



# DNA元素百科全书

——2012年生命科学研究的六大突破之一——

撰文·供图 奇云（淮南联合大学） 李大可（复旦大学）

2012年9月5日，由美国人类基因组研究所资助的一个规模巨大的国际性项目“DNA元素百科全书”计划公布首期成果：获得了迄今最详细的人类基因组分析数据。

这些新公布的“DNA元素百科全书”数据庞大，当天以30篇论文的形式同时发表在英国《自然》杂志等多份学术刊物上。其中，有6篇论文在《自然》杂志，24篇论文在《基因组研究》和《基因组生物学》杂志，另外一些附加论文在《科学》和《细胞》杂志发表。

这是“人类基因组计划”之后，国际科学界在基因研究领域取得的又一重大进展。2003年，科学家完成了由人类染色体中30亿个碱基对组成的核苷酸序列的测定，从而绘制出了人类基因组图谱。这项人类基因组计划的成功完成，预示着人类将会破译自身的遗传信息。然而，虽然基因组测序产生了数千兆字节的原始数据，却仍旧无法解释基因的运行方式。于是，继人类基因组计划之后，科学界的另一场盛宴——“DNA元素百科全书”

计划隆重登场。其目标就是鉴定人类基因组调控序列中所有的功能元件，编写一部人类DNA的百科全书。

启动于2003年的“DNA元素百科全书”计划，联合了来自英国、美国、西班牙、新加坡和日本的32个顶级实验室中442名科学家。研究支出1.85亿美元，花费了约300年的计算机时间，进行了1649项实验，使用了147种细胞、235种抗体，获得并分析了超过15兆兆字节（15万亿字节或称15TB）的原始数据。

由于印刷版杂志无法承载如此巨大的数据,“DNA元素百科全书”计划研究小组将研究数据在“DNA元素百科全书”计划门户网(<http://www.encodeproject.org/>)发布。

“DNA元素百科全书”计划还开辟了一种数据挖掘的新程序,称为“主题线程”。它将“DNA元素百科全书”计划的所有数据和在杂志发表的几十篇论文汇总在一起,可以依据主题,通过虚拟计算机进行数据的调阅、拼接、分析、整合和预测。这也为今后生命科学和复杂疾病的研究增添了一个新的、有力的工具。目前已经有100多篇发表的文章使用了“DNA元素百科全书”计划中的数据。虽然这些论文的作者并没有参与到“DNA元素百科全书”计划中,但是他们依然可以免费使用这些数据进行疾病调研。

“DNA元素百科全书”计划发现:人类基因组DNA,包括略少于21 000个蛋白质编码基因,8 800个小RNA分子,9 600个长非编码RNA分子,11 224个假基因。只有不到3%的基因组的基因,能够编码蛋白质。1972年,日本遗传学家大野乾给不能编码蛋白质的DNA片段起名为“垃圾DNA”。“DNA元素百科全书”计划的一个主要目的就是去分析那些“垃圾”的功能。经过“DNA元素百科全书”计

划的初步努力,现在科学家认为80%的基因组是有功能的,其余的20%也不是垃圾。这些称为“垃圾”的DNA实际上是一个庞大的控制面板,能调控数以万计基因的活性,并决定了一些基因的特质。如果没有这些开关的调控,基因将不能正常工作,而这些区域也许会使人类患上疾病。这个发现震惊了整个科学界!研究人员还发现,许多人类基因组的非编码部分包含蛋白可以绑定的区域,这将会影响附近和远处的基因表达。其他转录成RNA分子的非编码区从来没有翻译成蛋白。还有一些影响了DNA折叠和包装的方式。总之,这些区域并不是垃圾。

研究还发现,基因组DNA的76%转录成多种RNA。部分碱基,转变成RNA链,执行基因调控等功能。许多碱基在人类生物学中起作用,如有助于决定一个基因的起动的和终止。许多碱基仅落在化学修饰的位置,成为染色体的一些沉默片段。400万个位点是控制基因活性的开关,给每一类型细胞以独特的基因组标识。

“DNA元素百科全书”计划的成果意义非常重大。完成了人类基因组计划和国际单体型图计划,相当于让我们知道了基因组的DNA排序与个体之间的差异。而“DNA元

素百科全书”计划则是分析在DNA序列上的各种功能信息,例如了解除了编码蛋白的序列之外,在基因上或基因周边还有哪些序列促进或抑制一个基因的表达,以及这些功能元件各自以什么样的方式发挥作用。如果说人类基因组计划是将人类基因组这部天书上的文字一一辨识出来,那么“DNA元素百科全书”计划就是对这些文字进行注解。

由于“DNA元素百科全书”对于DNA上调控基因表达的体系做了详细和深入的解析,这些知识结合其他研究所获得的大量疾病相关基因,就可能帮助科学家针对这些基因的关键元件设计药物靶点,或者针对不同个体易感基因上功能元件的多态性设计个体化治疗方案,达到有的放矢的治疗目的。另外,很多基因产物,例如多种有分解作用的酶蛋白参加药物的代谢。了解了这些基因的功能元件,就可以根据这些元件上的遗传多样性,检测不同个体对于药物的反应和代谢能力,从而设计个体特异的用药方案,使药物治疗更加合理和准确有效。如今,编写的“DNA元素百科全书”计划初具雏形,但还有很多细节需要补充和完善。

(责编 桑新华)