

DOI: 10.3724/SP.J.1005.2013.00035

# 后生动物非编码保守元件

冯俊, 李光, 王义权

厦门大学生命科学学院, 细胞应激生物学国家重点实验室, 厦门 361005

**摘要:** 生物体基因组中除了编码序列之外, 还存在大量的非编码调控序列。比较基因组学研究发现: 脊椎动物、尾索动物、头索动物、果蝇、线虫等基因组中存在保守的非编码调控序列。这些非编码保守元件通常分布在与转录调控发育相关的基因上下游区域, 作为基因调控网络核心的一部分, 常常在基因表达过程中扮演转录增强子的角色。文章总结了近年来有关后生动物非编码保守元件的发现和主要特点, 并进一步就非编码保守元件在大规模基因组倍增之后的演化及其在生物躯体图式进化过程中的影响进行了综述。

**关键词:** 顺式调控; 非编码保守元件; 比较基因组学; 增强子; 躯体图式

## Research progress of conserved non-coding elements in metazoan

FENG Jun, LI Guang, WANG Yi-Quan

State Key Laboratory of the Cellular Stress Biology, School of life science, Xiamen University, Xiamen 361005, China

**Abstract:** In addition to protein coding sequence, the organism genome contains a significant amount of regulatory DNA. Comparative genomics reveals that the organism genomes of vertebrates, tunicate, cephalochordate, flies, and nematodes contain *cis*-regulatory elements with highly conserved non-coding elements (CNEs). CNEs that cluster around trans-dev genes are part of core gene regulatory networks (GRNs), and usually, they can act as transcriptional enhancers. In this review, we described the identification of CNEs and summarized their key properties across the metazoans, and then discussed the evolution of CNEs after large-scale genome duplication events and the role of CNEs in the evolution of animal body plan.

**Keywords:** *cis*-regulation; conserved non-coding elements; comparative genomics; enhancer; body plan

随着大规模基因组测序和比较基因组学的发展, 人们发现基因组中非编码保守调控序列(Conserved non-coding element, CNE)的数量远比想象的要多。这些原来被认为是垃圾DNA(Junk DNA)的序列很可能有重要的生物学功能, 参与多种生理生化过程。

例如 CNE 能作为一种“活跃分子”改变基因组整体 G+C 含量<sup>[1]</sup>, 为中期染色体带型提供了结构基础, 如 CpG 岛、DNA 环、G 或 R 带型为主的基质附着位点<sup>[2]</sup>; 同时作为 DNA 分子的重要组成部分, CNE 也是生物进化的物质基础之一<sup>[3]</sup>。

收稿日期: 2012-06-30; 修回日期: 2012-09-10

基金项目: 国家自然科学基金项目(编号: 30830023; No. 31071110)和教育部博士点基金(编号: 20110121120002)资助

作者简介: 冯俊, 硕士研究生, 专业方向: 动物发育遗传学。Tel: 0592-2184427; E-mail: junvon@126.com

通讯作者: 王义权, 教授, 博士生导师, 研究方向: 发育遗传与比较基因组学。E-mail: wangyq@xmu.edu.cn

网络出版时间: 2012-11-21 9:25:09

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20121121.0925.001.html>

进化上 CNE 具有高度的保守性, 与功能基因的 DNA 序列一样受纯净选择(Purifying selection)的作用<sup>[4]</sup>, 但不像非功能区域那样, 由于随机突变和漂移, 随着物种间系统演化距离变远而差别越大。因此, 借助一定的运算法则处理比对序列, 锚定演化速度较慢的序列, 能够使 CNE 从演化相对快的序列背景中突显出来。尽管 CNE 不产生在 RNA 或者蛋白水平上的功能性转录子, 但是在基因调控中可能发挥着重要的作用。因此非编码保守 DNA 近年来已经成为顺式调控研究领域的热点之一, 在越来越多的物种以及进化距离甚远的物种之间发现大量的 CNE。最新研究还发现 CNE 自身也在进化, 它们在动物躯体图式(Body plan)进化过程中扮演着重要的角色<sup>[5, 6]</sup>。

## 1 脊椎动物中的非编码保守序列

比较基因组学作为一种便捷的手段能够在庞大的基因组中对相对含量很少的非编码保守序列进行鉴别和功能注释。随着更多物种基因组数据的获得和比较基因组分析软件的开发, 人们发现了许多有潜在功能的非编码保守 DNA 序列。King 等<sup>[7]</sup>用已知的哺乳动物  $\beta$  珠蛋白( $\beta$ -globin)基因的顺式调控区域来评估几种比较基因组方法的效率, 发现几种比较基因组学方法能鉴别出 50%~60%的已知顺式调控区域。近来, Su 等<sup>[8]</sup>又对现在常用的 4 种鉴别顺式调控序列的方法(窗口集群法 Window clustering、概率建模法 Probabilistic modeling、系统发生足迹 Phylogenetic footprinting 和差别建模法 Discriminative modeling)进行了综合测评, 指出应根据不同的情况综合使用这些方法来弥补其中某一种方法的不足。然而, 如果顺式调控元件的序列与其反式作用因子的结构及其时空表达提供的信息不足, 对于有效地鉴别顺势调控模块(*cis*-regulatory modules)仍将是一个挑战<sup>[6]</sup>。

人类和河豚(*Fugu rubripes*)基因组的公开使得第一次直接比对进化距离较远的脊椎动物之间的 DNA 序列成为可能, 尽管人与河豚之间的分化时间约有 4.5 亿年, 但是研究发现在人和河豚的基因组中仍存在同线性(Synteny)高度保守的非编码序列(Highly conserved non-coding elements, HCNEs)<sup>[9]</sup>。这些保守元件总长 273 kb, 广泛分布于除 21 号和 Y

染色体外的所有染色体上, 长度在 93~740 bp 范围内, 序列一致性为 74%~98%。此外, 在脊椎动物基因组中, 保守的非编码调控序列只占非编码序列的 3%左右<sup>[10~12]</sup>。通过对人、小鼠和大鼠基因组的比较分析发现: 在直系同源区域中发现 481 个长度超过 200 bp、序列一致性达 100%的超保守元件(Ultra-conserved elements, UCEs), 这些保守元件广泛分布于基因组中, 几乎在哺乳动物、鸟类和鱼类中也都是保守的<sup>[13]</sup>。2003 年, Margulies 等<sup>[14]</sup>将 11 种非人脊椎动物基因组中与人染色体 7q31 上同源的 1.8 Mb 间隔序列进行比较, 也发现许多调控序列都含有高度保守序列。这些区域与所有编码区和部分非编码区有交叠, 可能与转录因子结合位点簇、非编码 RNA 转录子和其他一些功能元件相联系。但是受到可用于分析的基因组数量限制, 目前尚不能够在人类基因组中广泛识别这种多个物种均保守的序列。由于不同研究者采用的研究手段不尽相同, 如选择的参数(包括 CNE 的长度和相似百分数)改变, 赋予了这些非编码保守区很多名字, 例如 CNEs、CNS、deeply conserved elements、ultra-conserved、extremely conserved、extremely highly conserved sequences 和 hyper-conserved sequences 等<sup>[15]</sup>。

事实上, 非编码保守区的保守性比蛋白编码序列还要高, 这种保守性对于揭示物种间古老的关联可能十分有价值<sup>[9, 16, 17]</sup>。CNE 的核苷酸变化模式表明 CNE 受到极强的选择压力<sup>[12, 18]</sup>, 与中性进化的 DNA 序列相比, 估计某些 CNE 至少要减少了约 300 次的丢失<sup>[19]</sup>。因此, 存在于基因组中的 CNE 必定承担着极其重要的功能。然而, CNE 并不是均匀地分布在基因组中, 往往簇集在转录调控发育基因(Trans-dev genes)的周围<sup>[9, 17, 20]</sup>。CNE 和发育调控基因这种关联可以通过进一步分析它们在脊椎动物基因组中的倍增情况来加以验证<sup>[9, 17, 20, 21]</sup>。在基因组倍增之后, 尽管其他“旁观基因”(By-stander gene)通常会从含有 CNE 的调控发育复制区域丢失, 但是 CNE 却一直存在这些基因的周围<sup>[21, 22]</sup>。CNE 通常位于它的靶基因上下游约 1~2 Mb 的区域中<sup>[22]</sup>, 在小鼠和斑马鱼中进行功能验证时表现出转录增强子(Transcriptional enhancer)的活性<sup>[23, 24]</sup>。Woolfe 等<sup>[9]</sup>比对人及河豚基因组, 识别到近 1 400 个 CNEs, 利用斑马鱼胚胎在体内验证了 4 个相互没有关联的调

控发育基因(*Sox21*、*Pax6*、*Hlx9* 和 *Shh*)附近的 25 个 CNEs, 最终发现有 23 个元件在一个或多个组织中表现出增强子活性。Pennacchio 等<sup>[23]</sup>对人-河豚保守的 CNEs 或人-脊椎动物保守的 CNEs 共 167 个元件进行体内增强子活性验证, 发现 45%的元件在胚胎中发挥组织特异性增强子(Tissue specific enhancer)的作用。尽管如此, 我们至今还没有完全弄清楚 CNE 调控转录发育基因的选择性机制。

## 2 无脊椎动物中的非编码保守序列

尽管发育调控基因在整个后生动物(Metazoan)高度保守, 但是 CNE 在这些类群中却不是这样的情况。2005 年 Woolfe 等<sup>[9]</sup>用人和河豚中高度保守的非编码序列对数据库中无脊椎动物(包括海鞘、果蝇和线虫)基因组序列进行比对, 结果没有发现任何显著地同源性的序列。唯一的例外是, 在无脊椎的头索动物(Cephalochordate)文昌鱼(*Branchiostoma*)基因组中发现了脊椎动物与文昌鱼之间同源的 CNE<sup>[25-27]</sup>。最近, 我们实验室发现脊椎动物除了与头索动物存在同源的 CNE 之外, 与尾索动物(Urochordate)海鞘(*Ciona intestinalis*)和棘皮动物(Echinoderm)海胆(*Strongylocentrotus purpuratus*)之间也存在同源 CNE。此外无脊椎动物基因组自身也存在一些 CNEs, 这些发现主要来自果蝇(*Drosophila*)、线虫(*Caenorhabditis elegans*)和文昌鱼。

### 2.1 原口动物

线虫和果蝇作为原口动物(Protostome)中代表性的模式生物, 广泛应用于基因调控网络(Gene cis-regulatory networks, GRNs)研究, 已在它们的基因组中发现大量的 CNEs。Vavouri 等<sup>[28]</sup>比较了 3 种线虫(*Caenorhabditis elegans*, *C. briggsae*, *C. remanei*)基因组数据, 发现线虫基因组中存在数以千计的 CNE。Glazov 等<sup>[29]</sup>比对 2 种果蝇(*Drosophila melanogaster*, *D. pseudoobscura*)和一种冈比亚按蚊(*Anopheles gambiae*)的 *Hth* 基因组发现: CNE 受极强的纯净选择压力。此外, 有研究发现线虫的 CNE 与脊椎动物的 CNE 一样存在相似的核酸频率模式(Nucleotide frequency pattern), 尽管不同物种基因组的 AT 核苷酸含量变化很大, 但是 CNE 的 AT 核苷酸含量却惊人的相似<sup>[28, 30]</sup>。已在线虫中发现大量的 CNE

位于性染色体上<sup>[28]</sup>, 这种现象暗示性染色体上的发育调控基因在生殖细胞(Germ line)是沉默的, 而且这些基因中不包含管家基因(Housekeeping gene)<sup>[31]</sup>。

从进化角度来看, 调控相同功能基因的 CNE, 在脊椎动物和无脊椎动物中具有普遍相关性<sup>[28]</sup>, 人类与 CNE 相关的 156 个基因中, 有 40 个基因的直系同源基因在无脊椎动物中也与 CNE 相关。这些 CNEs 调控一组普遍的核心发育基因, 但不完全一致, 它们在决定各种不同的生物演化方式的基因调控网络中发挥着重要作用。在线虫中已鉴定的 CNEs 多半是一些已知的增强子<sup>[28]</sup>。Kuntz 等<sup>[32]</sup>对线虫 *Hox* 基因周围的 11 个非编码保守区进行线虫转基因功能验证, 发现其中 9 个是组织特异性的增强子(Tissue-specific enhancer)。此外, 对 4 种果实果蝇科(Tephritidae)果蝇的一个 *Even-skipped* 发育调控基因的上下游区域(Surrounding region)测序发现一些保守岛(Islands of conservation)上所含的保守序列也与发育调控基因的转录增强子相吻合<sup>[33]</sup>。

### 2.2 原索动物

原索动物(Protochordate)包括头索动物文昌鱼和尾索动物海鞘, 文昌鱼位于脊索动物门的基部, 而海鞘现在则被认为是最接近脊椎动物的无脊椎动物<sup>[34]</sup>。早期对脊椎动物和无脊椎动物基因组比较相信只有脊椎动物基因组中存在 CNE, 然而, Wang 等<sup>[25]</sup>通过对白氏文昌鱼(*Branchiostoma belcheri*)*Pax1/9* 基因相关的旁系同源域(Paralogon)的研究首次在非脊椎动物文昌鱼中发现与脊椎动物 CNE 同源的区域, 而文昌鱼和脊椎动物的进化距离已经超过 8 亿年<sup>[35]</sup>。由于文昌鱼基因组没有经历过脊椎动物起源时的两轮基因组倍增<sup>[36]</sup>, 是研究这些顺式调控元件古老功能的最佳模型, 此后又有多篇文昌鱼中 CNE 的研究论文报道。Amemiya 等<sup>[37]</sup>比对佛罗里达文昌鱼(*B. florida*)和 4 种脊椎动物的 *Hox* 基因簇, 发现该基因簇 3'端的保守性更强。随后, Pascual 等<sup>[38]</sup>对佛罗里达文昌鱼和欧洲文昌鱼(*B. lanceolatum*)的 *Hox* 基因簇比较也发现该基因簇的 3'端的保守性要强, 表明这些区域的保守性可能与 *Hox* 基因参与前后轴分化(Anterior-posterior axis patterning)和中枢神经系统(CNS)形成有关。2008 年佛罗里达文昌鱼基因组测序完成<sup>[39]</sup>, Holland 等<sup>[26]</sup>对人和佛罗里达文昌鱼的基

基因组进行全基因组比对发现 56 个 CNE, 进一步对文昌鱼 *Znf503/703* 基因的 CNE 和人的同源 CNE 在小鼠和文昌鱼胚胎中进行功能验证, 发现这些保守序列有 *Znf503/703* 基因组织特异性的增强子功能。然而, CNE 介导的表达谱并不与 *Znf503/703* 基因内源性表达谱完全一样, 这也许是因为 CNE 毕竟只有几十到几百个核苷酸的长度, 一些指导基因表达的功能已经丢失, 或者在进化过程中, 这些序列局部碱基的变化可导致基因功能上的分化。虽然全基因组比对能够更加广泛地发现保守序列, 但是对一些特殊的直系同源(Homologous)或者旁系同源(Paralogous)区域进行分析能够发现更多的保守序列。运用不同的计算方法, Hufton 等<sup>[27]</sup>对小鼠、河豚、斑马鱼和文昌鱼 44 个物种的基因组进行了再次比较, 从中发现一千多个最小长度为 45 bp 的进化保守的非编码元件(Phylogenetically conserved non-coding elements, PCNEs), 对其中的 42 个 PCNEs 在斑马鱼胚胎中进行功能验证, 发现大约 45% 的序列表现出增强子活性。这些结果表明增强子功能和转录调控机制的保守性长达 8 亿年之久。然而, 最近 Punnamoottil 等<sup>[40]</sup>发现文昌鱼 *Hox4* 基因的 CNE 在转基因斑马鱼中并没有表现出与 *Hox4* 基因一致的表达谱, 进一步对文昌鱼 *Hox4* 基因研究发现: 文昌鱼 *Hox4* 基因的调控区域存在一个 Cdx 类似的转录因子结合基序(Binding motif), 而转录因子 Cdx 对 *Hox4* 基因的表达起负调控作用<sup>[41]</sup>。这些结果给我们提供重要暗示, 尽管这些实验能够用来验证增强子的活性, 但是却不能辨别基因的负调控和非特异性表达之间的不同。

最近, 我们对白氏文昌鱼、海鞘、海胆(*Strongylocentrotus purpuratus*)以及脊椎动物的一些代表物种, 如人、狗(*Canis familiaris*)、小鼠、河豚(*Fugu rubripes*)、红原鸡(*Gallus gallus*)和爪蟾(*Xenopus tropicalis*)等的 *Pax* 基因家族进行生物信息学分析, 在海鞘和海胆中找到了与脊椎动物 *Pax* 基因转录调控同源的 CNE, 至此, 我们将 CNE 的保守性延伸至棘皮动物(或者说“全部脊索动物和棘皮动物”)。由于进化距离不同, 这些 CNE 在各物种之间的分布并不一样。与脊椎动物相比, 3 种无脊椎动物所含有的 CNE 明显要少, 这说明脊椎动物的许多 CNE 是基因组倍增之后产生的。有趣的是海鞘似乎比文昌鱼有更多的 CNE, 这可能从另一方面说明了海鞘比文昌

鱼更接近脊椎动物<sup>[34]</sup>。我们现正在针对这些保守元件的功能, 分别研究它们在文昌鱼和斑马鱼胚胎发育中的作用, 已发现这些序列都表现出极强的与 *Pax* 基因相关的组织特异性增强子的功能。我们拟通过研究这些 CNE 在脊椎动物系统(斑马鱼)和无脊椎系统(文昌鱼)中的表达谱是否发生功能分化来验证特定基因调控网络的演化等科学问题。

### 3 非编码保守元件的特点

#### 3.1 进化距离和保守性

CNE 一般是“寿命短”(Short lifetime)的, 并且绝大部分又是门系特异的(Phylum-specific)<sup>[4]</sup>。在脊椎动物与除海鞘、海胆、文昌鱼以外的无脊椎动物, 即原口类无脊椎动物比较中还没有找到过同源 CNE, 这些元件绝大部分单独地分布在脊椎动物、果蝇、线虫以及植物等基因组中<sup>[42]</sup>。大多数哺乳动物的 CNE 似乎在四足动物(Tetrapod)形成的早期才出现<sup>[43]</sup>, 它们在哺乳动物进化期间几乎没有变动, 这些 CNE (200 bp 范围内)的保守性能达到 100%, 而在鸡和哺乳动物之间的保守性也达到 95%<sup>[17]</sup>。这种保守性意味着功能上的重要性, 一些 CNE 可能由于其微小改变将对生命活动产生很敏感影响, 因此在纯净选择压力下保持着序列的一致, 此时可以将序列保守性作为序列功能性探测的一个标记。然而 CNE 的保守性不仅仅体现序列本身, 它与所调控基因之间的距离以及空间上也体现着保守性, 人们已发现哺乳动物 CNE 之间的距离与其它基因组元件之间的位置相比较要更保守, 这种保守性呈现双峰分布, 即一类 CNE 的间隔是高度保守的, 另一类 CNE 的间隔则与非哺乳类脊椎动物与人基因组大小比值相关<sup>[44]</sup>。

#### 3.2 组成

CNE 与编码区域和其他表达序列不同, 在其边缘或者侧翼区域有明显的特定核苷酸频率信号。无脊椎动物和植物 CNE 的 A/T 含量要比侧翼区域(Flanking regions)的 A/T 含量明显要高出大约为 6%<sup>[28, 30, 45]</sup>。脊椎动物 CNE 的 A+T 含量也比它所处周围的区域要高, CNE 的两端富含 A+T 的基序(Motif), 位于 CNE 两端的侧翼序列 A+T 含量陡降, 碱基组成的这一转变标记了侧翼 DNA 到 CNE 的转变<sup>[46]</sup>。CNE 的序列大多含有类似转录因子结合位点(Transcription

factor binding sites), 不含有 CpG 岛, 但是这些类似的结合位点只有非常少的部分能够归类到已知转录因子结合位点家族中<sup>[47]</sup>。

### 3.3 复制

许多 CNE 存在于基因组中不同的旁系同源基因(Paralogous genes)区域, 这些 CNE 被称之为“复制的非编码保守元件(Duplicated conserved non-coding elements, dCNEs)”。研究发现在全基因组倍增后, 一些旁系同源域保留了一些重要的顺式调控元件, 诸如增强子(Enhancer), 沉默子(Silencer), 绝缘子(Insulator)等<sup>[21, 45, 48, 49]</sup>。已发现 dCNE 大量存在于哺乳动物基因组中, 尽管继脊椎动物起源时的两轮基因组加倍后, 硬骨鱼类(Teleost)又发生了一次支系特异的基因组倍增, 但在它们基因组中这种 dCNEs 却很少见<sup>[21]</sup>。

### 3.4 在基因组中的分布

CNEs 在基因组中的分布不是随机的。虽然不同动物类群之间的 CNE 序列上缺少明显的同源性, 但它们在物种基因组中的分布却非常相似, 大都聚集在调控发育基因(Trans-devo gene)、管家基因和神经系统相关基因的周围<sup>[17, 20, 28, 50, 51]</sup>。在脊椎动物中, CNE 虽然均匀地分布于靶基因的上下游区域, 但是在远离基因编码区的调控区域密度却更高<sup>[52]</sup>。CNE 在占人类基因组 25% 的基因荒漠区(Gene deserts)中分布数量要比其他区域高出至少 4 倍<sup>[11, 53]</sup>。这些荒漠区的边缘, 存在许多已知与发育相关的基因, 如: *Otx2*、*Dach1*、*Sall1*、*Sox2*<sup>[54-57]</sup>, 这一现象提示了 CNE 在发育调控中的作用。

### 3.5 增强子功能

目前, 通过体内检测的方法发现后口动物中绝大部分 CNE 扮演着转录增强子的角色, 承担远距离或者近距离基因表达调控的任务<sup>[9, 15, 26, 27, 58]</sup>。我们将后口动物中一些 CNE 功能的研究做一归纳, 可以发现在不同物种中, 不同发育调控基因周围的绝大部分 CNE 作为增强子调控着相应基因的表达(表 1)。

## 4 非编码保守元件的演化

为什么基因组中这些 CNE 在经历如此漫长的进化距离后还这么保守? 研究发现其保守性并不是

因为这些区域突变率低或者重组过程中没有发生分离<sup>[12, 67]</sup>。事实上, 人与小鼠基因组的比较研究时就已经发现 CNE 的保守性与连锁不平衡(Linkage disequilibrium)没有关系, 说明自然选择对这些区域作用<sup>[68]</sup>。最近一项对脊椎动物非编码保守元件(保守非外显子元件, Conserved non-exonic elements, CNEEs)的研究发现基因调控元件进化经历了 3 次革新时期: 起源于 3~6 亿年的 CNEEs 与发育调控基因(Developmental regulatory gene)相关; 起源于 1~3 亿年的 CNEEs 与胞外信号受体基因(Extracellular signaling gene)相关; 起源于 0.5~1 亿年的 CNEEs 与翻译后修饰基因(Post-translational modification gene)相关<sup>[69]</sup>。这些结果表明 CNE 的保守性是由于其受到了高度适应的功能性约束(Highly adaptive functional constraint)。

现在已有模型用于解释这些非编码保守的顺式调控元件是如何进化的。由 Ohno<sup>[36]</sup>提出的经典基因复制模型认为, 基因复制为基因的新功能化(Neofunctionalization)、无功能化或假基因化(Nonfunctionalization or pseudogenization)提供了基础, 因为基因复制后产生的新复制子不受严格的选择压力。与这个模型一致的是顺式调控元件在基因组倍增的同时也发生复制, 同样也会经历这样的进化历程。对旁系同源基因保守非编码调控元件的研究发现, 复制的非编码保守元件(dCNEs)所调控的基因表达谱大多是重叠的<sup>[21, 70]</sup>。然而, 还有一些 dCNEs 的表达谱却不一样, 这表明 dCNE 序列微小的变化也会带来功能上的分歧(Functional divergence)。McEwen 等<sup>[21]</sup>认为一些脊椎动物的 CNE 家族能够追溯到古老的基因组复制事件中, 他们对哺乳动物和河豚中保守的约 2 300 个 CNE 重新鉴定发现了 124 个 dCNE 家族, 这些 dCNE 家族与一组旁系同源基因关联, 共同参与转录、发育和环境应答等生物学过程。尽管古老的基因组复制能够解释一小部分 CNE 的形成机制, 但 CNE 数量和基因家族数量之间的巨大差距说明 CNE 的演化应该还可能还有其他生物学机制。

目前, 人们认为哺乳动物中众多 CNE 家族成员中的一部分很可能是由转座因子(Transposable element)产生的。Doolittle 等<sup>[71]</sup>将转座因子赋予一种新的基因组功能, 猜想重复元件在基因组中发挥调控序列的功能, 丰富甚至创造了整个通路, 从而获得

表 1 后口动物一些非编码保守元件(CNE)的功能研究

基因	物种	用于比较的物种	表达的组织及细胞类型	测试的增强子	确定的增强子	反式作用因子结合位点	启动子	参考文献
<i>Znf503/703</i>	文昌鱼	人	脊索, 体节, 外胚层, 神经管	1	1	无	<i>FoxD</i>	[26]
<i>Sp5</i>	文昌鱼	河豚	神经, 心脏组织	1	1	无	<i>β-globin</i>	[27]
<i>SoxB2</i>	文昌鱼	河豚	神经管, 肌肉, 卵黄	1	1	无	<i>β-globin</i>	[27]
<i>Elav-like</i>	文昌鱼	河豚	神经管, 顶盖, 肌肉, 血岛, 卵黄	1	1	无	<i>β-globin</i>	[27]
<i>Six3/6</i>	文昌鱼	河豚	神经管, 卵黄, 脊髓与卵黄之间的区域	1	1	无	<i>β-globin</i>	[27]
<i>Msx</i>	文昌鱼	河豚	神经管, 脊索, 肌肉, 表皮, 尾芽, 脊髓与卵黄之间的区域	1	1	无	<i>β-globin</i>	[27]
<i>VariouS</i>	小鼠	人	前脑, 中脑, 后脑	1083	497	无	<i>β-globin</i>	[23]
<i>Sox3</i>	人	斑马鱼	大脑, 松果体, 底板, 内耳, 小脑	8	6	无	<i>Gata2</i>	[24]
<i>Sox10</i>	小鼠	鸡	听囊, 少突胶质细胞, 神经嵴, 外周神经系统, 肾上腺, 交感神经节	7	5	Sites for Sox/lef/Pax/Ap2 (EMSA)	<i>Hsp70</i>	[59]
<i>Shh</i>	斑马鱼	小鼠	胚盾, 下丘脑	3	3	无	<i>Gata2</i>	[60]
<i>Pax6</i>	斑马鱼	人	左右松果体缰, 顶板, 松果体, 中间松果体缰	8	6	无	<i>Gata2</i> , <i>Hsp70</i>	[24]
<i>Pax2,Pax8</i>	爪蟾	人	前肾, 听囊, 眼睛, 咽弓, 中后脑边界区域, 甲状腺	6	5	E-box, Fox, Gbx/Gsx, Meis/TGIF, STAT, Pbx-Hox	<i>β-globin</i> , <i>β-actin</i>	[61]
<i>Nkx2-5</i>	小鼠	人	心房, 心室, 大动脉, 胃末端区域, 舌头	3	3	<i>Gata/Smad (mut)</i>	<i>Hsp68</i>	[62]
<i>Mbp</i>	小鼠		少突胶质细胞	4	4	<i>Nkx (mut)</i>	<i>Hsp68</i>	[63]
<i>Hob2</i>	小鼠, 鸡	蝙蝠, 鸡	第四菱脑原节	1	1	<i>HoxB1, Prx, Prep1</i>	<i>β-globin</i>	[64]
<i>Hoxb4</i>	河豚	小鼠	第七八菱脑原节, 前中胚层, 神经管	3	3	无	<i>Hsp68</i> , <i>Hoxb4</i>	[65]
<i>Gata2</i>	小鼠	人	延髓泌尿生殖系统, 尾泌尿生殖系统	4	2	无	<i>Gata2</i>	[66]
<i>Sall1</i>	鸡	人	脊神经	5	1	无	<i>Thymidine kinase</i>	[56]
<i>Dach1</i>	小鼠	河豚	前脑, 中脑, 后脑, 视网膜, 肢芽, 神经管, 生殖堤	9	7	无	<i>Hsp68</i>	[55]

一种新的细胞功能。Bejerano 等<sup>[42]</sup>认为一类 CNE 来自四足动物分支时的 SINE(Short interspersed elements, 短散在元件)。SINE 是转座子的一类, 多拷贝 SINE 也表现为插入和取代等事件发生的副产物, 同时也与一些 CNE 的产生有联系。有关该转座因子产生的 CNE 家族的报道就是最好的例证<sup>[72]</sup>。基于这些例子, 可以推测许多其他 CNE 家族都可能起源于多细胞动物演化历史早期或者更早时期, 并通过转座元件进行分散, 对进化有利的插入被保持, 因此移动元件可能驱动了整个基因组特别是 CNE 的形成。

## 5 非编码保守元件在体轴进化过程中的作用

发育遗传学研究认为控制各生物类群发育的调节基因是基本相同的<sup>[73]</sup>, 这些基本的调节基因在核心发育过程中通过相互组合作用以形成动物的多样

性, 因为这些基因共同参与了不同动物体的核心发育过程, 它们的顺式调控元件也必然通过相互作用参与了动物形态多样性的形成<sup>[74, 75]</sup>。早在后生动物祖先中, 这些基因已经开始参与一些相对比较简单并控制体轴发育的基因调控网络。在动物早期进化过程中, 存在一个这些基因发生重装配(Re-wiring)的时期, 包括调控元件的复制、重组和革新(Innovation)等, 这些过程主要通过 DDC 模型(Duplication-degeneration-complementation model, 重复-冗余-互补模型<sup>[76]</sup>)和 DDI 模型(Duplication-degeneration-innovation model, 重复-冗余-革新模型<sup>[77]</sup>)来实现。顺式调控元件和基因在进化过程中的新功能化、亚功能化、无功能化或是假基因化使得基因调控网络产生多样化, 正是这些基因调控网络的多样化为各式各样的动物体轴形成提供了遗传基础。

由于进化过程中的选择压力以及发育调控的约束(Developmental constraint), 自寒武纪生物大爆炸(Cambrian explosion)之后核心基因调控网络高度保守, 这些对于生物体体轴的维持至关重要, 一旦这些核心基因调控网络遭到破坏将会出现灾难性的后果<sup>[74]</sup>。因此, 核心基因调控网络形成之后很难再发生改变。现在已有证据表明 CNEs 的突变与重组能够导致人类疾病<sup>[58, 78, 79]</sup>, 因为许多与重要调控基因相关的控制生物体早期体轴形成的顺式调控元件是这些进化缓慢的 CNEs<sup>[28]</sup>。保守的顺式调控元件(如 CNE)在生物体体轴稳定和 GRN 的保守过程中承担了十分重要的角色, 这些是由 CNEs 自身的特点决定的。在生物体进化早期特别是寒武纪生物大爆炸时期, 顺式调控元件与调节基因紧密地联系在一起, 这解释了为什么不同生物类群中的一类基因存在不同的转录增强子(为什么有不同的 CNE)。在寒武纪生物大爆炸之后动物体体轴的稳定以及基因调控网络的高度保守与许多顺式调控元件进化缓慢并且在不同物种中平行进化(Parallel evolution)是密切关联的。

## 6 展望

总之, CNE 是基因组中可能比编码序列具有更复杂生物学功能的序列。它们的保守性不仅体现在核苷酸水平, 在表观遗传学方面, 甚至在基因调控网络和调控机制方面也表现出高度的保守性。由于基因组 CNE 的丰富性、广阔性、保守性和功能性的特点, 使它们具有广泛的遗传作图、关联分析、基因作图和系统进化研究等方面的潜能。CNE 最初从脊椎动物基因组中发现, 随后又在无脊椎动物中的发现, 暗示了生物体形态多样性进化之谜。通过生物信息学和比较基因组学研究获得的大量 CNE, 其中许多已经进行间接的转基因功能验证。这些保守元件是否在无脊椎动物向脊椎动物过渡过程中发挥重要作用? 是否在生命活动以及疾病发生中还有其他重要功能? 进一步从整个基因调控网络角度来阐述非编码保守元件的功能将会有更多的发现。

### 参考文献(References):

[1] Zhang LG, Kasif S, Cantor CR, Broude NE. GC/AT-content spikes as genomic punctuation marks. *Proc Natl Acad Sci USA*, 2004, 101(48): 16855–16860.

- [2] Glazko GV, Koonin EV, Rogozin IB, Shabalina SA. A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet*, 2003, 19(3): 119–124.
- [3] Borsch T, Quandt D, Koch M. Molecular evolution and phylogenetic utility of non-coding DNA: applications from species to deep level questions. *Plant Syst Evol*, 2009, 282(3–4): 107–108.
- [4] Ishibashi M, Noda AO, Sakate R, Imanishi T. Evolutionary growth process of highly conserved sequences in vertebrate genomes. *Gene*, 2012, 504(1): 1–5.
- [5] Vavouri T, Lehner B. Conserved noncoding elements and the evolution of animal body plans. *Bioessays*, 2009, 31(7): 727–735.
- [6] Beaster-Jones L. Cis-regulation and conserved non-coding elements in amphioxus. *Brief Funct Genomics*, 2012, 11(2): 118–130.
- [7] King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res*, 2005, 15(8): 1051–1060.
- [8] Su J, Teichmann SA, Down TA. Assessing computational methods of cis-regulatory module prediction. *PLoS Comput Biol*, 2010, 6(12): e1001020.
- [9] Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJK, Cooke JE, Elgar G. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, 2005, 3(1): e7.
- [10] Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VVB, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghghi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, 2003, 424(6950): 788–793.
- [11] Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou MM, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. Evolutionarily conserved elements in ver-

- tebrate, insect, worm, and yeast genomes. *Genome Res*, 2005, 15(8): 1034–1050.
- [12] Drake JA, Bird C, Nemesh J, Thomas DJ, Newton-Cheh C, Reymond A, Excoffier L, Attar H, Antonarakis SE, Dermitzakis ET, Hirschhorn JN. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet*, 2005, 38(2): 223–227.
- [13] Frazer KA, Sheehan JB, Stokowski RP, Chen XY, Hosseini R, Cheng JF, Fodor SPA, Cox DR, Patil N. Evolutionarily conserved sequences on human chromosome 21. *Genome Res*, 2001, 11(10): 1651–1659.
- [14] Margulies EH, Blanchette M, NISC Comparative Sequencing Program, Haussler D, Green ED. Identification and characterization of multi-species conserved sequences. *Genome Res*, 2003, 13(12): 2507–2518.
- [15] Haeussler M, Joly JS. When needles look like hay: how to find tissue-specific enhancers in model organism genomes. *Dev Biol*, 2011, 350(2): 239–254.
- [16] Moghadam HK, Ferguson MM, Danzmann RG. Comparative genomics and evolution of conserved noncoding elements (CNE) in rainbow trout. *BMC Genomics*, 2009, 10(1): 278.
- [17] Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science*, 2004, 304(5675): 1321–1325.
- [18] Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D. Human genome ultraconserved elements are ultraselected. *Science*, 2007, 317(5840): 915.
- [19] McLean C, Bejerano G. Dispensability of mammalian DNA. *Genome Res*, 2008, 18(11): 1743–1751.
- [20] Sandelin A, Bailey P, Bruce S, Engström PG, Klos JM, Wasserman WW, Ericson J, Lenhard B. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics*, 2004, 5(1): 99.
- [21] McEwen GK, Woolfe A, Goode D, Vavouri T, Callaway H, Elgar G. Ancient duplicated conserved noncoding elements in vertebrates: A genomic and functional analysis. *Genome Res*, 2006, 16(4): 451–465.
- [22] Vavouri T, McEwen GK, Woolfe A, Gilks WR, Elgar G. Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. *Trends Genet*, 2006, 22(1): 5–10.
- [23] Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM. *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature*, 2006, 444(7118): 499–502.
- [24] Navratilova P, Fredman D, Hawkins TA, Turner K, Lenhard B, Becker TS. Systematic human/zebrafish comparative identification of *cis*-regulatory activity around vertebrate developmental transcription factor genes. *Dev Biol*, 2009, 327(2): 526–540.
- [25] Wang W, Zhong J, Su B, Zhou Y, Wang YQ. Comparison of *Pax1/9* locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. *Mol Biol Evol*, 2007, 24(3): 784–791.
- [26] Holland LZ, Albalat R, Azumi K, Benito-Gutiérrez È, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, Ferrier DEK, Garcia-Fernández J, Gibson-Brown JJ, Gissi C, Godzik A, Hallböök F, Hirose D, Hosomichi K, Ikuta T, Inoko H, Kasahara M, Kasamatsu J, Kawashima T, Kimura A, Kobayashi M, Kozmik Z, Kubokawa K, Laudet V, Litman GW, McHardy AC, Meulemans D, Nonaka M, Olinski RP, Pancer Z, Pennacchio LA, Pestarino M, Rast JP, Rigoutsos I, Robinson-Rechavi M, Roch G, Saiga H, Sasakura Y, Satake M, Satou Y, Schubert M, Sherwood N, Shiina T, Takatori N, Tello J, Vopalensky P, Wada S, Xu AL, Ye YZ, Yoshida K, Yoshizaki F, Yu JK, Zhang Q, Zmasek CM, de Jong PJ, Osoegawa K, Putnam NH, Rokhsar DS, Satoh N, Holland PWH. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res*, 2008, 18(7): 1100–1111.
- [27] Hufton AL, Mathia S, Braun H, Georgi U, Lehrach H, Vingron M, Poustka AJ, Panopoulou G. Deeply conserved chordate noncoding sequences preserve genome synteny but do not drive gene duplicate retention. *Genome Res*, 2009, 19(11): 2036–2051.
- [28] Vavouri T, Walter K, Gilks WR, Lehner B, Elgar G. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. *Genome Biol*, 2007, 8(2): R15.
- [29] Glazov EA, Pheasant M, McGraw EA, Bejerano G, Mattick JS. Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of *homothorax* mRNA splicing. *Genome Res*, 2005, 15(6): 800–808.
- [30] Walter K, Abnizova I, Elgar G, Gilks WR. Striking nucleotide frequency pattern at the borders of highly conserved vertebrate non-coding sequences. *Trends Genet*, 2005, 21(8): 436–440.
- [31] Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, 2003, 421(6920): 231–237.
- [32] Kuntz SG, Schwarz EM, DeModena JA, De Buysscher T, Trout D, Shizuya H, Sternberg PW, Wold BJ. Multigenome DNA sequence conservation identifies *Hox cis*-regulatory



- elements. *Genome Res*, 2008, 18(12): 1955–1968.
- [33] Peterson BK, Hare EE, Iyer VN, Storage S, Conner L, Papaj DR, Kurashima R, Jang E, Eisen MB. Big genomes facilitate the comparative identification of regulatory elements. *PLoS One*, 2009, 4(3): e4688.
- [34] Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, 2006, 439(7079): 965–968.
- [35] Blair JE, Hedges SB. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol*, 2005, 22(11): 2275–2284.
- [36] Ohno S. *Evolution by Gene Duplication*. Berlin, New York: Springer-Verlag, 1970.
- [37] Amemiya CT, Prohaska SJ, Hill-Force A, Cook A, Wasserscheid J, Ferrier DE, Pascual-Anaya J, Garcia-Fernández J, Dewar K, Stadler PF. The amphioxus *Hox* cluster: characterization, comparative genomics, and evolution. *J Exp Zool B Mol Dev Evol*, 2008, 310B(5): 465–477.
- [38] Pascual-Anaya J, D'Aniello S, Garcia-Fernández J. Unexpectedly large number of conserved noncoding regions within the ancestral chordate *Hox* cluster. *Dev Genes Evol*, 2008, 218(11–12): 591–597.
- [39] Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutiérrez EL, Dubchak I, Garcia-Fernández J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. The amphioxus genome and the evolution of the chordate karyotype. *Nature*, 2008, 453(7198): 1064–1071.
- [40] Punnamoottil B, Herrmann C, Pascual-Anaya J, D'Aniello S, Garcia-Fernández J, Akalin A, Becker TS, Rinkwitz S. *Cis*-regulatory characterization of sequence conservation surrounding the *Hox4* genes. *Dev Biol*, 2010, 340(2): 269–282.
- [41] Skromne I, Thorsen D, Hale M, Prince VE, Ho RK. Repression of the hindbrain developmental program by Cdx factors is required for the specification of the vertebrate spinal cord. *Development*, 2007, 134(11): 2147–2158.
- [42] Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, 2006, 441(7089): 87–90.
- [43] Stephen S, Pheasant M, Makunin IV, Mattick JS. Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol Biol Evol*, 2008, 25(2): 402–408.
- [44] Sun H, Skogerbø G, Chen RS. Conserved distances between vertebrate highly conserved elements. *Hum Mol Genet*, 2006, 15(19): 2911–2922.
- [45] Li XR, Tan LB, Wang LG, Hu SN, Sun CQ. Isolation and characterization of conserved non-coding sequences among rice (*Oryza sativa* L.) paralogous regions. *Mol Genet Genomics*, 2009, 281(1): 11–18.
- [46] Woolfe A, Goode DK, Cooke J, Callaway H, Smith S, Snell P, McEwen GK, Elgar G. CONDOR: a database resource of developmentally associated conserved non-coding elements. *BMC Dev Biol*, 2007, 7(1): 100.
- [47] Minovitsky S, Stegmaier P, Kel A, Kondrashov AS, Dubchak I. Short sequence motifs, overrepresented in mammalian conserved non-coding sequences. *BMC Genomics*, 2007, 8(1): 378.
- [48] Woolfe A, Elgar G. Comparative genomics using *Fugu* reveals insights into regulatory subfunctionalization. *Genome Biol*, 2007, 8(4): R53.
- [49] Tsang WH, Shek KF, Lee TY, Chow KL. An evolutionarily conserved nested gene pair—*Mab21* and *Lrba/Nbea* in metazoan. *Genomics*, 2009, 94(3): 177–187.
- [50] Taher L, Ovcharenko I. Variable locus length in the human genome leads to ascertainment bias in functional inference for non-coding elements. *Bioinformatics*, 2009, 25(5): 578–584.
- [51] Farré D, Bellora N, Mularoni L, Messeguer X, Albà MM. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol*, 2007, 8(7): R140.
- [52] Blanchette M, Bataille AR, Chen X, Poitras C, Laganière J, Lefebvre C, Deblois G, Giguère V, Ferretti V, Bergeron D, Coulombe B, Robert F. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res*, 2006, 16(5): 656–668.
- [53] Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L. Evolution and functional classification of vertebrate gene deserts. *Genome Res*, 2005, 15(1): 137–145.
- [54] Kimura-Yoshida C, Kitajima K, Oda-Ishii I, Tian E, Suzuki M, Yamamoto M, Suzuki T, Kobayashi M, Aizawa S, Matsuo I. Characterization of the pufferfish *Otx2 cis*-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. *Development*, 2004, 131(1): 57–71.
- [55] Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. *Science*, 2003, 302(5644): 413.
- [56] Izumi K, Aramaki M, Kimura T, Naito Y, Udaka T, Uchikawa M, Kondoh H, Suzuki H, Cho G, Okada Y, Takahashi T, Golden JA, Kosaki K. Identification of a prosencephalic-specific enhancer of *SALL1*: comparative genomic approach using the chick embryo. *Pediatr Res*,

- 2007, 61(6): 660–665.
- [57] Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H. Functional analysis of chicken *Sox2* enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev Cell*, 2003, 4(4): 509–519.
- [58] Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*, 2003, 12(14): 1725–1735.
- [59] Werner T, Hammer A, Wahlbuhl M, Bösl MR, Wegner M. Multiple conserved regulatory elements with overlapping functions determine *Sox10* expression in mouse embryogenesis. *Nucleic Acids Res*, 2007, 35(19): 6526–6538.
- [60] Ertzer R, Müller F, Hadzhiev Y, Rathnam S, Fischer N, Rastegar S, Strähle U. Cooperation of *sonic hedgehog* enhancers in midline expression. *Dev Biol*, 2007, 301(2): 578–589.
- [61] Ochi H, Tamai T, Nagano H, Kawaguchi A, Sudou N, Ogino H. Evolution of a tissue-specific silencer underlies divergence in the expression of *pax2* and *pax8* paralogues. *Nat Commun*, 2012, 3: 848.
- [62] Chi X, Chatterjee PK, Wilson W III, Zhang SX, Demayo FJ, Schwartz RJ. Complex cardiac *Nkx2-5* gene expression activated by noggin-sensitive enhancers followed by chamber-specific modules. *Proc Natl Acad Sci USA*, 2005, 102(38): 13490–13495.
- [63] Farhadi HF, Lepage P, Forghani R, Friedman HC, Orfali W, Jasmin L, Miller W, Hudson TJ, Peterson AC. A combinatorial network of evolutionarily conserved *myelin basic protein* regulatory sequences confers distinct glial-specific phenotypes. *J Neurosci*, 2003, 23(32): 10214–10223.
- [64] Maconochie MK, Nonchev S, Studer M, Chan SK, Pöppel H, Sham MH, Mann RS, Krumlauf R. Cross-regulation in the mouse *HoxB* complex: the expression of *Hoxb2* in rhombomere 4 is regulated by *Hoxb1*. *Genes Dev*, 1997, 11(14): 1885–1895.
- [65] Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Krumlauf R, Brenner S. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc Natl Acad Sci USA*, 1995, 92(5): 1684–1688.
- [66] Khandekar M, Suzuki N, Lewton J, Yamamoto M, Engel JD. Multiple, distant *Gata2* enhancers specify temporally and tissue-specific patterning in the developing urogenital system. *Mol Cell Biol*, 2004, 24(23): 10263–10276.
- [67] Sakuraba Y, Kimura T, Masuya H, Noguchi H, Sezutsu H, Takahashi KR, Toyoda A, Fukumura R, Murata T, Sakaki Y, Yamamura M, Wakana S, Noda T, Shiroishi T, Gondo Y. Identification and characterization of new long conserved noncoding sequences in vertebrates. *Mamm Genome*, 2008, 19(10–12): 703–712.
- [68] Kato M, Sekine A, Ohnishi Y, Johnson TA, Tanaka T, Nakamura Y, Tsunoda T. Linkage disequilibrium of evolutionarily conserved regions in the human genome. *BMC Genomics*, 2006, 7(1): 326.
- [69] Lowe CB, Kellis M, Siepel A, Raney BJ, Clamp M, Salama SR, Kingsley DM, Lindblad-Toh K, Haussler D. Three periods of regulatory innovation during vertebrate evolution. *Science*, 2011, 333(6045): 1019–1024.
- [70] Goode DK, Callaway HA, Cerda GA, Lewis KE, Elgar G. Minor change, major difference: divergent functions of highly conserved *cis*-regulatory elements subsequent to whole genome duplication events. *Development*, 2011, 138(5): 879–884.
- [71] Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 1980, 284(5757): 601–603.
- [72] Xie XH, Kamal M, Lander ES. A family of conserved non-coding elements derived from an ancient transposable element. *Proc Natl Acad Sci USA*, 2006, 103(31): 11659–11664.
- [73] Erwin DH, Davidson EH. The last common bilaterian ancestor. *Development*, 2002, 129(13): 3021–3032.
- [74] Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. *Science*, 2006, 311(5762): 796–800.
- [75] Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet*, 2009, 10(2): 141–148.
- [76] Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 1999, 151(4): 1531–1545.
- [77] Jiménez-Delgado S, Pascual-Anaya J, García-Fernández J. Implications of duplicated *cis*-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. *Brief Funct Genomic Proteomic*, 2009, 8(4): 266–275.
- [78] Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, Jamshidi N, Essafi A, Heaney S, Gordon CT, McBride D, Golzio C, Fisher M, Perry P, Abadie V, Ayuso C, Holder-Espinasse M, Kilpatrick N, Lees MM, Picard A, Temple IK, Thomas P, Vazquez MP, Vekemans M, Roest Crolius H, Hastie ND, Munnich A, Etchevers HC, Pelet A, Farlie PG, Fitzpatrick DR, Lyonnet S. Highly conserved non-coding elements on either side of *SOX9* associated with Pierre Robin sequence. *Nat Genet*, 2009, 41(3): 359–364.
- [79] Ott CE, Hein H, Lohan S, Hoogeboom J, Foulds N, Grunhagen J, Stricker S, Villavicencio-Lorini P, Klopocki E, Mundlos S. Microduplications upstream of *MSX2* are associated with a phenocopy of cleidocranial dysplasia. *J Med Genet*, 2012, 49(7): 437–441.