

14 分子生物学筛选

分子生物学的发展，不断推出新的技术，有助于筛选参与重要生物学功能的基因，从而进一步揭示分子机理。

可以通过提高基因表达或增加基因产物的功能筛选有特定功能的基因，既 gain of function (GOF) 的筛选。例如根据基因表达时间或空间来发现，或细节不同的“表达克隆”方法。也可以通过减少基因表达或降低基因产物的功能，筛选参与某种功能的基因，也既 loss of function (LOF) 的筛选。GOF 筛选有助于发现对功能充分的基因，LOF 发现对功能必要的基因。

本章以一些重要的例子来说明这些方法及其应用。

14.1 按表达时间找基因

根据基因表达的时间，加上功能检测，是寻找重要的基因的一种方法。

C3H 10T1/2 Cl8 是来自小鼠胚胎，为研究癌症而衍生的细胞系。1977 年，几位科学家用它筛选多种癌症化疗药物的作用，偶尔发现，胞嘧啶的衍生化合物 5-氮杂胞苷(5-azacytidine, aza) 处理后，出现长长的多核细胞(Benedict *et al.*, 1977)。已知 aza 有抑菌、抗癌、诱变等作用，当时已用于治疗白血病。南非科学家 Constantidines、Jones 和 Gevers 等证明，aza 处理 10T1/2 细胞后出现的是肌肉细胞(Constantidines, Jones and Gevers, 1977)。10T1/2 细胞和 3T3 细胞都可以被 aza 诱导产生，有时不仅肌肉细胞，而且还有脂肪细胞等(Taylor and Jones, 1979)。因为 aza 抑制甲基转移酶活性，推测是它抑制 DNA 甲基化而影响基因表达让细胞可以重新改变命运。

1986 年，美国西雅图 Fred Hutchinson 癌症研究中心的 Harold Weintraub (1945-1995) 实验室用这一模型研究决定肌肉细胞的基因。首先，Weintraub 的博士后 Andrew Lassar 等证明，aza 处理的 10T1/2 细胞不仅有肌肉细胞的形态特征，而且肌肉细胞应该表达的分子如肌球蛋白、乙酰胆碱受体、肌肉特异的肌动蛋白等也都表达了，所以是货真价实的肌肉细胞。之后，Lassar 等从 aza 处理 10T1/2 细胞后得到的肌肉细胞提取基因组 DNA，而这些 DNA 转染没有经过 aza 处理的 10T1/2 细胞，可以得到肌肉细胞。而未经 aza 处理的 10T1/2 细胞来的基因组 DNA 缺乏这一作用。从另外一种已知的肌肉细胞系 C2C12 来源的基因组 DNA，也有这一作用。他们从转染得到肌肉细胞的频率推测单个基因起了改变细胞命运的作用(Lassar, Paterson and Weintraub, 1986)。

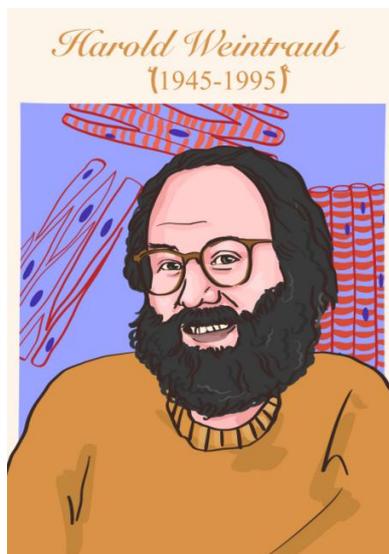


图 14-1 Weintraub

Weintraub 的研究生 Robert Davis 加入研究。1987 年，Davis、Lassar 和 Weintraub 报道，他们找到了将 10T1/2 成纤维细胞变成肌肉细胞的单一基因(Davis, Weintraub and Lassar, 1987)。他们寻找的过程，关键是

比较 aza 处理前后的基因差别。他们用来自 aza 处理后的 10T1/2 细胞的 mRNA，制备 ^{32}P 标记的 cDNA 探针(我们在此称为 A)，也用肌肉细胞系 C2C12 的 mRNA 制备 cDNA(我们在此称为 B)。Davis 等也用没有处理的 10T1/2 细胞的 mRNA 制备 cDNA(我们在此称为 C)。单链的 DNA 与双链的 DNA 早就有技术在羟基磷灰石柱(hydroxyapatite, HAP)上分离(Bernardi, 1965)。Davis 等杂交 A 和 C 后，过 HAP 柱，选择单链的 cDNA 留下，就是只在 aza 处理后 10T1/2 细胞变成的肌肉细胞特异的(我们在此称为 D)。多次过柱，以提高选择性。得到 10T1/2 细胞变成的肌肉特异的探针 D，再与 B 进行杂交，杂交后，过 HAP 柱，选择双链的 cDNA，也多次过柱选择双链，得到 A 和 B 共同的 cDNA。这样，最后拿到的就是 A 与 B 共同、而与 C 不同的 cDNA。因为其来自 poly(A)⁺ RNA (mRNA)，所以是不同的 mRNA。Davis 等同样也用 B 制备 ^{32}P 标记的 cDNA 探针，找其与 A 相同、与 C 不同的部分。

他们因此从 10^4 个 cDNA 中找到 92 个肌肉有、成纤维细胞没有的克隆，基因序列为 26 种。他们进一步加几个标准：在 10T1/2 细胞完全不表达，从肌肉母细胞分化到肌管过程表达不变化以排除肌肉终分化过程的基因、聚焦早期决定肌肉命运的基因等。符合所有这些标准的只有三个基因，他们分别称为 MyoA、MyoD 和 MyoH。这三个基因分别单独转染 10T1/2 细胞，只有 MyoD 可以让它们变成肌肉细胞。而且 MyoD 可以使多种细胞变成肌肉细胞(Davis, Weintraub and Lassar, 1987)。在鼠体内，MyoD 只在有肌肉细胞的组织和器官表达。机理上，MyoD 蛋白质是控制 mRNA 生成的转录因子(Lassar et al., 1989)。以后的研究证明 MyoD 类似的多个基因控制肌肉形成。这一研究成为细胞命运分子机理的典范。

14.2 按表达空间找基因

根据基因表达的空间，也可以寻找基因。

如第五章所述，脊椎动物如两栖类的胚胎早期，与卵母细胞一样，有动物极和植物极。两者不仅外观着色可以不同，而且如果受精后会成为不同的胚胎部分，植物极的细胞还向动物极的细胞发出分子信号，例如在动物极与植物极之间的边缘区产生中胚层及其衍生的组织细胞(Nieuwkoop, 1969; Nieuwkoop and Ubbels, 1972)。

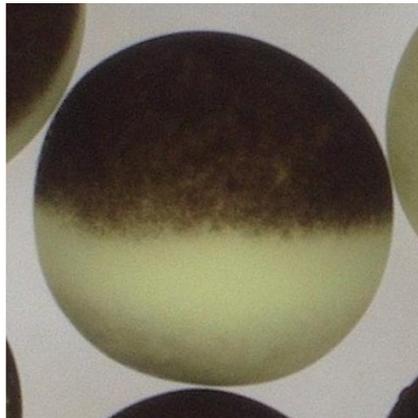


图 14-2 爪蟾卵母细胞的动物极和植物极

哈佛大学 Douglas Melton 实验室决定专门寻找卵细胞植物极和动物极特异的 mRNA。他们用卵细胞的 mRNA，反转录为 cDNA，制成含量 2×10^6 的 cDNA 文库(Rebagliati et al., 1985)。他们切下动物极和植物极，分别制备 mRNA，然后做 ^{32}P 标记的探针。用这些探针分别与卵细胞 cDNA 文库的模板杂交，寻找两极特异的 mRNA。他们找到 3 个动物极特异的 mRNA (分别命名为 An1、An2、An3) 和一个植物极特异的 mRNA (Vg1)。他们进一步分析 Vg1，发现它编码一个分泌性的蛋白质，属于转化分化因子(TGF- β)家族(Weeks and Melton, 1987)。

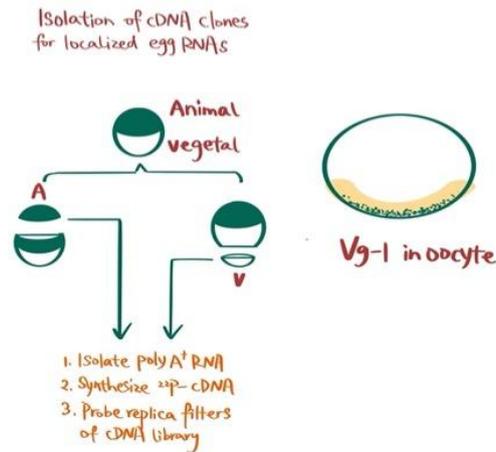


图 14-3 差异克隆获得 Vg-1，其 mRNA 分布

如第五章所述，胚胎发育的一个重要过程是形成腹背轴。两栖类的背轴含很多结构。虽然背轴在比较肠胚期才比较容易从外观看见，但其决定却在很早，例如爪蟾胚胎可能在受精过程因为精子进入位点就确定了其对侧为背侧(Gerhart *et al.*, 1989)。寻找可以决定背轴的基因的一个方法是到早期胚胎的背侧去找有背唇化功能的基因。已知用紫外线(UV)处理 1 细胞期的爪蟾胚胎可以导致胚胎的腹侧化(Gimlich and Gerhart, 1984)，而用氯化锂(LiCl)处理 1 细胞期的爪蟾胚胎可以导致胚胎的背侧化(Kao, Masui and Elinson, 1986)。美国伯克利加州大学的 Richard Harland 与博士后 William Smith 从不同胚胎提取 mRNA，注射到 UV 处理过的胚胎，检测它们能否让腹侧化的胚胎背侧化(形成背轴)(Smith and Harland, 1991)。他们比较了几种来源的 mRNA，发现作用最强的是 LiCl 处理过的胚胎来源的 mRNA(成功率 46%)，其次是正常胚胎来源的 mRNA(成功率 12%)，而腹侧化胚胎来源的 mRNA 具有的背侧化作用(成功率 8%)类似于没注射 mRNA 的对照(成功率 5%)。他们按分子量将作用最显著的 mRNA 分成几个组分，其中从组分 8 获得一个 mRNA 有背侧化作用，编码 Xwnt8，但当时已经有人发现同类基因(Smith and Harland, 1991)。他们从组分 9 获得一个全新的基因，编码一个分泌性蛋白质，他们命名为 Noggin (Smith and Harland, 1992)。Noggin 最早表达在肠胚刚刚开始形成时期，而且在背唇。以后也在神经管中。这是第一个发现特异表达在背唇的基因。功能上，它不仅能够背侧化早期胚胎，而且可以有神经诱导作用。

美国洛杉矶加州大学的 Eddy de Robertis 实验室专门用背唇制备 cDNA 文库，然后注射 mRNA，来寻找可以背侧化 UV 处理后胚胎的基因。他们发现 Chordin，它也先表达在背唇，以后在神经管，并且也有神经诱导作用(Sasai *et al.*, 1994)。

14.3 按功能特点寻找基因：蛙卵表达克隆

表达克隆，是典型的按照基因编码的蛋白质可能具有的功能，而克隆基因。

用于表达基因的细胞可以根据不同需要而定。用非洲爪蟾卵母细胞(以下简称卵母细胞)，是其中一种(Mowry, 2020)。

将其他生物来源的基因，体外转录为 mRNA 后，注射到爪蟾卵母细胞，可以观察到 mRNA 被翻译为蛋白质(Gurdon *et al.*, 1971)。但要能够用于表达克隆，例如某些神经递质或者激素的受体，不仅 mRNA 能够在卵母细胞表达、而且产物蛋白质能够上细胞膜，如果本身不能被检测到，就还需要其下游偶联的分子也存在，而能够在细胞外给予其激动剂时检测到反应。烟碱型乙酰胆碱受体 (nAChR) 本身是离子通道，可以上膜，形成可以被检测到的离子通道，用电鱼电器官获得的 mRNA 注射到卵母细胞，检测到了 nAChR 的反应 (Sumikawa *et al.*, 1981)。对哺乳类学习记忆至关重要的 NMDA 受体，就是用卵母细胞的表达克隆所找到的 (Moriyoshi *et al.*, 1991)。

五羟色胺有多种受体，其中 5-HT_{1c} 受体也是离子通道，是通过卵母细胞的表达克隆而发现的 (Lübbert *et al.*, 1987; Julius *et al.*, 1988)。从富含 5-HT_{1c} 的肿瘤制备 cDNA 文库，注射 mRNA 有 5-HT 引起的氯离子通道形成，然后继续筛选直到一个 cDNA。

神经肽 K 物质的受体不是离子通道，而是 GPCR，但其下游偶联的信号转导通路有可以被检测的反应，也可以用卵母细胞克隆 K 物质的受体 (Masu *et al.*, 1987)。

电流检测之外，还有其他方法检测功能。例如，转运蛋白可以转运特异的底物，如果有可以同位素标记的底物 (Hediger *et al.*, 1987)，或者现代化学分析仪器 (质谱等) 可以检测的底物，那么也可以用来表达克隆编码转运蛋白的基因。

14.4 按功能特点寻找基因：钙信号检测

科学研究可以是聚焦一个科学问题、自然现象；也可以是科学家偏好一种技术，利用技术寻找可以由技术帮助解决的问题。

钱永健 (Roger Tsien, 1952-2016) 以其化学和生物学背景，不断发明技术，用于生物学研究，带动技术发明逐渐成为生命科学的研究生涯。他发明钙离子检测的方法有特别的贡献。他先通过改造钙离子的螯合剂，使之在结合钙离子后发出荧光，从而用荧光强度变化检测活细胞内的钙离子浓度，有 fura-2 等一系列检测钙离子浓度的钙染料 (Tsien, 1980, 1981)。

在下村脩 (Osamu Shimomura, 1928-2018) (Shimomura, Johnson and Saiga, 1962; Morise, Shimomura, Johnson and Winant, 1974) 和 Douglas Prasher (1951-) (Prasher *et al.*, 1992) 工作的基础上，钱永健参与证明单个蛋白质 (绿色荧光蛋白，GFP) 可以因为结合钙离子后发出荧光 (Heim, Prasher and Tsien, 1994; Chalfie *et al.*, 1994)。

已知辣椒里面的分子辣椒素 (capsaicin) 可以增强痛敏感性、也可以造成热感。David Julius (1955-) 于 1980 年代在 Richard Axel 实验室做博士后的阶段用表达克隆寻找五羟色胺的受体 (Julius *et al.*, 1988)。到旧金山加州大学任教多年后，他继续用表达克隆，其中一个课题为寻找辣椒素的受体，认为找到辣椒素的受体可能有助于理解痛觉的机理。这一研究最后对感觉系统的分子机理有很大影响，且不限于外周感觉系统，虽然对理解痛觉影响相对较小。已知脊髓背根神经节 (dorsal root ganglia, DRG) 不仅传入多种外周感觉，而且对辣椒素有反应。英国科学家用同位素标记的离子证明辣椒素 (capsaicin) 可以诱导大鼠背根神经节 (dorsal root ganglia, DRG) 的神经元产生钙离子内流和钠离子内流、不诱导氯离子外流 (Wood *et al.*, 1988)。依据这一特征，Julius 带领实验室，用 DRG 的 mRNA 制备 cDNA 文库，然后转染到普通培养细胞 (常用的 HEK293 细胞)。起步估计用了含一万六千多种 cDNA 的库，不断分群并检测，最后拿到一个 cDNA，单独转染到 HEK 细胞就可以对辣椒素出现钙离子内流的反应 (Caterina *et al.*, 1997)。编码的蛋白质当时命名为 VR1，它不仅介导对辣椒素的反应，而且介导对热的反应，被加温所激活。

有趣的是，VR1 并非独一无二的蛋白质，而是属于一个已经知道蛋白质氨基酸序列的家族：Trp 通道。但是，在 Julius 发现 VR1 介导对辣椒素和热反应时，Trp 通道家族的功能其实不很清楚。1969 年，美国印第安纳州的 Purdue 大学生物系韩裔教授 Willam Pak (1932-2023) 用果蝇的视网膜电图 (electroretinogram, ERG) 研究果蝇的视觉、并寻找影响果蝇视觉的突变种 (Pak *et al.*, 1969)。同年，英国爱丁堡大学的科学家发现一种突变，根据 ERG 的变化，他们称突变果蝇为“瞬间感受器电位” (Transient Receptor Potential, Trp)，相应突变的基因也就是果蝇的 Trp 基因 (Cosens and Manning, 1969)。1989 年，美国加州大学伯克利分校 Gerald Rubin (1950-) 实验室的博士后 Craig Montell (1955-) 克隆了果蝇的 Trp 基因 (Montell and Rubin, 1989)。他们发现 Trp 基因编码的蛋白质有多重跨膜区，各由大约 19 个氨基酸组成的疏水段。因为它的突变影响果蝇视网膜对光反应，所以最容易推测的功能是离子通道。但证明它是离子通道花了相当长的时间。Trp 类似的基因后来发现于多种生物，包括人。但不仅果蝇的 Trp 本身并不感受光，其它的 Trp 具体起什么作用，如何起作用，也都不清楚。

Julius 实验室 1997 年的发现揭示了其直接功能，由此 Trp 的研究热火朝天，很多人加入 Trp 研究的行列，发现更多 Trp 家族成员，发现更多功能，如 Julius 实验室 2002 年用薄荷寻找冷敏感分子 (McKemy *et al.*, 2002)，结果发现是 Trp 家族另一成员，有多个 Trp 通道，温度变化的方向和范围不同的激活不同的 Trp 通道。还有科学家发现 Trp 通道感受压力，等等。



图 14-4 Tsien

视觉、嗅觉、味觉的第一级感受分子都是 GPCR。视觉的第一级是通过生物化学分离纯化最早在 19 世纪得到视紫红质。嗅觉的第一级也被发现是 G 蛋白偶联的通路，在发现对嗅觉分子敏感的腺苷酸环化酶 (Pace *et al.*, 1985)、嗅觉神经元特异的 G 蛋白 (Jones and Reed, 1989) 和环核苷酸激活的离子通道 (Dhallan *et al.*, 1990) 的基础上，美国哥伦比亚大学的 Richard Axel (1946-) 和他的博士后 Linda Buck (1947-) 通过多聚酶链反应 (PCR) 发现了小鼠嗅上皮的数以百计的 GPCRs (Buck and Axel, 1991)，以后证明它们是感受嗅觉的分子。味觉再模仿嗅觉也得到类似发现，但味觉的感受分子只有几个，分别对应不同的味觉 (甜、咸、苦、酸、鲜) (Hoon *et al.*, 1999; Chandrashekar *et al.*, 2009)。

确定 Trp 通道的功能，解决了其他常见感觉 (温、部分的压、部分的痛) 第一级的分子机理，有较大意义。

最初以为听觉也是用一种 Trp 通道，后来发现听觉的毛细胞还有其他结构完全不同的压力感受分子，如 TMC (Kurima *et al.*, 2002; Vreugde *et al.*, 2002; Jia *et al.*, 2020)。而外周和内脏的压力感受器也有多种，如美国加州的 Scripps 研究所的 Patapoutian 实验室发现的 Piezo (Coste *et al.*, 2010, 2012) 和 GPCR (Xu *et al.*, 2018)。貌似简单的压力，却在第一级的感受用了用了很多结构不同的分子。

14.5 按功能特点寻找基因：重组激活基因的发现

第七章我们讲到，体细胞 DNA 重组是免疫球蛋白多样性的最重要原因。

那么，发现重组酶就应该是理解其机理的关键。

重组酶发现于当时在 MIT 的分子生物学家 David Baltimore (1938-) 的实验室。Baltimore 两年完成自己的博士论文，32 岁发现反转录酶，37 岁与 Howard Temin 共享诺奖，1986 年发现 NF- κ B，1989 年发现重组激活基因 (RAG)。RAG 基因的发现是一系列实验的结果。

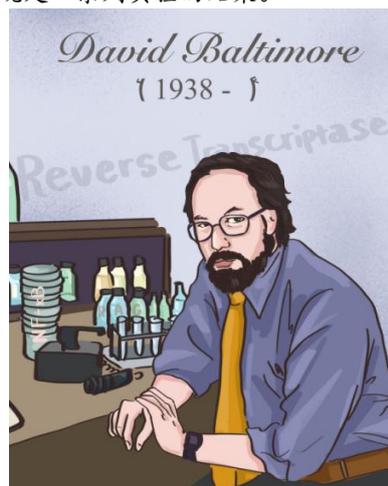


图 14-5 Baltimore

1984 年，Baltimore 的研究生 Susanna Lewis 设计了实验，用免疫球蛋白 κ 轻链基因的部分序列重构 V-J 重组。他们选取了鼠胚 V κ 和 J κ 及其附近的 DNA 序列，装在病毒里面。将这种病毒感染淋巴细胞样的体外培养细胞系 PD。他们在序列 V κ 和 J κ 之间，装了来自细菌的可以对黄嘌呤-鸟嘌呤磷酸核糖转移酶 (xanthine-guanine phosphoribosyl transferase, *gpt*)。因为预计重组导致其间 DNA 序列翻转，所以放到病毒的是反向的 *gpt* 序列。一旦发生 V κ 和 J κ 之间重组，其间的序列翻转，*gpt* 成为正向而可以表达 GPT 蛋白质，细胞能够抵抗霉酚酸。Lewis 等成功地观察到，她们的分子构造，转染入 PD 细胞系后，经过霉酚酸后得到的细胞株多数都是因出现了 V κ 和 J κ 之间 DNA 重组、*gpt* 序列翻转 (Lewis *et al.*, 1984)。

Baltimore 实验室的另一位研究生 David Schatz 用类似 Lewis 的方法，再构建了一个含 B 潮霉素抗性基因 (*hph*) 的病毒。用两种病毒，他们检验了多种 B 细胞前体细胞样的细胞系，多数能够重组 V κ 和 J κ 之间 DNA、翻转 *gpt* 或 *hph* 抗性基因序列，导致细胞对霉酚酸或 B 潮霉素有抵抗力。但如果用同样的病毒感染 3T3 成纤维细胞系，就不能发生 V κ 和 J κ 之间 DNA 重组 (Schatz and Baltimore, 1988)。Schatz 在实验室周

会上讲自己工作进展时，以色列来的博士后 Yoav Citri (1953-1995) 建议想办法激活 3T3 细胞中的 V (D) J 重组活性，这一建议影响了后面的研究 (Schatz and Baltimore, 2004)。

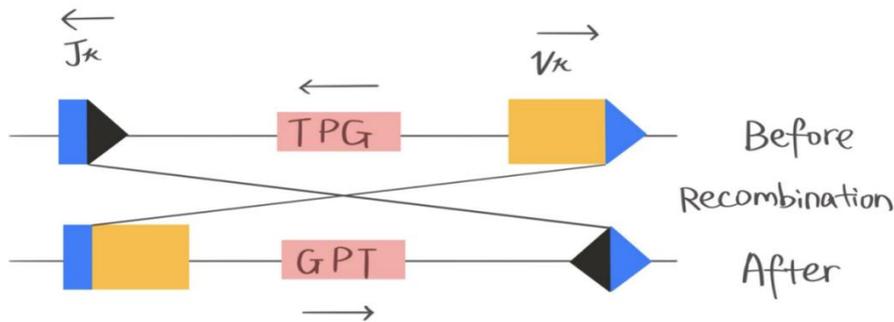


图 14-6 克隆 RAG 的实验设计

要在 3T3 细胞发生 V_{κ} 和 J_{κ} 之间 DNA 重组，当时一般的设计应该是把可重组 V_{κ} - J_{κ} 的细胞（前 B 细胞）的 mRNA 做成 cDNA 文库。文库的大量不同 cDNA 转染到 3T3 细胞，后者含有可以检测 V_{κ} 和 J_{κ} 之间 DNA 重组的 *gpt* 或 *hph*。在没有转染来自前 B 细胞 cDNA 的情况下，3T3 细胞会被霉酚酸或 B 潮霉素杀死。而如果有前 B 细胞的 cDNA 导致 V_{κ} 和 J_{κ} 之间 DNA 重组，使得其间的 *gpt* 或 *hph* 基因成为可以表达的方向，从而有 GPT 或 HPH 合成，这样的 3T3 细胞就能抵抗霉酚酸或 B 潮霉素。多数 3T3 细胞都被转染了与重组酶无关的 cDNA，这些 3T3 细胞会被霉酚酸或 B 潮霉素杀死，剩下可抵抗霉酚酸或 B 潮霉素的 3T3 细胞就很可能是表达了重组酶的 cDNA 从而导致 V_{κ} 和 J_{κ} 之间 DNA 的重组。

因为一种细胞表达很多 mRNA 从而可以得到很多不同的 cDNA，如果是一种 cDNA 编码重组酶，那么这种方法一般是可行的。但是，如果是多种 cDNA 编码的蛋白质才能引起 V_{κ} 和 J_{κ} 之间 DNA 重组，那么同时含有两种或多种起作用的 cDNA 的几率就会很小。Baltimore 当时在 Whitehead 研究所的同事 Robert Weinberg (1942-) 用分子生物学研究癌症，Weinberg 实验室用基因组 DNA 制备文库。将这种文库转染到 3T3 细胞时，每个 3T3 被转染的不会只是单个 cDNA，而是一段基因组 DNA，可能含相邻几个基因。

Schatz 把 B 淋巴瘤的基因组 DNA 转染到 3T3，基因组 DNA 接了一种抗性基因（组氨醇脱氢酶），首先用这种抗性基因针对的药物（组氨醇）选择被 B 淋巴瘤的基因组 DNA 转染了的 3T3 细胞。然后用霉酚酸选择发生了 V-J 重组的 3T3 细胞。他们从 1500 个抵抗组氨醇的 3T3 细胞克隆得到 1 个抗霉酚酸的 3T3 细胞株 (Schatz and Baltimore, 1988)。他们分析发现这一 3T3 细胞株确实出现了 V_{κ} 和 J_{κ} 之间 DNA 重组，使得其间的 *gpt* 表达，而不是其他原因对霉酚酸出现抵抗。但很多人不信，他们附近的实验室认为他们的结果不可靠、发表的是伪迹 (Schatz and Baltimore, 2004)。

Schatz 与研究生 Marjorie Oettinger 组成两人合作，每天倒班、想方设法克隆重组激活基因 (RAG)。他们用特定的寡核苷酸标记基因组的 DNA，从出现重组的细胞中通过用于标记的寡核苷酸取得基因组 DNA 片段，以此克隆了 RAG 的催化亚基 RAG-1 (Schatz, Oettinger and Baltimore, 1989)。RAG-1 基因编码含一千余氨基酸残基的蛋白质，其同源基因存在于有 V (D) J 重组的动物如人、鼠、兔、羊、马、鸡、蛙。在人和鼠能够进行 V (D) J 重组的细胞系，RAG-1 也有表达。RAG-1 表达的细胞和组织与 V (D) J 重组活性相关：RAG-1 表达于鼠的胸腺、骨髓和胚胎，而不在脾脏和脑表达。

在一派大好的形势下，有个异数：含 RAG-1 基因的 18kb DNA，催化 V (D) J 重组的活性居然与最初全基因组切的片段效率一样，比预计的低 100 到 1000 倍。当时他们的估计是也许因为所用的基因组 DNA 不够长，缺乏某些未知的转录控制元件和 34 个氨基酸 (Schatz, Oettinger and Baltimore, 1989)。但后来发现，得到全长的 cDNA 后，活性也不高 (Oettinger *et al.*, 1990)。他们经过几个月的迷惑不解，不断提各种可能。其中一个比较难以置信的可能性是 RAG-1 旁边碰巧还有一个基因，与它共同起作用 (Schatz and Baltimore, 2004)。在细菌中，已知相关功能的基因聚集在附近，而动物细胞的基因早已知道没有这样的规律，类似功能的基因在附近一般少见。

她们发现在原初有重组活性的 18kb 基因组 DNA 中，RAG-1 基因只占其中 6.6kb，另有 11kb，余下的基因组 DNA 中还有一个基因。她们进一步的工作证明这个基因可以提高 RAG-1 重组效率 1000 倍，因此她们称这一基因为 RAG-2 (Oettinger *et al.*, 1990)。以后知道，RAG-1 独立进化自转座酶 *Transib*，后来又捕获了 RAG-2 基因进一步增强其活性，之后 RAG-1 和 RAG-2 的序列和功能变化成为专门用于 Ig 基因重组的酶 (Carmona and Schatz, 2017)。

14.6 基因敲除或敲低

降低或敲除特定基因并观察后果是研究基因功能的重要方法。如果这种方法能够高效用于多个基因，可成为能够广泛应用于筛选。经过很多科学家的努力，现已发展出多种优秀的技术，使得针对单个基因的方法演变进行大规模筛选的技术。

分子生物学技术诞生的初期，MGH 的 Paul Zamecnik 实验室就探索了用“反义寡核苷酸”来抑制基因表达 (Zamecnik and Stephenson, 1978)。通过设计与靶基因转录的 mRNA 互补的反义寡核苷酸，并使它进入细胞后与 mRNA 结合，从而抑制基因的功能。最初，人们认为反义寡核苷酸直接抑制了 mRNA 的功能，后来发现实际上是细胞内有酶在降解这种双链 RNA。所用的寡核苷酸，初以 RNA (“反义 RNA”) 为主，后来有其他修饰的寡核苷酸。曾经有一段时间效率很低，多年工艺改造后，效率得到提高。

在细菌里面，按照 DNA 序列进行同源重组，很容易敲除目的基因。但在其他高等真核生物细胞中，同源重组的效率极低。美国犹他大学的 Mario Capecchi (1937-) 首先证明哺乳动物的细胞也有同源重组，尽管效率低下。然后他带领实验室发明可以筛选同源重组的正反选择方法，使得可以用小鼠胚胎干细胞 (ES) 进行同源重组。利用这种方法选出发生了重组的 ES 细胞，然后将这些细胞植入小鼠子宫，从而产生缺失目标基因的小鼠，这成为了第一个非常有效的哺乳动物基因敲除方法 (Thomas and Capecchi, 1987; Mansour, Thomas and Capecchi, 1988)。这一方法用于敲除单个基因很有效，但难以用于同时筛选很多基因。

RNA 干扰 (RNAi) 现象，是 Andrew Fire (1959-) 偶尔发现的。他的实验室用反义 RNA 研究线虫特定基因功能的时候，用正链 RNA 作为对照的时候，有时居然也有与反义 RNA 一样的“抑制作用” (Fire *et al.*, 1998)。他们注意到 1995 年另外一个实验室也观察到这样的结果，并如实报道了，但没有作为科学问题进一步研究 (Guo and Kemphues, 1995)。Fire 等的研究发现这不是针对少数几个 RNA 的现象，在线虫是普遍现象。以后吸引很多人研究 RNAi，搞清楚其特异的分子机理，而且知道在其他物种怎么可以高效地起作用。在线虫，甚至可以通过食物喂寡核苷酸，可以被摄取，而且到细胞里面起作用，所以可以用于筛选。衍生的技术也可以用于哺乳动物细胞的基因筛选。

“锌指核酸酶” (zinc finger nucleases, ZFN) 是美国霍普金斯大学印度裔科学家 Srinivasan Chandrasegaran 成功发明的技术。将核酸内切酶结构域 FokI 与特异 DNA 序列结合的锌指区融合，让锌指把核酸内切酶结构域带到特定 DNA，可以精确切割特定 DNA 序列，实现基因的敲除 (Kim, Cha and Chandrasegaran, 1996)。这一方法始于体外生化反应，进一步发展到可以用于动植物细胞 (Bibikova *et al.*, 2002)。

特殊生物可以含有特别的分子。地热环境中找到的生物含高温稳定的 DNA 多聚酶，可以被人们用于 PCR。同样引人注目的是黄单胞菌 (*Xanthomonas*) 中的转录激活因子，即所谓的“TALE” (transcription activator-like effectors)。这种蛋白识别 DNA 序列的方式非常有趣：它们主要依赖于 TALE 蛋白的 DNA 结合域，该域由一系列重复的氨基酸序列组成，每个重复大约由 34 个氨基酸构成 (Moscou and Bogdanove, 2009; Boch *et al.*, 2009)。每个 34 氨基酸重复单元中的第 12 和第 13 位氨基酸是决定 TALE 与特定 DNA 碱基配对特异性的关键，被称为“可变二核苷酸”，它们的特定组合可以识别 DNA 序列中的特定单个碱基。将可编程的 TALE 蛋白与 FokI 核酸酶的非特异性切割域融合，就形成了可以在特定 DNA 序列处引入双链断裂 (DSB) 的 TALENs (Transcription Activator-Like Effector Nucleases)，达到基因剔除的效果。然而这一基因编辑方法并不适用于基因功能的高通量筛选。

14.7 分子生物学筛选：CRISPR 系统

CRISPR-Cas9 系统，提供了迄今最有效的基因编辑方法。

日本和西班牙科学家最初在原核生物中发现 CRISPR: clustered regularly interspaced short palindromic repeats, 后来才意识到这些重复序列之间的间隔序列是细菌和古细菌对曾经入侵噬菌体的基因组整合和记忆，可以与 DNA 进行特异的碱基配对的结合。而与 CRISPR 相关 (Cas) 的 Cas9 是核酸内切酶，CRISPR/Cas9 组成的体系原本在细菌中起免疫作用。法国科学家 Emmanuelle Charpentier (1968-) 和美国科学家 Jennifer

Doudna (1964-) 合作发明作为通用于基因编辑的 CRISPR-Cas9 体系 (Jinek *et al.*, 2012), Cas9 在 sgRNA (single guide RNA) 引导下, 切割被 sgRNA 结合的 DNA 序列。CRISPR-Cas9 技术不仅在众多生物体中显示出高效的编辑能力, 而且众多科学家在不断提高其编辑性能并扩大其应用领域。

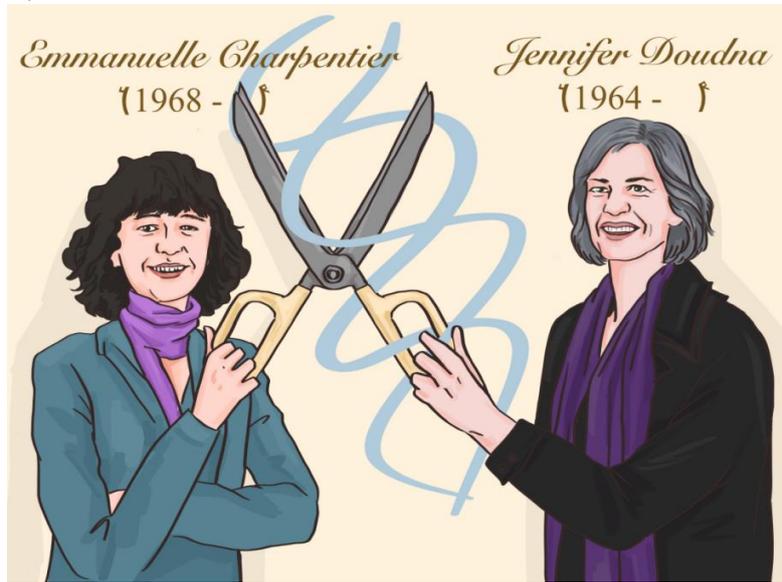


图 14-7 Charpentier 和 Doudna

对本章重要的是, 它可以衍生出筛选的方法。2014 年以下四个实验室同时发明的基于 CRISPR 系统的高通量基因功能筛选技术: MIT 的 Eric Lander (Wang *et al.*, 2014) 和张锋两个实验室 (Shalem *et al.*, 2014)、英国 Sanger 研究所的遊佐宏介实验室 (Koike-Yusa *et al.*, 2014) 和北京大学的魏文胜实验室 (Zhou *et al.*, 2014)。简而言之, 在用于筛选的细胞中通过慢病毒载体递送和表达 sgRNA 文库, 利用 CRISPR/Cas9 系统在多种癌细胞和多能干细胞中实现蛋白编码基因的功能性筛选。

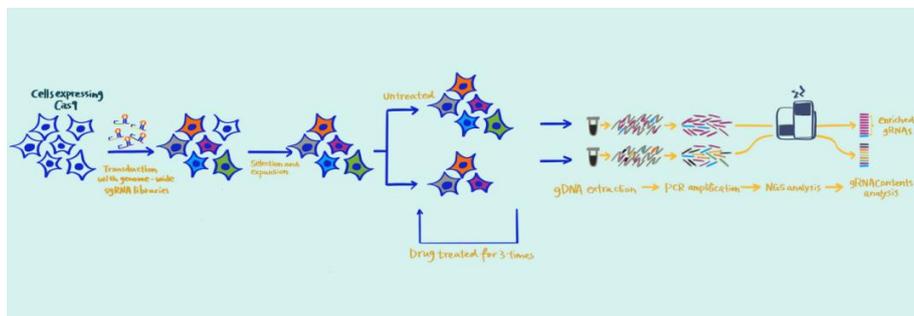


图 14-8 CRISPRi

基于 CRISPR 系统的高通量筛选可以有多种衍生技术, 除了基因敲除, 也可以进行基因表达的抑制筛选 (CRISPRi), 或者是基因表达的上调/激活筛选 (CRISPRa) (Gilbert *et al.*, 2013; Mali *et al.*, 2013; Perez-Pinera *et al.*, 2013; Maeder *et al.*, 2013)。除了对蛋白编码的基因进行功能筛选以外, 针对非编码 RNA 的高通量技术也被建立出来, 魏文胜实验室建立了基于大片段删除和 RNA 剪切破坏两种技术路线, 张锋和 UCSF 的 Jonathan Weissman 团队也分别报道了运用 CRISPRa 以及 CRISPRi 系统实现全基因组水平长链非编码 RNA (lncRNA) 的高通量功能性筛选。利用单碱基编辑器针对单氨基酸位点进行功能性研究的高通量技术平台也被陆续报道出来。

在培养细胞系之外, 能否在组织、器官甚至动物体内做筛选, 也是技术发明者们努力的目标。

参考文献

- Bernardi G, Chromatography of nucleic acids on hydroxyapatite. *Nature* 206:779-183。
- Benedict WF, Banerjee A, Gardner A, Jones PA (1977). Induction of morphological transformation in mouse C3H/10T $\frac{1}{2}$ clone 8 cells and chromosomal damage in hamster A(T₁)C1-3 cells by cancer chemotherapeutic agents. *Cancer Research* 37:2202-2208.
- Bibikova M, Golic M, Golic KG and Carroll D (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161:1169-1175.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A and Bonas J (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509-1512.
- Buck L and Axel R (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65:175-187.
- Carmona LM and Schatz DG (2017). New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *Federation of European Biochemical Societies Journal* 284:1590-1605.
- Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD and Julius D (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* 389:816-24.
- Chandrashekar JC, Yarmolinsky D, von Buchholtz L, Oka Y, Sly W, Ryba NJP and Zuker CS (2009). The taste of carbonation. *Science* 326:443-445.
- Constantinides PG, Jones PA and Gevers W (1977). Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. *Nature* 267:364-366.
- Cosens DJ and Manning A (1969). Abnormal electroretinogram from a *Drosophila* mutant. *Nature* 224: 285-287.
- Coste B, Mathur J, Schmidt M, Earley TJ, Ranade S, Petrus MJ, Dubin AE and Patapoutin A (2010). Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* 330:55-60.
- Coste B, Xiao B, Santos JS, Syeda R, Grandl J, Spencer KS, Kim SE, Schmidt M, Mathur J, Dubin AE, Nontal M and Patapoutian A (2012). Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* 483:176-181.
- Davis RL, Weintraub H and Lassar AB (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51:987-1000.
- Dhallan RS, Yao KW, Schrader KA and Reed RR (1990). Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* 347:184-187.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.
- Gerhart J, Danilchik M, Doniach T, Roberts S, Rowning B, Stewart R (1989). Intracellular signaling centers: their establishment and modes of action. *Development* 107:37-51.
- Gilbert LA, Larson MH, Morsut L, Liu Z, Brar, Torres, Stern-Ginossar N, Brandman, Whitehead EH, Doudna JA, Lim WA, Weissman JS and Qi LS (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154:442-451.
- Gimlich RL, Gerhart JC (1984). Early cellular interactions promote embryonic axis formation in *Xenopus laevis*. *Developmental Biology* 104:117-130.
- Guo S and Kemphues K (1995). *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 81:611-620.
- Gurdon JB, Lane CD, Woodland HR, Marbaix G (1971). Use of frog eggs and oocytes for the study of messenger RNA and its translation in living cells. *Nature* 233:177-182.
- Hediger MA, Coady MJ, Ikeda TS and Wright EM (1987). Expression cloning and cDNA sequencing of the Na⁺/glucose co-transporter. *Nature* 330:379-381.
- Hoon MA, Adler E, Lindemeier J, Battey JF, Ryba NJP and Zuker CS (1999). Putative mammalian taste receptors a class of taste-specific GPCRs with distinct topographic selectivity. *Cell* 96:541-551.
- Jia Y, Zhao Y, Kusakizako T, Wang Y, Pan C, Zhang Y, Nureki O, Hattori M and Yan Z (2020). TMC1 and TMC2 proteins are pore-forming subunits of mechanosensitive ion channels. *Neuron* 105:310-321.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816-821.
- Jones DT and Reed RR (1989). *G_{olf}*: an olfactory neuron specific-G protein involved in odorant signal transduction. *Science* 244:790-795.
- Jordt SE and Julius D (2002). Molecular basis for species-specific sensitivity to “hot” chili peppers. *Cell* 108:421-30.

- Julius D, MacDermott AB, Axel R and Jessell TM (1988). Molecular characterization of a functional cDNA encoding the serotonin 1c receptor. *Science* 241:558-564.
- Kao KR, Masui Y and Elinson RP (1986). Lithium-induced respecification of pattern in *Xenopus laevis* embryos. *Nature* 322:371-373.
- Kim Y-G, Cha J and Chandrasegaran S (1996). Hybrid restriction enzymes: zinc finger fusions to *FokI* cleavage domain. *Proceedings of the National Academy of Sciences USA* 93:1156-1160.
- Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MC and Yusa K (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology* 32:267-273.
- Kurima K, Peters LM, Yang Y, Riazuddin S, Ahmed ZM, Naz S, Arnaud D, Drury S, Mo J, Makishima T et al. (2002). Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function. *Nature Genetics* 30:277-284.
- Lassar AB, Paterson BM and Weintraub H (1986). Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* 47:649-656.
- Lassar AB, Buskin JN, Lockshon D, Davis RL, Apone S, Hauschka SD and Weintraub H (1989). MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatinine kinase enhancer. *Cell* 58:823-831.
- Lewis S, Gifford A and Baltimore D (1984). Joining of V_{κ} to J_{κ} gene segments in a retroviral vector introduced into lymphoid cells. *Nature* 308:425-428.
- Lübbert H, Hoffman BJ, Snutch TP, van Dyke T, Levine AJ, Hartig PR, Lester HA and Davidson N (1987). cDNA cloning of a serotonin 5-HT_{1C} receptor by electrophysiological assays of mRNA-injected *Xenopus* oocytes. *Proceedings of the National Academy of Sciences USA* 84:4332-4336.
- Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH and Joung JK (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* 10:977-979.
- Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L and Church GM (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology* 31:833-838.
- Mansour SL, Thomas KR and Capecchi MR (1988). Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* 336:348-352.
- Masu Y, Nakayama K, Tamaki H, Harada Y, Kuno M and Nakanishi S (1987). cDNA cloning of bovine substance-K receptor through oocyte expression system. *Nature* 329:836-838.
- McKemy DD, Neuhauser WM and Julius D (2002). Identification of a cold receptor reveals a general role for TRP channels in thermosensation. *Nature* 416:52-58.
- Montell C and Rubin GM (1989). Molecular characterization of the *Drosophila trp* locus: a putative integral membrane protein required for phototransduction. *Neuron* 2:1313-1323.
- Moriyoshi K, Masu M, Ishii T, Shigemoto R, Mizuno N and Nakanishi S (1991). Molecular cloning and characterization of the rat NMDA receptor. *Nature* 354:31-37.
- Moscou MJ and Bogdanove AJ (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501.
- Mowry KL (2020). Using the *Xenopus* oocyte toolbox. *Cold Spring Harbour Protocol* doi:10.1101/pdb.top095844.
- Nieuwkoop PD (1969). The formation of the mesoderm in urodelean amphibians. I. Induction by the endoderm. *Wilhelm Roux' Archive of Developmental Biology* 162:341-373
- Nieuwkoop PD, Ubbels GA (1972). The formation of the mesoderm in urodelean amphibians. IV. Qualitative evidence for the purely "ectodermal" origin of the entire mesoderm and of the pharyngeal endoderm. *Wilhelm Roux' Archive of Developmental Biology* 169:185-199.
- Pace U, Hanski E, Salomon Y and Lancet D (1985). Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* 316:255-258.
- Pak WL, Grossfield J and White NV (1969). Nonphototactic mutants in a study of vision of *Drosophila*. *Nature* 222:351-354.
- Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE and Gersbach CA (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* 10:973-976.
- Oettinger MA, Schatz DG, Gorka C and Baltimore D (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248:1517-1523.
- Rebagliati MR, Weeks DL, Harvey RP, Melton DA (1985). Identification and cloning of localized maternal RNAs from *Xenopus* eggs. *Cell* 42:769-777.

- Sasai Y, Lu B, Steinbeisser H, Geissert D, Gont LK and De Robertis EM (1994) Xenopus chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* 79:779-790.
- Schatz DG and Baltimore D (1988). Stable expression of immunoglobulin gene V(D)J recombinase activity by gene transfer into 3T3 fibroblasts. *Cell* 53:107-115.
- Schatz DG, Oettinger MA and Baltimore D (1989). The V(D)J recombination activating gene, RAG1. *Cell* 59:1035-1048.
- Schatz DG and Baltimore D (2004). Uncovering the V(D)J recombinase. *Cell* 116:S103-S106.
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG and Zhang F (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84-87.
- Smith WC and Harland RM (1991). Injected Xwnt-8 RNA acts early in Xenopus embryos to promote formation of a vegetal dorsalizing center. *Cell* 67:753-765.
- Smith WC and Harland RM (1992). Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. *Cell* 70:812-840.
- Sumikawa K, Houghton M, Emtage JS, Richards BM and Barnard EA (1981). Active multi-subunit ACh receptor assembled by translation of heterologous mRNA in Xenopus oocytes. *Nature* 292:862-864.
- Taylor S and Jones PA (1979). Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell* 17:771-779.
- Thomas KR and Capecchi MR (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51:503-512.
- Vreugde S, Erven A, Kros CJ, Marcotti W, Fuchs H, Kurima K, Wilcox ER, Friedman TB, Griffith AJ, Balling R et al. (2002). Beethoven, a mouse model for dominant, progressive hearing loss DFNA36. *Nature Genetics* 30:257-258.
- Wang T, Wei JJ, Sabatini DM and Lander ES (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343:80-84.
- Weeks DL and Melton DA (1987). A maternal mRNA localized to the vegetal hemisphere in Xenopus eggs codes for a growth factor related to TGF- β . *Cell* 51:861-777.
- Wood JN, Winter J, James IF, Rang HP, Yeats J and Bevan S (1988). Capsaicin-induced ion fluxes in dorsal root ganglion cells in culture. *Journal of Neuroscience* 8:3208-3220.
- Xu J, Mathur J, Vessières E, Hammack S, Nonomura K, Favre J, Grimaud L, Petrus M, Francisco A, Li J, Lee V, Xiang F-L, Mainquist JK, Cahalan SM, Orth AP, Walker JR, Ma S, Lukacs V, Bordone L, Bandell M, Laffitte B, Xu Y, Chien S, Henrion D and Patapoutian A (2018). GPR68 senses flo and is essential for vascular physiology. *Cell* 173:762-775.
- Zamecnik PC and Stephenson ML (1978). Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences USA* 75:280-284.
- Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y and Wei W (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509:487-491.

阅读

Schatz DG, Oettinger MA and Baltimore D (1989). The V(D)J recombination activating gene, RAG1. *Cell* 59:1035-1048.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816-821.