

建构“基因”:思想、技术与工具

杨 辰

(清华大学 科学史系, 北京 100084)

摘要:许多因素共同推动了科学对“基因”的认识。思想、技术和工具这三个相互联系的因素对基因科技的发展产生了关键的影响。从前遗传学时代,到经典遗传学时代,到分子生物学时代,再到后基因组学时代,“基因”的概念在连续演化的过程中,主要通过思想、技术与工具三者之间的互动被逐渐建构起来且具有在还原论与系统论两端之间波动的特点。未来基因科技的发展方向,很大程度上依然是这种建构过程的延伸。

关键词:基因;建构;还原论;系统论

中图分类号:N031 **文献标识码:**A

机械化、电气化、自动化分别造就了前三次工业革命,以生物技术和信息化为基础的第四次工业革命初现端倪。生物技术和信息技术的重要结合点之一是探究基因的本质。那么,基因究竟是什么?生物学家认为,基因是记录有遗传信息的DNA片段;化学家认为,基因是脱氧核糖核苷酸;哲学家认为,基因是生命的密码。那么,科学是怎样认识基因的,是什么塑造了科学对基因的看法?和众多科学发现类似,许多因素共同推动了基因科技的发展,而本文将着重关注思想、技术和工具这三个关键因素之间的互动对基因科技发展的影响。

从主体性的、内在的、抽象的维度来看,科学的发展就是科学家的科学思想发展过程。^[1]面对相同的实验现象,不同的科学家会做出完全不同的,有时甚至是截然相反的论断。即便是同一位科学家,对同一实验事实的解释,也有可能随着研究的深入产生变化。对于基因科技而言,思想的转变除了理论上的预见性推演之外,很重要是基于新的实验现象的发现,而能够观察到新的实验现象则有赖于技术和工具的革新。

以基因科技发展的标志性成果为节点来看,通常会将这一历程划分为四个阶段。第一阶段是漫长的前遗传学时代。基因一词诞生于1909年,当人们重新发现孟德尔(Gregor Johann Mendel)定律时,植物学家威廉·约翰森(Wilhelm Johannsen)将其中的遗传因子以“基因”二字命名。第二阶段是历经

半个世纪的经典遗传学时代,科学家用杂交的手段和模式生物“模糊”地观察性状的变化,摸索基因的本质。直到1953年,DNA神秘的面纱被双螺旋结构揭开,“基因”形象的分子化展开了第三阶段——分子生物学时代。2003年人类基因组计划的完成,则开启了第四阶段——后基因组学时代。这一阶段,人们对基因的研究已经超出了原有的分子化视野,一方面开始以基因组为单位从宏观层面关注它的功能,另一方面以碱基为单位从微观层面对其进行操控。本文将通过分析每个阶段中的典型科学事件,观察其中思想、技术和工具对研究的推动及三者之间的互动关系,以探究基因科技的发展机制。

一、基因:从概念走向信息

1. 孟德尔遗传规律的再发现

1900年,被埋没了35年之久的孟德尔遗传规律^[2],被荷兰植物生理学家和遗传学家雨果·德弗里斯(Hugo de Vries)^[3]、德国植物遗传学家卡尔·埃里克·科伦斯(Carl Erich Correns)^[4]、奥地利植物遗传学家埃里希·冯·切尔马克(Erich von Tschermak)^[5]和美国植物学家威廉·贾斯珀·斯皮尔曼(William Jasper Spillman)^[6]四人分别在研究月见草、玉米、豌豆和小麦的遗传规律过程中重新发现,这在

收稿日期:2024-7-3

作者简介:杨辰(1993—),女,湖北天门人,理学博士,清华大学科学史系博士后,主要研究方向:科技史、科技哲学。

遗传学史上被称为“戏剧性的再发现”。紧随其后,1909年约翰森出版了《精确遗传学的诸要素》^[7]一书,书中提出了基因、基因型、表现型等概念,将孟德尔提出的“遗传因子”命名为“基因”。从此,基因作为遗传的基本单位为遗传学界所公认和使用。但此时基因的概念并不清晰,人们只是将基因理解为生物体内若干代后都保持不变的某种物质。当时的一些科学家持有基因不可知论,并且认为基因的具体物质属性并不重要。在这种情况下,基因被视为一个抽象元素,通过模式生物杂交实验的可见和可量化结果对其进行捕捉。

2. 基因物质结构的探索

对基因物质结构的探索发端于科学思想上三个方向的转变,使得基因的研究进入经典遗传学时代。首先是两方面的发现使基因实体化。一方面,基因有了分子结构。赫尔曼·约瑟夫·穆勒(Hermann Joseph Muller)于1927年利用X射线诱导果蝇发生可遗传的突变,他将这一实验结果归因于基因分子结构的永久性改变。^[8]另一方面,基因的位置被确定。沃尔特·萨顿(Walter Sutton)和西奥多·博韦里(Theodor Boveri)率先将有性生殖过程中染色体的行为和基因传递的规律联系起来,催生了遗传的染色体学说的形成。托马斯·亨特·摩尔根(Thomas Hunt Morgan)和他的学生发现了基因的连锁与交换,并给出了第一个果蝇染色体连锁图,证明了基因在染色体上呈直线排列,同时确立了基因作为遗传基本单位的概念。^[9]第二个科学思想上的转变是不同学科思想开始在这一领域融合。基因的研究已不再是传统的生物学家的问题,它已经成为生物学、化学和物理学之间的交叉地带中亟待解决的问题。这种学科上的交叉并非只出现在基因科技上,而是当时科学界研究的大潮流。^[10]第三个思想上的转变则来自于小分子化学思想向高分子化学思想的转变,自此之后人们才正确地认识到以生物大分子面貌出现的“基因”。早在1869年,年轻的瑞士生物学家弗里德里希·米歇尔(Friedrich Miescher)报道了一种不能被蛋白消化酶破坏的物质并将其命名为核素,后来因其酸性特征,重新命名为核酸。当时提纯出的其实是DNA的粗提取物,还混有蛋白质。随着提纯技术的发展,科学家们虽然已经可以得到不含蛋白质的核酸样品,但也只将其视为细胞中的普通成分,没有人注意到它在生物体内具有重要的功能。1910年,人

们已经认识到DNA分子由四种碱基、脱氧核糖和磷酸组成,但正确地将这些组分链接起来又经历了40年的漫长研究之旅。仅针对DNA的研究而言,研究者在理论上有两条路可走——将DNA加以分解而研究其组分,或是将整个DNA分子视为一个整体进行研究。而起初有机化学家对DNA的研究采用了前者,提纯都采取了十分剧烈的分离方法,破坏了大分子的结构,这一困境直到高分子化学的兴起才取得进展。1944年,美国细胞学家奥斯瓦尔德·艾弗里(Oswald Avery)、科林·麦克劳德(Colin MacLeod)以及麦克林·麦卡蒂(Maclyn McCarty)在研究肺炎双球菌转化实验中,证明了遗传物质是DNA而不是蛋白质。^[11]1952年,阿尔弗雷德·赫尔希(Alfred Hershey)与玛莎·蔡斯(Martha Chase)合作,利用同位素示踪法再次确认DNA是遗传物质。^[12]

此时对于DNA结构的探索已到了“万事俱备,只欠东风”的阶段。作为当时的研究热点,势必会有多个研究小组开展着竞争。剑桥大学的沃森-克里克小组(即詹姆斯·沃森,James Watson;弗朗西斯·克里克,Francis Crick)最终拔得头筹并非偶然,除了科学家个人的勤奋与创造性想象力之外,还得益于优质的结晶样品、X射线衍射仪分辨率的提高以及模型的使用。^[13]而由莱纳斯·鲍林(Linus Pauling)创造的以键角和原子半径为基础的分子模型^[14]如今常常被人们忽略。手动搭建分子模型目前逐渐从科研一线退居幕后,较为广泛地出现在针对青少年的科普教育中。然而在当时,科学家通过亲手搭建分子模型这一实验方式,极大地促进了大分子结构的理解和阐明。如今,实体模型以分子模拟软件的形式出现,是研究中不可或缺的工具。除了单纯的结构预测,还可以设计实验方案或进行分子动力学模拟。至此,基因以清晰的分子形象实体化,既有化学元素组成,又有三维立体结构。在基因实体化的同时,作为遗传信息的承载者,自然界中海量的基因信息呈现在人们面前。

3. 人类基因组计划与基因组多样性

当一部分科学家在努力揭开DNA的神秘面纱之时,另一部分科学家已经开始思考基因如何编码遗传信息。人类基因组计划在测序技术出现之后应运而生,从而开启了分子生物学时代。该计划由美国科学家于1985年率先提出,1990年正式启动。2003年,美、英、法、德、日、中六国合作历时十余年、

累计投入约30亿美元,完成了人类基因组序列的绘制。^[15]弗朗西斯·柯林斯(Francis Collins)和克雷格·文特尔(John Craig Venter)向世人展示了人类基因组的两个基本完成版本。^[16]

柯林斯领导的国际研究采取的是系统而保守的测序方法。遗传学家肩负着绘制基因组遗传图和物理图的重任,并需要对这些图谱进行精准标记,以便随后测出的序列能按正确顺序拼接。而大规模测序工作的开展,则必须建立在遗传图和物理图已经绘制完成的基础之上,并在图谱中所有位点对应的克隆均已获得的情况下才能进行。文特尔领导的竞争性工作开始于1998年,使用当时最新的鸟枪法测定基因组序列,不需要依赖图谱,而是将整个基因组随机打断并克隆,然后对这些克隆进行随机测序,最后利用强大的计算机算力发现重叠序列并将这些序列拼接起来。^[17]然而归根结底,人类基因组计划的技术基础奠定于1977年,弗雷德里克·桑格(Frederick Sanger)发明了以荧光标记链终止法为原理的第一代测序技术。^[18]在这场关乎生命本质的科学远征中,研究者们不惜代价推进测序进程,实则暗含着一种技术理想主义的认识论预设:基因组序列作为决定生命现象的核心代码,其完整解析将自动导向对生命本质的全然揭示。

4. 后基因组学与基因编辑技术

随着测序成本的降低、测序通量的提高,研究生物大分子相互作用的技术和工具越来越丰富,一个更为复杂的基因图景出现在人们眼前,这也预示着后基因组学的到来。基因作为遗传蓝图的概念受到了巨大的冲击,一方面,科学家发现细胞中DNA的数量与物种的复杂性没有明确的关联,人类基因组所含有的基因数量在两万左右,番茄和水蚤的基因数量都比人类多出一万有余;另一方面,诸如DNA甲基化等翻译后修饰的表观遗传学改写着遗传信息,打破了遗传信息的传递仅依赖细胞核中DNA序列的经典基因概念。与此同时,在聚焦单一物种的基因测序研究之后,研究方向开始转向基因组多样性的系统测序,这使得人类可以在整个自然界的总体层面上认识基因组的多样性。万种脊椎动物基因组计划、地球微生物组计划、全球基因组计划相继提出并展开。一方面,这种“只读模式”的基因组学似乎是在为人类想要扮演造物主角色这一雄心勃勃的计划的合成生物学奠定基础;另一方面,体现出人们对基因从程序到数据库这一概念的转变。

除了更大规模、更高层次的测序计划以外,后基因组学时代的另一大特点是在分子化水平对基因组进行编辑。在基因概念尚未诞生之前,人类操纵生物体的各种尝试就一直在有意识或无意识地进行。广义的基因重组可分为三个水平:整体水平,如生物的有性杂交;细胞水平,如细胞融合等;分子水平,如将一个生物体的DNA插入另一个生物体的DNA中。分子水平的基因编辑得益于CRISPR/Cas9基因编辑系统的出现。^[19]这种涉及分离纯化、扩增、表达等一系列操作的过程,可以按照人类的意愿生产基因产物,通过插入、替换等方法改造基因,探究基因的结构和功能。基因编辑技术的蓬勃发展标志着人类进入了设计和创造新基因、新蛋白质和生物新性状的时代。

二、基因探寻过程中“脑”“手”和“器”的互动

就科学思想而言,前遗传学时代的科学家有关遗传的各种现象的主要知识来源是育种学家和医生,他们对遗传“偏差”的关注是为了寻找有用的新品种以及探寻预防和治疗疾病的途径。遗传学的历史渊源则是来自于早期人们对动物、植物选择育种的探索。即便是孟德尔本人,最初研究豌豆杂交品种时也并非为了系统地阐述遗传规律,而是为了说明杂交在生物进化中的作用。这一方面成为孟德尔杂交实验被埋没的原因之一,另一方面也隐含着一种可能,即当人们对遗传的关注点从“偏差”转为对生物体可遗传性状的关注之时,即为孟德尔杂交实验重见天日之时。

孟德尔的杂交实验背后由一套新的实验技术所支撑,这套实验技术具有高度简化和人为操纵的特点,这个特点得益于孟德尔用两年之久的预实验筛选稳定的、纯种豌豆作为实验对象。杂交不再被视为有性生殖的特例,而成为研究遗传过程和阐明生物遗传结构的普遍技术。模式生物这一被当今生物学家常用的实验工具也登上了舞台。在这一科学事件中,豌豆是模式生物的代表,后续生物学家使用的果蝇、玉米和噬菌体等模式生物均由此开端。模式生物具有双重身份,不仅是研究对象,也是“仪器”,使得科学家能够操纵生殖细胞内不可见的因素,从而利用可见性状来追踪这种操纵的影

响。杂交技术和模式生物在进一步阐明基因的形式结构、基因在染色体上的假定排列以及基因与生物体性状的多种关系方面,起到了举足轻重的作用。

思想上对可遗传性状的关注,让杂交技术和模式生物等新技术和新工具在一段时间内蓬勃发展。科学家们通过这种类似黑箱的模式探究基因与性状之间的关系,看似行走一片坦途之上,仿佛也越来越靠近“基因”的本质。但由于当时基因概念的模糊与不统一,因此一定程度上阻碍了对基因物质结构的进一步探索。

包括 DNA 在内的生物大分子的本质都笼罩在不可知论的乌云之下。以蛋白质为例,在 1951 年之前,人们甚至认为蛋白质没有确定的化学组成,直到桑格确定了牛胰岛素的 A 和 B 的完整氨基酸序列,而且生命的本质似乎与物理学的一个基本定律——热力学第二定律相悖。1943 年 2 月,物理学家埃尔温·薛定谔(Erwin Schrödinger)在都柏林三一学院发表了题为《生命是什么?》的系列演讲,提出了一个引人深思的问题:生命体的复杂之谜能否最终被还原为物理学或化学的原理来解释?思想上还原论观点的启发,加上化学家们在技术上的贡献,共同推动了分子生物学领域的蓬勃发展。从技术上来说,化学家的介入带来了体外实验技术的建立,科学家们得以在生物体外对 DNA 等生物大分子进行研究。同时,更加温和的分子提纯技术可以分离出结构未遭破坏的 DNA 分子,为下一步揭示 DNA 双螺旋结构做好了准备。这一阶段工具的发展助力提纯技术和表征技术的革新,超速离心机、X 射线衍射仪、放射性示踪仪等仪器大显身手。模式生物的定义随着小型化成员,即各类细菌的加入而扩大。它们易于保存在冰箱中,所需空间极小,在最佳培养条件下,其繁殖时间仅需半小时。这直接缩短了科学家们跟踪多代遗传过程的时间,大大加快了生物体从基因型到表型的实验流程。这一时期技术和工具的种类显著增加,逐渐形成了两种互动模式。

第一,当技术发生迭代时,工具必然随之改变以适应新技术,所形成的新方法将取代旧方法。以 DNA 测序技术为例。自 1977 年桑格及其团队开发出双脱氧测序法以来, DNA 测序技术已历经三代变革。一代测序技术中操作费时的胶分离和人力的依赖性导致速度慢且成本高,因此人类基因组的测定花

费了数年时间,耗费了数十亿美元才得以完成。于是,第二代测序技术的主要改进目标聚焦于自动化和流程升级,旨在降低成本并提升效率。这一改进始于 2005 年,涉及步骤优化与新仪器的应用。其核心原理是边合成边测序,借助大规模并行测序技术,实现了高通量测序。二代技术的发展使得单个人类基因组的测序成本降至数百美元。与第一、二代测序技术相比,第三代测序技术巧妙地绕开了 DNA 或 RNA 扩增这一步骤,直接对单个分子进行测序。在此过程中,单分子实时测序和纳米孔测序技术应运而生。这两种技术通过监测单个 DNA 分子在合成过程中或穿过纳米孔时所产生的信号变化,精准地读取 DNA 序列。目前,科学家正致力于开发多种工具,以更准确、更高效地检测单个 DNA 分子发出的信号。由于第三代测序技术在原理上的革新,使其能够借助简单的电子元件完成序列的读取。随着 DNA 测序技术从第一代逐步演进至第三代,在如今的实验室中已鲜少使用第一代测序技术。

第二,当技术没有更新而工具发生迭代时,也会产生新的方法,这种情况下,新旧方法之间不会相互取代。以 DNA 探测技术为例,常用的检测方法主要有两种:印记法和微阵列。印记法是将待测的未知 DNA 从凝胶转移到特定的膜上,然后利用标记探针进行杂交,以检测特定的 DNA 序列;而微阵列则是将已知序列的 DNA 片段固定在特殊材料上。随着固定材料优化,微阵列上固定化点的尺寸不断缩小,从最初仅能检测数十个点发展到如今在一张邮票大小的硅芯片上可以检测数十万甚至上百万个 DNA 点。印记法和微阵列的技术原理都是基于生物大分子的特异性识别,只是工具从凝胶基质转变为 DNA 硅芯片。这两种方法各有优势,适用于不同的实验目的,目前依然是生物实验室中常用的方法。而显微镜的迭代也属于这一模式,从光学显微镜到电子显微镜,均是利用波的性质来观测物体。因此,两种实验方法并非相互取代,而是各自适用于不同的研究领域和需求。

科学思想主要呈现为两种类型:一类是在原理上具有原创性的思想,例如薛定谔在都柏林三一学院所发表的强还原论观点。他试图从物理学和化学的角度来解释生命现象,这一跨学科的思想启发了科学家们探索生命现象的分子基础,从而推动了分子生物学这一新兴学科的诞生。这种原创性思想常常能够催生出一个全新的学科领域,或者开启一个崭新的时

代篇章。另一类科学思想则体现出技术上的预见性,典型如物理学家乔治·伽莫夫(George Gamow)提出的遗传密码三联体假说。^[20]1954年,DNA双螺旋结构被解析不到一年,伽莫夫便在《自然》杂志上发表题为《DNA分子进行的蛋白质合成》的文章,首次提出三联体密码假说。他指出,DNA的4种核苷酸(ACTG)通过特定组合编码蛋白质中的20种氨基酸。这种技术上具有预见性的思想为科学家们破译具体的科学问题指明了研究方向。

基因的分子化是伴随着遗传学的分子化而来的。这一阶段人们不仅仅关注基因,还关注生物在进化、遗传、发育以及细胞新陈代谢中的各种生物大分子和它们所组成的信息传递通路。在这个过程中,基因成为众多生物大分子中的一员。人们不再满足于解码单一的基因,而是将目光投向整个基因组。最有代表性的科学事件当属人类基因组计划的完成。这一时期实验工具及其概念本身都发生了巨大的变化。早期的分子生物学依赖于一系列“笨重”的分析仪器,这些仪器以客体的形式作用于细胞,从而了解它们的物理和化学特性。新一代仪器则不同,它们本身就是生物分子,如测序技术中用到的DNA聚合酶。这些生物分子能够复制、剪切和连接其他生物分子。在这些工具的帮助下,模式生物可以被赋予新的特性,如作为生物反应器克隆出大量具有治疗活性的分子。

基因的概念体现出多重性,既具体又抽象。一方面,基因被功能化和商品化。每个基因都有特定的、明确的功能,这直接导致基因成为技术的直接对象,在工具的加持下变得可操控。同时,基因具有了商品的特征。从20世纪90年代“人类基因工程”计划启动之日起,美国、日本和欧洲等国家和地区展开了一场激烈的基因专利争夺战。另一方面,基因是信息,带有“计划”和“程序”的概念。物种在进化过程中产生并储存在基因组中的遗传信息的总和体现出生命体自诞生之时就具备的计划性,而在生物体发育过程中以有序形式实现计划所依据的指令序列则表现出严格执行的程序性。面对数以万计碱基序列的基因,技术和工具的革新成为科学发展的主要动力,这一时期的科学思想主要聚焦于如何改进测序方法,以达到更低成本和更高通量。同时,技术和工具的研发主体发生变化,以营利为目的的私人公司开始主导测序方法的改进。这其中的原因之一是科学家的身份发生转变,他们不再仅仅是科研人员,还越来越多

地成为生物技术公司的代理人甚至所有者,并寻求多方投资,从而最大限度地发挥其产品的市场潜力。这种科学家身份的转变发生在诸多前沿科研领域,这使得科学研究被利益驱使的比重增大,对科学伦理提出了严峻挑战。

后基因组学时代,科学家们已经习惯评定生物体全基因组的序列、功能和表达,而其中技术和工具的发展不断拓展着人们对基因组“读、写、改”的能力。此时基因科技的研究版图,可大致分为四种类型。一是基因编辑技术的研究,聚焦CRISPR/Cas9编辑原理,从原理出发对编辑体系进行改造和优化,期望从中发现比CRISPR/Cas9更有潜力的编辑工具。二是围绕实验技术的研展开的研究,如拓展冷冻电镜的应用场景,突破光学极限的各类高分辨成像方法,各类基于不同物理性质,如光学、分子间作用力、磁场等单分子探测手段的研发,支撑组学研究的测序方法、高分辨质谱、分析组学数据的统计学工具和计算方法的开发。三是生物大分子研究相关的模拟研究,如AlphaFold结构预测模型的开发。四是以疾病为导向,研究从DNA开始的转录翻译等遗传全过程。这一方向可以用人类基因组计划作类比,目前的进度可以看作处于计划刚开始的阶段,测序手段较为缓慢且准确率低而成本高。对人体内乃至自然界中所有生物体内的各类蛋白质的形成以及作用的通路,科学家们目前仅了解冰山一角。这在大多数生物类博士生的毕业论文中表现得尤为明显,其研究主题往往聚焦于特定生物大分子之间的相互作用机制,或是某种技术方法的开发与应用。前三种类型的研究均以技术和工具的升级为核心,旨在为第四种类型的研究,即探索生命体奥秘本身提供支持。现阶段技术和工具有一定的成熟度,但还需继续迭代:基因编辑工具在特异性和脱靶效应方面的改进还任重道远,CRISPR工具箱在不断壮大;遗传组学、转录组学和翻译组学的研究方式也需要进一步开发;蛋白质、核酸等生物大分子之间的相互作用的表征手段还有待进一步拓展。这一阶段使用的工具呈现出小型化与大型化两极分化状态。小型化工具,如生物芯片等是科学家得以高通量地研究各类组学的重要支撑;而大型化主要体现在与“大科学”相配套的“大仪器平台”上。高通量测序仪、冷冻电镜等大型设备因其高额造价而以仪器共享平台的形式被实验室所共享。这一阶段作为高速科研助推器的人工智能算法和自动化实验室也不容忽视。2023年11月29日发表在*Nature*上的两篇文章分别报道了材料学AI——GNoME^[21]和自动化实

验室——A-Lab^[22],以及他们相互配合的潜力。材料学 AI 产生了 220 万个全新晶体结构,其中 736 个结构被独立实验所证实;而自动化实验室 A-Lab 在 17 天内完成了 355 次合成实验,成功合成了 58 个拟定化合物中的 41 个,成功率达到了 71%。人工智能算法和自动化实验室的出现,一方面使得成熟的技术和工具效率最大化,用机器替代大量重复性体力实验,效率更高且结果更稳定;另一方面也会加快技术和工具的更新换代,更好地服务于研究目标。然而,即便是在人工智能算法和自动化实验室对科研的加速推进下,技术和工具的进步也只能加快绘制出人们在已有的科学思想的指导下得到的科学图景。当然,有时可能会出现由于某些技术工具的研发“误打误撞”地打开了科学思想之外的科学图景——如显微镜的发明,但多数情况下,科学家们所能探索的科学之地都是在科学思想照亮的范围。就基因科技而言,围绕基因而展开的研究版图会随着科学家对基因这一概念的动态性的进一步把握而拓展。

技术和工具的不断优化使得研究变得前所未有的高效,但高效并不等于有效。未来的科学思想发展方向是思考当下识别生物大分子之间的相互作用的方式是否全面,如何对这些相互作用进行正确解读。各类 tag 标记、染料标记和非天然氨基酸修饰,是人们得以追踪生物大分子作用机制的重要手段,但已有多篇研究表明,目前每一种手段都有不适用的场景,如有些染料体积过大,会影响蛋白质正常的行为,从而使研究者得到的实验数据失真。生物大分子的模拟研究,则是最能显现科学思想重要性的一个方向。如今的计算机算力已经可以做到每秒百亿亿次,那么用何种模型、使用哪些数据库作为训练对象均需要科学思想的指导。目前即便是最好的模型,其有效性也还是需要实验数据作为支撑,并不能独立探索未知的领域。

技术和工具的发展也带来了相关的伦理问题。基因编辑技术的下一步发展不仅在于如何进行基因编辑,而且应当探讨是否要进行基因编辑。人类对基因序列的修改可能会如水中捞月般虚无缥缈。随着研究的深入,科学家们逐渐发现复杂的生命系统内的多基因调控性状是相当普遍的情形。即便是单基因调控性状,其在转录翻译等步骤也还有许多未知的调控环节共同决定着性状。同时,基因编辑的伦理问题也到了不容忽视的地步,在还未了解生命体全貌之时,贸然做出自以为有利的基因编

辑,是否会打开污染自然界基因库的“潘多拉之盒”?人类对于未知的后果是否有承担的能力?这些问题都有待人类在科学发展的进程中做出回答。

可见,在前遗传学时代,孟德尔通过杂交技术和细致的统计分析对豌豆这一模式植物进行研究,发现了遗传因子的分离和自由组合规律。这一过程不仅体现了技术进步如何促进新实验现象的发现(如性状的显隐性规律),也展示了从具体实验现象到抽象遗传规律的理论升华。在经典遗传学时代,科学家们聚焦于基因的化学本质,最终揭示了 DNA 的双螺旋结构,并提出中心法则。这一阶段的发展不仅通过 X 射线衍射技术和分子模型的应用,逐步揭示了 DNA 的结构,还确立了 DNA 作为遗传信息载体的地位,深化了人们对基因的结构和功能的理解。在分子生物学时期,各类工具的发明与改进为基因研究提供了基础,科学家们积累了大量关于基因和蛋白质的原始数据。在后基因组学时代,对海量生物数据的整体解析成为研究的核心任务。这一阶段,CRISPR-Cas 系统作为革命性的基因编辑工具,使得精准的基因操纵成为可能。与此同时,高通量测序技术的突破极大地提高了基因测序的效率和准确性。空间多组学仪器、质谱成像仪的应用,为各类组学研究提供了强有力的支撑。人工智能与大数据技术的结合,使得海量生物数据的分析和挖掘成为可能,推动了系统生物学的发展。

三、“基因”建构于还原论和系统论两端之间

纵观基因科技的发展历程,基因是从两个维度被建构的(见表 1)。一方面,“基因”的概念是逐渐演变的。这种变动体现在科学家们抽丝剥茧寻找真相的过程中。由前遗传学时代的“无中生有”,到经典遗传学时代的模糊概念,再到分子生物学时代的清晰分子结构,以及进一步到后基因组学时代的系统论整体把握。这种变动还体现在研究领域的不断拓展。前遗传学时代人们甚至不关心遗传的现象,而是更多地追寻自然界进化的奥秘;经典遗传学时代将人们的视野聚焦到各类模式生物的性状遗传的研究中;分子生物学时代则进入前所未有的微观世界,同时也让学科交叉在科研舞台上大放异彩;而后基因组学时代的科

学家必须以组学的研究方式来探究复杂的生物体系。另一方面,研究基因的技术和工具在整个发展历程中被创造并革新。孟德尔的杂交体系、适用于生物体的 X 射线衍射技术、高通量 DNA 测序技术、组学质谱等,都是在这一过程中诞生的,并

随着研究的发展同步迭代革新。由此可见,基因科技中不仅研究对象是被不断建构的,研究手段也是被不断建构的,正是这种建构性的演进才让基因科技具有蓬勃的生命力。

表 1 基因研究发展概述

阶段	研究突破	关键技术	典型工具	思想	主导要素	基因的概念
前遗传学时代	孟德尔遗传规律	杂交技术	模式生物 光学显微镜	系统论	技术(技术的进步推动了新实验现象的发现)	抽象性
经典遗传学时代	基因的化学本质 染色体与遗传的关系 DNA 双螺旋结构的发现 中心法则的提出	同位素标记技术 体外转化技术 基因突变技术 DNA 提纯技术 DNA 结晶技术 X 射线衍射技术	生物培养箱 超速离心机 X 射线衍射仪 分子建模工具	逐渐开始向还原论倾斜	思想(强还原论思想为研究指明方向,跨学科研究理念推动新技术和工具涌入这一领域)	具体性 分子化属性
分子生物学时代	人类基因组计划 结构生物学的突破	DNA 链终止法测序技术 DNA、RNA 合成技术 冷冻电镜技术 分子动力学模拟	凝胶电泳仪 高效液相色谱 质谱-色谱联用系统 初代测序仪 PCR 仪 核磁共振仪 冷冻电镜 超分辨显微镜	还原论顶峰	工具(各类仪器设备的快速发展推动了关于基因的原始数据积累)	规划性 程序性 功能化属性
后基因组学时代	基因编辑技术 人工智能与大数据	高通量测序技术 CRISPR 基因编辑技术 蛋白质鉴定与定量技术 生物大分子相互作用表征技术	生物芯片 第三代测序仪 实时定量 PCR 仪 CRISPR-Cas 系统 超高分辨率质谱成像系统 超级计算机 空间多组学仪器	还原论和系统论兼备	思想(海量数据的分析理解与进一步研究方向的确定,都需要思想的指引)	可编辑性 复杂性 数据库属性

(注:此表仅列举各阶段关键技术和典型工具。)

基因概念的变动是由科学思想带来的,但基因科技发展的四个阶段并非根据思想、技术和工具的发展来划分的,三者的发展显然在这一过程中体现出了不一样的脉络。科学思想在这一过程中体现出科学家群体思想从系统论到还原论,再到将还原论和系统论相互融合指导研究的一个渐变过程。还原论最早可以追溯到古希腊哲学家德谟克利特(Democritus)的《古希腊罗马哲学》,他提出“原子论”,认为世界由不可分割的最小微粒——原子构

成,所有现象都可以通过原子的运动和组合解释。系统论的起源则可以追溯到古希腊哲学家赫拉克利特(Heraclitus),他提出“对立统一”思想,认为宇宙是一个动态的整体,一切事物都处于不断变化和相互作用中。^[23]从孟德尔实验的再发现到 DNA 结构被解析之时,还原论开始逐渐占上风。在人类携手开展以人类基因组计划为首的各类基因组计划的时候,新一代的测序技术和各种初代的基因编辑技术,如锌指蛋白等纷纷登场,这意味着还原论的

比重在继续增大。一方面依赖高通量测序的基因组计划提供的“基因蓝图”把握了规划性,另一方面通过基因编辑的技术手段取得了可操作性。人们可以从基因的分子层面改造植物和生物。转基因作物、合成生物学中的细菌工厂等技术应用,昭示着还原论的思想到达顶峰。改造基因进入自然界的后果无从验证,生物体的复杂性和系统性被选择性地视而不见,生物研究仿佛进入了人们理想中“至简至美”的状态,即掌握基因就能掌握生物体。在强还原论思想的指导下,基因科技领域所取得的这些巨大成就,让大多数科学家不自觉地认为生命体内的谜团似乎可以通过原子和分子物理学解释,无需借助任何全新的理论。值得注意的是,引发人们思考这一问题的薛定谔本人却并不这么认为,他写道:“尽管就目前所知,生命体还没有摆脱‘物理规律’,但它有可能涉及我们迄今尚未发现的‘其他物理规律’……”^[24]在这一问题上,持此观点的并非他一人,量子力学大家尼耳斯·玻尔(Niels Bohr)等人也认为,要解释生命体的性质,或许需要借助新的物理学理论。事实也确实如此,科学家们很快便发现人类基因组计划提供的“基因蓝图”并没有想象中那样易懂,仅有1%的序列是可被遗传的,面对99%的序列解读,科学家们显得束手无策。就技术而言,在基因编辑技术中,即使是最易于操作的CRISPR/Cas9技术也相当不可控。随着发现DNA甲基化这类翻译后修饰的遗传性,基因的决定论开始被动摇,随之而来的表观遗传学、转录组学等一系列研究领域呈现出更为复杂的遗传翻译表达图景。科学思想中的还原论和系统论在面对诸多实验事实时开始相互融合,人们同时关注到微观和宏观的重要性,而未来基因科技的发展也将有赖于科学家对两个维度之间张力的正确把控。

后基因组学时代中的组学研究是系统论的显现。组学技术的特点是将基因组、表观基因组、转录组、蛋白组及代谢组等多维层次的信息进行有机整合,构建基因的调控网络,并全面探索和深层次理解各生物分子之间的调控及因果关系,从而准确阐释生命体的生物功能和生理机制。随着高通量测序技术、高分辨质谱技术及数据整合分析技术的发展,以多组学为特征的系统生物学研究实现了新突破,这一突破颠覆了生命科学研究的传统范式。传统的各自分离的单组学层面的生物化学反应和代谢途径模式,开始向对整体细胞系统进行大规模

研究的多组学整合分析模式转变。

而通过优化基因编辑技术,使其在分子水平上更加精准,则是还原论的体现。CRISPR/Cas9系统的前身——CRISPR序列,最早于1987年在大肠杆菌中被偶然发现,它提供了一种类似免疫的防御机制,保护细菌免受外来DNA的入侵,因其易于构建、编辑效率高等优点逐渐成为应用最为广泛的基因组编辑工具。然而人类并未“熟练操作”这把“剪刀”,当今的科学家们正致力于从特异性和精确度两方面对其进行优化,即只在目标靶点上实现所期望的基因编辑效果。

技术和工具的发展与思想的进步之间的互动犹如一首多声部交替出场的交响乐:思想的渐变显然是一个不均匀的过程,当其处于缓速变化时,技术和工具的发展往往对其进行推波助澜,承接起演奏“基因科技之歌”的主旋律任务。以人类基因组计划为例,前期人们以还原论的思想为指导,在成本高、测序精度有限的方式下破解基因奥秘。此时的研究领域分为三个方向:一是如何提高测序精度和通量,以节约时间,同时降低测序成本;二是对已经测序出的序列进行解读,分析其遗传信息;三是未雨绸缪的思想前瞻,可以预见测序终有一天会完成,那之后的研究方向是什么?很显然,前两个方向随着技术和工具的革新很快就到达了研究瓶颈,而第三个方向随着科学思想向系统论方向的转变而更上一层,迎来后基因组学时代。而当思想发生快速转变时,会造就新的研究热点。在研究需求的推动下,一方面新的技术和工具会应运而生,另一方面已有的技术和工具会进一步优化迭代。

前遗传学时代以孟德尔杂交实验的再发现为起点,见证了基因概念的诞生。经典遗传学时代主要围绕DNA化学成分的确和DNA双螺旋结构的发现,让基因的形象实体化。人类基因组计划拉开了分子生物学时代的序幕,基因概念中还原论的大背景下开始出现系统论的色彩。后基因组学时代则聚焦于表观遗传学、合成生物学,同时人工智能和自动化实验室的兴起也成为重要特征。这些学科和工具的出现为基因赋予了更丰富的内涵。

四、结 语

回顾基因科技发展史,基因在四个阶段不仅作为研究对象,还开拓着科学思想,引领着研究热点。

在进化论统治的时代,遗传性状未被重视之前,也就是基因概念尚未诞生之时,这一领域被束之高阁过一段时间。但自从基因这一概念横空出世之后,它便再也没有隐藏自己的锋芒。首先是经典遗传学时代,各学科交叉融合,一同揭开了基因的DNA形象本质。再到分子生物学时代,举全球之力,各国之间开展广泛合作,共同绘制人类基因组的蓝图。再到如今的后基因组学时代,科学家们不仅关注生物体内不同层面的组学研究,还将地球的整体生态纳入视野。研究领域不断拓展的背后是基因内涵不断丰富过程。

后基因组学时代中人工智能算法和自动化实验室的出现,更加凸显出科学思想的重要性。这两者广泛的应用前景会颠覆传统的“预测-分析-实验-改进-再实验”的科研模式,彻底解放科研工作者的双手。于是,从研究理论上进行独立思考将成为科研进展的主要推动力。这一趋势在带来无限机遇的同时,也引发多重挑战,至少体现在创新力培养、科研发现和数据污染三个方面。首先,实验训练作为培养创造型科研人才的重要环节,在从实践上帮助学生理解原理方面有着不可替代的作用,若完全依赖AI和机器,如何培养和体现人的创造性将成为问题。其次,过于规范的流程可能会遮蔽“异常”的实验现象,科研工作者处理和识别“异常”实验数据的能力也会下降,而很多科学发现都始于“异常”。最后,若将第一手的实验数据的获得都交给AI和机器,如何甄别真实的实验数据和伪造的实验数据将成为新的难题。而新实验范式并非万能,它只适用于相对成熟的技术和工具,面对新型技术的研发显得尤为无力。而越是前沿领域,则越倾向于使用尚未在学术界普及的新技术和新工具。

随着测序技术的发展,遗传信息不再仅仅是存储在基因组中、由细胞调用的生物学指令,同时也成为存储在基因测序数据库中、可供研究者直接检索与分析的数字化资源。这使得理论假设先于数据生成的研究模式,转向了以海量数据解析为驱动的新模式。理论驱动研究从明确的假设出发,通过实验验证因果,如设计基因敲除实验验证“基因A导致疾病B”。数据驱动研究则通过算法分析海量数据,从中发掘出如“基因A、蛋白质B与代谢物C的相互作用”的疾病关联信息,从而生成全新的假设和理论。这种研究模式的转变,反过来深刻重塑

了“基因”的概念。将基因组从一套严密的指令执行系统转变为一套极其敏感且充满反馈的响应系统,表明细胞能够调节基因表达以应对环境变化。点突变曾经一直被认为是生物变异性背后的基本机制,但现在看来,点突变只不过是基因组为应对环境挑战而进行修补的更广泛机制中的一个策略而已。基因中计划和程序的概念被削弱。基因变成了进化的产物,而非推动者。以基因敲除的研究为例,人们发现敲除基因的功能会被其他部分弥补。遗传网络显示出的冗余性,其实具有很强的缓冲能力。生物体与基因之间的关系被颠倒了:不是基因造就生物,而是生物造就基因。^[25]基因似乎从关键的决定因素变成了服务于生物体需求的普通组成部分。

在基因科技发展的不同阶段,思想、技术和工具三者均在某一时期起到了决定性的推动作用。孟德尔定律的再发现是由于人们从思想上对可遗传性状的关注,而DNA结构的发现中X射线晶体衍射技术功不可没。人类基因组计划之所以得以顺利开展,DNA聚合酶和高通量测序技术起着重要的支撑。然而,这些新技术和新工具之所以能够发挥关键作用,推动着基因科技不断向前突破,其中真正的指路明灯还是科学思想。一方面,物理、化学等学科为基因相关的研究注入新的观点与视角,建立在实验上的科研成果像不断扩大的聚光灯一般,让人们看到越来越广阔和清晰的基因科技的图景;另一方面,基因的概念逐渐从不统一且模糊到相对明确,从分子化再到系统化。这些转变都引领着科学家们开拓新的研究领域,在将技术和工具应用于研究的同时,又进一步对其进行优化和革新。在每一个阶段相互竞争的众多思想中,最能发挥工具效能的、具有最好的技术素养支持的思想往往得以胜出。思想、技术和工具之间这种相互支撑、相互适应与匹配的互动逻辑,推动着基因科技的前沿不断拓展。

参考文献

- [1] 林德宏. 科学思想史[M]. 南京: 南京大学出版社, 2020: 1-4.
- [2] Gregor M. Versuche über Pflanzen - Hybriden[J]. *Verhandlungen des Naturforschenden Vereines in Brünn*, 1866, 4: 3-47.
- [3] Vries H de. Sur la loi de Disjonction des Hybrides[J]. *Comptes Rendus Hebdomadaires des Séances de L'Académie des Sciences*,

- 1900, 130(13): 845–847.
- [4] Correns C. G. Mendel's Regel über das Verhalten der Nachkommen-schaft der Rassenbastarde[J]. *Berichte der Deutschen Botanischen Gesellschaft*, 1900, 18(4): 158–168.
- [5] Tschermak E von. Über künstliche Kreuzung bei *Pisum sativum*[J]. *Zeitschrift für das Landwirtschaftliche Versuchswesen in Österreich*, 1900, 3: 465–555.
- [6] Johnson L. Dr. W. J. Spillman's Discoveries in Genetics; An Evaluation of His Pre-Mendelian Experiments with Wheat[J]. *Journal of Heredity*, 1948, 39(9): 247–252.
- [7] Vernon H. *Elemente der Exakten Erblchkeitslehre*[M]. Jena: Gustav Fischer, 1909.
- [8] Carlson E A. *Genes, Radiation, and Society: The Life and Work of H. J. Muller* [M]. Ithaca, New York, London: Cornell University Press, 1981.
- [9] Morgan T H, Sturtevant A H, Muller H J, et al. *The Mechanism of Mendelian Heredity*[M]. New York: Henry Holt and Company, 1915.
- [10] 李文靖. 从化学键的本质到生命的本原: 鲍林与20世纪化学的两次跨学科综合[J]. *自然辩证法通讯*, 2023, 45(4): 1–8.
- [11] Dswald T, Avery M D, Colin M, et al. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types; Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type III[J]. *Journal of Experimental Medicine*, 1944, 79(2): 137–158.
- [12] Hershey A D, Chase M. Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage[J]. *Journal of General Physiology*, 1952, 36(1): 39–56.
- [13] Watson J D, Crick F H. Molecular Structure of Nucleic Acids; A Structure for Deoxyribose Nucleic Acid[J]. *Nature*, 1953, 171(4356): 737–738.
- [14] Pauling L, Corey R B. Configurations of Polypeptide Chains with Favored Orientations around Single Bonds; Two New Pleated Sheets [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1951, 37(11): 729–740.
- [15] Schmutz J, Wheeler J, Grimwood J, et al. Quality Assessment of the Human Genome Sequence[J]. *Nature*, 2004, 429(6990): 365–368.
- [16] Collins F, Galas D. A New Five-Year Plan for the U. S. Human Genome Project[J]. *Science*, 1993, 262(5130): 43–46.
- [17] Venter J C, Adams M D, Myers E W, et al. The Sequence of the Human Genome[J]. *Science*, 2001, 291(5507): 1304–1351.
- [18] Sanger F, Nicklen S, Coulson A R. DNA Sequencing with Chain-Terminating Inhibitors[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1977, 74(12): 5463–5467.
- [19] Mali P, Esvelt K M, Church G M. Cas9 as a Versatile Tool for Engineering Biology[J]. *Nature Methods*, 2013, 10(10): 957–963.
- [20] [美] 乔治·伽莫夫. 从一到无穷大[M]. 暴永宁, 译. 北京: 科学出版社, 2024: 313–314.
- [21] Merchant A, Batzner S, Schoenholz S S, et al. Scaling Deep Learning for Materials Discovery [J]. *Nature*, 2023, 624: 80–85.
- [22] Szymanski N J, Rendy B, Fei Y, et al. An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials [J]. *Nature*, 2023, 624: 86–91.
- [23] [古希腊] 赫拉克利特. 赫拉克利特著作残篇[M]. [加拿大] 罗宾森, 英译; 楚荷, 中译. 桂林: 广西师范大学出版社, 2007: 20.
- [24] Schrödinger E. *What is Life?* [M]. Cambridge: Cambridge University Press, 1992: 4.
- [25] Rheinberger H J, Müller-W S. *The Gene: from Genetics to Post-genomics* [M]. Chicago, London: The University of Chicago Press, 2017: 98.

Constructing the “Gene”: scientific Thought, Technologies and Tools

YANG Chen

(Department of the History of Science, Tsinghua University, Beijing 100084, China)

Abstract: The convergence of multiple factors has contributed to the advancement of scientific perception regarding “gene”. The concept of “gene” has been gradually constructed through the interaction of scientific thought, technologies and tools from the pre-genetic era to the era of post-genomics. Moreover, this process of constructing the concept of “gene” is characterized by fluctuations between reductionism and system theory. The future direction of gene technology will largely remain an extension of this constructive process.

Key words: gene; construct; reductionism; system theory

(本文责任编辑: 唐文佩 王慧斌)