细胞中，这些不稳定的转录产物易被 Nrd1-exosome-TRAMP 复合物迅速降解^[28-29]。异向转录基因对，按照各自的稳定性可以分为：stable-stable、unstable-stable、unstable-unstable^[24]。根据转录产物也可以分成下面几类：ncRNA-ncRNA 对、ncRNA-coding RNA、coding RNA-coding RNA。

研究认为异向转录基因拥有一个共同的启动子区域，该区域的染色体结构是裸露的，没有核小体结合 (nucleosome-free regions, NFRs)^[30-32]，可以允许转录调控因子 (transcription factors, TF) 顺利结合到该区域。而转录调控因子又可以进一步招募 RNA 聚合酶 II 以及其他的转录因子，促进转录起始复合物在该区域形成^[31,33-34]。异向转录基因对各自的转录起始位点分别位于该区域的两端^[4-5,34-36]。染色体重构因子可以通过影响该区域附近的核小体的排布，帮助染色体形成精确的组织结构，从而影响异向转录的形成 (详见 3.4 部分)。因此，这种开放的启动子区域是异向转录的一个重要特征。

ChIP 实验也发现，异向转录基因各自的起始位点与 RNAPII 的结合位点一致，并且显著富集 H3K4me3 信号^[15,23]，表明异向转录基因分别由各自的转录复合物负责转录。进一步研究发现，异向转录基因的启动子区域含有两个独立的核心启动

子。核心启动子是启动子区域的一段关键序列，这些序列通常在转录起始位点上 / 下游 50 bp，是转录起始复合物形成的关键位置，转录起始阶段的第一个步骤，因此被认为是基础转录元件^[37]，也是控制 RNA 聚合酶 II 向特定方向移动的关键元件^[38]。利用 GRO-seq 技术，Lam 等^[39] 研究人员发现，异向转录基因从各自的核心启动子开始转录 (核心启动子临近 NFRs 边缘)，并且异向转录基因对的核心启动子拥有相似的序列特征。同时两个核心启动子上都观察到了启动复合物的形成^[34,40]。在哺乳动物基因组中，对异向转录基因的启动子区域的序列分析发现，这些启动子上普遍存在 CpG 岛，但是很少有 TATA box，因此一些研究认为这样的特征有助于异向转录产物的形成^[7,38,41]。Core 等^[38] 用这个模型解释了为什么在果蝇 (*Drosophila melanogaster*) 中缺少这种异向转录，而在哺乳动物中广泛存在异向转录的现象。

综合上述结论，异向转录基因的启动模型具有如下特征 (图 2)：(1) 拥有一个共同的启动子区域和两个核心启动子；(2) 启动子区域的 DNA 裸露，没有核小体结合，是 DNase 超敏感区域；(3) 异向转录基因的起始位点处都有活跃转录的表观修饰信号，那些不稳定非编码 RNA 的上游基因区域上的延伸相关的表观修饰信号缺失；(4) 异向转录基因对的转录起始位点，通常相距约 250 bp^[24] 左右。

目前，关于人类基因组中基因的启动子是否本质上是异向转录这个问题，还存在争论。尽管大量研究认为哺乳动物的基因主要是异向转录的形式^[15,18,21,23]，但在 2015 年，Duttke 等^[35] 通过分析 HeLa-S3 细胞的 GRO-seq 数据和 DNaseI-seq 数据

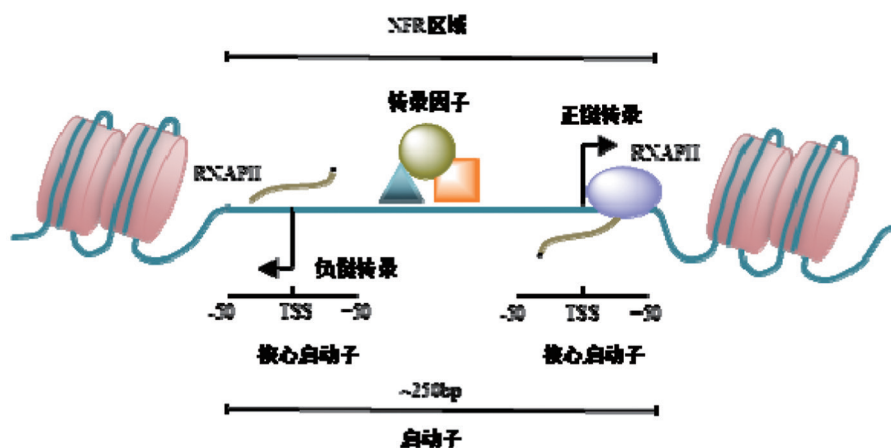


图2 异向转录单元模型

得出相反的观点: 人类基因的启动子本质上是单向转录的。他们的主要观点如下: (1) 基础的 RNA 聚合酶 II 转录复合物以及核心启动子本质上是单方向的; (2) 异向转录是染色体开放区域两端的核心启动子分别转录的结果; (3) 单向的启动子比较常见, 而且反方向上的核心启动子序列常常缺失; (4) 反向的转录的核心启动子具有独特的染色体信号。

他们利用 GRO-seq 和 DNase I 敏感位点的数据定义了 4 378 个只含有一个转录起始位点的启动子, 其中 51% 的启动子是单向转录的形式, 只含有一个转录起始位点簇, 且和基因的转录方向一样。49% 的基因启动子是异向转录或双向转录的形式 (他们把已注释的异向转录基因对单独归类为双向转录), 有两个核心启动子, 并具有独特的表观信号。

Andersson 等^[36]随后反驳了这一的观点, 他们利用了更多的高通量数据、DNaseI 超敏感位点 (DHSs, DNase I hypersensitive sites)、CAGE、RNA-seq、GRO-Seq、GRO-cap, 通过进一步分析 Duttke 等所定义的单向转录启动子, 发现在这些数据中都能观测到异向转录起始的信号, 而 Duttke 等之所以没有检测到是由于实验和数据处理的差异造成的。最终, Andersson 等认为 93% (3974/4285) 的启动子是异向转录的。

针对 Andersson 等^[36]的反驳, Duttke 等^[42]指出前者的数据统计存在问题, 从而夸大了异向转录的比例。但是, 尽管存在较大的争论, 对于异向转录启动子模型, 学术界还是基本达成了共识: 核心启动子很难单独行使功能, 需要附近的转录调控因子的结合来调控他们的活性。“启动子”常常包括了更上游的基因组区域, 这些区域可以帮助调控“核心启动子”的转录起始。这些启动子通常是核小体空白的区域, 包含的两个核心启动子分别转录, 从而形成所谓的异向转录 (或者双向转录), 如图 2 所示。而对于基因本质上是单向还是双向转录这个问题, 需要更多、更精确的数据, 乃至更新的实验

技术来进一步解释。

2 异向转录基因对的鉴定

早期的研究主要通过基因的转录起始位点之间的距离来鉴定异向转录基因^[15]。Seila 等^[23]检测了小鼠胚胎干细胞中短 RNA 的表达, 发现大量 (67%) 约 20 nt 长的 RNA 处于已知的编码基因的起始位点上游 1.5 kb 范围内, 称之为 TSSa RNAs。这些 RNA 的转录起始位点和已知的编码基因位点相距 100~300 nt 之间。Core 等^[15]提出了一个新的方法叫做 GRO-seq (global run-on sequencing), 该方法先提取细胞核, 然后抑制转录起始, 并利用 nuclear run-on assays (NRO) 延长初始的 RNAs。随后加入一种核苷酸类似物 5'-bromouridin 5'-triphosphate (BrUTP) 到这些初始 RNA 上, 并利用这种类似物的特异性抗体进行免疫共沉淀分离, 这些 NRO-RNA 被反转录然后进行测序。他们发现 77% 的活跃基因上游有显著的异向转录起始。这些异向转录基因对的起始位点相距在 100~250 bp 范围内。Ntini 等^[21]利用 RNA-seq 和 CAGE (cap analysis gene expression) 实验在 HeLa 细胞系中发现在已知基因的 TSS 上游 110 bp 存在反向转录产物。CAGE 方法^[43-45]首先提取 mRNA 的 5' 端开头的一小段序列 (20~21 nt) 作为 5' 端标签, 然后加上接头并逆转录为 cDNA, 再通过 PCR 得到扩增产物, 最后进行测序。测序结果进一步比对到参考基因组上就能找到准确的转录起始位点。该方法还被广泛应用于 FANTOM 研究计划^[46]。该方法的优点是通量高, 并且能够检测转录的水平。

此外还有一些高通量的方法被用来检测转录起始位点及异向转录基因 (见表 1)

这些方法是基于 TSS 位点之间的距离判断异向转录基因, 并没有明确提出异向转录基因的定义和鉴别标准。Duttke 等^[35]认为, 仅仅利用已经注释的基因的转录起始位点之间的距离判断异向转录

表1 鉴定转录起始位点的实验技术

缩写	全名	检测产物	文献
GRO-seq	Global run-on sequencing	Nascent RNA	[6, 15, 24]
NET-seq	Native elongating transcript sequencing read	Nascent RNA	[10, 47]
mNET-seq	Mammalian NET-seq	Nascent RNA	[48]
PRO-seq	Precision run-on sequencing	Nascent RNA	[49]
CAGE	Cap analysis gene expression	RNA 5'端	[50]
PEAT	A paired-end sequencing strategy	RNA 5'端	[51]
START-seq		Nascent RNA	[34]

基因是不准确的。根据异向转录的特点,他们首先利用 GRO-seq 数据定义基因的准确起始位点,然后用 DNase I 超敏感位点定义异向转录基因起始位点之间的染色体裸露区域(即 NFR),把距离在 ~250 bp 的基因定义为真正的异向转录基因。

此外,组蛋白修饰如 H3K4me2 和 H3K4me3 等信号也被用来判断异向转录基因对上游的基因是否存在活跃转录^[26-27,52]。综上所述,转录起始位点之间的距离、启动子的染色体开放状态、活跃转录信号都是准确鉴定异向转录的关键证据。

3 异向转录的调控因素

研究表明 RNA 聚合酶 II 在两个方向的核心启动子上的组装比例是几乎相等的^[15],然而最终形成稳定转录产物的通常仅来自于其中一个方向的转录^[53-54],那么是什么机制控制了两个转录方向的选择呢?

从转录的过程看, RNA 聚合酶 II 指导的转录过程主要有:染色质修饰、起始复合物的形成、转录延伸、转录终止。对于异向转录的调控也主要围绕这几个方面。

3.1 转录终止调控异向转录的方向

转录终止并降解是基因转录调控的有效方式之一,在真核生物中大部分的转录终止机制是通过识别 polyA 位点(PAS)进行 3' 端加尾实现的。PAS 位点是由一段保守的 AAUAAA 序列构成的,可以被多聚腺苷酸化复合物(CPAC)识别,并促进 RNA 聚合酶 II 的释放,从而终止转录过程。而另一方面,5' 端 U1 位点可以被 U1 snRNP (U1 small nuclear ribonucleoprotein) 识别,并可以抑制 PAS 介导的剪切,帮助 RNA 继续转录^[55-56]。

研究发现短片段 ncRNA (CUTs) 常常缺少 U1 snRNA 复合物识别位点,但是富集 poly(A) 序列(PAS)^[18,21]。这两个特征导致了 RNA 聚合酶 II 的释放,以及 RNA 多聚腺苷酸化的终止作用^[55,57]。实验进一步证明了上面的结论:将 PAS 克隆到报告基因,并在 HeLa 细胞中稳定表达,产生不稳定的 poly(A)+ RNA,而将 2GT/T 基序突变则会引起显著的 PAS 通读, AATAAA 突变则会消除 3' 端的加尾过程^[21]。这些不稳定的非编码 RNA 被 PAS 依赖的转录终止作用之后,进一步被降解^[22,58]。

在酵母中, RNA 聚合酶 II 通过 NRD 依赖的途径终止基因的转录^[59]。NRD 复合物由三个多肽组成: Nrd1、Nab3、Sen1, 它可以识别 NRD 结合位

点(NRD binding sites, NBS), 并通过和 CBP 作用招募核酸外切酶, 终止 RNA 的转录^[60]。NRD 依赖的转录终止产物随后被外切酶降解。编码基因的序列中 NBS 位点相比异向转录的非编码 RNA 少, 从而导致只有一个方向的转录正常进行^[28-29,61]。实验分析发现 Nrd1 缺失会导致异向转录的 RNA 增加 2 倍^[9]。

综上, 在哺乳动物中基因的转录终止依赖于 PAS 信号通路。而在酵母中, 转录终止受 NRD 信号通路介导。转录终止信号往往在异向转录基因对上的分布是不均衡的: 编码基因的 DNA 序列缺少 PAS (人) 或者 NBS (酵母) 信号, 而异向转录的非编码 RNA 却富集这些信号(图 3)。并且在人的基因组中, U1 识别位点在编码基因上更富集, 可以抑制 PAS 识别。

3.2 基因成环控制异向转录的方向

上文介绍的转录终止机制可以有效地降解和清理多余的异向转录产物。但是除此之外, 特定的基因环结构也可以控制编码基因转录的方向性。2004 年和 2005 年有研究发现, 转录起始因子和转录终止因子可以相互作用, 导致起始区域和终止区域互相靠近形成环状结构并控制 RNA 聚合酶 II 转录的方向^[62-63]。此外, 还有证据证明 CPAC、TFIIB、Ssu72 可以同时作用于基因转录起始区域和终止区域^[63-67]。2012 年, Tang-Wong^[8] 等在酵母细胞中发现基因内部可以通过成环的方式控制异向启动子的转录方向。突变 Ssu72-2 会阻止环的形成, 使得 RNA 聚合酶 II 可以重新结合到上游非编码 RNA 的启动子区域, 促进非编码 RNA 的转录。并且, 在 Ssu72-2 突变的细胞发现 605 种新的非编码 RNA (Ssu72 restricted transcripts, SRTs), 其中大部分的属于异向转录的产物。TFIIB、CPAC 因子、Pta1、Rna14、Rna15 都参与了基因的成环过程。

在高等生物中, 基因的成环过程还没有很深入的研究, 但是有证据表明在人和动物中也存在相似的机制。人的细胞中, TFIIB 可以和转录终止区域结合, 并和 3' 末端的转录终止因子相互作用^[68]。

基因成环机制提供了另一种控制基因转录方向的机制, 相比转录终止机制, 它可以指导 RNA 聚合酶 II 只向编码基因的方向转录, 但是目前还不清楚这种机制在基因组中是否普遍存在。

3.3 染色体修饰控制异向转录的方向

染色体的结构是由组蛋白和 DNA 构成的, 组蛋白的修饰可以特异地激活或抑制基因的转录, 并

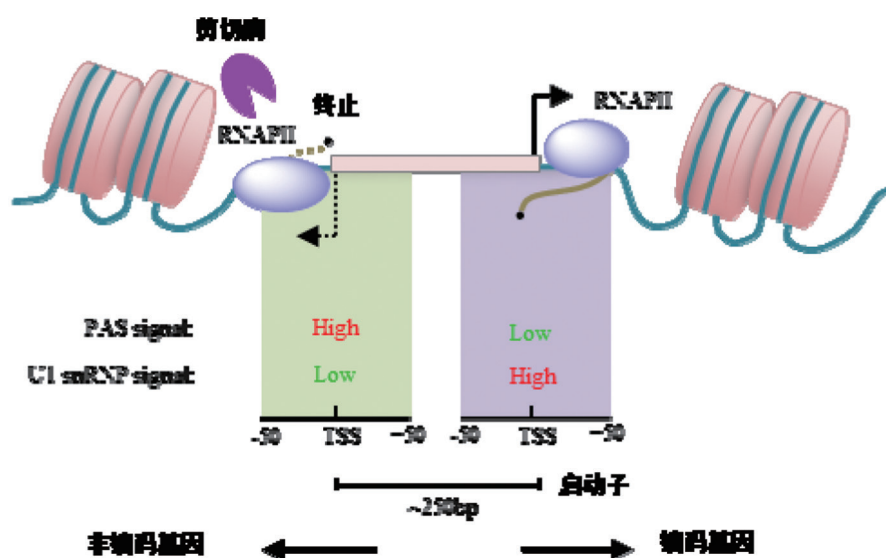


图3 PAS-U1 snRNP信号在异向转录基因对中的富集

且可以调控启动子的异向转录。研究表明异向转录的两个核心启动子上 (NFR 区域, 紧邻核小体) 的转录前起始复合物 (PIC) 的构成是一样的^[40]。在人的细胞中, H3K4me3 可以促进转录前起始复合物的形成^[23,69], 该信号所在的位置有转录起始复合物形成, 一般在异向转录基因对的两个核心启动子区域都能检测到该信号, 但是标记转录延伸的信号——H3K79me2 通常只出现在编码基因的区域^[23,70]。研究认为这些组蛋白修饰信号可能有利于染色体形成负超螺旋结构, 帮助 RNAPII 稳定结合并从转录起始继续走向转录延伸, RNAPII 在上游的停止有利于维持另一种染色体的环境, 阻止核小体在起始区域组装, 维持一种开放的染色体构象^[71]。

在酵母中, 研究推测 Ssu72 导致基因 3' 端 H4 的去乙酰化, H4 的去乙酰化可促进染色体形成更紧凑结构, 这种结构会限制转录因子的结合, 进而影响异向基因的转录^[8,72]。另外, Rpd3S 去乙酰化复合物可以抑制启动子的上游负链基因的转录起始^[10], 并且推测负链方向的转录可以促进正链的转录, 例如通过 DNA 解旋或者去除核小体。

3.4 核小体重构控制异向转录的方向

在酵母的基因组中, 有一种称为 Isw2 的染色体重构复合物, 它可以重排核小体的位置, 从而阻止上游的转录起始, 增强转录的方向性。通常情况下, Isw2 可以增加基因间区域的核小体的量, 让转录只能朝一个方向进行; 当 Isw2 活性缺失时, 会导致正链和负链都转录^[73-74]。

Marquardt 等^[11] 在酵母细胞中利用基因组替换

的实验筛选影响异向转录的基因, 他们发现染色体重构因子 CAF-I 复合物的突变会引起全基因组范围内异向的非编码 RNA 的转录, 说明 CAF-I 以及相关的因子可以抑制异向非编码 RNA 的转录。而 H3K56 的高度乙酰化、SWI/SNF 核小体重构复合物能促进核小体的翻转, 并帮助异向基因转录。

4 异向转录基因的功能

异向转录事件或者转录产物自身可以起到调控临近基因或者远处的基因活性的作用, 并且根据转录产物的长度、序列、稳定性的不同, 发挥着不同程度的功能。目前研究异向转录基因的功能主要有以下三类, 但并不确定哪种功能占据主导地位。现将不同的功能列举如下, 并举例说明。

4.1 驱动新基因的产生

异向转录现象在不同的物种中都有报道, 例如人^[2]、酵母^[75]、小鼠^[6], 甚至植物^[76] 的基因组中, 在人类基因组中发现的异向转录基因对在其他物种中也有一定程度的保守性。Wu 与 Sharp^[12] 提出模型认为, 在进化的过程中由于突变导致 U1-PAS 信号的得失导致新基因的产生或者变异。进化进程中, 基因的突变引起的基因上游 U1 信号获得和 PAS 信号缺失可以促进 RNA 聚合酶转录活性, 进而驱动新的基因产生, 新的基因又会进一步转移到基因组的其他区域。

4.2 调控邻近基因的表达

负链转录的非编码 RNA 可以顺式地作用于正链方向的基因, 影响其表达水平^[77-78]。这种调控

可以作用于多个水平：(1) 负链的非编码 RNA 的转录可以影响染色体内部的构象，进而影响邻近的基因表达。例如在酵母中，*fbp1*⁺ 区域上游的 ncRNA 转录会影响染色质逐渐形成开放构型，如果插入转录终止子阻止这些 ncRNA 的转录，则会导致染色体构象停止改变^[79]；还可以通过解旋作用使得启动子区域的染色质开放，促进旁边基因的转录。(2) 反链的非编码 RNA 还可以通过招募因子改变染色体的修饰，进而影响编码基因的转录，例如 CCND1 基因的启动子上游的一个非编码 RNA 可以招募 RNA 结合蛋白 TLS，并抑制 CCND1 上的 CBP 和 p300 组蛋白的乙酰转移酶的活性，从而抑制 CCND1 的表达^[80]。(3) 反向的非编码 RNA 转录产物还可以通过抑制启动子区域的转录因子结合，或者和临近基因竞争转录因子，从而抑制临近基因的转录，例如 TPI1 基因和其上游的不稳定非编码 RNA 受同样的转录激活因子的调控，二者因此竞争相同的聚合酶和转录因子^[4,77]。(4) 通过形成环状的结构调控上游基因的转录（见 3.2 部分）。最近，Luo 等^[81]的研究表明在小鼠多能细胞中，EVX1 基因上游的异向转录基因 *Evx1as* 可以结合到染色体调控位点促进染色体成环构象，进而调控 *Evx1as* 的表达。

4.3 产生增强子RNA

异向转录可产生具有增强子功能的 RNA，叫做增强子 RNA (enhancer RNA, eRNA)^[24,78,82-83]。这类 RNA 通常比较短，且没有 polyA，可以通过帮助招募或者增强通用转录复合物与靶向基因启动子结合的稳定性，进而激活或者启动转录^[84-88]。Pnueli 等^[88]发现一种异向转录产生的 eRNA，被称为 CgaeRNA，可以特异性地促进小鼠的促性腺细胞和甲状腺细胞中的绒毛膜促性腺激素 α (chorionic gonadotropin alpha, Cga) 基因的转录活性。主要作用方式是要通过组蛋白修饰、染色体重构和 DNA 成环作用，促进和稳定自身以及其他转录起始因子和 Cga 基因的启动子区域的结合。

5 展望

尽管在小鼠及人的胚胎干细胞中已经对全基因组的异向转录进行了鉴定，但是由于 lncRNA 的表达具有高度的细胞特异性^[89-90]，因此，在其他组织或细胞中可能存在新的活跃的异向转录，或者不一样的表达模式。目前，尚无研究对异向转录基因的细胞和组织特异性进行系统鉴定。其次，尽管大部

分的异向转录 lncRNA-mRNA 存在共调控的表达模式，但是在细胞分化过程中，外界刺激或者不同的组织和细胞中这种协调表达可能受到调控，但是影响其转录（或者特异性转录）的上游调控因子尚不清楚。最后，这些异向转录的非编码 RNA 可以调控相邻的编码基因的表达，但是尚无针对二者相互调控关系的系统性的研究，并且哪种调控方式占主要，是否在生物体中普遍存在，这些问题目前尚未得到较好的回答。

对于异向转录是不是哺乳动物基因组中一个普遍、本质的特性依然存在争论^[35-36]。但是争论的关键问题实质上是对于准确鉴定基因转录的起始位点，以及更精确的数据分析和鉴定。因此，要想准确回答这个问题，需要进一步通过实验和生物信息的方法来支持。

越来越多的证据表明 lncRNA 几乎参与所有重要的生物过程，lncRNA 和其他因子一起构成了全新的介导遗传信息传递方式和表达调控网络，在生物进化过程中具有重要的意义。进一步整合已有资源以及系统分析异向转录 lncRNA 的生成和功能相关的调控机制，有利于深入探讨 lncRNA 和生命现象及重大疾病的关系。

[参 考 文 献]

- [1] Danino YM, Even D, Ideses D, et al. The core promoter: at the heart of gene expression. *Biochim Biophys Acta*, 2015, 1849: 1116-31
- [2] Trinklein ND, Aldred SF, Hartman SJ, et al. An abundance of bidirectional promoters in the human genome. *Genome Res*, 2004, 14: 62-6
- [3] Koyanagi KO, Hagiwara M, Itoh T, et al. Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene*, 2005, 353: 169-76
- [4] Neil H, Malabat C, d'Aubenton-Carafa Y, et al. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 2009, 457: 1038-42
- [5] Xu ZY, Wei W, Gagneur J, et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 2009, 457: 1033-U7
- [6] Sigova AA, Mullen AC, Molinie B, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci USA*, 2013, 110: 2876-81
- [7] Grzechnik P, Tan-Wong SM, Proudfoot NJ. Terminate and make a loop: regulation of transcriptional directionality. *Trends Biochem Sci*, 2014, 39: 319-27
- [8] Tan-Wong SM, Zaugg JB, Camblong J, et al. Gene loops enhance transcriptional directionality. *Science*, 2012, 338: 671-5

- [9] Schulz D, Schwalb B, Kiesel A, et al. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*, 2013, 155: 1075-87
- [10] Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 2011, 469: 368-73
- [11] Marquardt S, Escalante-Chong R, Pho N, et al. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell*, 2014, 157: 1712-23
- [12] Wu X, Sharp PA. Divergent transcription: a driving force for new gene origination? *Cell*, 2013, 155: 990-6
- [13] Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Nat Genet*, 2012, 44: 212-21
- [14] Lercher MJ, Urrutia AO, Hurst LD. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet*, 2002, 31: 180-3
- [15] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 2008, 322: 1845-8
- [16] Adachi N, Lieber MR. Bidirectional gene organization: a common architectural feature of the human genome. *Cell*, 2002, 109: 807-9
- [17] Takai D, Jones PA. Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. *Mol Biol Evol*, 2004, 21: 463-7
- [18] Almada AE, Wu XB, Kriz AJ, et al. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, 2013, 499: 360-U141
- [19] Andersson R, Refsing Andersen P, Valen E, et al. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun*, 2014, 5: 5336
- [20] Flynn RA, Almada AE, Zamudio JR, et al. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci USA*, 2011, 108: 10460-5
- [21] Ntini E, Jarvelin AI, Bornholdt J, et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol*, 2013, 20: 923-8
- [22] Preker R, Nielsen J, Kammler S, et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, 2008, 322: 1851-4
- [23] Seila AC, Calabrese JM, Levine SS, et al. Divergent transcription from active promoters. *Science*, 2008, 322: 1849-51
- [24] Core LJ, Martins AL, Danko CG, et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet*, 2014, 46: 1311-20
- [25] Kapranov P, Cheng J, Dike S, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 2007, 316: 1484-8
- [26] Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, 129: 823-37
- [27] Guenther MG, Levine SS, Boyer LA, et al. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 2007, 130: 77-88
- [28] Arigo JT, Eyler DE, Carroll KL, et al. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell*, 2006, 23: 841-51
- [29] Thiebaut M, Kisseleva-Romanova E, Rougemaille M, et al. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell*, 2006, 23: 853-64
- [30] Yuan GC, Liu YJ, Dion MF, et al. Genome-scale identification of nucleosome positions in *S.cerevisiae*. *Science*, 2005, 309: 626-30
- [31] Albert I, Mavrich TN, Tomsho LP, et al. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature*, 2007, 446: 572-6
- [32] Mavrich TN, Jiang C, Ioshikhes IP, et al. Nucleosome organization in the *Drosophila* genome. *Nature*, 2008, 453: 358-62
- [33] Raisner RM, Hartley PD, Meneghini MD, et al. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, 2005, 123: 233-48
- [34] Scruggs BS, Gilchrist DA, Nechaev S, et al. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol Cell*, 2015, 58: 1101-12
- [35] Duttke SH, Lacadie SA, Ibrahim MM, et al. Human promoters are intrinsically directional. *Mol Cell*, 2015, 57: 674-84
- [36] Andersson R, Chen Y, Core L, et al. Human gene promoters are intrinsically bidirectional. *Mol Cell*, 2015, 60: 346-7
- [37] Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev*, 2002, 16: 2583-92
- [38] Core LJ, Waterfall JJ, Gilchrist DA, et al. Defining the status of RNA polymerase at promoters. *Cell Rep*, 2012, 2: 1025-35
- [39] Lam MT, Cho H, Lesch HP, et al. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 2013, 498: 511-5
- [40] Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 2012, 483: 295-301
- [41] Lepoivre C, Belhocine M, Bergon A, et al. Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics*, 2013, 14: 914
- [42] Duttke SH, Lacadie SA, Ibrahim MM, et al. Perspectives on unidirectional versus divergent transcription. *Mol Cell*, 2015, 60: 348-9
- [43] Takahashi H, Kato S, Murata M, et al. CAGE (cap analysis of gene expression): a protocol for the detection of

- promoter and transcriptional networks. *Methods Mol Biol*, 2012, 786: 181-200
- [44] Kodzius R, Kojima M, Nishiyori H, et al. CAGE: cap analysis of gene expression. *Nat Methods*, 2006, 3: 211-22
- [45] de Hoon M, Hayashizaki Y. Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. *Biotechniques*, 2008, 44: 627-8, 630, 632
- [46] Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol*, 2015, 16: 22
- [47] Nguyen T, Fischl H, Howe FS, et al. Transcription mediated insulation and interference direct gene cluster expression switches. *Elife*, 2014, 3: e03635
- [48] Nojima T, Gomes T, Grosso AR, et al. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*, 2015, 161: 526-40
- [49] Kwak H, Fuda NJ, Core LJ, et al. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 2013, 339: 950-3
- [50] Carbajo D, Magi S, Itoh M, et al. Application of gene expression trajectories initiated from erbb receptor activation highlights the dynamics of divergent promoter usage. *PLoS One*, 2015, 10: e0144176
- [51] Ni T, Corcoran DL, Rach EA, et al. A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat Methods*, 2010, 7: 521-7
- [52] Kim TH, Barrera LO, Zheng M, et al. A high-resolution map of active promoters in the human genome. *Nature*, 2005, 436: 876-80
- [53] Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol*, 2007, 14: 103-5
- [54] Pelechano V, Chavez S, Perez-Ortin JE. A complete set of nascent transcription rates for yeast genes. *PLoS One*, 2010, 5: e15442
- [55] Berg MG, Singh LN, Younis I, et al. U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, 2012, 150: 53-64
- [56] Kaida D, Berg MG, Younis I, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, 2010, 468: 664-8
- [57] Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes Dev*, 2011, 25: 1770-82
- [58] Preker R, Almvig K, Christensen MS, et al. PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res*, 2011, 39: 7179-93
- [59] Mischo HE, Proudfoot NJ. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta*, 2013, 1829: 174-85
- [60] Vasiljeva L, Buratowski S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell*, 2006, 21: 239-48
- [61] Marquardt S, Hazelbaker DZ, Buratowski S. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription*, 2011, 2: 145-54
- [62] O'Sullivan JM, Tan-Wong SM, Morillon A, et al. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*, 2004, 36: 1014-8
- [63] Ansari A, Hampsey M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, 2005, 19: 2969-78
- [64] Singh BN, Hampsey M. A transcription-independent role for TFIIB in gene looping. *Mol Cell*, 2007, 27: 806-16
- [65] Calvo O, Manley JL. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol Cell*, 2001, 7: 1013-23
- [66] Medler S, Al Husini N, Raghunayakula S, et al. Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor 1 subunits required for gene looping. *J Biol Chem*, 2011, 286: 33709-18
- [67] Sun ZW, Hampsey M. Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection *in vivo*. *Mol Cell Biol*, 1996, 16: 1557-66
- [68] Wang Y, Fairley JA, Roberts SG. Phosphorylation of TFIIB links transcription initiation and termination. *Curr Biol*, 2010, 20: 548-53
- [69] Lauberth SM, Nakayama T, Wu XL, et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell*, 2013, 152: 1021-36
- [70] Min IM, Waterfall JJ, Core LJ, et al. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev*, 2011, 25: 742-54
- [71] Naughton C, Corless S, Gilbert N. Divergent RNA transcription: a role in promoter unwinding? *Transcription*, 2013, 4: 162-6
- [72] Shahbazian MD, Grunstein M. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem*, 2007, 76: 75-100
- [73] Whitehouse I, Rando OJ, Delrow J, et al. Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, 2007, 450: 1031-5
- [74] Yadon AN, Van de Mark D, Basom R, et al. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Mol Cell Biol*, 2010, 30: 5110-22
- [75] Cohen BA, Mitra RD, Hughes JD, et al. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*, 2000, 26: 183-6
- [76] Dhadi SR, Krom N, Ramakrishna W. Genome-wide comparative analysis of putative bidirectional promoters from rice, *Arabidopsis* and *Populus*. *Gene*, 2009, 429: 65-73
- [77] Wei W, Pelechano V, Jarvelin AI, et al. Functional consequences of bidirectional promoters. *Trends Genet*, 2011, 27: 267-76
- [78] Lee JY, Khan AA, Min H, et al. Identification and characterization of a noncoding rna at the mouse *Pena* locus. *Mol Cells*, 2012, 33: 111-6

- [79] Hirota K, Miyoshi T, Kugou K, et al. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature*, 2008, 456: 130-4
- [80] Wang X, Arai S, Song X, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*, 2008, 454: 126-30
- [81] Luo S, Lu JY, Liu L, et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell*, 2016, 18: 637-52
- [82] Kim TK, Hemberg M, Gray JM, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 2010, 465: 182-7
- [83] De Santa F, Barozzi I, Mietton F, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*, 2010, 8: e1000384
- [84] Li W, Notani D, Ma Q, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 2013, 498: 516-20
- [85] Mousavi K, Zare H, Dell'Orso S, et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell*, 2013, 51: 606-17
- [86] Melo CA, Drost J, Wijchers PJ, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell*, 2013, 49: 524-35
- [87] Hsieh CL, Fei T, Chen Y, et al. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci USA*, 2014, 111: 7319-24
- [88] Pnueli L, Rudnizky S, Yosefzon Y, et al. RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α -subunit gene. *Proc Natl Acad Sci USA*, 2015, 112: 4369-74
- [89] Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*, 2011, 25: 1915-27
- [90] Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*, 2012, 22: 1775-89