

六、重组 DNA 和基因工程

基因是早年遗传学家提出的亚微观上的理论上的物体，用它来解释特征如何由双亲传给后代。人们曾设想由基因构成染色体，在细胞分裂时作为中心的染色体较大，能在显微镜下观察到。现在普遍相信 DNA 结构携带着基因的信息；因此 DNA 包含着遗传密码。如果科学家能够合成 DNA，他们当然也能改变其结构，从而控制遗传密码。

当有一并非由遗传得到的个体特征出现，而它又能作为一个遗传因子及时传给下一代时，这就发生了突变；即改变了生殖细胞中核苷酸的结构。已知几种能源，例如 辐射，能够产生突变。这是有道理的，因为某些类型的能量可以破坏某些键，它们随后又按另一种顺序重新形成。

如果科学家能够控制遗传密码，他们是否能控制遗传病，如镰刀形红细胞贫血症、痛风、几种形式的糖尿病、或智力发育痴呆症？如果对 DNA 的精细结构及构造这些结构时的酶活性的了解不断增加，可以有理由在结构与总体性质之间的某些详细关系终将揭晓。如果这一点能够实现，人们就可能制造某些化合物，当把它们引入活细胞时，将能抵抗或阻滞遗传症状。

以上，就是重组 DNA 和基因工程的诱人所在。

(一) 重组 DNA

基因重组现象早在摩尔根等研究果蝇遗传时就被发现了。重组 DNA 则是 70 年代初，适应基因的碱基序列分析和调节控制等基本研究的需要，建立起来的新的精细的生物技术。它着重探讨用酶切的方法，特别是用 60 年代末陆续分离出的各种限制性内切酶，将一定的外源 DNA(即所要研究或利用的 DNA 片段)在试管中同载体(如大肠杆菌中呈小圈形 DNA 的质粒)相结合成重组 DNA 分子，引入生物体，使其产生能表达引入 DNA 片段功能的新片段。

限制性内切酶(Restriction enzymes)是由瑞士微生物学家阿尔伯(Arber/Werner, 1929—)提出的。早在 20 世纪 50 年代，朱塞佩·伯塔尼(Giuseppe Bertani)就报告一种他称为“寄主控制修饰”(Host-Controlled Variation)的现象。在这种现象中，在一种寄主身上成功地生长的噬菌体(侵染细菌的病毒)都很难在其它种的细菌上生长。1962 年阿尔伯提出，细菌具有非常专门的酶，这些酶能切割噬菌体的脱氧核糖核酸(DNA)而将入侵的噬菌体消灭。这种所谓限制性内切酶的存在是由 H·O·史密斯(Smith)后来证实的。

正如阿尔伯所指出的那样，原来这种酶总是非常准确地在同一特属的位置向后者发起攻击。正是这种特性使限制内切酶具有上述重要性，如果 DNA 的链是如此准确地在特定的已知位置被切割的话，那么只需要一种能按所需要的结合把这些链联结起来的力，遗传工程就变成了现实。因为发现限制内切酶可使 DNA 链“发粘”，而且使它容易与某些“发粘”的链结合，所以对于分子生物学家来说这就意味着遗传工程最终将成为一种可行的主张。

史密斯(Smith Hamilton Othanel, 1931—)为美国分子生物学家，1971 年起任约翰斯·霍普金斯微生物学教授。1970 年，史密斯成功地证实了阿尔伯用以解释噬菌体中“寄主——控制变异现象”的假设。他鉴别出从流感嗜血杆菌中提取出的一种酶，后来被叫做 Hind^{III}，它能在特定的位点切割脱氧

核糖核酸 DNA 分子。这是第一个被识别出的所谓限制性内切酶——到 1978 年已发现 80 多种限制性内切酶。这种酶用来对基因作受控切割，它使 70 年代以来蓬勃兴起的遗传工程技术成为可能。

1970 年史密斯鉴定出第一种限制性内切酶，许多微生物学家都清楚地认识到这项技术最终可用于基因作图。美国分子生物学家、约翰斯·霍普金斯大学微生物学教授内森斯 (Nathans Daniel, 1928—) 立即开始对引起肿瘤的 SV₄₀ 病毒进行研究。到 1971 年他已能够证明该病毒可劈成 11 个分离的和特殊的碎片。在第二年，他确定了这些碎片的顺序。这样就打开了绘制基因图的道路。这也有助于脱氧核糖核酸 (DNA) 重组技术的发展。

以上三人，阿尔伯可谓限制性内切酶之催生者，史密斯为限制性内切酶之接生者，而内森斯则为限制性内切酶之启用者。正是为了这项工作，他们共同分享了 1978 年诺贝尔生理学或医学奖。

美国分子生物学家、华盛顿大学教授伯格 (Berg Paul, 1926—) 根据内森斯和史密斯的技术，研究在特定部位切割基因的方法，并研究此后以不同的方式将其重新组合的方法。这样就开创了脱氧核糖核酸 (DNA) 重组技术。它意味着在老的基因、老的病毒以及老的细菌处可以形成新的基因、新的病毒和细菌——有新性的新的生物。它也意味着可以设计出能够合成对人有用的物质 (如胰岛素) 的微生物，或其特性可供人利用的微生物——例如以石油废物为生的，或能从海水中富集某些矿物的微生物。

但是，它还意味着可能形成一些具有新的致病力的微生物；动物和人因为对它们没有天然免疫力，会被它们感染而发生致死的疾病。于是伯格在《科学》(science, 1974.7.24) 发表了“伯格信件”。这封信引来了许多著名分子生物学家的回信。在这封信中他对无控制重组脱氧核糖核酸 (DNA) 的实验的遗传危险性提出了警告。伯格说，用特别的酶，从一个有机体那里切出一段 DNA，再把它们嵌入另一个有机体的 DNA 中已变为可能。例如在所有的实验室都可以找到的无害微生物埃希氏 (Escherich) 大肠杆菌中可以嵌入一致肿瘤病毒 SV₄₀ 的活性 DNA，并可能在人群中传播，会造成无法预料的后果。伯格随后提出，完全自愿地推迟某些类型的实验，而对多数其他的实验则进行严格的控制。在加利福尼亚州阿西诺马举行一次国际性会议。随后在 1976 年由美国国立卫生研究院公布了一个严格的准则。能达到和保持这种一致很大程度是由于正直而有权威的伯格要求的结果。很有讽刺意味的是，正因为伯格从事于基因拼接技术研究，并使得率先进行 DNA 重组成为可能，他获得了 1980 年诺贝尔化学奖。

但是，此后发现，对危险估计过高了，现在已放松了一些控制。无独有偶，在伯格发出信件的 20 多年后的今天，随着用体细胞克隆羊多莉的问世，又在国际上出现了轩然大波。有人担心用此技术可以克隆人，甚至于可以克隆出希特勒来。似乎英国科学家维尔穆特克隆出来不是一只羊而是一个人。其实“克隆羊”只是突破了利用胚胎细胞进行细胞核移植的传统方式，离“克隆人”还有一段距离。无可否认，原则上是可以“克隆人”的。但人是社会动物，充其量无非是克隆其“臭皮囊”而已，而人的社会意识形态则是永远无法复制的。

(二) 基因工程

基因工程是应用重组 DNA 的技术，达到改造生物品种的目的，为农业、医药等目的服务。第一个基因工程成功的事例是：1976 年底在美国使大肠杆菌产生人脑分泌的生长抑制素。生长抑制素是含有 14 个氨基酸的多肽链，由此可以推导出含有相应的遗传密码、计 42 个核苷酸的 DNA 片段。用人工合成方法合成这个 DNA 片段，再把 DNA 片段组成重组 DNA，引入大肠杆菌，大肠杆菌就变成能产生人脑分泌的生长抑制素的新品种。1978 年美国科学家又用类似方法，使大肠杆菌变成产生人胰岛素的工厂。目前这些工作都是把比较简单的高等生物基因引入细菌。至于把细菌的基因引入高等生物的细胞内，使之表达预期的性能，或者在不同的高等生物细胞之间的基因工程，仍是大量研究的课题。例如，仅有少数细菌和简单的藻类才具有能将空气中的氮固定成氨的“固氮”基因。如果能把这种基因转移给农作物，如小麦、水稻、玉米、棉花等，使这些农作物能主动利用空气中的氮，就可以用不着既耗费资金又污染环境的大量氮肥工厂。因此，固氮基因的研究是吸引不少科学家去研究的重大课题，植物抗病基因的基因工程也吸引着科学家们进行研究，如抗棉虫棉花已经取得了重大经济效益。总之，基因工程的成果，繁花似锦，此处难以详述，只能择其要者。

1. 人类基因组计划

人类基因组计划可能是生命科学领域迄今最为浩大的工程，也是人类科学探索主流从认识外在世界的物理科学转向认识人体自身的生命科学的一个奠基工程。这项计划的推进，不仅冲击传统的医学、制药等行业，而且对社会、伦理、法律诸方面都产生影响。

现代自然科学已使人类成为这一星球真正的主宰，也使人类的健康踏上了新的台阶。人类根治了天花，战胜了霍乱等瘟疫，控制了麻疯病等不治之症。

可是，与现代自然科学在太空、信息、武器、交通等方面的辉煌成绩相比，人类对自身的认识和保护却不如人意：全球 20% ~ 50% 的人每天备受各种慢性病的折磨；我国 11% 的人患有高血压，4.2% 的人有不同程度的残疾，2.5% 的人智力低下，1/6 的孩子患近视、色盲，肿瘤、心血管等主要死因已成为驱除不掉的幽灵，艾滋病等“天外横祸”使人们对未知灾难有了新的恐惧……

人类为健康付出了昂贵的代价。美国一年的医药费在 7000 亿美元以上；美国国家医学研究院的经费居国家各研究预算之首，超过 100 个亿。我国“八五”期间每年的医疗费用，也超过 1000 亿元，接近全国财政总收入的 20%。

人类的医学已经有了过去难以想象的进步，但是，相对于人体奥秘来说，我们知道的还太少。人类基因组计划（HGP, Human Genome Project）就是人类认识自身，揭开人体奥秘，奠定 21 世纪医学、生物学发展基础的一个庞大工程。

“人类的真谛”

最早提出这一设想的是美国生物学家、诺贝尔奖获得者杜伯克（Dulbecco）。1986 年 3 月 7 日，他在美国《科学》杂志（Science）上发表了一篇题为“癌症研究的转折点——人类基因组的全序列分析”的短文，指出包括癌症在内的人类疾病的发生，都与基因直接或间接有关。因此，我们面临两种选择：要么大家各自研究自己感兴趣的基因；要么从整体上研究人类的基因组，分析整个人类基因组的序列（Sequencing the whole

genome)。

杜伯克阐述了后一策略的必要性与艰巨性。他说“这一计划的意义，可与征服宇宙的计划媲美。我们也应该以征服宇宙的气魄来进行这一计划。”“这样的工作是任何一个实验室难以承担的，它应该成为国家级项目，并成为国际性项目，人类的DNA序列是人类的真谛。这个世界上发生的一切事情，都与这一序列息息相关。”

事实上，生物学技术、特别是DNA克隆和测序等技术的进展，为研究人类基因组提供了可能性；同时，人类本身的众多遗传变异，也为研究自身提供了珍贵的材料。专家们认识到，人类基因组研究的理论与技术进展，将更直接、更广泛、更迅速地用于解决其它生物基因组的问题，揭示生命现象的本质；同时，它将推动生物高新技术的发展，产生重大的经济效益。因此，人类基因组计划一经提出，便被广泛接受，并成为国际性的重大计划。

各国纷起响应

美国从1984年小范围讨论开始，到1990年国会正式批准，美国在这项计划中走在各国前列。它的总体规划是：拟在15年内至少投入30亿美元，进行对人类全基因组的分析。其主要内容包括：人类基因组的基因图构建与序列分析；人类基因的鉴定；基因组研究技术的建立；人类基因组研究的模式生物；信息系统的建立。此外，还有人类基因组研究的社会、法律与伦理问题，交叉学科的技术训练，技术的转让，研究计划的外延等。

与此同时，欧洲一些国家也把人类基因组计划列为国家级项目。意大利的国家研究委员会在1987年组织了15个(现发展到30个)实验室开始了人类基因组研究。英国自1989年2月开始，特点是全国协调、资源集中。全国协调与资助由帝国癌症研究基金会与国家研究委员会(ICRP-NRC)共同负责，并建立了英国人类基因组资源中心。

法国的国家人类基因组计划于1990年6月宣布开始，由科学研究部委托国家医学科学院制定，主要特点是注重整体基因组(相对于美国的“单个染色体策略”)、cDNA和自动化。

此外，欧洲共同体在1990年6月通过了欧洲人类基因组研究计划，主要资助23个实验室，重点用于资源中心的建立与运转。

日本在美国的推动下，于1990年开始人类基因组计划，该计划由多个部(省)分头资助；建立“以日本人为材料的”基因组文库与cDNA文库；发展先进技术(如激光显微切割装置)。

此外，加拿大、以色列、瑞典、芬兰、荷兰、挪威、澳大利亚、新加坡、前苏联及前东德等也都开始了不同规模、各有特色的人类基因组研究。德国在1995年也加入了这项研究计划。

全球化

人类只有一个基因组。人类基因组的研究成果应该成为人类共同享有的财富。1988年4月，在科学家的倡导下，国际人类基因组组织(HUGO)成立，它代表全世界从事人类基因组研究的科学家，协调全球范围的人类基因组研究，被誉为“人类基因组的联合国”。

1988年10月，联合国科教文组织(UNESCO)成立了人类基因组委员会。1990年在莫斯科召集了以发展中国家为主体的人类基因组会议，1992年在巴西召开了第一次南北基因组会议。

发展中国家以积极的态度参与这一国际活动。印度、巴西、墨西哥、智

利、肯尼亚等国都以不同方式参与这一全球范围的合作与竞争。

“HGP 精神”

全球性人类基因组计划的实施与全球化，借助了信息革命的技术成果。建立了遍布全球的完整的数据库与信息网络。遗传图、物理图、序列图的构建能得以提前、完整地实现，最成功地体现了信息高速公路的优越性。人类基因组计划也是自然科学史上第一次将物质的实验试剂变成信息的试剂。研究人类基因组的最重要的试剂——DNA 探针与引物，可以通过国际互连网络直接索取相应的 DNA 序列，并据以在自己的实验室进行合成。

从实施人类基因组计划伊始，具有远见的科学家即号召各国政府重视这一项目，号召全世界科学家共同参与，建议所有的进展、所有的数据随时公布于众，让全世界免费享用。在实施过程中，各国科学家精诚合作，共享材料，共享数据，共同攻关。这在人类自然科学史上是史无前例的。人类基因组计划与另两个有全球性意义的项目，即曼哈顿原子弹计划和阿波罗登月计划相比，显示了人类的谐同与进步。

人类基因组计划在启动伊始，便重视其可能对社会、法律、伦理方面的冲击，特别注意这一方面的研究，并形成主流意见。HUGO 的几个重要声明，充分体现了现代自然科学求真、求善及对社会的责任感。

1996 年 3 月，多国科学家小组宣布已绘制出迄今最完整的人类基因图谱。人类基因组已经取得的成果，可以用 4 张图来概括。

2. 人类基因图谱

遗传图

遗传图又称连锁图，它是以具有多态性（在一个遗传“位点”上具有多个等位基因）的遗传标记为“路”，以遗传学距离（即在产生精子或卵子的减数分裂事件中，两个位点之间进行交换、重组的百分率 cM）为图距，反映基因遗传效应的基因组图。因此，建立人类遗传图的关键是必须有足够的高度多态的遗传标记。

STR 作为遗传标记使人类基因组的连锁分析与遗传制图发生了革命性的变化。1996 年初，已经建立了有 6000 多个以 STR 为主体的遗传标记，平均分辨率（两个标记之间的平均距离）为 0.7cM 的人类基因遗传图。

人类遗传图的科学意义与实际意义是不可估量的。以 6000 多个遗传标记作为“路”能够把人的基因组分成 6000 多个区域，只要找到某一表现型的基因与某一标记邻近的证据，就可以把这一基因定位于这一标记所界定的区域。

在疾病的遗传分析中，可以摒弃疾病发生中漫长的、多因子参与的、复杂的生理、生化过程，把疾病看成是一个或多个基因决定的表现型，然后根据“疾病位点”与选定的“遗传标记”之间的遗传学距离，即两个位点之间可能发生“遗传重组”的机率，确定疾病这一表现型基因在基因组中的位置。

物理图

人类基因组的物理图是以一个“物理标记”作为“路标”，以 Mb、kb、bp 作为图距的基因组图。这一物理标记，以前是一个序列未知的 DNA 片段（DNA 探针），现在是基于 DNA 探针的序列 STS（序列标记位置）。迄今，已经有了 15000 个 STS，平均图距（即分辨率）已达 200kb。与遗传图一样，现有的物理图已把人类庞大的基因组分成具有界标的 15000 个小区域。

物理图还有一个内容，是要以 DNA 的克隆片断连接成相互重叠的“片段

重叠群 (contig)”。这些片段都含有同一个 STS 的对应序列。

以 STS 为路标的物理图与遗传图相互参照,可以把遗传学的信息转化为物理学信息(如某一区域的大小为多少 cM,可基本“折算成某一区域大小为多少 kb”),而“片段重叠群”则提供了该区域可以操作的基因组材料,即相互重叠、覆盖这一区域的 DNA 片段,用这些片段为实验操作材料,可进行这一区域基因组研究或寻找这一区域的基因。

序列图

人类基因组的核苷酸序列图是分子水平的最高层次的、最详尽的物理图。测定总长 1 米、由 30 亿核苷酸组成的人类基因组序列图,是人类基因组计划中最为明确、最为艰巨的定时、定量、定质的硬任务,人类基因组计划在开始启动时,估计至少要 30 亿美元的投入,需 15 年完成。

近几年来,由于策略的成熟,寡核苷酸人工合成与自动测序仪的改进,测定速度比以前提高了好几个数量级。美国圣路易华盛顿大学与英国伦敦 Sanger 中心预计,仅凭他们两个实验室的经济、技术实力与经验,即使没有新的技术突破,也有望在 3 年~5 年内完成人类基因组 90% 以上的测序任务。

转录图

迄今认为所有生物的性状,包括疾病都是由结构或功能蛋白质决定的。而已知的所有蛋白质都是由 RNA 聚合酶指导合成的带有多 A (多聚腺苷酸)尾巴的 mRNA (信使 RNA) 编码的。因此,人类基因组的转录图或 cDNA (包括 cDNA 片段,即 EST) 可表达序列“标记”图,就是人类基因组图的雏形。

在人类基因组中,只有 2%~3% 的序列直接为蛋白质编码,在人体每一种特定组织中只有 10% 的基因表达,即只有不足 1 万个不同类型的 mRNA (只有在胎儿脑组织中,可能有 30%~60% 的基因表达)。根据 mRNA 的特点,可能用一种通用性引物以自动或手工测序仪,一次测定 mRNA 双端尾侧的几百个 bp。这一测定 EST 序列的策略已被好几个实验中心采用。现在国际数据库中 EST 的数量以每日 1000 多个的速率增长,至 1996 年 3 月,至少已有 60 万个,序列总长近 100Mb。

要完成人类基因组的转录图,还需要将 EST 在人的基因组中定位。EST 的大规模快速分离,给自己制造了“瓶颈”:越来越多的 EST 急待定位。现在直接定位一个 EST 的成本约为 140 美元~170 美元。但是,通过 EST 来源的基因组片段来定位 EST 已取得重大进展,约有 70% 可通过与基因组片段的序列比较而间接地定位。由相互重叠的 EST 组成的重叠片段群已达 3 万多个,它们在基因组中的位置已大致或接近确定。

3. 美科学家绘制出 X 染色体的高精度图谱

根据 1997 年最近一期《基因组研究》杂志报道,美国科学家已绘制出迄今为止最完整的 X 染色体图谱,为基因专家们在 X 染色体上寻找致病基因提供了便利条件。

X 染色体是一个巨大的 DNA 分子,上面带有 1.6 亿个碱基对。该染色体的数量可决定人的性别,女性携带两个 X 染色体,男性携带一个 X 染色体、一个 Y 染色体。以大卫·施莱辛格博士为首的 25 位华盛顿大学医学院的生化专家用了 10 年时间绘制出的这张图谱比以往任何一张染色体图谱都更为详细。图谱虽然没有排列 X 染色体上的任何基因,但科学家们在 X 染色体上每隔 7.5 万对碱基就做了一个标记。如果把 X 染色体比做一条 2100 英里长的路,则施莱辛格等人的工作就相当于在这条路上每隔 1 英里设置了一个路

标。

施莱辛格等人还发现，X 染色体上的基因并非是均匀排列的，图谱上只有 5 个部分基因密集程度很高，而其他部分基因数目很少；另外，研究人员还在这张图谱上标出了哪些部位的基因代代相传时变化不大、哪些基因遗传给后代时可能会出现变异。

这张图谱可以帮助其他研究人员在每个“路标”之间将 DNA 分子准确定位，还可以帮助研究人员在 X 染色体上寻找致病基因。该图谱标明了 X 染色体上基因分布非常集中的区域，如果在这些区域搜集目标基因，则成功的几率就会大大提高。