

第2章 基因和行为

所有行为都是由基因和环境的相互作用塑造的。简单动物的典型模式化行为受到环境的影响，而人类高度进化的行为则受到基因所决定的先天特性的制约。基因并不直接控制行为，但是基因编码的核糖核酸和蛋白质在不同的时间和多个层面上发挥作用，影响大脑。基因指定了组装大脑的发育程序，对神经元、神经胶质细胞和突触的特性至关重要，这些特性使得神经回路能够正常运作。稳定遗传的基因代代相传，它们创造了一种机制，使得新的体验能够在学习过程中改变大脑。

在本章中，我们将探讨基因如何影响行为。我们首先概述基因确实影响行为的证据，然后回顾分子生物学和遗传传递的基本原理。接着，我们将提供遗传影响行为的记录方式的实例。通过对蠕虫、苍蝇和老鼠等动物的研究，我们对基因如何调节行为有了深刻的理解，这些动物的基因组是可以进行实验操作的。从对人类大脑发育和功能的分析中，已经发现了许多基因与人类行为之间的有力联系。尽管研究人类的复杂特征存在巨大的挑战，但最近的进展已经开始揭示神经发育和精神障碍（如孤独症、精神分裂症和双相情感障碍）的遗传风险因素，这为阐明基因、大脑和行为之间的关系提供了另一个重要的途径。

2.1 了解分子遗传学和遗传可能性对研究人类行为至关重要

许多人类精神疾病和神经系统疾病都含有遗传成分。与一般人群相比，患者的亲属更有可能患上这些疾病。遗传因素在群体特征中所占的比例被称为遗传力。遗传力最有力的证据来自双胞胎研究，这一方法最初由弗朗西斯·高尔顿在1883年提出。同卵双胞胎源自一个受精卵，该受精卵在受精后不久分裂为2个；这样的单卵双胞胎共享所有基因。相比之下，异卵双胞胎由2个不同的受精卵发育而来；这些双卵双胞胎（如普通兄弟姐妹一样）平均共享大约一半的遗传信息。如图2.1.1A所示，多年来的系统比较研究显示，同卵双胞胎在神经和精神特征上的相似性（一致性）往往高于异卵双胞胎，这为这些特征的遗传成分提供了证据。

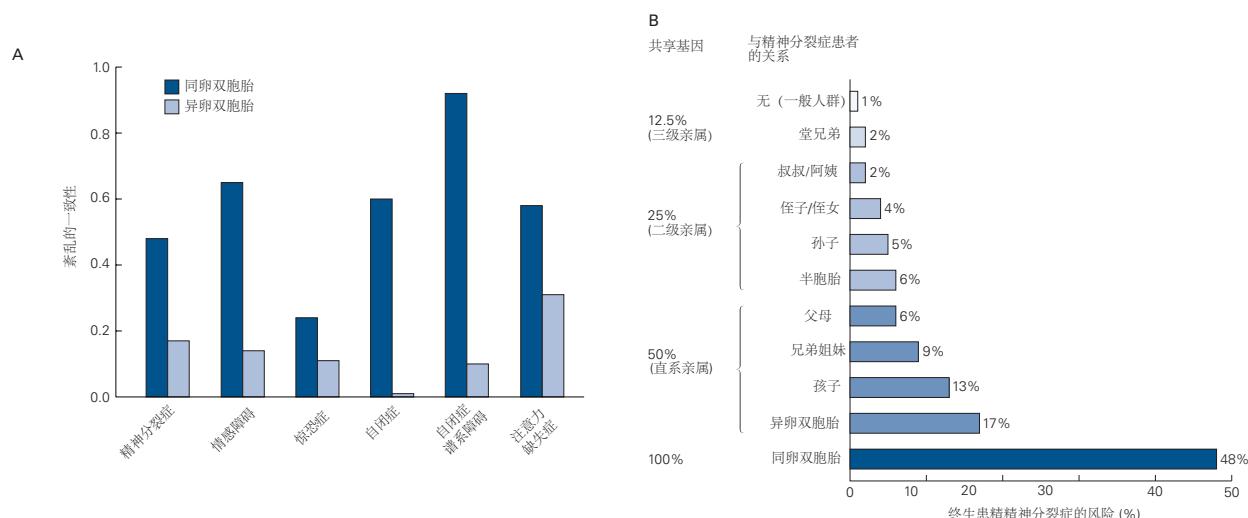


图 2.1.1: 精神疾病的家族风险提供了遗传性的证据。A. 同卵双胞胎之间精神疾病的相关性比异卵双胞胎之间的相关性大得多。同卵双胞胎几乎共享所有基因，并且有很高（但不是100%）共享疾病状态的风险。异卵双胞胎共享50%的遗传物质。0表示没有相关性（2个随机人的平均结果），而1.0分表示完全相关^[5]。B. 精神分裂症患者的近亲患精神分裂症的风险更大。就像异卵双胞胎一样，父母和孩子以及兄弟姐妹共享50%的遗传物质。如果只有一个基因导致精神分裂症，那么患者的父母、兄弟姐妹、孩子和异卵双胞胎的风险应该是相同的。家庭成员之间的差异表明更复杂的遗传和环境因素在起作用^[6]。

在双胞胎研究模型的一个变体中，明尼苏达双胞胎研究调查了那些在生命早期就被分开并在不同家庭中长大的同卵双胞胎。尽管他们的成长环境有时存在巨大差异，但这些双胞胎仍然共享着对相同精神疾病的易感性，

甚至倾向于拥有相似的性格特征，比如外向性。这项研究提供了大量证据，表明遗传变异对正常的人类差异有所贡献，不仅仅是对疾病状态。

人类疾病和行为特征的遗传力通常远低于 100%，这表明环境在形成疾病或特征中扮演着重要角色。如图 2.1.1B 所示，来自双胞胎研究的许多神经学、精神病学和行为特征的遗传力估计值约为 50%，但特定特征的遗传力可能更高或更低。尽管对同卵双胞胎和其他亲属关系的研究支持了人类行为具有遗传成分的观点，但这些研究并未告诉我们哪些基因是重要的，更没有说明特定基因如何影响行为。这些问题通过对实验动物的研究得到了解答，在这些研究中，遗传和环境因素被严格控制，同时现代的基因发现方法也正在引领我们系统地、可靠地识别出脱氧核糖核酸序列和结构中的特定变异，这些变异促成了人类精神病学和神经学的表现型。

2.2 对基因组结构和功能的理解正在不断演变

分子生物学和遗传学是现代对基因理解的核心领域。在这里，我们总结了这些领域的一些关键概念；本章末尾的词汇表 2.10 定义了常用的术语。

基因由脱氧核糖核酸组成，而脱氧核糖核酸是从一代传递到下一代的遗传物质。在大多数情况下，每个基因的精确复制体都被提供给生物体中的所有细胞，并通过脱氧核糖核酸复制提供给后代。该一般规则的罕见例外（即引入种系或体细胞脱氧核糖核酸并在疾病风险中发挥重要作用的新生突变）将在后面讨论。脱氧核糖核酸由两条链组成，每条链都有一个由脱氧核糖和磷酸构成的主链，并且连接着一系列 4 个亚基：核苷酸腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶。如图 2.2.1 所示，两条链配对，一条链上的腺嘌呤总是与互补链上的胸腺嘧啶配对，鸟嘌呤与胞嘧啶配对。这种互补性确保了脱氧核糖核酸复制过程中脱氧核糖核酸的准确复制，并驱动脱氧核糖核酸转录成称为转录本的核糖核酸长度。由于基因组几乎完全是双链的，因此碱基或碱基对可以作为测量单位互换使用。包含一千个碱基对的基因组片段称为 1 千碱基（1 kb）或 1 千碱基对（1 kbp），而一百万个碱基对则称为 1 兆碱基（1 Mb）或 1 兆碱基对（1 Mbp）。核糖核酸与脱氧核糖核酸的不同之处在于核糖核酸是单链的，具有核糖而不是脱氧核糖骨架，并使用核苷碱基尿苷代替胸腺嘧啶。

在人类基因组中，大约有 2 万个基因编码蛋白质产物，这些蛋白质产物是通过将线性信使核糖核酸序列翻译成由氨基酸组成的线性多肽（蛋白质）序列而产生的。如图 2.2.2 所示，一个典型的蛋白质编码基因由一个编码区和非编码区组成，编码区被翻译成蛋白质。编码区通常排列在称为外显子的小编码区段中，它们被称为内含子的非编码区隔开。在翻译成蛋白质之前，内含子从信使核糖核酸中删除。

许多功能性核糖核酸转录物不编码蛋白质。实际上，在人类基因组中，已经鉴定出的非编码转录本超过 4 万个，相比之下，蛋白质编码基因大约有 2 万个。这些基因包括核糖体核糖核酸和转运核糖核酸，它们是信使核糖核酸翻译机制的重要组成部分。其他非编码核糖核酸包括长链非编码核糖核酸，任意定义为长度超过 200 个碱基对，不编码蛋白质但可以在基因调控中发挥作用；指导信使核糖核酸剪接的多种类型的小非编码核糖核酸，包括小核核糖核酸；和与特定信使核糖核酸中的互补序列配对以抑制其翻译的微小核糖核酸。

身体中的每个细胞都包含每个基因的脱氧核糖核酸，但仅将基因的特定子集表达为核糖核酸。被转录成核糖核酸的基因部分由非编码脱氧核糖核酸区域所环绕，这些区域可能被其他蛋白质结合（包括转录因子），以调节基因表达。这些序列基序包括启动子、增强子、沉默子和绝缘子，它们一起允许核糖核酸在正确的时间在正确的细胞中准确表达。启动子通常位于待转录区域的起始位置附近；而增强子、沉默子和绝缘子可能位于与被调控基因有一定距离的地方。每种类型的细胞都有其独特的一套脱氧核糖核酸结合蛋白，这些蛋白与启动子和其他调控序列相互作用，以调节基因表达和细胞特性。

大脑表达的基因数量比身体的任何其他器官都要多，而且在大脑内，不同的神经元群表达不同的基因组。由启动子、其他调控序列以及与之相互作用的脱氧核糖核酸结合蛋白控制的选择性基因表达，使得有限数量的基因能够在大脑中产生大量不同的神经元细胞类型和连接。

尽管基因决定了神经系统的初始发育和特性，但个体的经历和特定神经回路中的活动可以改变基因的表达。通过这种方式，环境影响被整合进神经回路的结构和功能中。遗传学研究的主要目标之一是阐明：单个基因如何影响生物过程、基因网络如何相互影响活动、基因如何与环境相互作用。

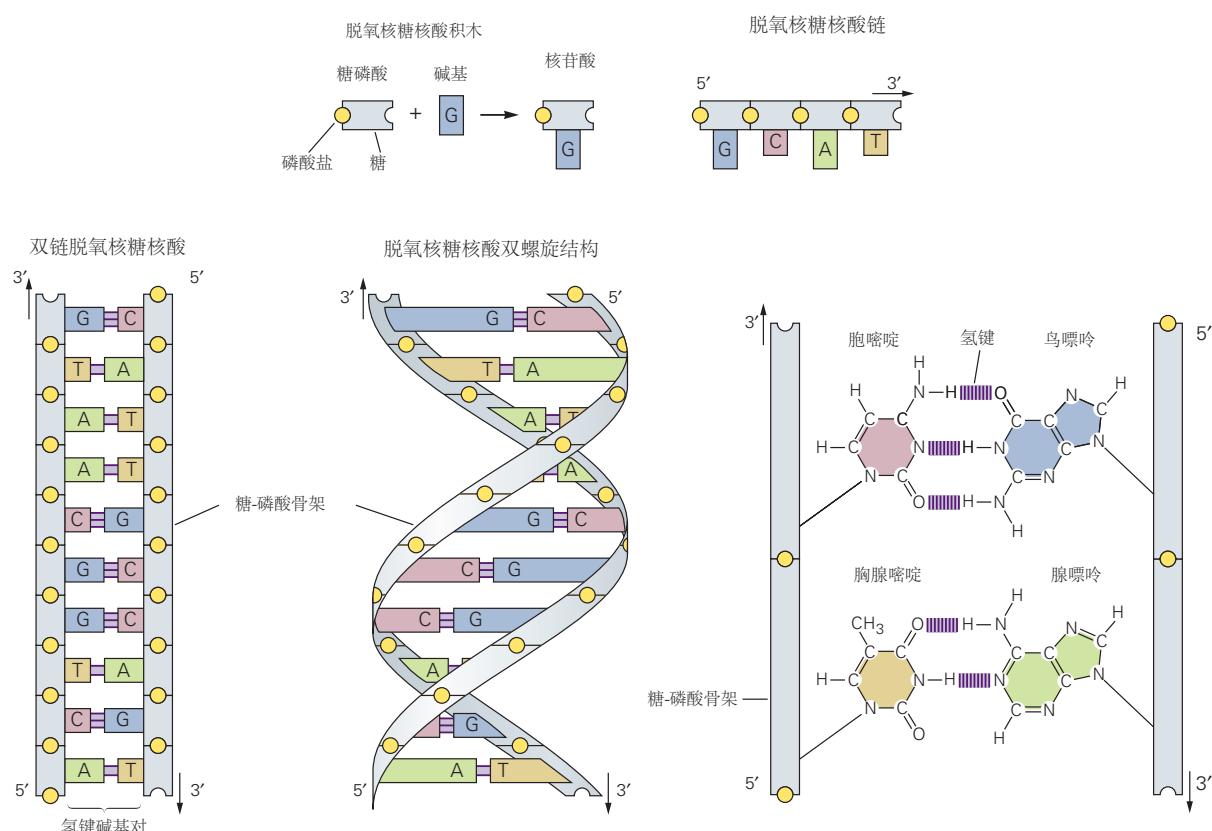


图 2.2.1: 脱氧核糖核酸的结构。4 种不同的核苷酸碱基，腺嘌呤、胸腺嘧啶、胞嘧啶和鸟嘌呤，组装在双链脱氧核糖核酸螺旋的糖磷酸主链上^[7]。

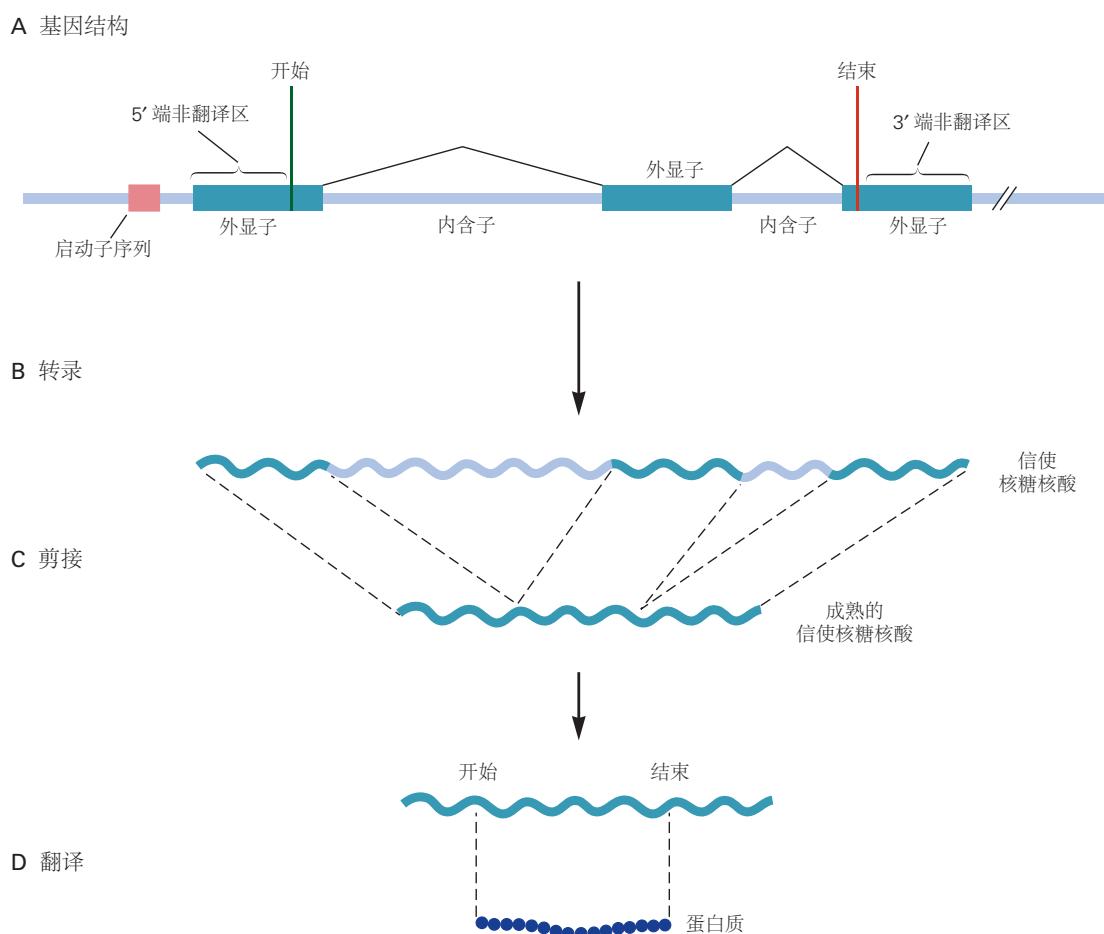


图 2.2.2: 基因结构和表达。A. 一个基因由被非编码区（内含子）隔开的编码区（外显子）组成。它的转录受非编码区调节，例如经常在基因开始附近发现的启动子和增强子。B. 转录导致产生包括外显子和内含子的初级单链核糖核酸转录物。C. 剪接从未成熟的转录物中去除内含子，并将外显子连接成为成熟的信使核糖核酸，后者从细胞核输出。D. 成熟信使核糖核酸的翻译产生蛋白质产物。

2.2.1 基因排列在染色体上

细胞中的基因有序地排列在称为染色体的长而线性的脱氧核糖核酸上。人类基因组中的每个基因都可重复地位于特定染色体上的特征位置（基因座），这个遗传“地址”可以用来将生物学特征与基因效应相关联。大多数多细胞动物（包括蠕虫、果蝇、小鼠以及人类）都是二倍体；每个体细胞携带两套完整的染色体，一套来自母亲，另一套来自父亲。

人类有大约 2 万个基因，但只有 46 条染色体：如图 2.2.3 所示，22 对常染色体（存在于男性和女性中的染色体）和两条性染色体（女性有两条 X 染色体，男性有一条 X 染色体和一条 Y 染色体）。每位父母都向二倍体后代提供每对常染色体中的一条。每位父母也向女性（XX）后代提供一条 X 染色体，但 XY 染色体的男性从他们的母亲那里继承了单一的 X 染色体，从父亲那里继承了单一的 Y 染色体。伴性遗传是由摩尔根在 1910 年通过对果蝇的研究首次发现的。这种与单个 X 染色体相关的伴性遗传模式在人类遗传学研究中具有极其重要的意义，某些 X 连锁遗传病通常只在男性中观察到，但遗传自母亲，并通过儿子传递。

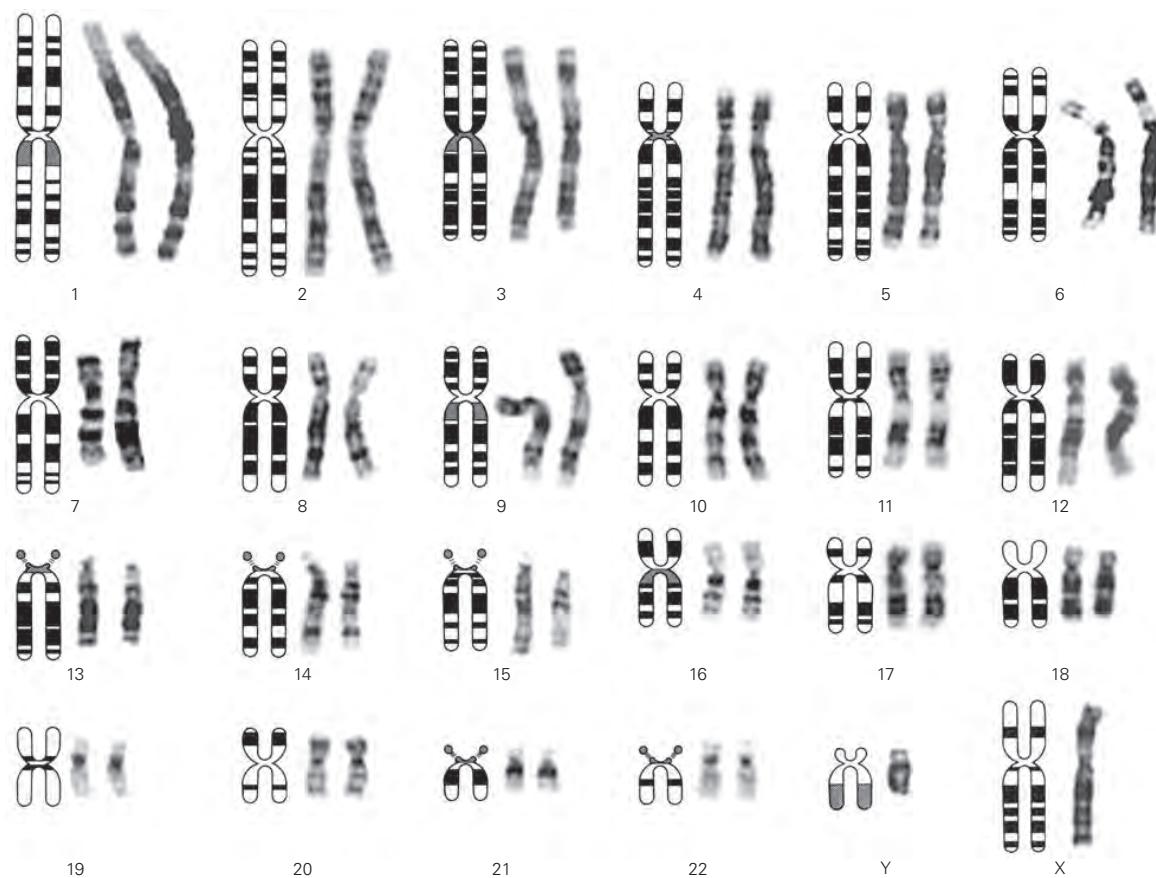


图 2.2.3：中期正常人类染色体的图解说明了每条染色体的独特形态。特征大小和特征亮区和暗区允许染色体彼此区分^[8]。

除了染色体上的基因外，生物体的极少数基因通过线粒体传递，线粒体是执行代谢过程的细胞质内的细胞器。所有孩子的线粒体都来自卵细胞，因此是由母亲传给孩子的。某些人类疾病，包括一些神经肌肉退行性疾病、一些智力障碍的形式以及一些耳聋形式，是由线粒体脱氧核糖核酸中的突变引起的。

2.3 基因型和表现型之间的关系通常很复杂

个体中特定常染色体基因的 2 个副本称为等位基因。如果 2 个等位基因相同，则称个体在该位点是纯合的。如果等位基因因为突变而不同，个体在该位点上就是杂合的。男性对于 X 染色体上的基因是半合子。一个群体可以拥有一个基因的许多等位基因；例如，影响人类眼睛颜色的单个基因 *OCA2*，可以有不同的等位基因，它们编码蓝色、绿色、淡褐色或棕色的色调。由于存在这种变异，区分生物体的基因型（其遗传构成）和表现型（其外观）非常重要。广义上讲，基因型是构成个体基因组的全套等位基因；狭义上，它指的是某个基因的特定等位基因。相比之下，表现型是对整个生物体的描述，是生物体基因型在特定环境下表达的结果。

如果突变表现型只有在基因的 2 个等位基因都发生突变时才表现出来，那么这种表现型就被称为隐性。这可能发生在个体是突变等位基因的纯合子，或者他们是所谓的复合杂合子，即在他们的每条染色体上的相应基因中携带一个不同的有害等位基因。隐性突变通常是由功能性蛋白的丢失或减少引起的。在人类和实验动物中，突变性状的隐性遗传是常见的。

如果突变表现型是由一个突变等位基因和一个野生型等位基因的组合产生的，那么这种表现型特征和突变等位基因就被称为是显性的。一些突变是显性的，因为 50% 的基因产物不足以产生正常表现型（单倍体不足）。其他显性突变导致异常蛋白的产生，或者导致野生型基因产物在不适当的时间或地点表达；如果这种异常表达与正常蛋白产物相抗衡，它就被称为显性负效应突变。

当考虑到拥有同一基因的一个正常（野生型）等位基因和一个突变等位基因的后果时，基因型和表现型之间的区别就变得明显了。在包括孤独症和癫痫在内的一系列神经发育障碍的基因发现方面的最新进展表明：人类基因组对单倍体不足的敏感性比之前认为的要高。然而，尽管一个基因的 2 个副本的完全失活通常有可预测的效果，但单倍体不足的严重程度和表现形式在不同个体之间有很大的差异，这种现象被称为可变、部分或不完全外显率。

干扰人类发育、细胞功能或行为的遗传变异是一个从常见的等位基因（也称为多态性）到罕见变异的连续体，前者通常对生物学和行为有较小的个体影响，而后者可能有较大的生物学效应（文本框 2.1）。虽然这些分类是有用的概括，但在一些重要的案例中，常见的多态性会带来很大的疾病风险；载脂蛋白 E 基因的一种常见变异，存在于 16% 的人口中，导致迟发性阿尔茨海默病的风险增加 4 倍。虽然这些分类是有用的概括，但在一些重要案例中，常见的多态性可能带来较高的疾病风险；例如，*APOE* 基因的一种常见变异，在人群中的占比为 16%，可导致晚发性阿尔茨海默病的风险增加 4 倍。

文本框 2.1 (突变：遗传多样性的起源)

尽管脱氧核糖核酸复制通常是以高保真度进行的，但被称为突变的自发错误确实会发生。突变可能是由脱氧核糖核酸中嘌呤碱基和嘧啶碱基的损伤、脱氧核糖核酸复制过程中的错误以及减数分裂过程中发生的重组引起的。

编码区内单个脱氧核糖核酸碱基（也称为点突变）的变化分为 5 大类：

1. 无声突变会改变碱基，但不会导致编码蛋白质发生明显变化。
2. 错义突变是一种点突变，导致蛋白质中的一个氨基酸被另一个氨基酸取代；利用信息学和经验证据，这些突变被越来越多地分为至少 2 个亚类：损害蛋白质功能的突变和功能中性的突变。
3. 无义突变是指其中特定氨基酸的编码区内的密码子（三重核苷酸）被终止密码子取代，从而导致蛋白质产物缩短（截短）。

4. 经典剪切位点突变改变了指定外显子/内含子边界的核苷酸。

5. 框移突变是指其中核苷酸的小片段插入或缺失改变了阅读框架，导致产生截短或异常的蛋白质。

在目前的文献中，属于后 4 类的突变（包括破坏性错义突变）通常被称为可能的基因破坏突变。

在实验遗传学研究中，当生物体暴露于化学诱变剂或电离辐射时，突变的频率会大大增加。化学诱变剂倾向于诱导点突变，涉及单个脱氧核糖核酸碱基对的变化或几个碱基对的缺失。电离辐射可诱导大量插入、缺失或易位。

在人类中，点突变在卵母细胞和精子中以较低的自发率发生，导致孩子出现突变，但父母双方都没

有，称为新生突变。每一代，整个基因组（约 30 亿个碱基对）都会发生 70 至 90 个单碱基变化，其中一个碱基对平均会导致蛋白质编码基因的错义突变或无义突变。父亲年龄较大的孩子新发点突变的数量增加，而母亲年龄较大的儿童染色体异常的频率增加。

随着 2001 年人类基因组测序和检测基因变异的高分辨率方法的不断提高，现在也很清楚，点突变并不是人类之间唯一的序列差异。某些序列可能在染色体上缺失或重复多次，因此在不同的个体中可能具有不同数量的拷贝。超过 1 个碱基和少于 1000 个碱基对的变化被称为插入缺失突变。当这种变异包含超过 1000 个碱基对时，它们被称为拷贝数变异。

任何基因变异对疾病或综合症的贡献都可以被称为简单的（或孟德尔的）或复杂的。一般来说，一个简单的或孟德尔突变是一个足以赋予表现型而没有额外遗传风险的突变。这并不意味着每个有突变的人都会表现出完全相同的表现型。然而，特定疾病等位基因和表现型之间通常存在高度可靠的关系，这种关系接近一对一的关系（如镰状细胞性贫血或亨廷顿舞蹈症）。

相反，复杂的遗传疾病是指遗传风险因素改变了疾病的概率，但并不是完全的因果关系。这种遗传贡献可能涉及罕见的突变、常见的多态性或两者兼有，并且通常是非常异质的，多个不同的基因和等位基因具有增加风险或发挥保护作用的能力。大多数复杂的疾病也与环境有关。



2.4 基因在进化中得以保存

2001 年，人类基因组的近乎完整核苷酸序列被公布，同时许多动物基因组的完整核苷酸序列也已被解码。这些基因组之间的比较得出了一个令人惊讶的结论：独特的人类物种并非源自于独特新人类基因的发明。

人类和黑猩猩在生物学和行为上存在显著差异，但它们共享 99% 的蛋白质编码基因。此外，如图 2.4.1 所示，人类约 2 万个基因中的大多数也存在于其他哺乳动物中（如小鼠），而且超过一半的人类基因与无脊椎动物，如蠕虫和苍蝇中的基因非常相似。这一惊人的发现揭示了，人类与其他动物共有的古老基因以新的方式被调控，从而产生了人类特有的属性，例如产生复杂思想和语言的能力。

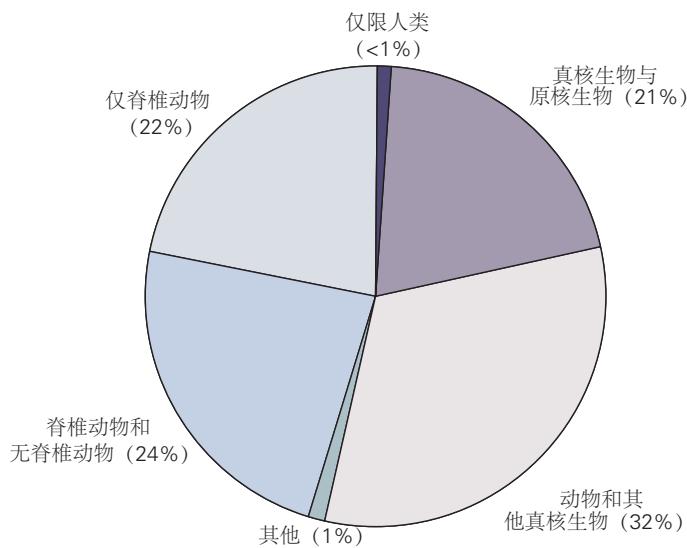


图 2.4.1：大多数人类基因都与其他物种的基因有关。不到 1% 的人类基因是人类特有的；其他基因可能为所有生物、所有真核生物、仅动物或仅脊椎动物共享^[9]。

由于基因在进化过程中的这种保守性，对一种动物的研究结果通常可以应用于其他具有相关基因的动物，这在人类实验不可行时尤其重要。例如，编码与人类基因相似的氨基酸序列的小鼠基因，通常与相应的人类直系同源基因具有相似的功能。

如图 2.4.2 所示，大约一半的人类基因的功能已从其他生物体的直系同源基因中得到证实或推断。人类、苍蝇甚至单细胞酵母共有的一组基因编码了参与中间代谢的蛋白质；脱氧核糖核酸、核糖核酸和蛋白质的合成；细胞分裂；和细胞骨架结构、蛋白质运输和分泌。

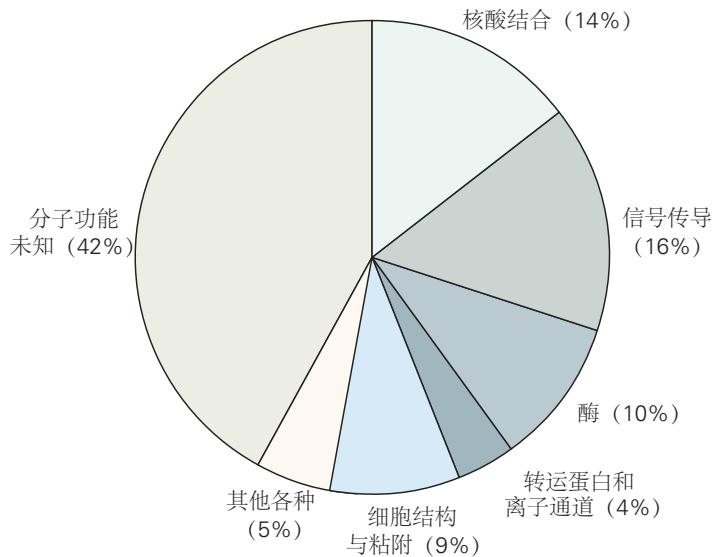


图 2.4.2: 26,383 个人类基因的预测分子功能^[10]。

从单细胞生物到多细胞动物的进化伴随着与细胞间信号传导和基因调控有关的基因数量的扩展。多细胞动物（例如蠕虫、苍蝇、小鼠和人类）的基因组通常编码数千个跨膜受体，比单细胞生物中存在的多得多。这些跨膜受体在发育过程中的细胞间通信、神经元之间的信号传递以及作为环境刺激的感应器中发挥作用。多细胞动物的基因组还编码 1 千种或更多不同的脱氧核糖核酸结合蛋白，这些蛋白调节其他基因的表达。

人类的许多跨膜受体和脱氧核糖核酸结合蛋白与其他脊椎动物和无脊椎动物中的特定直系同源基因有关。通过统计动物共有的遗传基因，我们可以推断出，哺乳动物和无脊椎动物之间的基本分子通路，包括神经发育、神经传递、电兴奋性和基因表达，存在于它们的共同祖先中。此外，对动物和人类基因的研究显示：人脑中最重要的基因往往是那些在动物系统发育过程中最为保守的基因。哺乳动物基因与其无脊椎动物对应基因之间的差异，通常是由哺乳动物中的基因复制或基因表达和功能的微妙变化引起的，而不是完全新基因的创造。

2.5 可以在动物模型中研究行为的遗传调控

由于人类与动物基因之间存在进化的保守性，通过动物模型研究基因、蛋白质和神经回路之间的关系（这些都是行为的基础）可能会为我们提供对人类中这些关系的深刻见解。在基因功能研究中，已经成功应用了 2 种重要的策略。

在经典的遗传分析中，首先对生物体进行化学或辐射诱变以诱导随机突变，然后筛选出影响特定行为（例如睡眠）的可遗传变化。这种方法不会对涉及的基因类型产生偏见；它是一种对所有可能引起可检测变化的突变进行的随机搜索。通过遗传追踪可遗传的变化，可以识别出在突变生物体中被改变的个别基因。因此，在经典遗传学中，发现的路径是从表现型到基因型，从生物体到基因。相反，在反向遗传学中，会针对一个特定感兴趣的基因进行改变，制造出转基因动物，并研究这些基因被改变的动物。这种策略既有针对性又有偏见，研

究从特定基因开始，发现的路径是从基因型到表现型，从基因到生物体。

这 2 种实验策略及其更微妙的变化构成了实验遗传学的基础。通过经典和反向遗传学的基因操作是在实验动物中进行的，而不是在人类身上进行的。

2.5.1 转录振荡调节苍蝇、小鼠和人类的昼夜节律

西摩·本泽和他的同事在 1970 年左右启动了关于基因对行为影响的首次大规模研究。他们采用随机诱变和经典遗传分析方法，识别出影响黑腹果蝇学习行为和先天行为的突变：包括昼夜节律（每日）、求偶行为、运动、视觉感知和记忆（文本框 2.2 和文本框 2.3）。这些诱导的突变极大地推进了我们对基因在行为中作用的理解。

文本框 2.2 (在实验动物中产生突变)

苍蝇的随机诱变

果蝇行为的遗传分析是在个体基因发生突变的果蝇身上进行的。突变可以通过化学诱变或插入诱变进行，这些策略可以影响基因组中的任何基因。类似的随机诱变策略被用于秀丽隐杆线虫、斑马鱼和小鼠中产生突变。

化学诱变，例如用甲磺酸乙酯，通常会在基因中产生随机点突变。当被称为转座子的可移动脱氧核糖核酸序列随机插入其他基因时，就会发生插入突变。

果蝇中最广泛使用的转座元件是 P 元件。P 元件可以被修饰为携带眼睛颜色的遗传标记，这使得它们在遗传杂交中易于追踪，并且它们也可以被修饰以改变它们所插入的基因的表达。

为了引起 P 元件转位，携带 P 元件的果蝇菌株被杂交到不携带的果蝇菌株。这种遗传杂交会导致所产生的后代中 P 元件的不稳定和移位。P 元件的动员导致其在随机基因中转移到新的位置。

小鼠的靶向诱变

哺乳动物基因分子操作的进展已经允许用突变版本精确替换已知的正常基因。产生突变小鼠菌株的过程涉及 2 种不同的操作。如图 2.5.1f 所示，染色体上的基因被称为胚胎干细胞的特殊细胞系中的同源重组所取代，修饰后的细胞系被整合到胚胎的生殖细胞群中。

必须首先克隆感兴趣的基因。该基因发生突变，然后将选择性标记（通常是耐药性基因）引入突变片段中。然后将改变的基因引入胚胎干细胞，并分离出包含改变基因的细胞克隆。对每个克隆的脱氧核糖核酸样本进行测试，以鉴定突变基因已整合到同源（正常）位点而不是其他一些随机位点的克隆。

当已经鉴定出合适的克隆时，在胚泡阶段（受精后 3 至 4 天）将细胞注射到小鼠胚胎中，此时胚胎由大约 100 个细胞组成。然后，这些胚胎被重新引入一只雌性体内，这只雌性已经为植入做了激素准备，并被允许足月。得到的胚胎是干细胞系和宿主胚胎之间的嵌合混合物。

小鼠的胚胎干细胞有能力参与发育的各个方面，包括生殖系。注射的细胞可以成为生殖细胞，并将改变后的基因传递给后代小鼠。这项技术已被用于在对神经系统发育或功能至关重要的各种基因中产生突变。

限制基因敲除和调控转基因表达

为了提高基因敲除技术的实用性，已经开发出将缺失限制在特定组织中或动物发育过程中特定点的细胞上的方法。区域限制的一种方法利用 Cre/loxP 系统。Cre/loxP 系统是来源于 P1 噬菌体的位点特异性重组系统，其中噬菌体酶 Cre 重组酶催化 34bp loxP 识别序列之间的重组，这些序列通常不存在于动物基因组中。

loxP 序列可以通过同源重组插入胚胎干细胞的基因组中，使得它们位于感兴趣基因（称为 floxed 基因）的一个或多个外显子的侧翼。当干细胞被注射到胚胎中时，人们最终可以培育出一只小鼠，其中感兴趣的基因是游离的，并且仍然在动物的所有细胞中发挥作用。

然后可以产生第二批转基因小鼠，其在神经启动子序列的控制下表达 Cre 重组酶，该序列通常在受限的大脑区域中表达。通过将 Cre 转基因小鼠系与具有感兴趣的游离基因的小鼠系杂交，该基因将仅在表达 Cre 转基因的那些细胞中被删除。

在图 2.5.2A 所示的例子中，编码 N-甲基-D-天冬氨酸谷氨酸受体 NR1（或 GluN1）亚基的基因被 loxP 元件侧翼，然后在钙/钙调蛋白依赖性蛋白激酶 2 启动子的控制下与表达 Cre 重组酶的小鼠系杂交，该启动子通常在前脑神经元中表达。如图 2.5.2B 所示，在这个特定的细胞系中，表达偶然局限于海马体阿蒙角 1 区，导致该脑区 NR1 亚基的选择性缺失。因为钙/钙调蛋白依赖性蛋白激酶 2 启动子只在出生后激活基因转录，所以这种策略可以最大限度地减少早期发育变化。

除了基因表达的区域限制外，对基因表达时间的控制给研究者提供了额外的灵活性，并可以排除在成熟动物表现型中观察到的任何异常是转基因产生的发育缺陷结果的可能性。这可以在小鼠身上通过构建一种可以用药物开启或关闭的基因来实现。

一个是从创建 2 个小鼠开始。品系 1 携带一种特殊的转基因，该转基因受启动子四环素响应元件的控制，而四环素响应元件通常只在细菌中发现。这个启动子本身不能启动基因；它需要被特定的转录调节因子激活。因此，第二组小鼠表达第二种转基因，该转基因编码一种杂交转录因子四环素激活因子，该因子识别并结合四环素响应元件启动子。四环素激活因子的表达可以置于小鼠基因组中的启动子的控制下，该启动子通常仅在特定类别的神经元或特定大脑区域中驱动基因转录。

当这 2 种小鼠交配时，一些后代将携带 2 种转基因。在这些小鼠中，四环素激活因子与四环素响应元件启动子结合并激活下游转基因。四环素激活因子转录因子之所以特别有用，是因为当它与某些抗生素（如四环素）结合时，它会失活，从而通过给小鼠服用抗生素来调节转基因表达。还可以产生表达被称为反向四环素激活因子的四环素激活因子突变形式的小鼠。这种反式激活剂不会与四环素响应元件结合，除非动物被喂食多西环素。如图 2.5.3 所示，在这种情况下，转基因总是被关闭，除非给药。

核糖核酸干扰和规律成簇的间隔短回文重复改变基因功能

最后，可以通过现代分子工具靶向基因来灭活基因。一种这样的方法是核糖核酸干扰，它利用了真核细胞中大多数双链核糖核酸被常规破坏的事实；即使只有一部分是双链的，整个核糖核酸也会被破坏。通过引入一个短核糖核酸序列，人工地使选定的信使核糖核酸变成双链，研究人员可以激活这一过程，降低特定基因的信使核糖核酸水平。

文本框 2.3 (神经解剖学导航术语)

转基因在苍蝇和老鼠中的引入

如图 2.5.4 所示，通过将脱氧核糖核酸注射到新受精卵的细胞核中，可以在小鼠体内实验性地引入基因。在一些注射的卵子中，新基因或转基因被整合到其中一条染色体上的随机位点。由于胚胎处于单细胞阶段，整合的基因被复制并最终进入动物的所有（或几乎所有）细胞，包括种系。

通过将用于产生色素的基因注射到从白化病菌株获得的蛋中而拯救的外壳颜色标记基因来说明基因掺入。有色素斑块的小鼠表明脱氧核糖核酸的成功表达。通过测试注射动物的脱氧核糖核酸样本，证实了转基因的存在。

在苍蝇身上也使用了类似的方法。将待注射的脱氧核糖核酸克隆到可转座元件（P 元件）中。当注射到胚胎中时，脱氧核糖核酸被插入生殖细胞核的脱氧核糖核酸中。P 元件可以被工程化以在特定时间和特定细胞中表达基因。转基因可以是恢复突变体功能的野生型基因，或者是改变其他基因表达或编码特异性改变的蛋白质的设计基因。

我们对昼夜节律行为控制的遗传基础有着特别完整的认识。动物的昼夜节律将某些行为与大约 24 小时的周期联系起来，这个周期与太阳的升起和落下同步。昼夜节律调节的核心是一个内在的生物钟，它以 24 小时为周期振荡。由于这个时钟具有内在的周期性，即使在没有光照或其他环境影响的情况下，昼夜节律行为也能持续存在。

生物钟可以被重置，这样昼夜循环的变化最终会导致内在振荡器的相应变化，这是任何经历时差响应的旅行者都熟悉的现象。时钟通过眼睛传递给大脑的光信号来重置。最终，时钟驱动特定行为的效应路径，如睡眠和运动。

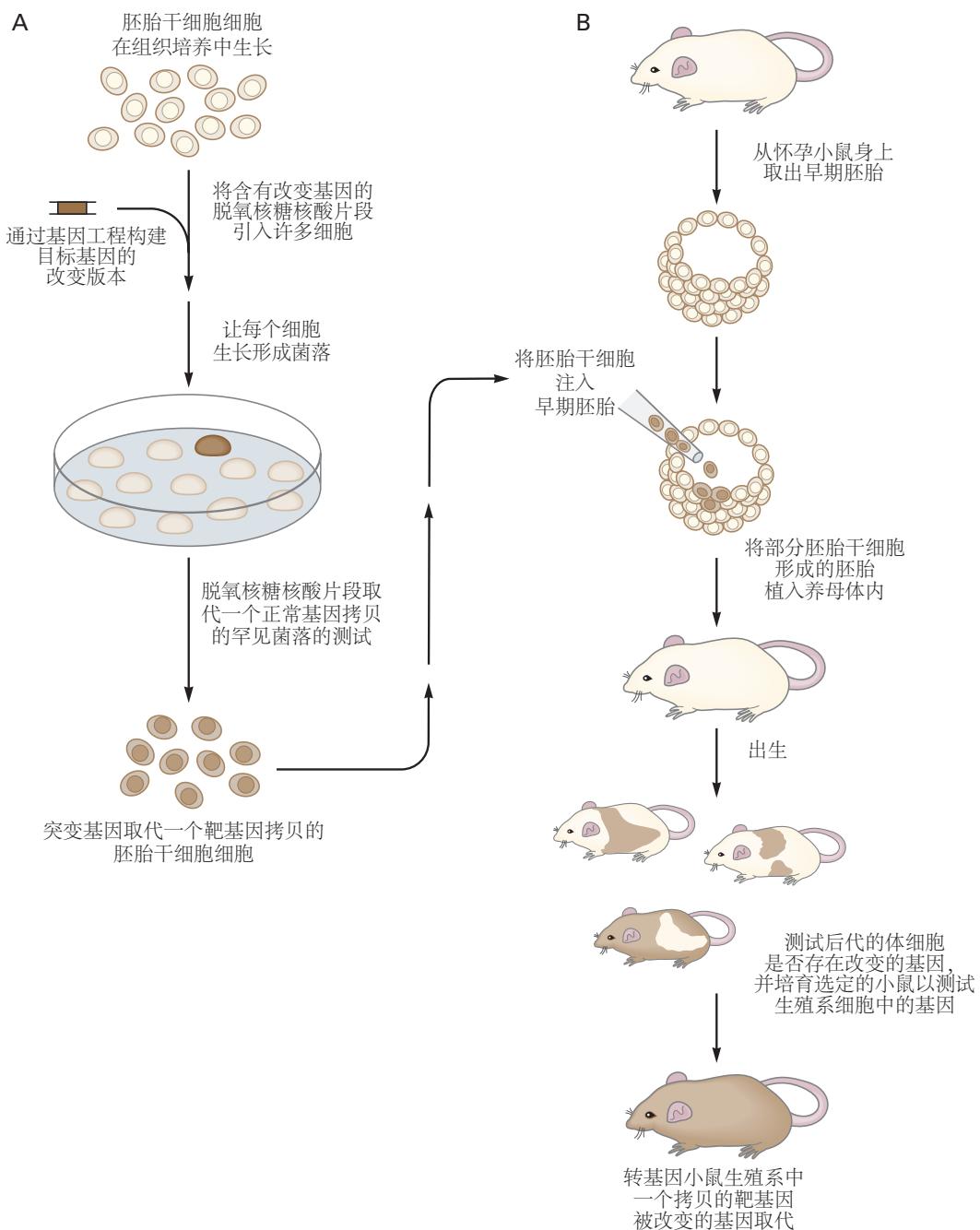


图 2.5.1: 创造突变小鼠菌株^[7]。A. 产生具有特定靶向突变的小鼠干细胞。B. 利用改变的胚胎干细胞创造转基因小鼠。

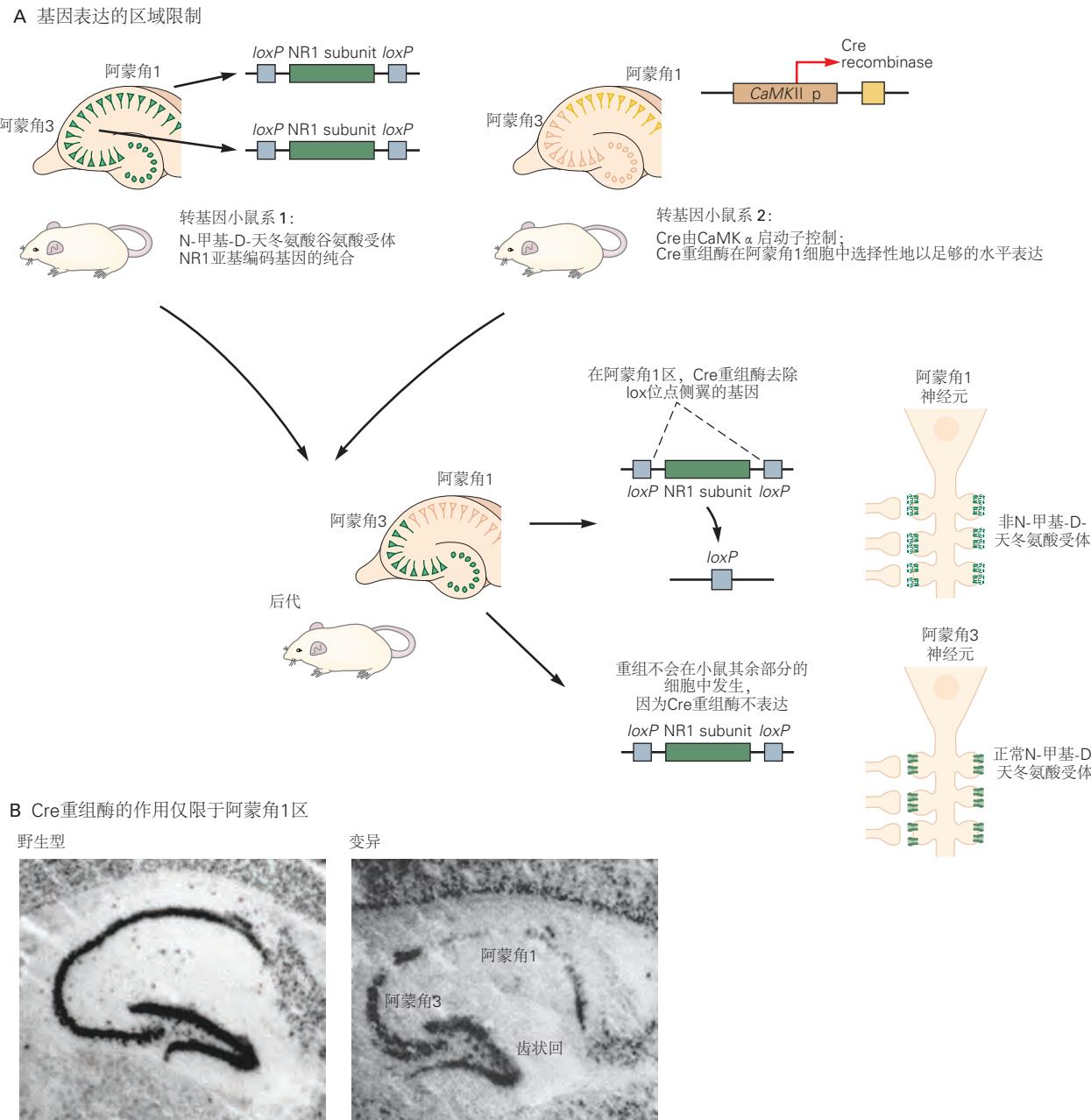


图 2.5.2: 用于选择性区域基因敲除的 Cre/loxP 系统。A. 培育了一种小鼠系，其中编码 N-甲基-D-天冬氨酸受体 NR1 亚基的基因两侧有 loxP 遗传元件（转基因小鼠系 1）。然后将这些所谓的“floxed NR1”小鼠与第二系小鼠杂交，其中编码 Cre 重组酶的转基因置于细胞类型或组织类型特异性转录启动子的控制下（转基因小鼠系 2）。在本实施例中，来自钙/钙调蛋白依赖性蛋白激酶 2a 基因的启动子用于驱动 Cre 基因的表达。在携带 Cre 重组酶转基因的 Floxy 基因纯合的子代中，仅在驱动 Cre 表达的启动子活性的细胞类型中，通过 Cre 介导的 loxP 重组，Floxy 的基因将被删除。B. 原位杂交用于检测野生型和突变小鼠海马切片中 NR1 亚基的信使核糖核酸，这些小鼠含有 2 个固定的 NR1 等位基因，并在钙/钙调蛋白依赖性蛋白激酶 2a 启动子的控制下表达 Cre 重组酶。在突变小鼠中，NR1 的信使核糖核酸表达（暗染色）在海马体阿蒙角 1 区显著减少，但在阿蒙角 3 和齿状回保持正常^[11]。

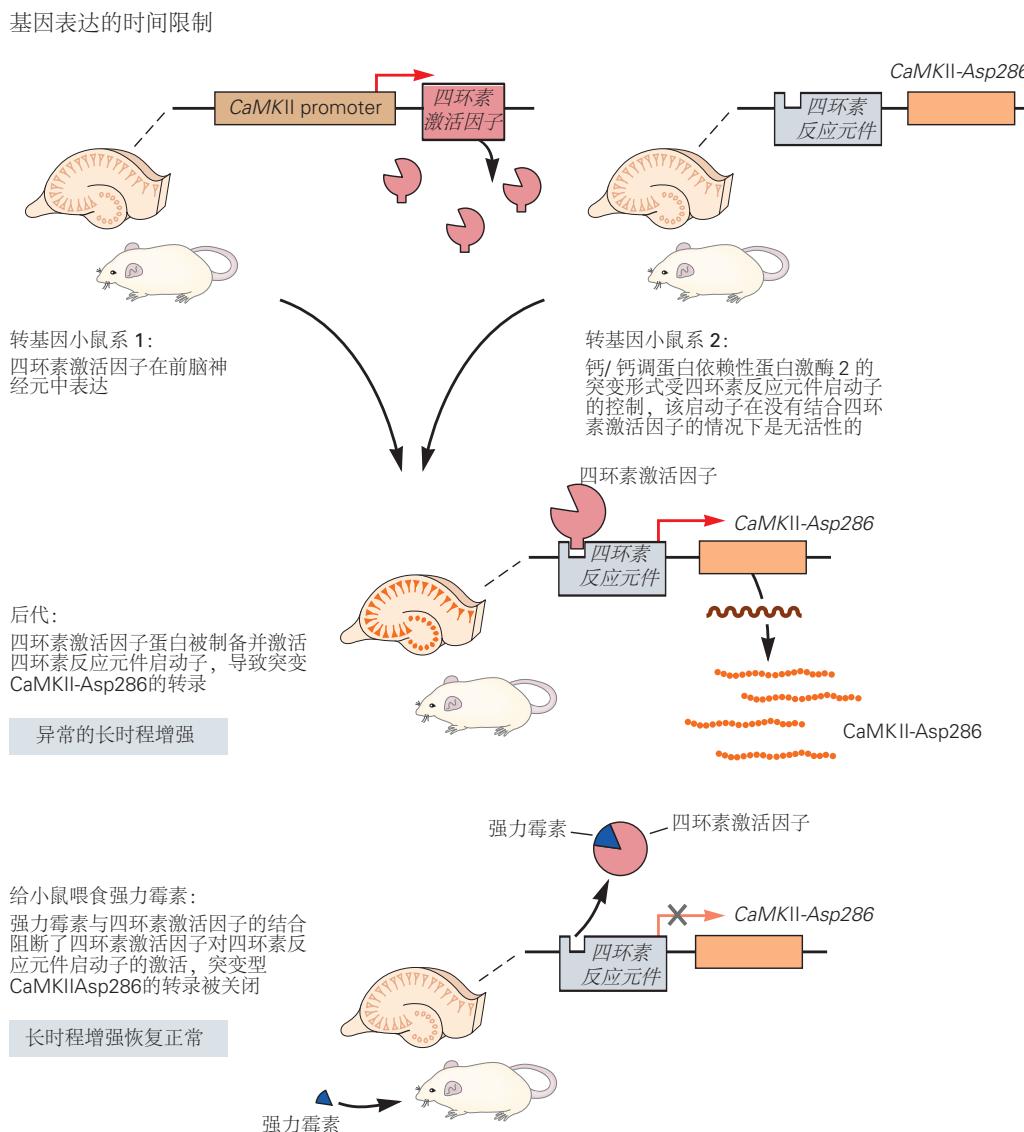


图 2.5.3: 四环素系统用于转基因表达的时间和空间调控。培育了 2 个独立的转基因小鼠系。一个系在钙/钙调蛋白依赖性蛋白激酶 2a 启动子的控制下表达四环素激活因子，这是一种结合了识别细菌四环素响应元件操纵子的细菌转录因子的工程蛋白。第 2 条线包含一个感兴趣的转基因（这里编码一种组成型活性形式的钙/钙调蛋白依赖性蛋白激酶 2），它使激酶在没有 Ca^{2+} 的情况下持续活性，其表达受四环素响应元件的控制。当这 2 个系交配时，后代以仅限于前脑的模式表达四环素激活因子蛋白。当四环素激活因子蛋白与四环素响应元件结合时，它将激活感兴趣的下游基因的转录。给后代服用四环素（或多西环素）与四环素激活因子蛋白结合，并引起构象变化，导致该蛋白与四环素响应元件解除结合，阻断转基因表达。通过这种方式，小鼠将在前脑中表达钙/钙调蛋白依赖性蛋白激酶 2-Asp286，并且可以通过向小鼠施用多西环素来阻断这种表达^[12]。

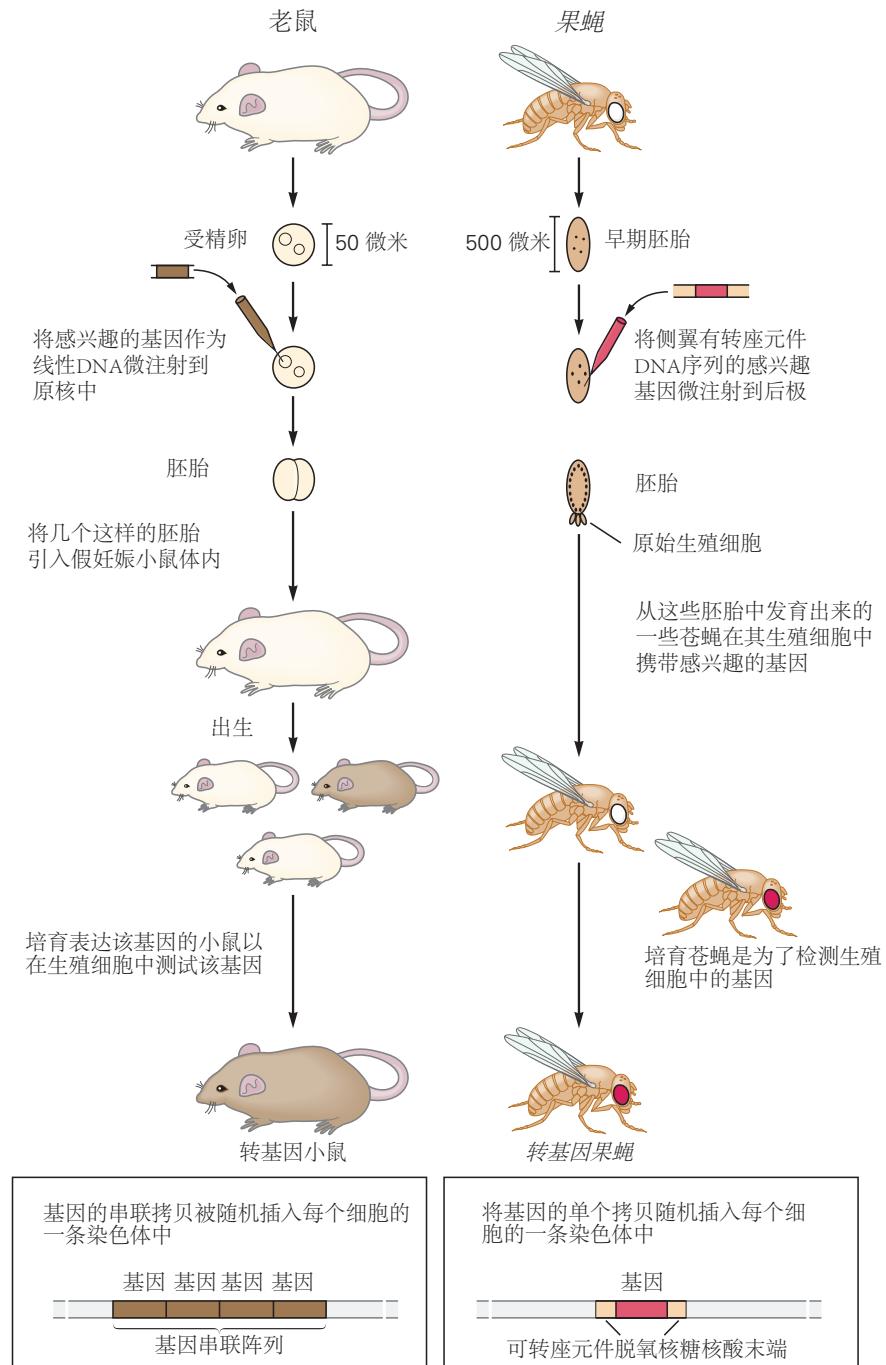


图 2.5.4: 产生转基因小鼠和苍蝇。在这里，注射到小鼠体内的基因会导致毛色的变化，而注射到苍蝇体内的基因则会导致眼睛颜色的变化。在这 2 个物种的一些转基因动物中，脱氧核糖核酸被插入不同细胞的不同染色体位点（见底部的插图）⁷¹。

本泽的团队筛选了数千只突变果蝇，寻找那些因为基因突变而无法遵循昼夜节律的罕见果蝇。这项工作首次揭示了生物钟的分子机制。周期或每个基因的突变影响果蝇内部时钟产生的所有昼夜节律行为。

有趣的是，如图 2.5.5 所示，每个突变都可以通过多种方式改变生物钟。心律失常的节律基因突变果蝇在任何行为中都没有表现出明显的内在节律，缺乏节律基因的所有功能，因此节律基因对节律行为至关重要。那些保留了基因某些功能的突变导致了异常的节律。长日等位基因产生了 28 小时的行为周期，而短日等位基因产生了 19 小时的周期。因此节律基因不仅是时钟的重要组成部分，它实际上是一个计时员，其活动可以改变时钟运行的速率。

除了昼夜节律行为的改变外，节律基因突变体没有其他重大的不利影响。这一观察很重要，因为在发现之前，许多人质疑是否可能存在动物生理需要不需要的真正“行为基因”。节律基因似乎确实是这样一种“行为基因”。

节律基因是如何计时的？蛋白质产物周期蛋白是一种转录调节因子，可影响其他基因的表达。周期蛋白的水平全天受到监管。清晨，周期蛋白及其信使核糖核酸较低。在一天中，周期蛋白信使核糖核酸和蛋白质积累，在黄昏后和夜间达到峰值水平。然后在下一个黎明前下降。这些观察为昼夜节律之谜提供了答案，即一个中枢调节器在一天中出现和消失。但它们也不令人满意，因为它们只是将问题往后推了一步，即什么是周期蛋白循环？这个问题的答案需要发现额外的时钟基因，这些基因在果蝇和老鼠身上都有发现。

受到果蝇昼夜节律筛查成功的鼓舞，约瑟夫·高桥在 1990 年代开始了在小鼠上进行的类似但更为劳动密集的遗传筛选。他筛选了数百只突变小鼠，寻找昼夜运动周期发生改变的罕见个体，并发现了一个他称之为时钟基因突变。如图 2.5.6 所示，当时钟基因突变的纯合小鼠被转移到黑暗中时，它们最初会经历极长的昼夜节律周期，然后完全失去昼夜节律。因此，时钟基因似乎可以调节昼夜节律的 2 个基本特性：昼夜节律周期的长度和在没有感觉输入的情况下节律性的持久性。这些特性在概念上与果蝇中节律基因基因的特性相同。

小鼠时钟基因与果蝇中的节律基因基因一样，编码一个转录调节器，其活动在一天中波动。小鼠时钟蛋白和果蝇周期蛋白也共享一个称为 PAS 结构域的结构域，这是转录调节因子子集的特征。这一观察表明，相同的分子机制（PAS 结构域转录调节的振荡）控制着果蝇和小鼠的昼夜节律。

更重要的是，对果蝇和小鼠的平行研究表明，相似的转录调节因子组会影响这 2 种动物的生物钟。在克隆小鼠时钟基因后，克隆了一个果蝇昼夜节律基因，发现它与小鼠时钟的关系比节律基因。在另一项研究中，研究人员通过反向遗传学鉴定出一种与苍蝇节律基因相似的小鼠基因，并将其失活。突变小鼠与苍蝇节律基因突变体一样，存在昼夜节律缺陷。换句话说，苍蝇和小鼠都使用时钟基因和节律基因来控制它们的昼夜节律。一组基因（而不是一个基因）是昼夜节律的保守调节器。

这些基因的特性揭示了昼夜节律的分子机制，并戏剧性地展示了这些机制在果蝇和小鼠之间的相似性。在果蝇和小鼠中，时钟蛋白都是一种转录激活因子。它与一个伴侣蛋白一起，控制着决定运动活动水平等行为的基因的转录。时钟蛋白及其伙伴还刺激节律基因基因的转录。然而，如图 2.5.7 所示，周期蛋白抑制时钟蛋白刺激每个基因表达的能力，因此随着周期蛋白的积累，每个转录下降。24 小时周期的出现是因为周期蛋白的积累和激活在节律基因转录后延迟了许多小时，这是周期蛋白磷酸化、周期蛋白不稳定性以及与其他循环蛋白相互作用的结果。

节律基因、时钟基因和相关基因的分子特性产生了昼夜节律所必需的所有特性。

1. 昼夜节律基因转录随 24 小时周期变化：夜间周期蛋白活性高；时钟蛋白白天的活性较高。
2. 昼夜节律基因是相互影响信使核糖核酸水平的转录因子，产生振荡。时钟蛋白按转录激活，周期蛋白抑制时钟蛋白功能。
3. 昼夜节律基因还控制其他基因的转录，这些基因进而影响许多下游响应。例如，在果蝇中，神经肽基因 pdf 控制着运动活动水平。
4. 这些基因的振荡可以通过光来重置。

2017 年诺贝尔生理学或医学奖授予了杰弗里·霍尔、迈克尔·罗斯巴什和迈克尔·扬，以表彰他们对这一分子时钟机制的详细阐释。

相同的遗传网络控制着人类的昼夜节律。患有晚期睡眠相位综合症的人有较短的 20 天周期，并且表现出极端的早睡早起的“晨型人”表现型。路易斯·普塔切克和傅颖慧发现，这些个体在人类 per 基因上存在突变。这

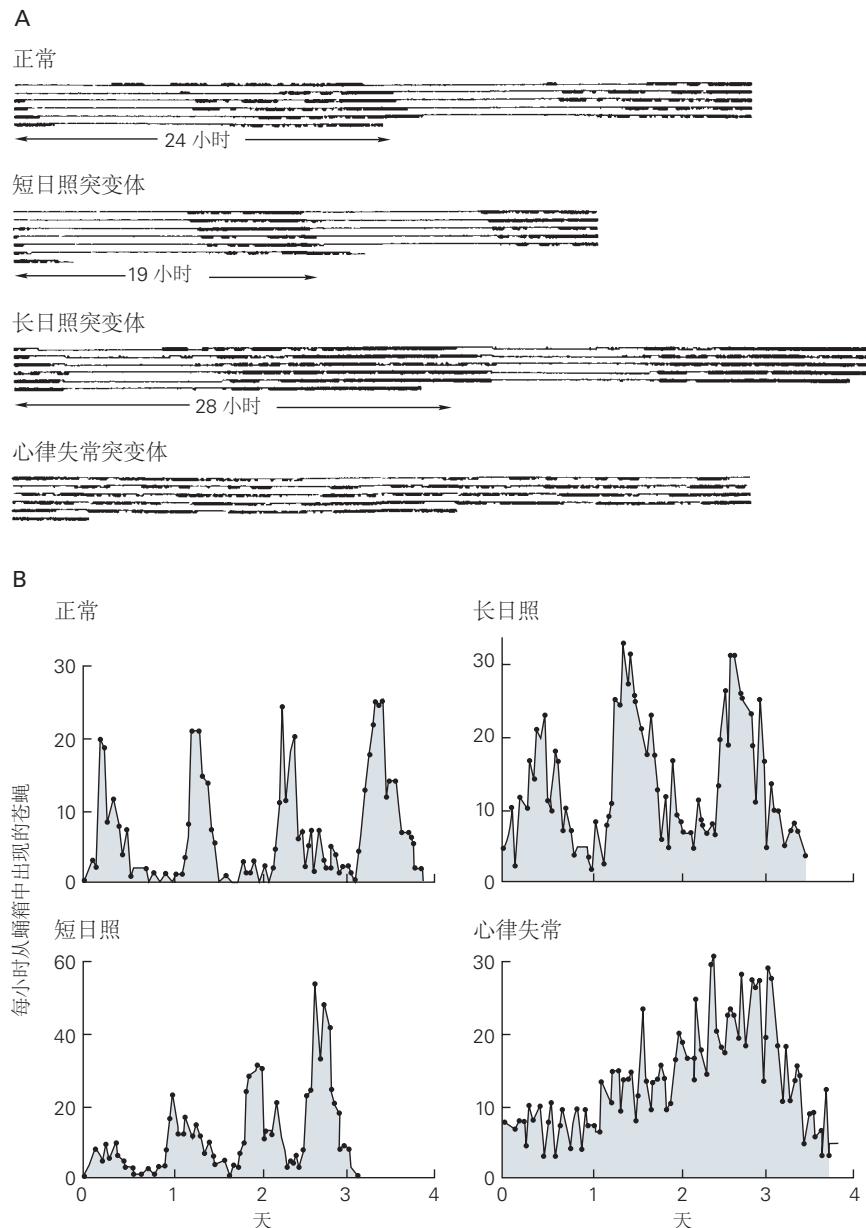


图 2.5.5: 一个基因控制着果蝇行为的昼夜节律。周期或每个基因的突变会影响果蝇内部时钟调节的所有昼夜节律行为^[13]。A. 正常果蝇和节律基因突变体的 3 种菌株的运动节律：短日照、长日照和心律失常。将苍蝇从 12 小时光照和 12 小时黑暗的循环转变为持续黑暗，然后在红外光下监测活动。记录中的粗段表示活动。B. 正常的成年苍蝇种群以周期性的方式从它们的蛹壳中出现，即使在持续的黑暗中也是如此。这些图显示了在持续黑暗的 4 天期间每小时出现的苍蝇数量（4 个种群中的每一个）。出现没有任何可辨别的节律的心律失常突变种群。

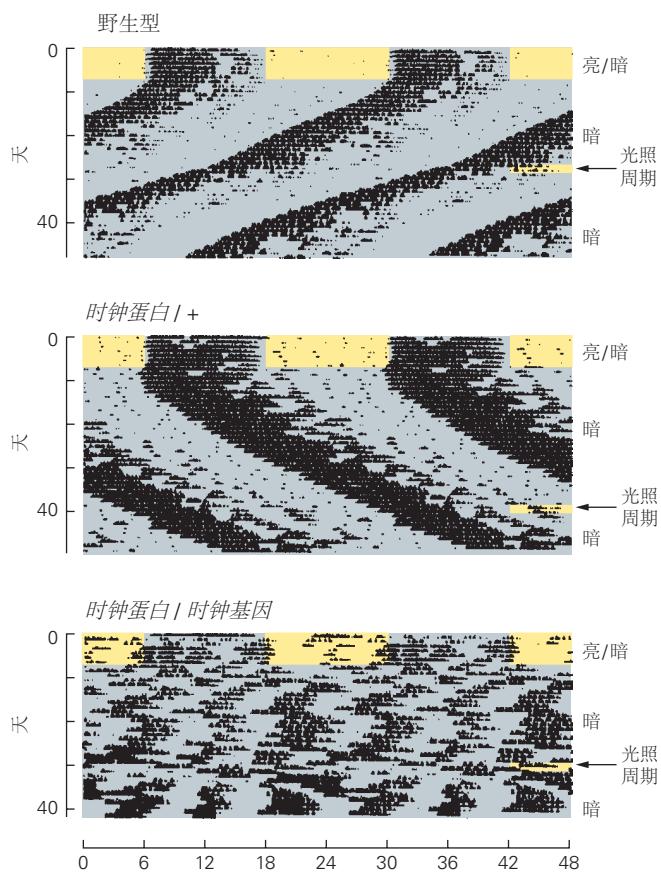


图 2.5.6: 时钟基因对小鼠昼夜节律的调节。记录显示了 3 种动物的运动活动期：野生型、杂合型和纯合型。所有动物在前 7 天保持 12 小时的亮暗循环，然后转移到恒定黑暗。之后，他们被暴露在 6 小时的光照周期中以重置节律。野生型小鼠的昼夜节律具有 23.1 小时的周期。杂合时钟/+ 小鼠的周期为 24.9 小时。纯合子时钟/时钟小鼠在转移到恒定黑暗时经历昼夜节律性的完全丧失，并在光照期后短暂表达 28.4 小时的节律^[14]。



图 2.5.7: 驱动昼夜节律的分子事件。控制生物钟的基因受 2 种核蛋白周期蛋白和非周期蛋白的调节。这些蛋白慢慢积累，然后相互结合形成二聚体。一旦它们形成二聚体，它们就会进入细胞核并关闭包括它们自己在内的昼夜节律基因的表达。它们通过抑制刺激节律基因和非周期蛋白基因转录的时钟蛋白和 CYCLE 来实现。周期蛋白非常不稳定；其中大部分降解得如此之快，以至于它永远没有机会抑制每个转录的时钟依赖性。周期蛋白的降解受不同蛋白激酶的至少 2 种不同磷酸化事件的调节。当周期蛋白与非周期蛋白结合时，周期蛋白受到保护免于降解。随着时钟蛋白驱动越来越多的节律基因和非周期蛋白基因表达，足够的周期蛋白和非周期蛋白最终积累到两者可以结合并稳定彼此，此时它们进入细胞核，在那里它们自身的转录受到抑制。结果，节律基因和非周期蛋白基因的信使核糖核酸水平下降；此后，周期蛋白和非周期蛋白水平下降，时钟蛋白可以（再次）驱动节律基因和非周期蛋白基因信使核糖核酸的表达。在白天，非周期蛋白被光调节的信号通路（包括隐花色素）降解，因此周期蛋白/非周期蛋白复合物仅在夜间形成。时钟蛋白诱导周期蛋白和非周期蛋白表达，但被周期蛋白和非周期蛋白抑制。

些结果表明，从昆虫到人类，行为基因是保守的。晚期睡眠时相综合症在睡眠章节（第 44 章）中进行了讨论。

2.5.2 蛋白激酶的自然变异调节果蝇和蜜蜂的活性

在之前描述的昼夜节律的遗传学研究中，随机诱变被用来识别生物过程中的感兴趣基因。所有正常个体都有节律基因、时钟基因和相关基因的功能拷贝；只有在经过诱变处理后，才产生了不同的等位基因。关于基因在行为中作用的另一个更微妙的问题是：哪些遗传变化可能导致正常个体间的行为差异。玛尔拉·索科洛夫斯基及其同事的研究带来第一个与物种正常个体行为变化相关基因的确定。

果蝇幼虫在活动水平和运动方式上存在差异。如图 2.5.8 所示，一些称为漫游者的幼虫会长距离移动。其他被称为“守家者”的幼虫则相对静止。从野外分离的果蝇幼虫可能是“漫游者”或“守家者”，这表明这些是自然变异，而非实验室诱导的突变。这些特征是可遗传的；“漫游者”父母会有“漫游者”后代，“守家者”父母则会有“守家者”后代。



图 2.5.8：果蝇漫游者和保姆幼虫在享用酵母时的觅食行为不同^[15]。A. 漫游型幼虫从一块地移动到另一块地，而保姆型幼虫在一块地上停留很长时间。当在一块地里觅食时，漫游者幼虫比保姆幼虫移动得更多。仅在琼脂上，漫游者和保姆幼虫的移动速度相等。B. 当漫游者在一片食物中觅食时，它们的踪迹长度比保姆幼虫长（踪迹长度是在 5 分钟内测量的）。这种觅食行为的差异映射到一个单一的蛋白激酶基因，该基因在不同的苍蝇幼虫中的活性不同。

索科洛夫斯基利用不同野生果蝇之间的杂交来研究“漫游者”和“守家者”幼虫之间的遗传差异。这些杂交实验表明，两者之间的差异在于一个主要基因，即觅食基因座。觅食基因编码信号转导酶，一种由细胞代谢物环鸟苷-3,5-单磷酸盐激活的蛋白激酶。因此，这种自然行为变异源于信号转导通路的调节变化。许多神经元功能受蛋白激酶调节，例如由觅食基因编码的环鸟苷-3,5-单磷酸盐依赖性激酶。像蛋白激酶这样的分子在将短期神经信号转化为神经元或神经回路的长期特性变化中尤为重要。

为什么果蝇野生种群中会保留信号酶的变异性，这些种群通常既包括“漫游者”也包括“守家者”？答案是环境的变化创造了对替代行为平衡选择的压力。拥挤的环境有利于“漫游者”幼虫，它们在竞争对手之前更有效地移动到新的、未被利用的食物来源；而稀疏的环境则有利于“守家者”幼虫，它们更彻底地利用当前的食物来源。

觅食基因也存在于蜜蜂中。蜜蜂在其生命周期的不同阶段表现出不同的行为；一般来说，年轻的蜜蜂担任护理工作，而年长的蜜蜂则成为觅食者，离开蜂巢。觅食基因在活跃的觅食蜜蜂的大脑中以高水平表达，而在

更年轻和更静止的护士蜜蜂中以低水平表达。幼蜂中环鸟苷-3,5-单磷酸盐信号传导可能导致它们过早地进入觅食阶段；这种变化可能是由环境刺激或蜜蜂年龄的增长所引起的。

因此，同一个基因控制着2种不同昆虫行为的变异，但以不同的方式。在果蝇中，行为变异表现在不同的个体之间；而在蜜蜂中，则表现在同一个个体的不同年龄阶段。这种差异说明了重要调控基因如何被不同物种的不同行为策略所利用。

2.5.3 神经肽受体调节几种物种的社会行为

行为的许多方面都与动物与其他动物的社会互动有关。社会行为在不同物种间差异很大，但在一个物种内，它具有很大的先天成分，这部分是由遗传控制的。在秀丽隐杆线虫这种圆虫中，研究者分析了一种简单的社会行为形式。这些动物生活在土壤中，以细菌为食。

不同的野生型菌株在进食行为上表现出显著差异。如图 2.5.9 所示，来自标准实验室株的动物倾向于独居，它们分散在细菌食物形成的草坪上，并且不与彼此互动。而其他株系的动物则展现出群居进食模式，会加入由数十或数百只动物组成的大型进食群体。这些株系之间的差异是遗传的，因为这2种进食模式都是稳定遗传的。



图 2.5.9：秀丽隐杆线虫的进食行为取决于编码神经肽受体的基因的活性水平。在一个菌株中，个体蠕虫在隔离状态下吃草（左），而在另一个菌株，个体聚集在一起觅食。这种差异可以通过神经肽受体基因中的单个氨基酸取代来解释^[16]。

群居蠕虫和独居蠕虫之间的差异是由单个基因中的单个氨基酸替换引起的，这个基因是一个涉及神经元间信号传递的大型基因家族的成员。这个神经肽受体-1基因编码一种神经肽受体。神经肽因其在协调神经元网络中的行为方面的作用而闻名。例如，海蜗牛的神经肽激素可以激发与产卵这一单一行为相关的一系列复杂动作和行为模式。哺乳动物的神经肽与进食行为、睡眠、疼痛以及许多其他行为和生理过程有关。神经肽受体的突变能够改变社会行为，这表明这类信号分子对于产生行为和个体间差异的形成都非常重要。

神经肽受体也与哺乳动物社会行为的调节有关。神经肽催产素和加压素刺激哺乳动物的亲和行为，例如配对结合和父母与后代的结合。在小鼠中，催产素对社会认知是必需的，即识别熟悉的个体。催产素和加压素在草原田鼠中已被深入研究，这些啮齿动物会形成持久的配对关系来抚养它们的幼崽。在交配期间，雌性草原田鼠大脑中释放的催产素会促进配对关系的形成。同样，在交配期间，雄性草原田鼠大脑中释放的加压素也会促进配对关系和父性行为的形成。

配对关系的强度在不同哺乳动物物种间有显著差异。雄性草原田鼠与雌性形成长期的配对关系，并帮助它们抚养后代，因此被描述为一夫一妻制。然而，与之密切相关的雄性山地田鼠则广泛交配，并不参与父性行为。这些物种中雄性行为的差异与大脑中血管加压素受体表达的差异相关。如图 2.5.10 所示，在草原田鼠中，血管加压素受体在特定大脑区域（腹侧苍白球）中以高水平表达。而在山地田鼠中，这一区域的表达水平要低得多，尽管在其他大脑区域表达水平较高。

催产素和加压素及其受体的重要性已通过在小鼠上进行的反向遗传学研究得到证实和扩展，这些研究比对

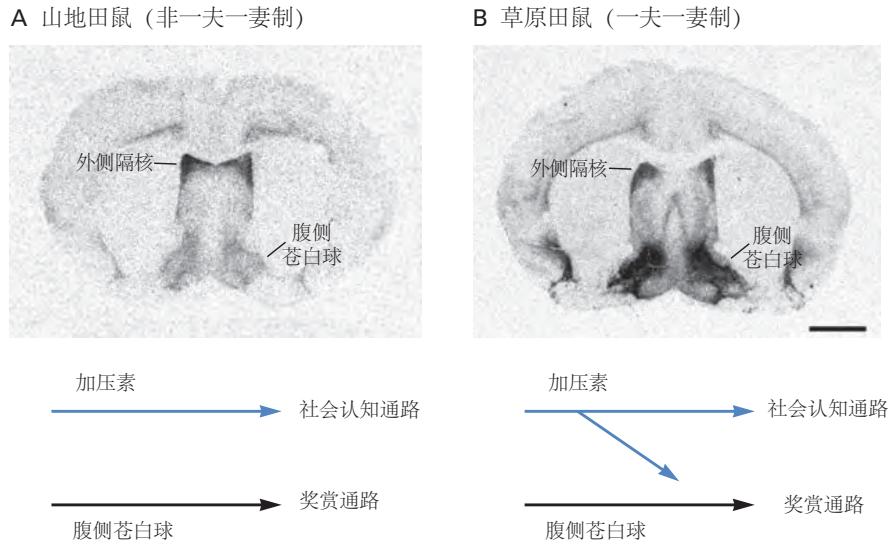


图 2.5.10: 血管加压素受体在 2 种亲缘关系密切的啮齿动物中的分布^[17]。A. 受体在山地田鼠的外侧隔核中表达较高，但在腹侧苍白球中表达较低，这不会形成配对键。B. 在一夫一妻制草原田鼠的腹侧苍白球中表达很高。受体在腹侧苍白球中的表达使加压素将社会认知通路与奖励通路联系起来。

田鼠的研究更容易进行遗传操作。将草原田鼠的血管加压素受体基因引入行为更像山地田鼠的雄性小鼠，增加了血管加压素受体在腹侧苍白球中的表达，并增加了雄性小鼠对雌性的亲和行为。因此，不同物种间加压素受体表达模式的差异可能促成了社会行为的差异。

对不同啮齿动物中加压素受体的分析提供了对基因和行为在进化过程中发生变化的机制的深入了解。因此，腹侧前脑 V1a 加压素受体表达模式的进化变化改变了神经回路的活动，将腹侧前脑的功能与交配激活的加压素分泌神经元的功能联系起来。结果，社会行为发生了改变。

催产素和加压素在人类社会行为中的作用尚不明确，但配对结合和抚养幼崽在哺乳动物中的核心作用暗示这些分子可能在我们人类中也发挥作用。

2.6 人类遗传综合症的研究为社会行为的基础提供了初步见解

2.6.1 人类脑部疾病是基因与环境相互作用的结果

在人类中发现的第一个神经系统疾病的基因清楚地展示了基因与环境在决定认知和行为表现型时的相互作用。苯丙酮尿症于 1934 年由挪威的阿斯比约恩·佛林描述，影响大约每 15,000 名儿童中的一名，导致严重的认知功能障碍。

患有这种疾病的儿童有 2 个编码苯丙氨酸羟化酶的苯丙酮尿症基因的异常拷贝，该基因编码的苯丙氨酸羟化酶是一种将氨基酸苯丙氨酸转化为酪氨酸的酶。这种突变是隐性的，因此杂合子携带者没有症状。缺乏正常功能的 2 个基因副本的儿童会从饮食蛋白质中累积高水平的苯丙氨酸，这反过来又导致产生干扰神经元功能的有毒代谢物。苯丙氨酸如何具体地对大脑产生不利影响的生化过程尚未完全理解。

苯丙酮尿症表现型（智力障碍）是基因型（纯合子苯丙酮尿症基因突变）和环境（饮食）相互作用的结果。因此，苯丙酮尿症的治疗简单有效：可以通过低蛋白饮食来预防发育迟缓。苯丙酮尿症基因功能的分子和遗传分析已使受影响个体的生活得到显著改善。自 1960 年代初以来，美国对新生儿苯丙酮尿症进行了强制性检测。通过在症状出现之前识别出患有遗传性疾病的儿童并调整他们的饮食，可以预防疾病发展的许多方面。

本书后面的章节描述了很多单基因特征的例子（例如苯丙酮尿症），这些例子增进了我们对大脑功能和功能障碍的理解。从这些研究中浮现出一些主题。例如，一些罕见的神经退行性疾病（如亨廷顿病和脊髓小脑性共济失调）是由蛋白质中谷氨酸残基的病理性显性扩张引起的。这些多谷氨酸重复疾病的发现突显了未正确折叠

和聚集的蛋白质对大脑的潜在危险。发现癫痫发作可以由离子通道的多种突变引起，使人们意识到这些疾病主要是神经元兴奋性障碍。

2.6.2 罕见的神经发育综合症为社会行为、知觉和认知的生物学提供了见解

在儿童时期显现的神经和发育障碍突显了遗传学在人类大脑功能中的重要性和复杂性。基因影响特定认知和行为回路的早期证据来自对一种称为威廉斯综合症的罕见遗传病的研究。患有此疾病的个体通常表现出正常的语言能力和极高的社交性；在发育早期，他们缺乏儿童在陌生人面前常有的羞怯。与此同时，他们在空间处理方面严重受损，整体智力存在障碍，并有很高的焦虑率（尽管很少有社交焦虑症）。

与孤独症谱系障碍等其他疾病相比，威廉综合症的损伤模式表明，语言和社交技能可以与大脑的其他一些功能区分开来。在孤独症儿童中，与语言相关的大脑区域受损，但在威廉综合症中活跃或加重。相比之下，威廉综合症患者的一般智力和空间智力受损程度超过一半的孤独症谱系障碍儿童。

威廉综合症是由染色体区域第 7 号染色体 1 区 1 带 2 亚带 3 次亚带的杂合缺失引起的，通常包含约 1.5 Mb 和 27 个基因。对这一缺陷最简单的解释是：该区段内基因的表达水平降低，因为该区域的每个基因只有 1 个副本而非 2 个副本。影响社交沟通和空间处理的确切基因尚不清楚，但它们非常引人关注，因为它们可能提供对人类行为遗传调控的洞察。

孤独症谱系障碍研究的最新发现进一步强调了遗传变异与社会和智力功能之间的复杂关系，这一关系最初由威廉综合症阐明。在过去十年中，基因组技术的进步允许使用高通量方法筛选基因组中的染色体结构变异，并且分辨率远高于光学显微镜（见文本框 2.1）。2007 年和 2008 年的开创性研究表明：患有孤独症谱系障碍的个体比未受影响的个体更频繁地携带新生拷贝数变异。这些发现促成了一些首次发现特定基因组区段对常见形式的综合症（即没有综合症特征证据的孤独症谱系障碍，也称为特发性或非综合症性孤独症谱系障碍）有贡献。

2011 年，两项同时进行的大规模研究对一个非常明确定义的群体中的去新拷贝数变异进行了研究，发现在威廉综合症中被删除的确切相同区域在个体中显著增加了孤独症谱系障碍的风险。然而，在这些案例中，是罕见的重复（该区域的一个额外副本），而不是缺失，显著增加了社会障碍的风险。这些发现表明，相同一组基因的丢失和增加可能导致对比鲜明的社会表现型（尽管两者通常都会导致智力障碍），进一步支持了认知和行为功能领域是可分离的，但可能共享重要的分子机制。

脆性 X 综合症是另一种儿童神经发育障碍，可以深入了解认知功能的遗传学；与威廉综合症不同，它已被映射到 X 染色体上的单个基因。脆性 X 综合症的表现各不相同。受影响的儿童可能有智力障碍、社交认知差、社交焦虑高和重复行为；大约 30% 的脆性 X 综合症男孩符合孤独症谱系障碍的诊断标准。脆性 X 综合症还与更广泛的特征相关，包括身体特征（例如拉长的脸和突出的耳朵）。

已证明脆性 X 综合症是由基因突变引起的，该基因突变会降低一种名为脆性 X 智力迟钝蛋白的基因表达。由于该基因位于 X 染色体上，当男性的一份拷贝发生突变时，他们会失去该基因的所有表达。脆性 X 智力迟钝蛋白蛋白在神经元中调节信使核糖核酸向蛋白质的翻译，这一过程本身受神经元活动调节。在神经元中受调节的翻译是学习所需的关键组成部分。因此，翻译水平上的脆性 X 缺陷会上升到影响神经元功能、学习和更高层次的认知过程。有趣的是，大部分与孤独症谱系障碍和精神分裂症风险增加相关的其他基因都受脆性 X 智力迟钝蛋白调节。

另一种遗传基础广为人知的孟德尔疾病是雷特综合症（在第 62 章中详细讨论）。雷特综合症是一种 X 连锁的进行性神经发育障碍，是女性智力障碍的最常见原因之一。这种疾病几乎总是局限于女性，因为典型的雷特突变在发育中的男性胚胎中通常是致命的，而男性胚胎只有一条 X 染色体。受影响的女孩通常发育到 6 至 18 个月大时，她们无法获得语言能力，智力功能出现退化，并且显示出强迫性的、不受控制的手部扭动，而非有目的的手部运动。此外，患有雷特综合症的女孩通常会表现出一段时间的社交互动明显受损，这可能与孤独症谱系障碍无法区分，尽管人们认为社交功能在以后的生活中很大程度上得到了保留。胡达·佐格比和她的同事发现，这种综合症的主要原因是甲基 CpG 结合蛋白 2（MeCP2）基因的突变。脱氧核糖核酸中特定 CpG 序列的甲基化会改变附近基因的表达，而 MeCP2 的一项既定功能是它结合甲基化脱氧核糖核酸作为调节信使核糖核酸转录过程的一部分。

罕见综合症还提供了一些对精神分裂症遗传底物的初步见解（第 60 章）。例如，正如罗伯托·什普林茨恩及其同事在 1978 年首次描述的那样，22q11 染色体缺失会导致广泛的身体和行为症状，包括精神病，现在通常被称为腭心面综合症、迪格奥尔格综合症或 22q11 缺失综合症。由于与相同缺失相关的表现型范围极其广泛，什普林茨恩的最初描述遭到了一些怀疑。现在人们普遍认为，22q11 缺失是与精神分裂症和儿童期发病的精神分裂症相关的最常见的染色体异常。此外，同一区域的染色体丢失已被发现与孤独症的个体风险有关。迄今为止，该区域内负责精神病表现型的特定基因尚未被明确确定。此外，来自孤独症文献的最新证据表明，很可能是这一区间内多个基因的组合，每个基因都赋予相对较小的个体效应，是社会残疾表现型的原因。

2.7 精神疾病涉及多基因特征

如前所述，与神经退行性和精神疾病的总体负担相比，单基因综合症较为罕见。因此，人们可能会质疑，如果罕见疾病只占疾病总负担的一小部分，那么研究这些罕见疾病的理由何在。原因在于，这些罕见病状可以为我们提供对更常见、更复杂疾病形式所涉及的生物学过程的深刻见解。例如，人类遗传学的一个显著成就就是发现了导致早发性阿尔茨海默病或帕金森病的罕见遗传变异。拥有这些严重罕见变异的个体只占所有患有这些疾病的个体的一小部分，但对这些罕见疾病变异的识别揭示了在更广泛的患者群体中也被干扰的细胞过程，从而指向了普遍的治疗途径。同样，对雷特综合症、脆性 X 综合症和其他神经发育障碍背后的病理生理机制的探索已经导致了对精神综合症合理药物开发的一些初步尝试。

在本章的其余部分，我们将扩展对 2 种复杂神经发育和精神表现型的遗传学讨论：孤独症谱系障碍和精神分裂症。与前面讨论的罕见孟德尔例子相比，这些疾病的常见形式的遗传学确实更为多样化、多变和异质，涉及不同个体中的许多不同基因，以及在组合中提供易感性的多个风险基因。此外，对于这 2 种诊断，尽管对遗传贡献的支持是显著的，但同样也有充分的证据表明环境因素的贡献。

对这些疾病的理解进展来自于快速发展的基因组技术和统计方法的结合、开放数据共享的文化，以及足够大的患者群体的整合，这些群体提供了足够的能力来检测非常罕见的高渗透性等位基因以及携带小风险增量的常见遗传变异。重要的是，最近在理解这 2 种综合征方面的成功，为追求它们的生物学后果和这些遗传风险因素所传达的分子、细胞和回路层面的病理生理学提供了坚实的基础。

2.7.1 孤独症谱系障碍遗传学的研究进展凸显了罕见突变和新生突变在神经发育障碍中的作用

孤独症谱系障碍是一系列不同程度的发育综合症，影响大约 2% 至 3% 的人口，特征是社交沟通障碍以及模式化的兴趣和重复行为。男性明显占优势，平均而言，受影响的男孩是女孩的 3 倍。根据定义，孤独症谱系障碍的临床症状，在生命的前 3 年出现，尽管在生命的最初几个月内，受影响和未受影响儿童之间的高度可靠差异通常就可以识别出来。

受影响个体之间存在相当大的表现型变异性，这导致了孤独症谱系障碍的广泛诊断分类。此外，这些个体比一般人群更频繁地出现癫痫发作和认知问题，并且经常在适应功能方面有严重障碍。然而，许多孤独症患者并没有受到如此深远的影响，并能过上非常成功的生活。

如图 2.1.1A 所示，孤独症具有很强的遗传成分，这可能是它成为首批屈服于现代基因发现工具和方法的遗传复杂神经精神综合症之一的原因。孤独症谱系障碍具有更广泛的意义，因为它为理解典型的人类行为（语言、复杂智力和人际互动）提供了洞察力。重要的是，孤独症谱系障碍中观察到的社会沟通缺陷可以与正常智力和其他领域的典型功能共存，这表明大脑在某种程度上是模块化的，具有可以独立变化的不同认知功能。

虽然孤独症谱系障碍的综合征形式只占所有病例的一小部分，但在更常见的所谓“特发性”或“非综合征”形式的障碍中，首次发现也表明罕见突变扮演了重要角色，这些突变具有重大的生物学效应。例如，在 2003 年，对极少数具有孤独症特征的女性 X 染色体上缺失区域的基因进行测序，导致在基因神经连接蛋白 4X 中发现了罕见的功能丧失突变，一种在兴奋性神经元中编码突触粘附分子的基因，在几个受影响的男性家庭成员中发现。不

久之后，对一个大型家族进行的连锁分析显示，患有智力障碍和孤独症谱系障碍的家族成员都携带有 NLGN4X 的功能丧失突变。

染色体结构中的新生亚显微缺失和重复可能会显著增加个体患孤独症谱系障碍的风险。这些拷贝数变异聚集在基因组的特定区域，确定特定的风险区间。使用这种方法的最早报告表明，染色体 16p11.2 的新生拷贝数变异，虽然仅存在于大约 0.5% 至 1% 的受影响个体中，但具有很大（大于 10 倍）的孤独症谱系障碍风险。随后的研究现在已经确定了十几个或更多具有风险的新拷贝数变异，包括第 16 号染色体 1 区 1 带 2 亚带、第 1 号染色体 2 区 1 带、第 15 号染色体 1 区 1 带 1 亚带 3 次亚带和第 3 号染色体 2 区 9 带；第 22 号染色体 1 区 1 带、第 22 号染色体 1 区 3 带（删除基因 SHANK3）和第 2 号染色体 1 区 6 带（删除基因 NXRN1）的缺失；和第 7 号染色体 1 区 1 带 2 亚带 3 次亚带（威廉综合症区域）的从头重复。

有趣的是，尽管这些 CNVs 对孤独症谱系障碍具有很高的风险，但包括精神分裂症和双相情感障碍在内的其他精神疾病的研究也发现，许多相同的区域同样增加了这些疾病的风险。此外，通过基因型（例如，第 16 号染色体 1 区 1 带 2 亚带缺失和重复）确定的个体研究发现了多种相关的行为表现型，从特定语言障碍到智力障碍再到精神分裂症。这种“一对多”现象对于阐明精神疾病的特定病理生理机制以及从基因发现到治疗的概念化提出了重要的挑战。

新生稀有的拷贝数变异会增加孤独症谱系障碍和其他发育障碍的风险，这一广泛且可复制的发现立即引发了一个问题，即单个基因中的稀有新生突变是否可能带来类似的风险。事实上，低成本、高通量脱氧核糖核酸测序技术的发展，最初集中在基因组的编码部分，导致了被认为可能破坏受影响个体基因功能（可能的基因破坏突变）的大量新生突变的鉴定。这些突变在无关个体中反复出现，现在已经被用作识别孤独症谱系障碍特定风险基因的手段。

对孤独症谱系障碍新生突变的大规模研究现在已经确定了 100 多个相关基因，其中约 45 个达到了统计学上的最高度置信。这些基因具有广泛的已知功能，但分析揭示了参与突触形成和功能以及转录调控的基因在统计学上显著过量。此外，编码核糖核酸的风险基因数量超过预期，这些核糖核酸是脆性 X 智力低下蛋白和/或在早期大脑发育中活跃蛋白质的目标。

2.7.2 精神分裂症基因的鉴定凸显了罕见风险变异和常见风险变异的相互作用

精神分裂症影响了大约 1% 的年轻人，导致思维障碍和情绪退缩的模式，严重损害了生活。如图 2.1.1B 所示，它具有很强的遗传性，并且还具有与发育中胎儿的压力相关的强大环境成分。第二次世界大战后荷兰饥荒时期出生的儿童，在多年后患精神分裂症的风险大大增加；而在 1960 年代风疹病毒大流行期间，母亲在怀孕期间感染风疹病毒的儿童，也面临更高的风险。

基因以及环境都对精神分裂症有贡献。与孤独症一样，人类基因组的测序、常见变异的全基因组基因分型和拷贝数变异检测的廉价方法开发，以及整合了非常大的患者群体，所有这些都促进了精神分裂症遗传学的重大转变。

首先，与早先提到的孤独症谱系障碍的发现基本平行，到 2000 年代初期，罕见的新发拷贝数变异开始与精神分裂症的风险有关。一小部分病例与具有较大风险的染色体异常有关，例如，第 22 号染色体 1 区 1 带的缺失。这些染色体异常与孤独症谱系障碍相关的位点完全或几乎完全重叠，但在这些位点上的缺失和重复的风险分布似乎并不相同。例如，虽然第 16 号染色体 1 区 1 带 2 亚带区域的重复和缺失都与孤独症谱系障碍和精神分裂症有关，但该区域的重复更可能导致精神分裂症，而缺失则更常见于孤独症谱系障碍和智力障碍。

关于精神分裂症，过去 15 年来最重要的进展是全基因组关联研究的出现。与之前描述的基于假设的候选基因研究不同，全基因组关联研究依赖于同时检测基因组中每个基因的多态性。这种无假设的方法，当与大型群体一起使用并适当校正多重比较时，在医学上已成为一种高度可靠和可重复的策略，用于识别常见疾病中的常见风险等位基因。

涉及近 4 万个病例和 113,000 个对照的全基因组关联研究已经确定了 108 个精神分裂症的风险位点。这组中任何个体遗传变异的影响都非常适度，通常导致风险增加不到 25%。此外，全基因组关联研究中检测的许多遗传多态性映射到基因组的非编码区域。因此，虽然已经确定了 108 个风险基因座，但尚不完全清楚哪些基因

对应于所有这些风险变异。在某些情况下，变异与单个基因的距离足够近，可以合理推断出这种关系；在其他情况下，这仍有待确定。

与精神分裂症风险相关的基因为确定该疾病背后的生物学提供了起点。例如，自 20 世纪 90 年代后期以来，有证据表明一个叫做主要组织相容性的区域与精神分裂症风险有关。因此，在精神分裂患者群体中，主要组织相容性区域具有人类基因组任何部分中最强的全基因组关联研究信号。群体中的大量患者使详细研究成为可能，将主要组织相容性区域中的这种强大的风险关联信号解析为 3 个不同的基因座（可能是 3 个不同的基因）。在这 3 个基因座中，一个编码补体 C4 因子的基因对疾病风险具有强烈且明确的影响。史蒂文·麦卡罗和他的同事表明，补体 C4 基因座代表拷贝数变异的自然病例，健康个体在他们拥有的基因拷贝数方面存在很大差异，并且 C4A 等位基因的表达水平与精神分裂症风险的增加相关。随后的后续研究表明，敲除 C4 基因的小鼠在发育过程中存在突触修剪缺陷，这表明人类 C4A 过量可能导致突触修剪过度的假设，这一过程长期以来一直受到精神分裂症文献的关注。

这一发现是将基因组学与增加疾病风险的可能生物学机制联系起来的重要证明。即便如此，具有最高风险 C4 单倍型且没有精神分裂症家族史的个体会由于该等位基因而平均受影响的几率从 1% 增加到大约 1.3%。为了有一个规模感，有一个一级亲属患有精神分裂症会使风险增加大约 10 倍。这一有希望的开端及其局限性反映了遗传学家和神经生物学家现在面临的挑战，即从成功的常见变异基因发现转向详细阐述导致人类病理学的特定机制。

除了确定许多特定的风险位点外，精神分裂症的全基因组关联研究还反复发现许多常见等位基因的小个体效应加起来会增加风险。这些结果为研究总体的基因型-表现型关系提供了另一个强大的途径。事实上，已经很清楚，个体携带的风险等位基因数量可以对发展该疾病的风险产生显著（且累加）的影响。例如，那些在所谓的多基因风险评分（一个与个体携带的加性遗传风险总量相关的汇总统计数据）中处于最高十分位数的人，与普通人群相比，风险增加了 8 到 20 倍。尽管累积效应的生物学尚不清楚，但这一观察结果为研究与疾病轨迹和治疗响应相关的一系列有趣问题奠定了基础，并且几乎肯定会重新激发结合神经影像学和基因组学的研究。这些后期类型的研究，类似于早期的常见变异发现努力，由于研究选定的、生物学上合理的候选基因的固有局限性，一直存在可靠性差的问题。

最后，类似于在孤独症谱系障碍中采用的高通量测序方法，也开始在精神分裂症中产生结果。具体来说，寻找罕见和新发风险等位基因的外显子组测序已经取得了一些成功。然而，与孤独症谱系障碍相比，此类研究需要更大的群体来确定可能的基因破坏突变的统计学显著风险，这表明这些类型变异的总体影响大小在精神分裂症中可能要小得多。迄今为止，这些研究已经鉴定出一些风险基因，并涉及关键的神经生物学通路。特别是，最近的外显子组研究指出了活性调节的细胞骨架复合体中分子的重要性，以及包含组蛋白 H3 赖氨酸 4 特异性甲基转移酶的基因组结构域，与精神分裂症发病机制相关。

2.8 神经精神疾病遗传基础的观点

基因影响行为的许多方面。即使是分开抚养的人类双胞胎，在性格特质和精神疾病方面也展现出显著的相似性。家畜和实验室动物可以被培育以展现特定的、稳定的行为特征；此外，对神经发育和精神疾病的广泛遗传变异的贡献也正在不断被发现。

一系列并行的进步带来了一个理解基因、大脑和行为之间关系的绝佳机会。可以用于操作和研究模型系统的技术已经发生了革命性的变化。同时，在确定人类神经精神疾病的遗传风险因素方面也取得了显著进展。尽管这个领域在这个过程中仍处于早期阶段，但已经出现了多个成功基因发现及其应用于深入生物学理解的有价值例子。

最近对神经发育和精神疾病遗传学研究的许多引人注目的发现之一是，不同诊断范畴之间的遗传风险存在重叠。生物学不遵循分类诊断标准可能并不令人惊讶，但考虑该领域如何追踪这些效应并提出新的治疗策略，无疑是一个巨大的概念挑战。

此外，值得注意的是，对于许多其他尚未看到先前提到进展类型的精神疾病，计算方法很简单：更大的投资和更大的样本量将带来更深入的洞察力。例如，最近对图雷特综合症和强迫症新生突变的研究清楚地表明，鉴

定高置信度风险基因的限速因子是亲子三人组测序的可用性。同样，重度抑郁症的全基因组关联研究直到最近才达到足以确认具有统计学意义相关常见变异的样本量。这些研究包括了数十万人，并且毫不奇怪地发现了具有非常小个体影响的风险等位基因。

最后一点突出了一个观点，即对于行为、发育和精神疾病的基因组学来说，没有一种方法能适用于所有情况。从模型系统的研究到罕见孟德尔疾病的阐明，再到解析对常见疾病有贡献的常见和罕见变异，当今可用的工具和机会是前所未有的。未来几年应该能够深入理解精神疾病和神经发育障碍的生物学，并可能带来能够帮助患者及其家庭的治疗方法。

2.9 要点

1. 脆性 X 综合症、雷特综合症和威廉综合症等罕见遗传综合症为复杂人类行为的分子机制提供了重要见解。此外，虽然仍有大量工作要做，但对这些综合症的研究已经挑战了相关认知和行为缺陷是不可改变的观念，并证明了广泛的模型系统在阐明保守生物学机制方面的效用。

2. 人类基因组测序、高通量基因组检测的发展，以及计算和方法学的进步，已经深刻改变了我们对人类行为和精神疾病遗传学的理解。包括精神分裂症和孤独症在内的几种典型疾病已经取得了显著的进展，导致鉴定出数十个明确的遗传风险基因区域和染色体区域。

3. 在过去的 10 年里，精神病遗传学和基因组学领域的成熟揭示了预先指定候选基因研究的局限性。这些类型的研究现在已经被全基因组扫描所取代，这些扫描既包括常见等位基因也包括罕见等位基因。结合严格的统计框架和共识统计阈值，这些研究正在产生高度可靠和可重复的结果。

4. 目前，累积的证据表明，包括常见和罕见、遗传和新生、生殖细胞和体细胞、序列和染色体结构变异在内的所有类型的遗传变异，都构成了复杂行为综合症的基础。然而，这些不同类型的遗传变化对于特定疾病的影响程度各不相同。

5. 人类行为遗传学的最新进展中一个引人注目的发现是：具有不同症状和自然病史的综合症之间存在遗传风险的重叠。理解相同的突变如何在不同个体中导致高度多样化的表现型结果，将是一个重大的未来挑战。

6. 常见精神疾病的研究结果表明，遗传异质性非常高。这一点，加上迄今为止已确定的风险基因的生物学多效性，以及人类大脑发育的动态性和复杂性，都预示着在从对风险基因的理解转向对行为的理解方面，我们面临着重要的挑战。同样，目前，阐明风险基因的生物学与揭示行为综合症的病理生理学之间存在重要的区别。

2.10 术语表

等位基因。人类携带两组染色体，分别来自父母。例如由于单核苷酸多态性，两组中的等价基因可能存在差异。等位基因是特定基因的两种（或多种）形式之一。

着丝粒。染色体含有一个紧密的区域，称为着丝粒，姐妹染色单体（复制后形成的每个染色体的 2 个精确副本）在此连接。

克隆。生成特定脱氧核糖核酸片段的足够副本，以便对其进行测序或以其他方式研究的过程。

互补脱氧核糖核酸。利用一种名为逆转录酶的酶，由信使核糖核酸分子合成的脱氧核糖核酸序列。互补脱氧核糖核酸可用于实验测定信使核糖核酸在内含子（非蛋白质编码部分）被剪接后的序列。

基因的保守性。存在于两个不同物种中的基因被称为保守基因，来自不同物种的两个基因被称为直系同源基因。保守性可以通过测量两个序列在 碱基（核糖核酸或脱氧核糖核酸）或氨基酸（蛋白质）水平上的相似性来检测。相似性越高，两个序列的保守性就越高。

拷贝数变异。特定基因区域的缺失或重复，导致个体某些基因的拷贝数多于或少于通常的 2 个。拷贝数变异在某些神经系统和精神疾病中有所发现。

规律成簇的间隔短回文重复。一种酶核糖核酸系统，其中酶切割与核糖核酸向导匹配的靶序列；核糖核酸向导可被改造以识别细胞内一个或多个目标基因序列，从而进行突变。

常染色质。基因组中富含基因的区域（另见异染色质）。

真核生物。细胞具有复杂的内部结构，包括细胞核。动物、植物和真菌都是真核生物。

基因组。生物体的完整脱氧核糖核酸序列。

基因型。个体携带的一组基因；通常指个体在基因组特定区域的特定等位基因对（基因的替代形式）。

单倍型同一染色体上紧密连锁（即可能一起遗传）的等位基因（基因的替代形式）或序列变异的特定组合。

异染色质。基因组中紧凑、基因贫乏的区域，富含简单序列重复。

内含子和外显子。基因转录为连续序列，但产生的信使核糖核酸分子中只有部分片段包含编码蛋白质产物的信息。这些片段称为外显子。外显子之间的区域称为内含子，在核糖核酸产物形成之前由核糖核酸剪接而成。

长臂和短臂。着丝粒两侧的区域称为臂。由于着丝粒不在染色体的中心，因此一条臂比另一条臂长。

信使核糖核酸。蛋白质并非直接从基因组脱氧核糖核酸合成。相反，核糖核酸模板（前体信使核糖核酸）由基因序列构建而成。该核糖核酸随后以各种方式进行加工，包括剪接。剪接后的核糖核酸最终成为蛋白质合成的模板，被称为信使核糖核酸。

突变。与某些参考状态相比，基因组发生的改变。突变并不总是有害的。

表现型。生物体的可观察属性和物理特征。

多态性。基因组中存在一个区域，该区域在群体个体间存在差异。要被称为多态性，一个变异必须存在于群体中相当多的个体中。

原核生物。单细胞生物，内部结构简单，无细胞核。细菌和古细菌都是原核生物。

蛋白质组。由基因组编码的完整蛋白质组。

基因重组。卵子和精子形成过程中，脱氧核糖核酸在等价染色体对之间交换的过程。基因重组会导致后代的染色体与父母的染色体不同。

限制性核酸内切酶。一种在特定短序列处切割脱氧核糖核酸的酶。不同类型的限制性内切酶切割的序列不同。

核糖核酸干扰。一种通过将与目标信使核糖核酸互补的小核糖核酸引入细胞来降低特定基因功能的方法。目标信使核糖核酸与小核糖核酸配对会导致内源性信使核糖核酸的破坏。

单核苷酸多态性。由单个核苷酸改变引起的多态性。单核苷酸多态性常用于基因图谱研究。

剪接。从转录的核糖核酸中去除内含子（非编码部分）的过程。外显子（蛋白质编码部分）也可以被去除。根据去除的外显子，相同的初始核糖核酸或基因可以产生不同的蛋白质。以这种方式产生的不同蛋白质是剪接变体或选择性剪接。

转录。将基因复制成信使核糖核酸的过程。这是将基因转化为蛋白质的第一步，尽管并非所有转录本都能产生蛋白质。

转录组。从基因组转录出来的一整套核糖核酸。

翻译。利用信使核糖核酸序列合成蛋白质的过程。信使核糖核酸作为模板，携带氨基酸的转移核糖核酸分子在其上排列。氨基酸连接在一起，形成蛋白质链。