

What is all that “junk” doing in our genomes?

“垃圾”DNA的奥秘

张常^{①②}, 王新文^③, 王亮^④, 高山^{①②*}

① 中国科学院苏州生物医学工程技术研究所, 苏州 215163;

② 中国科学院生物医学检验技术重点实验室, 苏州 215163;

③ 第四军医大学口腔医学院黏膜科, 西安 710032;

④ 中国科学院微生物研究所病原微生物与免疫学重点实验室, 北京 100101

* 联系人, E-mail: gaos@sibet.ac.cn

2016-03-23 收稿, 2016-05-17 修回, 2016-05-18 接受, 2016-08-19 网络版发表

国家自然科学基金(81472827)、中国科学院“百人计划”和中国科学院重点部署项目(KFZD-SW-204)资助

摘要 依据DNA双螺旋和遗传信息的中心法则, 只有编码蛋白的DNA才被认为是具有功能的。然而人类基因组大小与蛋白质数量矛盾以及生物体进化与基因组大小的不相关性(即C值悖论), 导致科学家对非编码DNA概念的提出, 戏称这些DNA为“垃圾”DNA。随着第二代测序技术发展以及数据分析能力的提升, 大量研究表明, 这些“垃圾”DNA很多都可以在生理与疾病状态下特异性转录。同时, 疾病全基因组关联研究显示, 大量与疾病相关单核苷酸多态性位点位于非编码区, 这共同证明了它们是有功能的。进一步研究表明, 人类基因组中至少75%的序列是转录的, 并将这些转录且不编码蛋白质的产物称为非编码RNA。本研究组系统总结了3大类非编码RNA包括微小RNA、长链非编码RNA和环状RNA的定义和分类以及在疾病中的功能, 对这些的非编码RNA的研究将会为疾病的治疗以及诊断方法开启新纪元。

关键词 “垃圾”DNA, 编码DNA, 非编码RNA, 长链非编码RNA, 环状RNA, 肿瘤

生命的孕育、生长、繁衍和死亡等一切生命现象是不以人的意志为转移的, 它们皆由基因控制。早在1860年, 遗传学家孟德尔通过豌豆(*Pisum sativum* Linn)实验提出了遗传定律。1953年, 詹姆斯·沃森(James Watson)和弗朗西斯·克里克(Francis Crick)发现并提出了DNA双螺旋结构模型, 标志着分子生物学时代的开启。随后, 遗传学“中心法则”的建立、补充和完善使人们认识了遗传信息是如何传递的。这个传递的过程包括核酸之间的双向传递和核酸到蛋白质的单向传递, 这两种传递方式的结合最终实现遗传信息在生物体内部以及世代间的传递。这一过程的发现暗示, 生命体基因组的绝大部分应该是编码蛋白质的DNA。

1 “垃圾”DNA的起因

1964年, 德国科学家弗里德利希·福格尔^[1](Friedrich Vogel)在*Nature*撰文称: 若根据当时估算的人类基因组大小, 按照“中心法则”和基因组都由蛋白质编码基因组成的原则, 那么一个人的基因数量有可能多达670万个, 很显然这个数字大得让人颤栗, 当时对基因组大小认识显示, 人类不能够制造出670万种蛋白质。同时, 普遍认为自然界中的生物体遵循从简单到复杂、从低等到高等进化的模式, 理论上生物体具有越来越复杂的生命活动, 遗传信息也必然会增加, 那就意味着基因组也会变大, 然而, 事实却并非如此。基因组的大小通常用C值(C-value)来表示, 研究发现, C值与生物体的复杂程度并不呈现

引用格式: 张常, 王新文, 王亮, 等. “垃圾”DNA的奥秘. 科学通报, 2016, 61: 3079–3084

Zhang C, Wang X W, Wang L, et al. The mystery of “junk” DNA (in Chinese). Chin Sci Bull, 2016, 61: 3079–3084, doi: 10.1360/N972016-00381

线性关系^[2]。例如,变形虫(*Amoeba proteus*,一种单细胞真核生物)的C值大约是人的C值200倍。于是,科学家将这种C值和生物的结构或组成的复杂性不匹配的现象称为“C值悖论”(C-value paradox)^[3]。

虽然当时科学家无法解释这一现象,但就生命体中存在非编码DNA这一观点达成共识。1972年,日本遗传学家大野乾(Susumu Ohno)^[4]提出了“垃圾”DNA (“junk”DNA)这一概念,用来描述基因组中不能编码蛋白质的DNA序列。这似乎对进化提出挑战,按常理进化都是“精准”演化,用则进,废则退。但“垃圾”DNA的存在,暗示大自然在基因组演化做得并不完美,具有一定的随意性和不确定性。20世纪90年代初期,由美国科学家发起,并由英国、法国、德国、日本和中国科学家相继先后加入,共同参与实施了人类基因组计划。经过科学家不断的努力,到21世纪初期,完成了人类基因组“天书”的解读。结果发现,人类的基因总数不超过25000^[5],在基因组中,能够编码蛋白的DNA只占到其中的1%~1.5%,再除去3%左右的调控编码基因的调控元件DNA,剩余的95%以上的DNA序列是没有任何功能的“垃圾”DNA。显然,基因组就像在一大片“垃圾”DNA组成的荒漠里,零星地点缀着编码基因及调控元件DNA。试想一下,当人们坐在电影院欣赏一部时长100 min的电影,结果95 min都是在播广告,仅有的5 min故事也被肢解在其中,人们还会津津有味坐在那里欣赏这个故事吗?摆在面前的两大问题:是真的利用如此之少的基因构成如此复杂的生命个体?基因组中“垃圾”DNA片段真的没有用吗?显然这两个问题是相互验证的。这好比人类对宇宙的认识,所认知的宇宙只占了宇宙的极少一部分,而大部分是由无法探测和研究的物质组成,这部分未知东西被称为“暗物质”,这些未知功能的“垃圾”DNA也被戏称为基因组中的“暗物质”。

2 “垃圾”DNA的解读

随着人类基因组计划的完成,美国国家人类基因组研究院(National Human Genome Research Institute, NHGRI)决定对人类基因组进行一番彻底的搜寻,看看其中究竟有多少DNA是没有意义、没有功能的“垃圾”DNA。得益于测序技术的进步,第二代测序技术极大地提高了测序通量,可以一次性完成从数十万到数百万的DNA分子测序,使得对一个物种

的基因组和转录组深度测序变得方便易行,为“垃圾”DNA的解读提供了技术的支撑。所以,在2003年启动的ENCODE (The Encyclopedia of DNA Elements)计划^[6]和2006年的癌症基因组计划(The Cancer Genome Atlas, TCGA)^[7]为这些非编码DNA片段的研究提供了极大的帮助。

众所周知,从统计学意义上来说,人与人之间的蛋白质编码基因有99%的相似性,但是,个体的“垃圾”DNA却有着显著的差异,这能够较好地解释为何基因组编码部分大体上相似,每个人却都存在不同。也就是说,是“垃圾”DNA让每个人都变得独一无二。实验发现,在这些“垃圾”DNA中存在大量微小的基因开关,控制着基因在细胞中的功能,构成基因调控网络里很重要的一环,这个环节异常,会导致疾病发生,所以这些解读证实人类基因组大部分是有功能的^[8]。很多过去被认为无功能的DNA,实际上有可能转录为RNA,或者作为转录因子结合位点,或者是DNA化学修饰的靶点,以多种形式对真正基因的表达起到了关键的调节作用。这些研究彻底颠覆了传统的以蛋白质编码基因为中心的基因组学观念,并将人类带入了一个全新的基因组时代,甚至有科学家将该计划的一系列研究结果比喻为人类基因组的Google地图。从下面几个小例子来探讨这些“垃圾”DNA的重要性。

首先,不少疾病的全基因组关联研究(genome-wide association study, GWAS)鉴定出许多DNA单核苷酸多态性(single nucleotide polymorphisms, SNPs)和这些疾病有一定的关联性,但这些SNPs将近90%位于传统上所谓的“垃圾”DNA区域里^[9]。ENCODE项目检查了5000多个这类SNPs,发现这些位点在一种或多种转录因子调节的区域。而且这些位点还与脱氧核糖核苷酸酶I(deoxyribonuclease I, DNase I)超敏位点相关联^[10],说明这些区域不是传统理解的沉默区。这只是ENCODE项目寻求方法以证明不存在所谓“垃圾”DNA的一个例子。另外,有科学家将人和小鼠(*Mus musculus*)的基因组序列进行比对,发现在非编码DNA中,有5%的序列是高度保守的。选取其中的一部分序列在大尺度上进行同源分析,发现这部分序列中超过25%的部分存在于10种以上哺乳动物的DNA序列中,保守性甚至比同源的编码蛋白质基因要强^[11]。这就意味着这些序列经历了几亿年演变而基本没有发生差异变化。这种保守性说明它们

对于物种来说有着非常重要的作用。同时,在人类基因组中,在基因区和基因之间都包含大量的重复序列,其含量占到基因组含量的30%~50%^[12],而且重复序列也与人类疾病的发生密切相关。例如,研究人员通过比对良性和恶性神经鞘瘤的全基因组DNA甲基化发现重复序列是低甲基化的,这些重复序列的异常甲基化与疾病的发生有一定的关联性^[13]。另外,Alu重复序列属于短散布核元件,在人类基因组中所占比例大约10%^[14]。研究发现,该序列为新的外显子的主要来源,并且是灵长类特异性的反转录转座子,实验表明,与肌肉营养失调有关基因*SEPN1* (*selenoprotein N1*),其中一个外显子来源于Alu^[15]。

更为重要的是,这些研究证实人类基因组75%都具有转录活性,且转录后产物具备一定的生物学功能^[5]。把这些转录而不编码蛋白的产物统称非编码RNA (non-coding RNA, ncRNA)。ncRNA包括核糖体RNA (ribosome RNA, rRNA)、转运RNA (transfer RNA, tRNA)、小核RNA (small nuclear, snRNA)、核仁小分子RNA (small nucleolar RNA, snoRNA)、微小RNA (micro RNA, miRNA)、长链非编码RNA (long non-coding RNA, lncRNA)、环状RNA (circular RNA, circRNA)等。当前,ncRNA的研究热点主要聚焦在这几大类,一类就是研究比较清楚的miRNA,一类是刚被重视起来的lncRNA,还有一类就是新崛起的circRNA。

2.1 miRNA

miRNA是一种内源性的小RNA,成熟的miRNA大小长约20~25 bp,不具备编码蛋白质的能力,但是能够调控编码基因的表达。这些成熟的miRNA是由初级转录产物经过一系列核酸酶的剪切加工而成的,随后组装进RNA诱导的沉默复合体(RNA-induced silencing complex, RISC),通过碱基互补配对的原则识别靶mRNA的5'或3'非翻译区域,从而引起靶mRNA降解或抑制其翻译活性^[16]。

miRNA广泛存在于生物体内,在序列结构上具有高度的保守性,但表达方式具有时空和组织特异性。1993年,科学家从秀丽隐杆线虫(*Caenorhabditis elegans*)体内鉴定出第一个miRNA: *lin-4*,证明其通过反式作用调控编码基因*lin-14*的活性,从而影响线虫的发育^[17]。随后,越来越多的miRNA被发现,并被证实参与各种生物学进程。例如,细胞生长、发育、

凋亡、代谢、激素信号及分化等。进而人们发现,miRNA的表达水平与多种癌症相关,该方向目前已成为肿瘤研究领域的热点。研究表明,miRNA基因主要位于染色体的脆性位点上,而染色体中的脆性位点区域更易引发癌症的滋生^[18]。大部分的miRNA在行使其生物学功能时扮演着抑癌基因或类似癌基因的角色。例如,Let-7在多种肿瘤中被证实也具有抑癌基因的作用,可以抑制癌基因*RAS* (rat sarcoma virus)^[19], *MYC* (v-myc avian myelocytomatosis viral oncogene homolog)^[20], *HMG A2* (high mobility group A2)^[21]等的活性。MiR-21作为一种类似癌基因的miRNA^[22],在多种癌症中过度表达以维持肿瘤的生长,还可以靶向一些肿瘤抑制基因和入侵基因,在增加肿瘤的增殖和迁移能力方面发挥重要作用。

2.2 lncRNA

lncRNA通常是指长度大于200个核苷酸的一类非编码RNA。基于lncRNA在与邻近其基因的相对位置和转录方向,可以将其分为5类:正义lncRNA、反义lncRNA、内含子lncRNA、基因间lncRNA、双向lncRNA。lncRNA与miRNA前体一样,都具有细胞和组织特异性,能够形成二级结构,并且转录后需要经过加工和修饰等特点。和mRNA相比,lncRNA的表达水平相对较低,而且在物种间的保守性较差。lncRNA已被证实在生物体内主要通过对表观遗传、转录调控及转录后加工等环节的干预发挥功能,在各种疾病发病过程中发挥重要作用。例如,在癌症中,lncRNA的异常表达影响癌细胞的增殖、凋亡和侵入等。同时,lncRNA的功能与表达的丰度存在很大关联,这有可能使其成为一种癌症诊断和预后评估的理想标志物^[23]。

目前,有几个lncRNA已经被选为人类癌症检测的生物标志物。其中新型前列腺抗原3 (prostate cancer antigen 3, PCA3)作为诊断前列腺癌的生物标志物已获美国食品药品监督管理局(Food and Drug Administration, FDA)批准。尤其把传统的前列腺特异抗原(prostate-specific antigen, PSA)和PCA3检测结合起来,展示了更高的准确率^[24]。另外,HOX转录反义RNA(HOTAIR)是一个位于HOX基因家族反义链的lncRNA^[25]。研究表明,HOTAIR在肺癌、结直肠癌、乳腺癌等多种常见癌症的引发和恶化中起重要作用,在癌组织中表达量均出现不同程度的异常。HOTAIR

通过对组蛋白进行表观遗传的修饰,从而导致了 *WIF-1* (wnt inhibitor factor 1), *PTEN* (phosphatase and tensin homolog deleted on chromosome ten)等基因的表达调控异常,影响了 *WNT* (wingless int1), *AKT* (protein kinase B)等信号通路,进而使肿瘤细胞获得侵袭转移、逃避生长抑制、抵抗细胞凋亡等特性。因此, *HOTAIR*可作为一种分子标记或靶点,在恶性肿瘤的早期诊断、疗效判断、预后预测等方面具有广阔的临床应用前景。

另外,研究还发现,在预测的 *lncRNA* 序列内部发现了具有编码能力的微肽。例如,道格拉斯·安德森 (Douglas Anderson) 和本杰明·尼尔森 (Benjamin Nelson) 等人先后发现了两个在骨骼肌特异表达的 *lncRNA* 序列中,分别包含一段可以编码46个氨基酸的微肽 *myoregulin*^[26] 和34个氨基酸的肽 *dwarf open reading frame* (*DWORF*)^[27], 它们对骨骼肌的生理具有重要的调控作用。

2.3 circRNA

circRNA 是一种内源性的、非线性的非编码RNA, 与传统的线性RNA不同, *circRNA* 由反向剪接产生, 具有封闭的环状结构, 能够抑制核酸外切酶或 *RNase R* 的降解, 并且缺乏5'端的帽子结构和3'端的 *poly(A)* 尾。*circRNA* 主要来自编码基因的外显子, 可能仅有

一个外显子构成, 也可能由多个外显子构成。*circRNA* 在组织中广泛表达, 具有发育阶段特异性和物种的保守性。

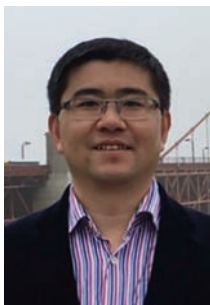
早在19世纪80年代, *circRNA* 就被科学家所发现^[28], 但是由于中心法则的盛行, 不具备编码能力的环状RNA在当时根本不受重视。进入21世纪, 得益于非编码RNA研究的兴起和深度测序技术以及生物信息学的快速发展, 才使得 *circRNA* 渐渐浮出水面。目前发现, 一些 *circRNA* 在细胞质中可充当 *miRNA* 海绵分子^[29]。例如, *ciRS-7* 主要在脑中尤其是小脑中大量表达, 它包含超过70个 *miR-7* 的结合位点, 通过结合 *miR-7* 抑制其活性, 从而影响 *miR-7* 靶基因的活性, 而 *miR-7* 又直接或间接调控癌症相关的信号通路中致癌相关蛋白活性^[30]。*circRNA* 还可作为RNA结合蛋白 (RNA binding protein, RBP) 的隔绝子, 或核内翻译的调控子, 是基因表达调控网络的重要参与者^[31]。

伊萨克·牛顿早就说过, “自然不行徒劳之举, 少已够用, 多则何益”。非编码RNA既然随着物种的进化并没有消失, 肯定是“天生我才必有用”。而且现在越来越多的科学研究已经证明了各类非编码RNA存在的生理意义及重要的应用价值。随着科学技术的不断进步、发展和科学家坚持不懈的努力, 我们期待真正理解、掌握并最终运用生命这部“天书”, 为人类生命科学研究作贡献。

参考文献

- 1 Vogel F. A preliminary estimate of the number of human genes. *Nature*, 1964, 201: 847
- 2 Thomas C A Jr. The genetic organization of chromosomes. *Annu Rev Genet*, 1971, 5: 237–256
- 3 Eddy S R. The C-value paradox, junk DNA and ENCODE. *Curr Biol*, 2012, 22: R898–R 899
- 4 Ohno S. So much “junk” DNA in our genome. *Brookhaven Symp Biol*, 1972, 23: 366–370
- 5 Djebali S, Davis C A, Merkel A, et al. Landscape of transcription in human cells. *Nature*, 2012, 489: 101–108
- 6 ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. *Science*, 2004, 306: 636–640
- 7 Cancer Genome Atlas Research Network, Weinstein J N, Collisson E A, et al. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat Genet*, 2013, 45: 1113–1120
- 8 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012, 489: 57–74
- 9 Schaub M A, Boyle A P, Kundaje A, et al. Linking disease associations with regulatory information in the human genome. *Genome Res*, 2012, 22: 1748–1759
- 10 Thurman R E, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. *Nature*, 2012, 489: 75–82
- 11 Dermitzakis E T, Reymond A, Scamuffa N, et al. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science*, 2003, 302: 1033–1035
- 12 Faulkner G J, Kimura Y, Daub C O, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet*, 2009, 41: 563–571
- 13 Feber A, Wilson G A, Zhang L, et al. Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. *Genome Res*, 2011, 21: 515–524
- 14 Schmid C W. Alu: A parasite's parasite? *Nat Genet*, 2003, 35: 15–16

-
- 15 Lin L, Shen S, Tye A, et al. Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet*, 2008, 4: e1000225
 - 16 Bartel D P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 2004, 116: 281–297
 - 17 Lee R C, Feinbaum R L, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 1993, 75: 843–854
 - 18 Calin G A, Sevignani C, Dumitru C D, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA*, 2004, 101: 2999–3004
 - 19 Johnson S M, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. *Cell*, 2005, 120: 635–647
 - 20 Sampson V B, Rong N H, Han J, et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res*, 2007, 67: 9762–9770
 - 21 Park S M, Shell S, Radjabi A R, et al. Let-7 prevents early cancer progression by suppressing expression of the embryonic gene *HMG2*. *Cell Cycle*, 2007, 6: 2585–2590
 - 22 Si M L, Zhu S, Wu H, et al. MiR-21-mediated tumor growth. *Oncogene*, 2007, 26: 2799–2803
 - 23 Yarmishyn A A, Kurochkin I V. Long non-coding RNAs: A potential novel class of cancer biomarkers. *Front Genet*, 2015, 6: 145
 - 24 Bussemakers M J, van Bokhoven A, Verhaegh G W, et al. Dd3: A new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res*, 1999, 59: 5975–5979
 - 25 Hajjari M, Salavat A. HOTAIR: An oncogenic long non-coding RNA in different cancers. *Cancer Biol Med*, 2015, 12: 1–9
 - 26 Anderson D M, Anderson K M, Chang C L, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, 2015, 160: 595–606
 - 27 Nelson B R, Makarewich C A, Anderson D M, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*, 2016, 351: 271–275
 - 28 Sanger H L, Klotz G, Riesner D, et al. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci USA*, 1976, 73: 3852–3856
 - 29 Hansen T B, Jensen T I, Clausen B H, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013, 495: 384–388
 - 30 Lasda E, Parker R. Circular RNAs: Diversity of form and function. *RNA*, 2014, 20: 1829–1842
 - 31 Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 2013, 495: 333–338



高山

牛津大学医学肿瘤学博士，中国科学院苏州生物医学工程技术研究所研究员，博士生导师。中国科学院“百人计划”获得者和江苏省“六大人才高峰”入选者。现任中国科学院生物医学检验技术重点实验室副主任和生物标志物研究中心主任。长期从事肿瘤的分子机制和生物标志物的研究，以及新型生物技术研发，注重基础医学与转化医学相结合，逐渐在这些领域形成了具有特色和影响力的系统性研究体系。

The mystery of “junk” DNA

ZHANG Chang^{1,2}, WANG XinWen³, WANG Liang⁴ & GAO Shan^{1,2}

¹ Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China;

² Key Laboratory of Bio-medical Diagnostics, Chinese Academy of Sciences, Suzhou 215163, China;

³ Department of Oral Medicine, School of Stomatology, The Fourth Military Medical University, Xi'an 710032, China;

⁴ CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

The basic genetic law of life is built based on DNA double helix model and the central dogma of molecular biology (i.e. DNA-RNA bidirectional transcription and RNA-protein translation), in which DNA and most of RNA carry coding information, and proteins make up the structure of the body and carry out most of biological functions. This describes the flow of genetic information within and between individual and protein as the functional molecules in life, so DNA that codes for protein (known as exon) is functional in view of these rules. In this review, we provide an overview of the origin of “junk” DNA and further discuss how “junk” DNA functions in depth. With a glimpse on landscape of human genome, only very small fractions are protein coding DNA in our book of life. By contrast, the large fractions are non-coding DNA, which cannot be translated into proteins and have been assumed that such DNA do not contain any information nor have function. Also it has been found that the genome size of organism does not correlate well with the complexity of organism, suggesting large amounts of non-coding DNA exist in lower organism. Such non-coding DNA in organism genome are regarded as uselessness and commonly referred to as “junk” DNA. However, with the advent of next generation sequencing technologies and ability to improvement of analyzing data, these provide the possibility to systematically understand so called “junk” DNA. First, genome-wide association studies have successfully identified many single nucleotide polymorphisms (SNPs) underlying susceptibility to diseases; however, the majority of SNPs locate in non-coding region of genome. Moreover, the parts of non-coding DNA are highly conserved between human and mice. All of these suggest non-coding DNA are functional in some way. Second, it has been revealed that about 75 percent of our genome is actually transcribed. Such transcripts that do not code any protein are termed as non-coding RNAs (ncRNAs). These ncRNAs, such as canonical transfer and ribosomal RNAs, as well as the recently identified microRNAs (miRNAs), long non-coding RNAs (lncRNAs) circular RNAs (circRNAs) etc, have been shown to play the important physiological function in organism. Also the deregulation of these ncRNAs has been found to have relevance not only to tumorigenesis, but also to neurological, cardiovascular, developmental and other diseases. Here we further discuss the rapidly advancing fields of miRNA, lncRNA and circRNA in detail. We summarize their production, gene structure and organization in the genome and diverse functions. Although miRNA has been well studied in last decade, we are still in early step of understanding the nature and extent of the involvement of other ncRNAs in physiology and disease. This will shed light on great advances in therapeutic strategies and diagnostic approaches based on the understanding on the molecular mechanisms of ncRNAs.

“junk” DNA, coding DNA, non-coding RNA, long non-coding RNA, circular RNA, tumor

doi: 10.1360/N972016-00381