

Equipe de tradução

Adriana de Freitas Schuck Bizarro (Cap. 12)

Farmacêutica. Mestre em Biologia do Desenvolvimento pela Universidade de Barcelona.

Andréia Escosteguy Vargas (Iniciais e Caps. 4, 5 e 9)

Doutora em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (UFRGS).  
Pos-graduada do Laboratório de Proteômica e Engenharia de Proteínas do Instituto Carlos Chagas/Fundação Oswaldo Cruz (FIOCRUZ), Curitiba, PR.

Ardaia Breda (Glossário, Índice, Caps. 7, 8 e 11)

Pesquisadora do Departamento de Bioquímica da Texas A&M University, Ph.D. em Biologia Celular e Molecular pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

Cláudia Paiva Nunes (Caps. 21 e 22)

Pesquisadora no LANA-GRO-RS. Doutora em Bioquímica e Biologia Molecular pelo Departamento de Bioquímica da UFRGS.

Cristiano Bizarro (Cap. 3)

Professor adjunto da PUCRS. Doutor em Biologia Celular e Molecular pela UFRGS. Pós-Doutor em Biologia Celular e Molecular pela UFRGS. Pós-Doutor em Biologia Celular e Molecular pela Universitat de Barcelona, UB, Espanha. Pós-Doutor em Biologia Celular e Molecular pela PUCRS.

Dalana Remd (Caps. 16 e 24)

Farmacêutica. Mestre em Biologia Celular e Molecular pela PUCRS.

Doutoranda do Programa de Pós-Graduação em Biologia Celular e Molecular da PUCRS e vinculada ao Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF).

Denise Cantarelli Machado (Caps. 14 e 23)

Professora da Faculdade de Medicina e pesquisadora do Instituto de Pesquisas Biomédicas da PUCRS. Mestre em Genética pela UFRGS. Doutora em Imunologia pela University of Sheffield, Inglaterra.

Gaby Renard (Glossário, Índice, Caps. 13, 17, 18 e 21)

Pesquisadora da Quatro G Pesquisa & Desenvolvimento Ltda., TECNOPUC. Mestre e Doutora em Ciências Biológicas: Bioquímica pela UFRGS.

Paulo Luiz de Oliveira (Caps. 1, 2 e 10)

Biólogo. Professor titular aposentado do Departamento de Ecologia do Instituto de Biociências da UFRGS. Mestre em Botânica pela UFRGS. Doutor em Ciências Agrárias pela Universität Hohenheim, Stuttgart, República Federal da Alemanha.

Rosane Sheibe (Caps. 6, 19 e 20)

Doutora em Biologia Molecular pela University of Sheffield, Inglaterra.

Valnês da Silva Rodrigues Junior (Cap. 15)

Pesquisador do Centro de Pesquisas em Biologia Molecular e Funcional da PUCRS. Mestre em Biologia Celular e Molecular pela UFRGS. Doutor em Farmacologia Bioquímica e Molecular pelo Programa de Pós-Graduação em Medicina e Ciências da Saúde da PUCRS.

B615

Biologia celular e molecular / Harvey Lodish ... [et al.] ;  
[tradução: Adriana de Freitas Schuck Bizarro ... et al.] ;  
revisão técnica: Ardaia Breda, Gaby Renard. - 7. ed. -  
Porto Alegre : Artmed, 2014.  
xxiv, 1210 p. : il. color. ; 28 cm.

ISBN 978-85-8271-049-4

1. Biologia. 2. Biologia celular. 3. Biologia molecular.  
I. Lodish, Harvey.

CDU 576

Catálogo na publicação: Ana Paula M. Magnus - CRB 10/2052

Harvey Lodish  
Arnold Berk  
Chris A. Kaiser  
Monty Krieger  
Anthony Bretscher  
Hilde Ploegh  
Angelika Amon  
Matthew P. Scott

# BIOLOGIA CELULAR E MOLECULAR

7ª EDIÇÃO

Revisão técnica desta edição

Ardaia Breda

Pesquisadora do Departamento de Bioquímica da Texas A&M University, Ph.D. em Biologia Celular e Molecular pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

Gaby Renard

Pesquisadora da Quatro G Pesquisa & Desenvolvimento Ltda., TECNOPUC. Mestre e Doutora em Ciências Biológicas: Bioquímica pela Universidade Federal do Rio Grande do Sul (UFRGS).



2014

- Wood, R. D., M. Mitchell, and T. Lindahl. Human DNA repair genes. *Mutat. Res.* 573:275-283.
- Yoshida, K., and Y. Miki. 2004. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci.* 95:866-871.
- Virus: parasitas do sistema genético celular
- Flint, S. J., et al. 2000. *Principles of Virology: Molecular Biology, Pathogenesis, and Control*. ASM Press.
- Hull, R. 2002. *Mathews' Plant Virology*. Academic Press.
- Hull, A. 1999. The tobacco mosaic virus particle: structure and assembly. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 354:531-535.
- Kriple, D. M., and J. M. Howley, eds. 2001. *Fields Virology*. Lippincott Williams & Wilkins.
- Konberg, A., and T. A. Baker. 1992. *DNA Replication*, 2d ed. W. H. Freeman and Company. Good summary of bacteriophage molecular biology.



A planta é um platemínio ("verme achatado") de vida livre com uma incrível capacidade de regeneração. Se a cabeça e a cauda de uma planta adulta forem cortadas, o verme de imediato irá regenerar essas estruturas (conforme mostrado no quadro superior esquerdo). O papel de genes específicos no processo de regeneração pode ser estudado pela repressão da expressão gênica por RNA de interferência (RNAi) antes do corte da cabeça e da cauda. Os oito quadros restantes mostram a variedade de defeitos de regeneração observados após o RNAi de diferentes genes: *responsivus*, pela regeneração dos apêndices por RNAi; da esquerda para a direita, conseguindo pelo quadro superior central, *sno*; *snrhd*, *β-calatrina-1*, antígeno de carcinoma, *POU23*, *rodlerin*, *Novel*, *calbid* e *pml*. (Cortesia de Peter Reddien/NIH, Whitehead Institute.)

## SUMÁRIO

5.1	Análise genética de mutações para identificação e estudo de genes	172
5.2	Clonagem e caracterização do DNA	182
5.3	Uso de fragmentos de DNA clonados para estudo da expressão gênica	198

**N**a área de biologia celular molecular, reduzida a seus elementos mais básicos, busca-se um entendimento acerca do comportamento biológico das células em termos de mecanismos químicos e moleculares subjacentes. Com frequência, a investigação de um novo processo molecular enfoca a função de uma determinada proteína. Existem três perguntas fundamentais que os biólogos celulares geralmente fazem sobre uma proteína nova recém-descoberta: qual é a função da proteína no contexto da célula viva, qual é a função bioquímica da proteína purificada, e onde a proteína está localizada? Para responder a essas questões, pesquisadores empregam três ferramentas de genética molecular: o gene que codifica a proteína, uma linhagem celular ou organismo mutante que não possui a proteína funcional, e uma fonte de proteína purificada para estudos bioquímicos. Neste capítulo, serão considerados vários aspectos de duas estratégias experimentais básicas para a obtenção de todas as três ferramentas (Figura 5-1).

A primeira estratégia, geralmente chamada de *genética clássica*, começa com o isolamento de um mutante que

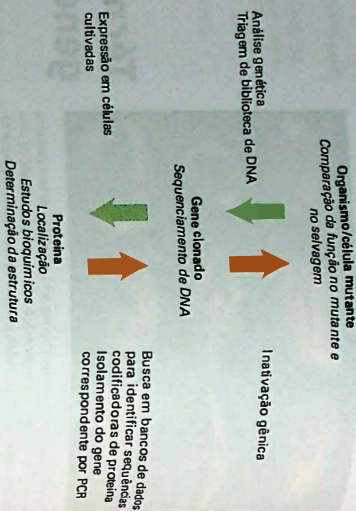
## Técnicas de genética molecular

5.4	Localização e identificação de genes de doenças humanas	206
5.5	Inativação da função de genes específicos em eucariotos	212

parece ser diferente em algum processo de interesse. Métodos genéticos são então utilizados para identificar e isolar o gene afetado. O gene isolado pode ser manipulado a fim de que produza grandes quantidades da proteína para experimentos bioquímicos e prope sondas para estudo de onde e quando a proteína codificada é expressa em um organismo. A segunda estratégia segue essencialmente os mesmos passos da abordagem clássica, mas na ordem inversa, começando com o isolamento de uma proteína de interesse ou sua identificação com base na análise da sequência genômica de um organismo. Uma vez isolado, o gene correspondente pode ser alterado e então reinsertado em um organismo. Em ambas as estratégias, pelo exame das consequências fenotípicas das mutações que inativam um gene em particular, os geneticistas conseguem correlacionar os conhecimentos acerca de sequência, estrutura e atividade bioquímica da proteína codificada à sua função no contexto de uma célula viva ou organismo multicelular.

Um componente importante em ambas as estratégias para estudo de uma proteína e sua função biológica é o isolamento do gene correspondente. Assim, serão discutidas

**FIGURA 5-1** Visão global de duas estratégias para relacionar função, localização e estrutura de produtos gênicos. Um organismo mutante é o ponto de partida para a estratégia genética clássica (setas verdes). A estratégia reversa (setas laranjas) geralmente começa com a identificação de uma sequência codificante de proteína pela análise de bancos de dados de sequências do genoma. Em ambas as estratégias, o gene real é isolado a partir de uma biblioteca de DNA ou pela amplificação específica da sequência gênica a partir do DNA genômico. Uma vez isolado, o gene clonado pode ser usado para produzir a proteína codificada em sistemas de expressão bacterianos ou eucarióticos. Alternativamente, um gene clonado pode ser inativado por uma das várias técnicas e usado para gerar células ou organismos mutantes.



várias técnicas pelas quais os pesquisadores podem isolar, sequenciar e manipular regiões específicas do DNA de um organismo. Após, será discutida uma variedade de técnicas muito utilizadas para análise de onde e quando um determinado gene é expresso e em que parte da célula sua proteína está localizada. Em alguns casos, o conhecimento da função de uma proteína pode levar a avanços médicos significativos, e o primeiro passo no desenvolvimento de tratamentos para uma doença hereditária é identificar e isolar o gene afetado, que será descrito aqui. Finalmente, serão discutidas técnicas que eliminam a função da proteína normal a fim de analisar o papel da proteína na célula.

**5.1 Análise genética de mutações para identificação e estudo de genes**

Conforme descrito no Capítulo 4, a informação codificada na sequência de DNA dos genes especifica a estrutura – e, portanto, a estrutura e a função – de todas as moléculas de proteína de uma célula. O poder da genética como ferramenta para estudo de células e organismos está na habilidade dos pesquisadores em alterar seletivamente todas as cópias de apenas um tipo de proteína em uma célula fazendo uma alteração no gene para aquela proteína. Análises genéticas de mutantes deficientes em um determinado processo podem revelar (a) novos genes necessários para que o processo ocorra, (b) a ordem na qual os produtos gênicos atuam no processo e (c) se as proteínas codificadas por diferentes genes interagem umas com as outras. Antes da análise de como estudos genéticos desse tipo podem fornecer informações acerca do complicado mecanismo de um processo celular ou de desenvolvimento, primeiramente serão explicadas alguns termos genéticos básicos utilizados ao longo de nossa discussão.

As diferentes formas, ou variantes, de um gene são designadas alelos. Os geneticistas geralmente se referem à numerosos variantes genéticos de ocorrência natural que existem nas populações, sobretudo em populações humanas, como alelos. O termo mutação é geralmente reservado para situações nas quais um alelo tem origem re-

centemente recente, como após o tratamento de um organismo com uma substância mutagênica, agente que causa uma alteração hereditária na sequência de DNA.

Estritamente, o conjunto de alelos para todos os genes carregado por um indivíduo é seu genótipo. Entretanto, esse termo também é usado de maneira mais limitada para referir apenas aos alelos de um determinado gene ou genes sob investigação. Em organismos eucarióticos, o termo selvagem geralmente se refere ao genótipo-padrão utilizado como referência em experimentos de reprodução. Assim, o alelo normal, não mutado, via de regra é chamado de selvagem. Por conta da enorme variação alélica que ocorre naturalmente nas populações humanas, o termo *selvagem* em geral se refere a um alelo que está presente em uma frequência muito mais alta do que qualquer outra das possíveis alternativas.

Geneticistas fazem uma importante distinção entre o *genótipo* e o *fenótipo* de um organismo. O *fenótipo* refere-se a todos os atributos físicos ou traços de um indivíduo que são consequência de um dado genótipo. Na prática, entretanto, o termo *fenótipo* com frequência é usado para indicar as consequências físicas que resultam apenas dos alelos que estão sob investigação experimental. Características fenotípicas facilmente observadas são fundamentais para a análise genética de mutações.

**Alelos mutantes recessivos e dominantes geralmente apresentam efeitos opostos sobre a função gênica**

Uma diferença genética fundamental entre organismos experimentais é se carregam apenas um conjunto de cromossomos ou duas cópias de cada um deles. Os primeiros são chamados de haploides, os últimos de diploides. Organismos multicelulares complexos são de mosca-da-fruta, camundongos, seres humanos) são diploides, enquanto muitos organismos unicelulares e plantas são haploides. Alguns organismos, em particular a levedura *Saccharomyces cerevisiae*, podem existir nos estados haploide ou diploide. As células normais carregam dois cromossomos, ambos plantas e animais, portanto, mais de duas cópias de cada cromossomo e, dessa forma, são designadas poliploides. Além disso, células cancer-

genas carregam como células diploides mas ao longo do processo de transformação em células tumorais podem ganhar cópias extras de um ou mais cromossomos e assim são designadas de aneuploides. Entretanto, nossa discussão sobre técnicas e análises genéticas se refere a organismos diploides, incluindo leveduras diploides.

Embora vários alelos diferentes de um gene possam ocorrer em diferentes organismos de uma população, qualquer organismo diploide irá carregar duas cópias de cada gene e assim poder ter no máximo dois alelos diferentes. Um indivíduo com dois alelos diferentes é heterozigoto para um gene, enquanto um indivíduo que carrega dois alelos idênticos é homozigoto para um gene. Um alelo mutante recessivo é definido como aquele no qual ambos os alelos precisam estar mutados a fim de que se observe o fenótipo mutante; isto é, o indivíduo precisa ser homozigoto em relação ao alelo mutante para apresentar o fenótipo mutante. Em contrapartida, as consequências fenotípicas de um alelo dominante podem ser observadas em um indivíduo heterozigoto que carrega um alelo mutante e outro selvagem (Figura 5-2).

O fato de um alelo mutante ser recessivo ou dominante fornece informações valiosas acerca da função do gene afetado e da natureza da mutação. Alelos recessivos geralmente resultam de uma mutação que inativa o gene afetado, levando à *perda de função* parcial ou completa. Tais mutações recessivas podem remover parte de um gene ou o gene inteiro do cromossomo, romper a expressão do gene, ou alterar a estrutura da proteína codificada, alterando sua função. Alelos dominantes geralmente são consequência de uma mutação que causa algum tipo de *ganho de função*. As mutações dominantes podem aumentar a atividade da proteína codificada, conferir-lhe uma nova função, ou levar a um padrão de expressão espacial ou temporalmente inadequado.

Mutações dominantes em certos genes, no entanto, estão associadas a uma perda de função. Por exemplo, alguns genes são *haploinsuficientes*: a remoção ou inativação de um dos dois alelos do gene leva a um fenótipo mutante porque não há produzido gênio suficiente. Em outros exemplos raros, uma mutação dominante em um alelo pode levar a uma alteração estrutural na proteína que interfere na função da proteína selvagem codificada pelo outro alelo. Esse tipo de mutação, chamada de *dominante-negativa*, produz um fenótipo semelhante aquele produzido por uma mutação de perda de função.

Alguns alelos podem apresentar ambas as propriedades recessiva e dominante. Nesses casos, declarações sobre a dominância ou recessividade de um alelo devem especificar o fenótipo. Por exemplo, o alelo do gene da hemoglobina humana designado *Hb<sup>s</sup>* apresenta mais de uma consequência fenotípica. Indivíduos que são homozigotos para este alelo (*Hb<sup>s</sup>/Hb<sup>s</sup>*) possuem a doença debilitante anemia falciforme, mas indivíduos heterozigotos (*Hb<sup>s</sup>/Hb<sup>A</sup>*) não apresentam a doença. Portanto, *Hb<sup>s</sup>* é recessivo para a doença anemia falciforme. Por outro lado, indivíduos heterozigotos (*Hb<sup>s</sup>/Hb<sup>A</sup>*) são mais resistentes à malária do que os homozigotos (*Hb<sup>A</sup>/Hb<sup>A</sup>*), revelando que *Hb<sup>s</sup>* é dominante para a resistência à malária.

Um agente comumente utilizado para induzir mutações (mutagênese) em organismos experimentais é o etilmetanosulfato (EMS). Embora esse agente mutagênico possa alterar a sequência de DNA de várias maneiras, um de seus efeitos mais comuns é modificar quimicamente bases de guanina no DNA, levando à conversão de um par de bases G-C para um par A-T. Essa alteração na sequência de um gene, que envolve apenas um único par de base, é conhecida como mutação de ponto. Uma mutação de ponto *silenciosa* não causa alteração na sequência de aminoácidos ou na atividade da proteína codificada por um gene. Entretanto, consequências fenotípicas perceptíveis devidas a alterações na atividade de uma proteína podem surgir a partir de mutações de ponto que resultam na substituição de um aminoácido por outro (mutação de sentido trocado), na introdução de um códon de parada (mutação *sem sentido*), ou em uma alteração na fase de leitura de um gene (mutação de *mudança de fase de leitura*). Como alterações na sequência de DNA, que levam à diminuição na atividade da proteína são muito mais prováveis do que aquelas que levam ao aumento ou à alteração qualitativa da atividade da proteína, a mutagênese geralmente produz muito mais mutações recessivas do que dominantes.

**A segregação de mutações em experimentos de reprodução revela sua dominância ou recessividade**

Geneticistas exploram o ciclo de vida normal de um organismo para testar a dominância e a recessividade dos alelos. Para analisar este processo, precisa-se primeiro revisar o tipo de divisão celular que origina os gametas (espermatozoides e ovulos em plantas e animais superiores). Enquanto as células do corpo (somáticas) da maioria dos organismos multicelulares se dividem por mitose, as células germinativas que originam os gametas sofrem meiose. Assim como as células somáticas, as células ger-

GENÓTIPO	FENÓTIPO	GENÓTIPO	FENÓTIPO
DIPLOIDE	Salvagem	DIPLOIDE	Mutante
DIPLOIDE	Salvagem	DIPLOIDE	Mutante
DIPLOIDE	Salvagem	DIPLOIDE	Mutante

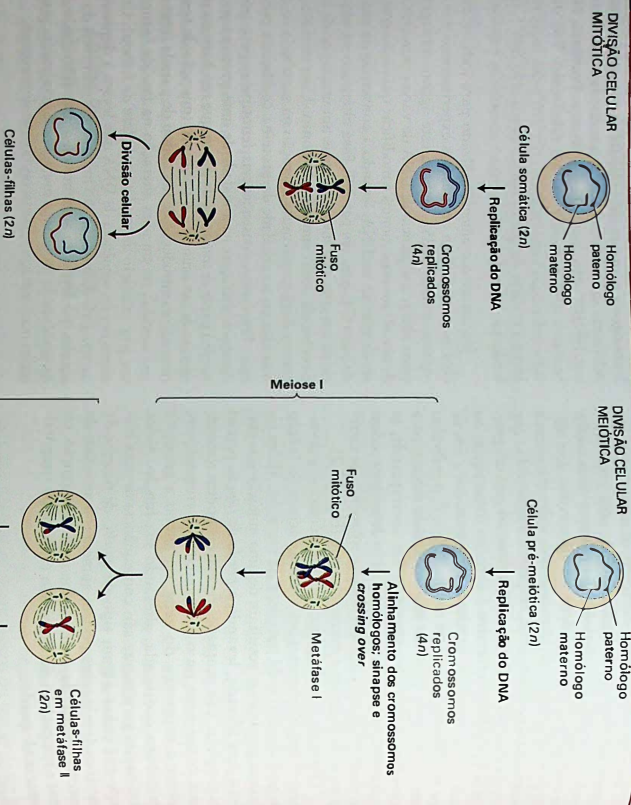
**FIGURA 5-2** Efeitos de alelos mutantes dominante e recessivo no fenótipo de organismos diploides. Uma única cópia de um alelo dominante é suficiente para produzir o fenótipo mutante, enquanto ambas as cópias de um alelo recessivo precisam estar presentes para

provocar um fenótipo mutante. Mutações recessivas geralmente causam uma perda de função; mutações dominantes geralmente causam um ganho de função ou uma função alterada.

minutárias pré-meióticas são diploides, contendo dois homólogos de cada tipo morfológico de cromossomo. Os dois homólogos que constituem cada par de cromossomos

somos homólogos descendem de genitores diferentes, portanto, seus genes podem existir em diferentes formas alélicas. A Figura 5-3 retrata os principais eventos das

**ANIMAÇÃO EM FOCO: MIOSE**



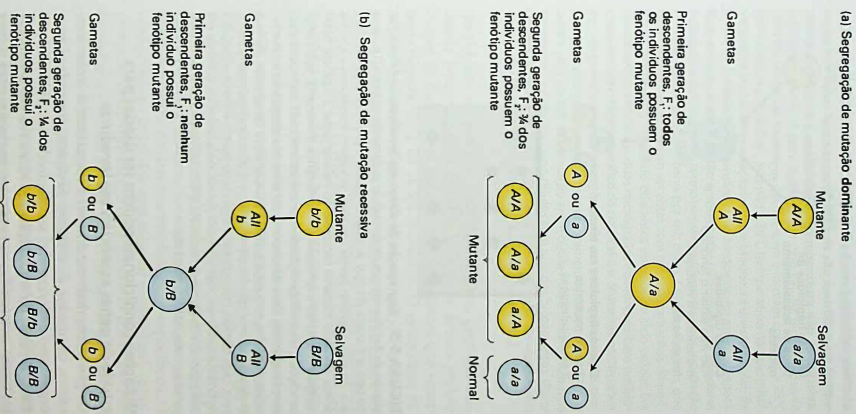
**FIGURA 5-3 Comparação entre mitose e meiose.** Tanto as células somáticas quanto as germinativas pré-meióticas possuem duas cópias de cada cromossomo (2n), um materno e outro paterno. Na mitose, os cromossomos replicados, cada um composto por duas cromátides-irmãs, alinham-se no centro da célula de forma que ambas as células-filhas recebem um homólogo materno e outro paterno de cada tipo morfológico de cromossomo. Durante a primeira divisão meiótica, no entanto, cada cromossomo replicado párea com o seu respectivo homólogo no centro da célula; este pareamento é chamado de *sinapse*, e o *crossing over* entre cromossomos homólogos

faixa evidente neste estágio. Um cromossomo replicado de cada tipo morfológico vai para cada célula-filha. As células resultantes sofrem uma segunda divisão sem passar por replicação do DNA, com as cromátides-irmãs de cada tipo morfológico de cromossomo segregando para as células-filhas. Na segunda divisão meiótica, o alinhamento de pares de cromossomos homólogos na metafase I é aleatório em relação aos cromossomos maternos e paternos em cada célula-filha.

divisões celulares mitótica e meiótica. Na mitose, a replicação do DNA é sempre seguida pela divisão celular, gerando duas células-filhas diploides. Na meiose, um ciclo de replicação de DNA é seguido por duas divisões celulares separadas, gerando quatro células haploides (1n) que contêm apenas um cromossomo de cada par de homólogos. A distribuição, ou segregação, dos cromossomos homólogos replicados para as células-filhas durante a primeira divisão meiótica é aleatória; isto é, homólogos maternos e paternos segregam independentemente, gerando células-filhas com diferentes misturas de cromossomos paternos e maternos.

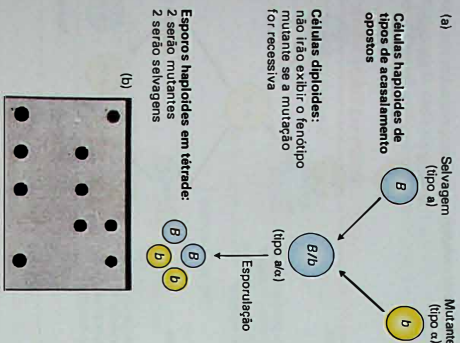
Como uma forma de evitar complexidade indesejada, geneticistas geralmente procuram começar experimentos de reprodução com linhagens que sejam homogotas para os genes sob investigação. Nestas linhagens *puras*, cada indivíduo irá receber o mesmo alelo de cada genitor e, portanto a composição dos alelos não mudará de uma geração para a outra. Quando uma linhagem pura mutante é acasalada com uma linhagem pura selvagem, toda a primeira geração de descendentes (F<sub>1</sub>) será heterozigota (Figura 5-4). Se a geração F<sub>1</sub> exibir o traço mutante, então o alelo mutante é dominante e o alelo selvagem é recessivo. O cruzamento posterior entre indivíduos da F<sub>1</sub> também revelará diferentes padrões de herança de acordo com a dominância ou recessividade da mutação. Quando indivíduos da F<sub>1</sub> heterozigotos para um alelo dominante forem cruzados entre si, três quartos da geração F<sub>2</sub> resultante exibirá o traço mutante. Em contrapartida, quando indivíduos da F<sub>1</sub> heterozigotos para um alelo recessivo forem cruzados entre si, apenas um quarto da geração F<sub>2</sub> resultante exibirá o traço mutante.

Conforme observado, a levedura *S. cerevisiae*, um importante organismo experimental, pode existir em estado haploide ou diploide. Nesses eucaríotes unicelulares, cruzamentos entre células haploides podem determinar se um alelo mutante é dominante ou recessivo. Células de levedura haploides, que carregam uma cópia de cada cromossomo, podem ser de dois tipos de acasalamento, conhecidos como  $\alpha$  e  $\beta$ . Células haploides do tipo de acasalamento oposto podem se acasalar e produzir diploides  $\alpha\beta$ , as quais possuem duas cópias de cada cromossomo. Se uma nova mutação com um fenótipo perceptível for isolada em uma linhagem haploide, a linhagem mutante pode ser acasalada com uma linhagem selvagem do tipo de acasalamento oposto para produzir diploides  $\alpha\beta$  que são heterozigotos quanto ao alelo mutante. Se estes diploides exibirem o traço mutante, então o alelo mutante será dominante, mas se os diploides forem como os selvagens, então o alelo mutante será recessivo. Quando os diploides  $\alpha\beta$  são submetidos a condições de privação de nutrientes, as células sofrem meiose, originando quatro esporos haploides, dois do tipo  $\alpha$  e dois do tipo  $\beta$ . A esporulação de uma célula diploide heterozigota gera dois esporos carregando o alelo mutante e dois carregando o alelo selvagem (Figura 5-5). Sob condições apropriadas, os



**FIGURA 5-4 Padrões de segregação de mutações dominante e recessiva em cruzamentos entre linhagens puras de organismos diploides.** Todos os descendentes na primeira geração (F<sub>1</sub>) são heterozigotos. Se o alelo mutante for dominante, os indivíduos de F<sub>1</sub> não exibem o fenótipo mutante, como mostrado na parte (a). Se o alelo mutante for recessivo, os indivíduos de F<sub>1</sub> não exibem o fenótipo selvagem, como mostrado na parte (b). O cruzamento entre os heterozigotos de F<sub>1</sub> também produz diferentes proporções de segregação para os alelos dominante e recessivo na geração F<sub>2</sub>.

esporos de levedura irão germinar, produzindo linhagens haploides vegetativas de ambos os tipos de acasalamento.



**FIGURA 5-5 Segregação de alelos em levedura.** (a) Células haploides de *Saccharomyces* de tipos de acasalamento opostos (i.e., uma de tipo a e outra do tipo a) podem cruzar e produzir uma diploide *aa*. Se uma haploide portar um alelo selvagem dominante e a outra portar um alelo mutante recessivo do mesmo gene, a diploide heterozigota resultante expressará a característica dominante, sob certas condições, uma célula diploide formada uma tetrade com quatro esporos haploides. Dois dos esporos da tetrade expressarão a característica recessiva, e os outros dois, a característica dominante. (b) Se o fenótipo mutante não for viável sob condições de crescimento restritas, cada tetrade e/ou representada aqui como quatro esporos separados verticalmente e crescidos em colônias em meio mínimo, consiste em dois esporos viáveis e dois inviáveis. (Parte (b) reproduzida de B. Senger et al., 1998, *EMBO J* 17:21196)

**Mutações condicionais podem ser usadas para estudo de genes essenciais em leveduras**

Os procedimentos usados para identificar e isolar mutantes, chamados de *tragens genética*, dependem de o organismo experimental ser haploide ou diploide e, se for o último, de a mutação ser recessiva ou dominante. Genes que codificam proteínas essenciais para a vida estão entre os mais interessantes e importantes a serem estudados. Uma vez que a expressão fenotípica das mutações em genes essenciais leva à morte do indivíduo, são necessárias *tragens* genéticas inteligentes para isolar e manter os organismos com mutações letais.

Em células de levedura haploides, genes essenciais podem ser estudados pelo uso de *mutações condicionais*. Entre as mutações condicionais mais comuns estão as mutações de sensibilidade à temperatura, que podem ser isoladas em bactérias e eucariotos inferiores mas não em eucariotos de sangue quente. Por exemplo, uma única mutação de sentido trocado pode fazer a proteína mutante ressaltante ter uma estabilidade térmica reduzida, de forma que a proteína seja totalmente funcional sob uma temperatura (p. ex., 23°C), mas comece a desnaturar e seja inativa

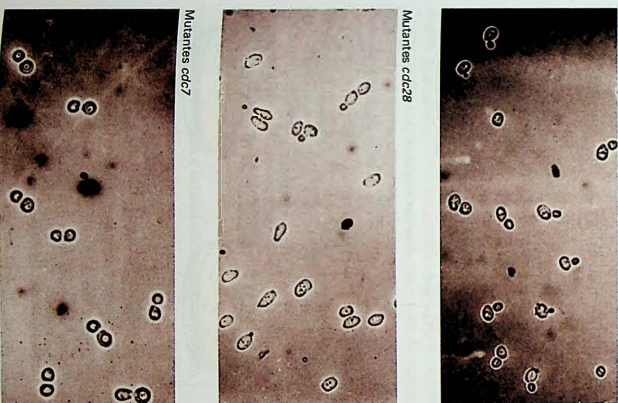
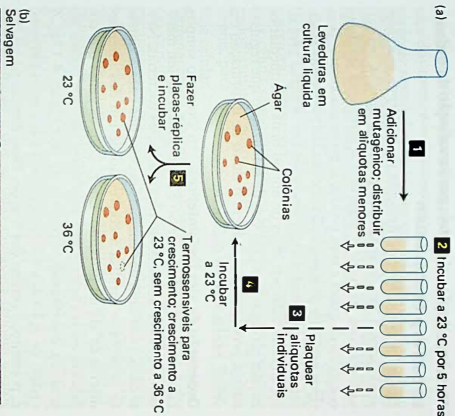
sob outra temperatura (p. ex., 36°C), enquanto a proteína normal seria totalmente estável e funcional em ambas as temperaturas. Uma temperatura na qual se observa o fenótipo mutante é chamada de *não permissiva*; temperatura *permissiva* é aquela na qual não se observa o fenótipo mutante, embora o alelo mutante esteja presente. Portanto, linhagens mutantes podem ser mantidas em uma temperatura permissiva e então subcultivadas a uma temperatura não permissiva para a análise do fenótipo mutante.

Um exemplo de *tragem* particularmente importante para mutantes termosensíveis de levedura *S. cerevisiae* vem dos estudos de L. H. Hartwell e colaboradores no final dos anos 1960 e início dos 1970. Os pesquisadores propuseram-se a identificar genes importantes para a regulação do ciclo celular durante o qual uma célula sintetiza proteínas, replica seu DNA e então sofre mitose, cada célula-filha recebe uma cópia de cada cromossomo. O crescimento exponencial de uma única célula de levedura por 20 a 30 divisões celulares forma uma colônia de leveduras visível em meio agar sólido. Uma vez que mutantes com bloqueio completo do ciclo celular não conseguiriam formar uma colônia, mutantes condicionalmente necessários ao estudo de mutações que afetam esse processo celular básico. Para rastrear tais mutantes, os pesquisadores primeiramente identificaram células de levedura mutadas que podiam crescer normalmente a 23°C, mas que não formavam colônias quando submetidas a 36°C (Figura 5-6a).

Uma vez que os mutantes termosensíveis haviam sido isolados, análises posteriores revelaram que alguns eram de fato defeituosos na divisão celular. Em *S. cerevisiae*, a divisão celular ocorre por meio de um processo de brotamento, e o tamanho do broto, que é facilmente visualizado por microscopia de luz, indica a posição da célula no ciclo celular. Cada um dos mutantes que não conseguia crescer a 36°C foi examinado por microscopia após várias horas sob temperatura não permissiva. O exame de vários mutantes termosensíveis diferentes revelou que cerca de 1% exibia um bloqueio distinto no ciclo celular. Estes mutantes foram então designados *mutantes cdc* (ciclo de divisão celular). É importante salientar que os mutantes de levedura não apenas falhavam em crescer, como fariam se carregassem uma mutação que afetasse o metabolismo celular geral. Em vez disso, se cresciam normalmente durante parte do ciclo celular, sob temperatura não permissiva, os mutantes de interesse ficavam presos em um determinado estágio do ciclo celular, de forma que várias células eram vistas neste estágio (Figura 5-6b). A maioria das mutações *cdc* haploides são cruzadas com haploides selvagens, os diploides heterozigotos resultantes não são nem termosensíveis nem defeituosos em ciclo celular.

**Mutações letais recessivas em diploides podem ser identificadas por endogamia e mantidas em heterozigotos**

Em organismos diploides, os fenótipos resultantes de mutações recessivas podem ser observadas apenas em



**FIGURA EXPERIMENTAL 5-6 Leveduras haploides portadoras de mutações letais termosensíveis são mantidas em temperatura permissiva e analisadas em temperatura não permissiva.** (a) *Tragem* genética para mutantes de ciclo celular termosensíveis (*cdc*) em leveduras. Leveduras que crescem e formam colônias a 23°C (temperatura permissiva), mas não a 36°C (temperatura não permissiva) podem portar uma mutação letal que bloqueia a divisão celular. (b) Esporos de células termosensíveis são bloqueados em estágios específicos do ciclo celular. Aqui são apresentadas micrografias de leveduras selvagens e dois mutantes termosensíveis diferentes após incubação em temperatura não permissiva por seis horas. Células selvagens, que continuam a crescer, podem ser vistas com todos os tamanhos diferentes de brotamento, refletindo diferentes estágios do ciclo celular. Em contrapartida, células nas duas micrografias inferiores exibem um bloqueio em um estágio específico do ciclo celular. Mutantes *cdc28* param em um ponto anterior à emergência de um novo brotamento e, portanto, aparecem como células sem brotamentos. Mutantes *cdc7*, que param pouco antes da separação entre célula-mãe e broto (célula-filha emergente), aparecem como células com grandes brotamentos. (Parte (a) ver L. H. Hartwell, 1967, *J. Bacteriol.* 93:1662; parte (b) reproduzida de L. M. Heledorf e L. H. Hartwell, 1974, *J. Mol. Biol.* 84:445)

geralmente apenas um alelo de um gene, gerando mutantes heterozigotos, *tragens* genéticas devem incluir etapas de endogamia para gerar descendentes que sejam homozigotos para os alelos mutantes. O geneticista H. Muller desenvolveu um procedimento geral e eficiente para a realização desses experimentos de endogamia na mosca-da-fruta *Drosophila*. Mutações letais recessivas em *Drosophila* e outros organismos diploides podem ser mantidas em indivíduos heterozigotos, e suas consequências fenotípicas, analisadas nos homozigotos.

A abordagem de Muller foi usada com grande efeito por C. Nüsslein-Volhard e E. Wieschaus, que rastream sistemáticamente mutações letais recessivas, afetando a embriogênese em *Drosophila*. Embriões homozigotos mortos, carregando mutações letais recessivas identificadas por essa *tragem*, foram examinados ao microscópio em busca de defeitos morfológicos específicos. O conhecimento atual dos mecanismos moleculares subjacentes ao desenvolvimento de organismos multicelulares fundamenta-se, em grande parte, no quadro detalhado do desenvolvimento embrionário revelado pela caracterização de mutantes de *Drosophila*.

**Testes de complementação determinam se diferentes mutações recessivas ocorrem em um mesmo gene**

Na abordagem genética para estudo de um determinado processo celular, pesquisadores geralmente isolam várias mutações recessivas que produzem o mesmo fenótipo. Um teste comum para determinar se as mutações ocorrem no mesmo gene ou em genes diferentes explora o fenômeno da complementação genética, isto é, o restabelecimento do fenótipo selvagem pelo acasalamento de dois mutantes diferentes. Se duas mutações recessivas, *a* e *b*, estiverem no mesmo gene, então um organismo diploide heterozigoto para ambas as mutações (i.e., carregando um alelo *a* e um alelo *b*) exibirá o fenótipo mutante porque nenhum dos alelos fornecerá uma cópia funcional do gene. Em contrapartida, se as mutações *a* e *b* estiverem em genes *individuais*, então os heterozigotos carregando uma uni-

ca cópia de cada alelo mutante não exibição o fenótipo mutante porque um alelo selvagem de cada gene também estaria presente. Neste caso, diz-se que as mutações *complementam* uma à outra. A análise de complementação não pode ser realizada em mutantes dominantes porque o fenótipo conferido pelo alelo mutante e exibido mesmo na presença de um alelo selvagem do gene.

A análise de complementação de um conjunto de mutantes com o mesmo fenótipo pode distinguir os genes individuais em um conjunto de genes funcionalmente relacionados, os quais precisam funcionar todos para produzir um dado traço fenotípico. Por exemplo, a triagem para mutações *cdc* em *Saccharomyces* descrita anteriormente gerou vários mutantes recessivos termosensíveis que apareciam presos no mesmo estágio do ciclo celular. Para determinar quantos genes haviam sido afetados pelas mutações, Hartwell e colaboradores realizaram testes de complementação em todas as combinações de pares de mutantes *cdc* seguindo o protocolo geral delineado na Figura 5-7. Os testes identificaram mais de 20 diferentes genes *cdc*. A caracterização molecular subsequente dos genes *cdc* e de suas proteínas, conforme descrito em detalhe no Capítulo 20, forneceu um quadro para a compreensão acerca de como a divisão celular é regulada em organismos desde as leveduras até os seres humanos.

**Mutantes duplos são úteis para avaliação da ordem na qual as proteínas atuam**

Baseados em análises criteriosas de fenótipos mutantes associados a um determinado processo celular, pesquisadores com frequência deduzem a ordem na qual um conjunto de genes e seus produtos proteicos atuam. Dois tipos gerais de processos são adequados para essas análises: (a) vias biossintéticas nas quais um material precursor é convertido por meio de um ou mais intermediários em um produto final; (b) vias de sinalização que regem outros processos e envolvem o fluxo de informação em vez de intermediários químicos.

**Ordenação de vias biossintéticas** Um exemplo simples do primeiro tipo de processo é a biossíntese de um metabólito como o aminoácido triptofano nas bactérias. Neste caso, cada uma das enzimas necessárias para a síntese de triptofano catalisa a conversão de um dos intermediários da via no próximo. Em *E. coli*, os genes que codificam estas enzimas são adjacentes uns aos outros no genoma, constituindo o óperon *trp* (ver Figura 4-13a). A ordem de ação dos diferentes genes para estas enzimas, e, portanto, a ordem das reações bioquímicas na via, foi inicialmente deduzida a partir dos tipos de compostos intermediários acumulados em cada mutante *c*. No caso

de vias sintéticas complexas, entretanto, a análise fenotípica dos mutantes defeituosos em uma única etapa pode fornecer resultados ambíguos que não permitem a ordenação conclusiva das etapas. Mutantes duplos defeituosos em duas etapas da via são particularmente úteis ao ordenamento de tais vias (Figura 5-8a).

No Capítulo 14, foi discutido o uso clássico da estratégia de duplo-mutante para ajudar a elucidar a via secretora. Nessa via, as proteínas a serem secretadas pela célula se movem de seu sítio de síntese no retículo endoplasmático rugoso (RE) para o aparelho de Golgi; depois, para vesículas secretoras e, por fim, para a superfície celular.

**Ordenação de vias de sinalização** Conforme será visto em capítulos posteriores, a expressão de muito genes eucarióticos é regulada por vias de sinalização iniciadas por hormônios extracelulares, fatores de crescimento, ou outros sinais. Tais vias de sinalização podem incluir numerosos componentes, e a análise de duplo-mutantes com frequência fornece conhecimento sobre as funções e interações desses componentes. O único pré-requisito para a obtenção de informações úteis a partir deste tipo de análise é que as duas mutações tenham efeitos opostos no produto final da mesma via regulada. Mais comumente, uma mutação repõe a expressão de um determinado gene-reporter mesmo quando o sinal está presente, enquanto outra mutação resulta na expressão do gene-reporter mesmo quando o sinal está ausente (i.e., expressão constitutiva). Conforme ilustrado na Figura 5-8b, dois mecanismos reguladores simples são consistentes com tais mutantes individuais, mas o fenótipo duplo-mutante pode distinguir entre eles. Esta abordagem geral possibilitou aos geneticistas o delineamento de muitas das etapas-chave em uma variedade de vias reguladoras, o que tornou possível a realização de ensaios bioquímicos mais específicos.

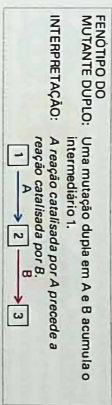
A técnica difere da análise de complementação porque ambos os mutantes, dominante e recessivo, podem ser submeteridos à análise de duplo-mutante. Quando duas mutações recessivas são testadas, o duplo-mutante criado deve ser *homozigoto* em ambas as mutações. Além disso, mutantes dominantes podem ser submeteridos à análise de duplo-mutante.

**Supressão genética e letalidade sintética podem revelar a interseção ou a redundância de proteínas**

Dois outros tipos de análise genética podem fornecer pistas adicionais sobre como proteínas que atuam no mesmo processo celular interagem umas com as outras na célula viva. Ambos os métodos aplicáveis a muitos organismos experimentalmente, envolvem o uso de duplo-mutantes nos quais os efeitos fenotípicos de uma mutação são alterados pela presença de uma segunda mutação.

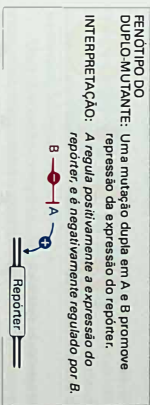
**Mutações supressoras** O primeiro tipo de análise fundamenta-se na *supressão genética*. A fim de entender esse fenômeno, suponha que duas mutações de ponto levem a alterações estruturais em uma proteína (A) que diminui

(a) Análise de uma rota biossintética. Uma mutação em A acumula o intermediário 1. Uma mutação em B acumula o intermediário 2.



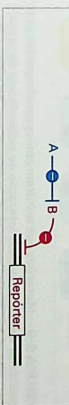
(b) Análise de uma via de sinalização

Uma mutação em A promove expressão da expressão do reporter. Uma mutação em B promove expressão constitutiva do reporter.



**FENÓTIPO DO DUPLO-MUTANTE:** Uma mutação dupla em A e B promove repressão da expressão do reporter.

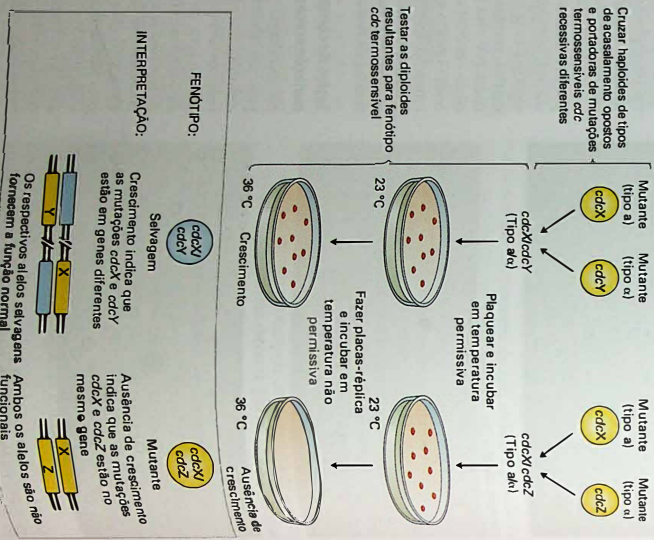
**INTERPRETAÇÃO:** A regula positivamente a expressão do reporter, e B negativamente regula por B.



**FIGURA 5-8** A análise de mutantes duplos geralmente pode ordenar as etapas de rotas biossintéticas ou de sinalização. Quando mutações em dois genes diferentes afetam o mesmo processo celular, mas apresentam fenótipos bastante distintos, o fenótipo do mutante duplo com frequência revela a ordem na qual os dois genes devem atuar. (a) No caso de mutações que afetam a mesma rota biossintética, um mutante duplo acumulará o intermediário imediatamente anterior à etapa catalisada pela proteína que atua inicialmente no organismo selvagem. (b) A análise de mutante duplo de uma via de sinalização é possível se duas mutações tiverem efeitos opostos na expressão de um gene-reporter. Neste caso, o fenótipo observado no mutante duplo fornece informações sobre a ordem na qual as proteínas atuam e se elas são reguladoras positivas ou negativas.

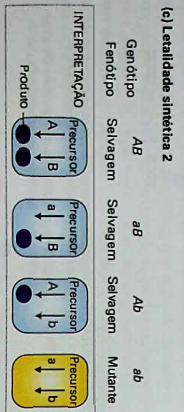
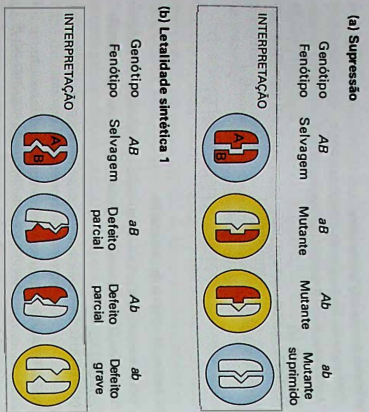
nam sua habilidade para se associar a outra proteína (B) envolvida no mesmo processo celular. Da mesma forma, mutações na proteína B levam a pequenas alterações estruturais que inibem sua habilidade de interagir com a proteína A. Considere, além disso, que o funcionamento normal das proteínas A e B depende de sua interação. Teoricamente, uma alteração estrutural específica na proteína A poderia ser suprimida por alterações compensatórias na proteína B, permitindo que as proteínas mutantes interagiram. Nos raros casos em que tais mutações supressoras ocorrem, linhagens portadoras de ambos os alelos mutantes seriam normais, enquanto linhagens portadoras de apenas um ou outro alelo mutante teriam um fenótipo mutante (Figura 5-9a).

**FIGURA EXPERIMENTAL 5-7** A análise de complementação determina se mutações recessivas estão no mesmo gene ou em genes diferentes. Testes de complementação em leveduras são realizados pelo cruzamento entre células haploides  $\alpha$  e  $\sigma$  portando diferentes mutações recessivas para gerar células diploides. Na análise de mutações *cdc*, pares de linhagens *cdc* termosensíveis diferentes foram sistematicamente cruzadas, e as diploides resultantes foram testadas para crescimento em temperaturas permissivas e não permissivas. Neste exemplo hipotético, os mutantes *cdcX* e *cdcY* se complementam e, portanto, possuem mutações em genes diferentes, enquanto os mutantes *cdcX* e *cdcZ* possuem mutações no mesmo gene.



**INTERPRETAÇÃO:** Crescimento indica que as mutações *cdcX* e *cdcY* estão em genes diferentes. Ausência de crescimento indica que as mutações *cdcX* e *cdcZ* estão no mesmo gene.

**FENÓTIPO:** Os respectivos alelos selvagens funcionam a função normal.



**FIGURA 5-9** **Mutações que resultam em supressão genética ou letalidade sinérgica revelam interação ou redundância de proteínas.** (a) A observação de que mutantes duplos com duas proteínas deletivas (A e B) apresentam um fenótipo selvagem, mas de que mutantes simples apresentam fenótipo mutante, indica que a função de cada proteína depende da interação entre elas. (b) A observação de que mutantes duplos possuem um defeito fenotípico mais grave do que mutantes simples também é evidência de que duas proteínas (a, b) subunidades de um heterodímero) devem interagir para funcionar normalmente. (c) A observação de que um mutante duplo é inviável, mas se que os mutantes simples correspondentes possuem o fenótipo selvagem, indica que duas proteínas atuam em rotas redundantes para produzir um produto essencial.

A observação de supressão genética em linhagens de *Leveduras* portadoras de um alelo de actina mutante (*act1-1*) e uma segunda mutação (*sac6*) em outro gene forneceu evidências iniciais para uma interação direta *in vivo* entre as proteínas codificadas pelos dois genes. Estudos bioquímicos posteriores mostraram que as duas proteínas – Act1 e Sac6 – de fato interagem na construção de estruturas de actina funcionais dentro da célula.

**Mutações sintéticas letais** Outro fenômeno, chamado *letalidade sintética*, produz um efeito fenotípico oposto àquele da supressão. Neste caso, o efeito deletivo de uma mutação é exacerbado (em vez de suprimido) por uma segunda mutação em um gene relacionado. Uma situação na qual tais mutações sintéticas letais podem ocorrer é ilustrada na Figura 5-9b. Neste exemplo, uma proteína heterodimérica é parcialmente, mas não completamente, inativada por mutações em uma das subunidades não idênticas. Entretanto, em mutantes duplos portadores de mutações específicas nos genes que codificam ambas as subunidades, ocorre pouca interação entre as subunidades, resultando em efeitos fenotípicos severos. Mutações sintéticas letais também podem revelar genes não essenciais cujas proteínas atuam em vias redundantes para produzir um componente celular essencial. Como representado na Figura 5-9c, se a proteína não for inativada por uma mutação, a outra via poderá fornecer o produto necessário. Entretanto, se ambas as vias forem inativadas ao mesmo tempo, o produto essencial não poderá ser sintetizado, e os mutantes duplos serão inviáveis.

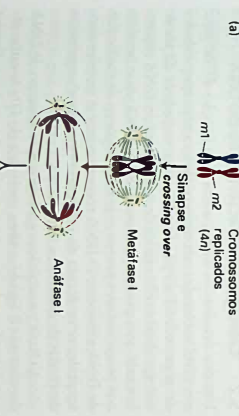
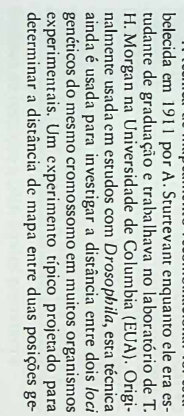
**Genes podem ser identificados pelo mapeamento da sua posição no cromossomo**

A discussão anterior sobre análise genética ilustra como um geneticista pode obter informações sobre a função gênica por meio da observação dos efeitos fenotípicos produzidos pela junção de diferentes combinações de alelos mutantes na mesma célula ou organismo. Por exemplo, combinações de diferentes alelos do mesmo gene em um organismo diploide podem ser usadas para determinar se uma mutação é dominante ou recessiva, ou se duas mutações recessivas diferentes estão no mesmo gene. Além disso, combinações de mutações em genes diferentes podem ser usadas para determinar a ordem da função gênica em uma via ou para identificar relações funcionais entre genes, como supressão ou aumento sintético. De maneira geral, todos esses métodos podem ser vistos como testes analíticos baseados na *função gênica*. Agora será considerado um tipo de análise genética fundamentalmente diferente, baseado em *posição gênica*. Estudos projetados para determinar a posição de um gene em um cromossomo, geralmente chamados de estudos de mapeamento genético, podem ser usados para identificar o gene afetado por uma determinada mutação ou para determinar se duas mutações estão no mesmo gene.

Em muitos organismos, estudos de mapeamento genético contam com as trocas de informação genética que ocorrem durante a meiose. Como representado na Figura 5-10a, a recombinação genética ocorre antes da primeira divisão meiótica nas células germinativas, quando os cromossomos replicados de cada par homólogo alinham-se uns com os outros. Neste momento, seqüências de DNA homólogas em cromátides maternos e paternos podem ser trocadas entre elas, em um processo conhecido como *crossing over*. Hoje sabe-se que os *crossings* resultantes entre cromossomos homólogos formam ligações estruturais importantes para a segregação adequada dos pares de cromátides homólogos em pólos opostos durante a primeira divisão meiótica (para discussão, ver Capítulo 19).

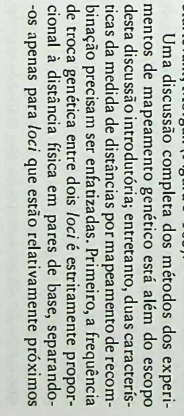
Considere duas mutações diferentes, cada uma herdada de um dos genitores, que estão localizadas por trás de uma da outra no mesmo cromossomo. Dois tipos diferentes de gametas podem ser produzidos, caso o

*crossover* ocorra ou não entre as mutações durante a meiose. Se não ocorrer *crossover* entre elas, gametas com uma ou outra mutação, serão produzidos. Se um *crossover* ocorrer entre as duas mutações, gametas conhecidos como *tipos recombinantes* serão produzidos. Neste exemplo, cromossomos recombinantes conterão ambas as mutações ou nenhuma delas. Os sítios de recombinação ocorrem mais ou menos ao acaso ao longo dos cromossomos; assim, quanto mais próximos dois genes estiverem, menor a probabilidade de que ocorra recombinação entre eles durante a meiose. Em outras palavras, *quanto mais próximos estiverem dois genes do mesmo cromossomo, menos frequente será a recombinação entre eles*. Dois genes do mesmo cromossomo suficientemente próximos, de maneira que há menor produção de gametas recombinantes do que parentais, são considerados *geneticamente ligados*.



**FIGURA 5-10** **A recombinação durante a meiose pode ser usada para mapear a posição dos genes.** (a) Um indivíduo portador de duas mutações, designadas *m1* (anarelado) e *m2* (vermelho), que estão no mesmo cromossomo, e a parceira do mesmo cromossomo, é ilustrado. Se ocorrer *crossing over* em um intervalo entre *m1* e *m2*, antes da primeira divisão meiótica, então dois gametas recombinantes serão produzidos, um deles portará ambas *m1* e *m2*, enquanto o outro não portará nenhuma mutação. Quanto maior a distância entre duas mutações em uma cromátide, maior a chance de que sejam separadas por re-

combinação e maior a proporção de gametas recombinantes produzidos. (b) Em um experimento de mapeamento típico, uma linhagem heterozigota para dois genes diferentes é cruzada. A frequência de gametas parentais ou recombinantes produzidos por esta linhagem pode ser determinada a partir do fenotipo dos descendentes em um cruzamento-teste com uma linhagem homocigota recessiva. A distância de mapa genético em centimorgan (cM) é dada pela porcentagem dos gametas que são recombinantes.



A distância genética entre A e B pode ser determinada pela frequência de gametas parentais e recombinantes:

$$\text{Distância genética em cM} = 100 \times \frac{\text{gametas recombinantes}}{\text{total de gametas}}$$

Considera-se dois genes ligados A e B com alelos recessivos a e b. Cruzamento de dois mutantes para gerar uma linhagem duplamente heterozigota:

$$A/aB/b \times a/aB/B$$

↓

$$\begin{matrix} A & B & a & B \\ a & b & a & b \end{matrix}$$

Cruzamento do duplo heterozigoto para testar a linhagem:

$$\begin{matrix} A & B & a & B \\ a & b & a & b \end{matrix} \times \begin{matrix} A & B & a & B \\ a & b & a & b \end{matrix}$$

Tipos parentais                      Tipos recombinantes

(p. ex., a menos de cerca de 10 CM). Para loci ligados que estão mais afastados do que isso, uma medida de distância genética de troca genética tende a subestimar a distância física devido à possibilidade de dois ou mais *crossovers* ocorrerem dentro de um intervalo. No caso limitante no qual o número de tipos recombinantes iguala o número de tipos parentais, os dois loci consideram-se estar distantes no mesmo cromossomo ou em cromossomos diferentes; nestes casos, os loci são considerados *não ligados*.

Um segundo conceito importante necessário à interpretação de experimentos de mapeamento genético em diferentes tipos de organismos é que, embora a distância genética seja definida da mesma forma para diferentes organismos, a relação entre a frequência de recombinação (i.e., distância de mapa genético) e a distância física varia entre os organismos. Por exemplo, uma frequência de recombinação de 1% (i.e., uma distância genética de 1 cM) representa uma distância física de cerca de 2,8 quilobases em leveduras comparada a uma distância de aproximadamente 400 quilobases em *Drosophila* e cerca de 780 quilobases em humanos.

Um dos principais usos dos estudos de mapeamento genético é para localização do gene afetado por uma mutação de interesse. A presença de várias características genéticas diferentes já mapeadas, ou marcadores genéticos distribuídas ao longo de um cromossomo permite que a posição de uma mutação não mapeada seja determinada pelo estudo de sua segregação em relação a estes genes marcadores durante a meiose. Assim, quando tais marcadores estiverem disponíveis, mais precisamente se pode mapear uma mutação. Na Seção 5.4, será visto como os genes afetados em doenças hereditárias humanas são identificados usando-se tais métodos. Um segundo uso geral dos experimentos de mapeamento é determinar se duas mutações diferentes estão no mesmo gene. Se duas mutações estiverem no mesmo gene, elas irão apresentar forte ligação em experimentos de mapeamento; se estiverem em genes diferentes, geralmente serão não ligadas ou exibirão ligação fraca.

**CONCEITOS-CHAVE da Seção 5.1**

**Análise genética de mutações para identificação e estudo de genes**

- Organismos diploides portam duas cópias (alótipos) de cada gene, enquanto organismos haploides possuem apenas uma cópia.
- Mutações recessivas levam à perda de função, que é mascarada se um alelo normal do gene estiver presente. Para que o fenótipo mutante se manifeste, ambos os alelos devem apresentar a mutação.
- Mutações dominantes levam a um fenótipo mutante na presença de um alelo normal do gene. Os fenótipos associados com mutações dominantes geralmente representam um ganho de função, mas no caso de alguns genes resultam de uma perda de função.

• Na meiose, uma célula diploide sofre uma replicação do DNA e duas divisões celulares, gerando quatro células haploides nas quais os cromossomos maternos e paternos e seus alelos associados são aleatoriamente distribuídos (ver Figura 5-3).

• Mutações dominantes e recessivas exibem padrões de segregação característicos em cruzamentos genéticos (ver Figura 5-4).

• Em leveduras haploides, mutações termosensíveis são particularmente úteis para a identificação e para o estudo de genes essenciais à sobrevivência.

• O número de genes funcionalmente relacionados envolvidos em um processo pode ser definido pela análise de complementação (ver Figura 5-7).

• A ordem na qual os genes atuam em uma via de síntese pode ser deduzida a partir do fenótipo de mutantes duplos defeituosos em duas etapas do processo afetado.

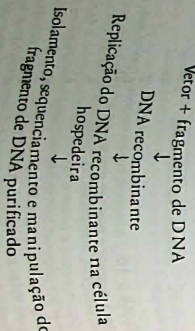
• Interações funcionalmente significativas entre proteínas podem ser deduzidas a partir dos efeitos fenotípicos de mutações supressoras alelo-específicas ou mutações sintéticas letais.

• Experimentos de mapeamento genético utilizam o *crossing over* entre cromossomos homólogos durante a meiose para medir a distância genética entre duas mutações diferentes no mesmo cromossomo.

**5.2 Clonagem e caracterização do DNA**

Estudos detalhados da estrutura e da função de um gene em nível molecular necessitam de grandes quantidades do gene individual purificado. Uma variedade de técnicas, geralmente chamadas de *tecnologia do DNA recombinante*, são usadas na clonagem de DNA, que permite aos pesquisadores preparar grande número de moléculas de DNA idênticas. O DNA recombinante é simplesmente qualquer molécula de DNA composta por sequências derivadas de diferentes fontes.

A chave para a clonagem de um fragmento de DNA de interesse é ligá-lo a uma molécula de DNA-vetor capaz de replicar dentro de uma célula hospedeira. Depois que uma única molécula de DNA recombinante, composta por um vetor mais um fragmento de DNA inserido, é introduzida em uma célula hospedeira, o DNA grande número de moléculas de DNA idênticas. O esquema básico pode ser resumido da seguinte forma:



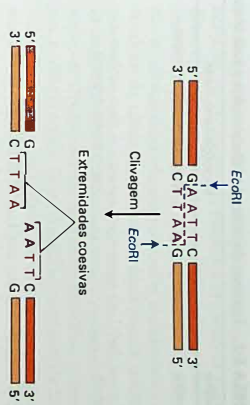
Embora os pesquisadores tenham concebido numerosas variações experimentais, este diagrama de fluxo indica as etapas essenciais na clonagem de DNA. Nesta seção, primeiramente serão descritos métodos para isolamento de uma sequência de DNA específica a partir de um grande conjunto de outras sequências de DNA. Este processo geralmente envolve a clivagem do genoma em fragmentos e a colocação de cada fragmento em um vetor de forma que todo o DNA possa ser propagado como moléculas recombinantes em células hospedeiras independentes. Enquanto muitos tipos diferentes de vetores existem, nossa discussão enfocará principalmente vetores plasmidiais e células hospedeiras de *E. coli*, muito utilizados. Várias técnicas podem ser empregadas para se identificar a sequência de interesse nesta coleção de fragmentos de DNA, conhecida como biblioteca de DNA. Uma vez que um fragmento de DNA específico tenha sido isolado, é geralmente caracterizado pela determinação da sequência exata de nucleotídeos da molécula. A seção é finalizada com uma discussão sobre a reação em cadeia da polimerase (*polymerase chain reaction*, PCR). Esta poderosa e versátil técnica pode ser utilizada de diversas maneiras para gerar grandes quantidades de uma sequência específica e manipular o DNA em laboratório. Os vários usos dos fragmentos de DNA clonados serão discutidos em seções subsequentes.

**Enzimas de restrição e DNA-ligasas permitem a inserção de fragmentos de DNA em vetores de clonagem**

Um dos principais objetivos da clonagem de DNA é a obtenção de pequenas regiões do DNA de um organismo que constituem genes específicos. Além disso, apenas moléculas de DNA relativamente pequenas podem ser clonadas em qualquer um dos vetores disponíveis. Por essas razões, as longas moléculas de DNA que compõem o genoma de um organismo devem ser divididas em fragmentos que possam ser inseridos no vetor de DNA. Dois tipos de enzimas – enzimas de restrição e DNA-ligasas – facilitam a produção das moléculas de DNA recombinantes.

**Clivagem de moléculas de DNA em pequenos fragmentos** Enzimas de restrição são endonucleases produzidas por bactérias que geralmente reconhecem sequências específicas de 4 a 8 pb, chamadas *sítios de restrição*, e então clivam ambas as fitas de DNA no local. Sítios de restrição são geralmente pequenas sequências *palindrômicas*; isto é, a sequência do sítio de restrição é a mesma em cada fita de DNA quando lida na direção 5' – 3' (Figura 5-11).

Para cada enzima de restrição, as bactérias também produzem uma *enzima de modificação*, que protege o próprio DNA de uma bactéria hospedeira da clivagem por meio da modificação do DNA do hospedeiro em, ou próximo a, cada sítio de clivagem em potencial. A enzima de modificação adiciona um grupo metil a uma ou duas bases, geralmente no sítio de restrição. Quando um grupo metil está presente, a endonuclease de restrição é impedida de clivar o DNA. Com as endonucleases de restrição,



**FIGURA 5-11 Clivagem de DNA pela enzima de restrição EcoRI.** Essa enzima de restrição de *E. coli* faz cortes não uniformes na sequência palindrômica específica de 6 pb apresentada, gerando fragmentos com extremidades de fita simples complementares e coesivas. Várias outras enzimas de restrição também produzem fragmentos com extremidades coesivas.

a enzima de metilação forma um sistema de restrição-modificação que protege o DNA do hospedeiro ao mesmo tempo em que destrói DNA estranho (p. ex., DNA de bacteriófago ou DNA adquirido durante transformação) ao clivá-lo em todos os sítios de restrição no DNA.

Várias enzimas de restrição fazem cortes não uniformes nas duas fitas de DNA em seus sítios de reconhecimento, gerando fragmentos que possuem uma cauda de fita simples em ambas as extremidades, chamadas de *extremidades coesivas* (ver Figura 5-11). As caudas nos fragmentos gerados em um dado sítio de restrição são complementares àquelas de todos os outros fragmentos gerados pela mesma enzima de restrição. Sob temperatura ambiente, as regiões de fita simples podem parear transtoriamente com aquelas nos outros fragmentos de DNA gerados a partir da mesma enzima de restrição. Algumas enzimas de restrição, tais como *AluI* e *SmaI*, clivam ambas as fitas de DNA no mesmo ponto do sítio de restrição, gerando fragmentos com extremidades “cegas” nas quais todos os nucleotídeos nas extremidades dos fragmentos estão pareados com nucleotídeos na fita complementar.

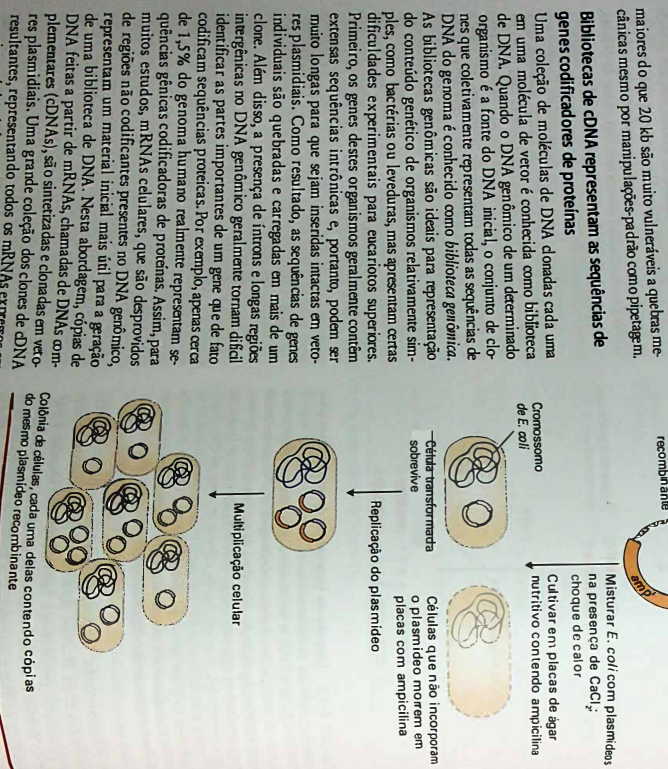
O DNA isolado de um organismo individual possui uma sequência específica, que contém por acaso um conjunto específico de enzimas de restrição. Assim, uma dada enzima de restrição cliva o DNA de uma determinada fonte em um conjunto de fragmentos reprodutíveis. A frequência com a qual uma enzima de restrição cliva o DNA, e, portanto, o tamanho médio dos fragmentos de restrição resultantes, depende muito do tamanho do sítio de reconhecimento. Por exemplo, uma enzima de restrição que reconhece um sítio de 4 pb cliva o DNA em média uma vez a cada 4<sup>4</sup>, ou 256, pares de bases, enquanto uma enzima que reconhece uma sequência de 8 pb cliva o DNA em média uma vez a cada 4<sup>8</sup> pares de bases (cerca de 65 kpb). Enzimas de restrição foram purificadas a partir de várias centenas de diferentes espécies de bactérias, permitindo que moléculas de DNA sejam clivadas em um grande número de diferentes sequências correspondendo aos sítios de reconhecimento das enzimas (Tabela 5-1).





**ANIMAÇÃO DE TÉCNICAS: Clonagem em plasmídeos**

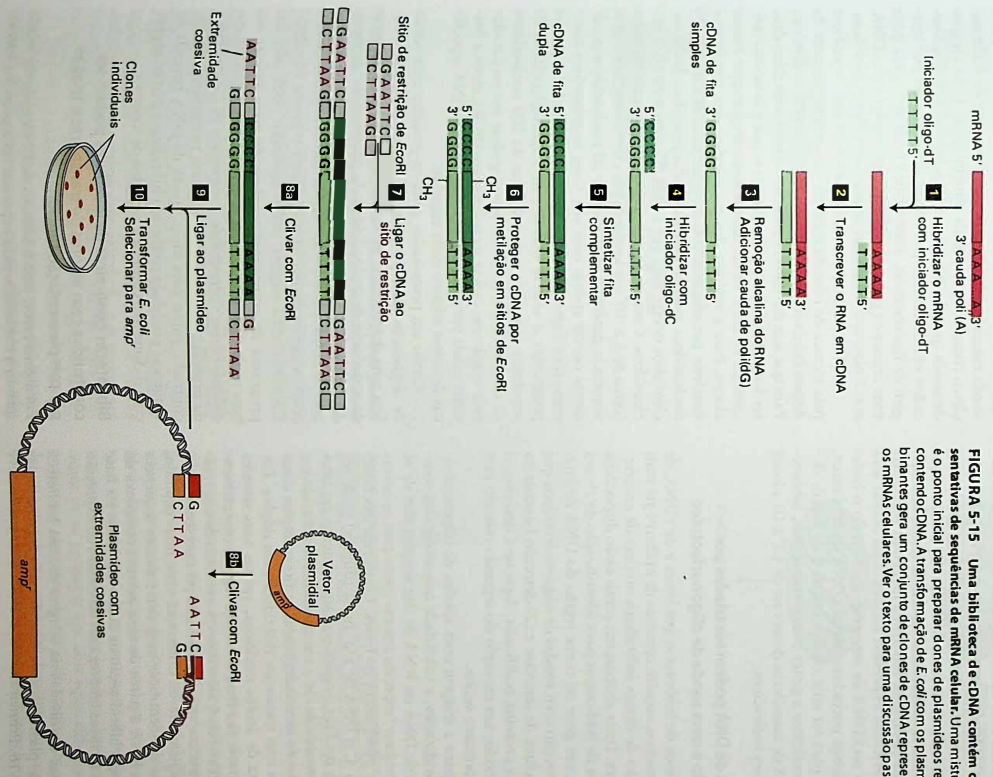
**FIGURA EXPERIMENTAL 5-14. A clonagem de DNA em um vetor plasmidial permite a amplificação de um fragmento de DNA. Um fragmento de DNA a ser clonado é primeiramente inserido em um vetor plasmidial contendo um gene de resistência à ampicilina (*amp<sup>r</sup>*), como aquele mostrado na figura 5-13. Apenas as poucas células transformadas são selecionadas para a amplificação de plasmídeos, sobrevivendo em meio contendo ampicilina. Nos selos de transformação o DNA replica sozinho para as células-filhas, resultando na formação de uma colônia resistente à ampicilina.**



**CDNAs preparados por transcrição reversa de RNAs celulares podem ser donados para gerar bibliotecas de cDNA**

A primeira etapa na preparação de uma biblioteca de cDNA é isolar o mRNA total do tipo celular ou tecido de interesse. Devido à sua cauda de poli(A), os mRNAs são facilmente separados de rRNAs e tRNAs, mas precisam ser ligados a uma matriz de vidro ou de celulose para serem ligados a uma matriz. O procedimento geral para preparar uma biblioteca de cDNA a partir de uma mistura de

No preparo de cDNAs de fita dupla para clonagem, pequenas moléculas de DNA fita dupla contendo o sítio de reconhecimento para uma determinada enzima de restrição são ligadas a ambas as extremidades dos cDNAs, utilizando-se DNA-ligase do bacteriófago T4 (Figura 5-15, etapa 2). Conforme observado, esta ligase pode unir moléculas de DNA de fita dupla com "extremidades cegas" desprovidas de extremidades coesivas. As moléculas



**FIGURA 5-15. Uma biblioteca de cDNA contém cópias representativas de seqüências de mRNA celular. Uma mistura de mRNAs é o ponto inicial para preparar clones de plasmídeos recombinantes contendo cDNA. A transformação de E. coli com os plasmídeos recombinantes gera um conjunto de clones de cDNA representando todos os mRNAs celulares. Ver o texto para uma discussão passo a passo.**

RNAs celulares está delineado na Figura 5-15. A enzima transcriptase reversa, encontrada em retrovírus e utilizada na síntese de uma fita de DNA complementar a cada molécula de mRNA, começando por um iniciador (primeira etapa 1). As moléculas híbridas cDNA-mRNA resultantes são convertidas em vários tipos de moléculas de cDNA de fita dupla correspondentes a todas as moléculas de mRNA da preparação original (etapas 2 a 4). Cada cDNA de fita dupla contém uma região oligo-dC-oligo-dG de fita dupla em uma das extremidades e uma região de fita dupla oligo-dT-oligo-dA na outra. A metilação do cDNA protege-o de subseqüente clivagem por enzimas de restrição (etapa 6).

rados e covalentemente unidos pela DNA-ligase. (Figura 5-15, etapa 9). As moléculas de DNA resultantes são transformadas em células de *E. coli* para gerar clones individuais, e cada clone porta um cDNA derivado de um único mRNA.

Como genes diferentes são transcritos em taxas muito diferentes, os clones de cDNA correspondentes a genes abundantemente transcritos estarão representados muitas vezes em uma biblioteca de cDNA, enquanto cDNAs correspondendo a genes pouco expressos serão raríssimos ou mesmo ausentes. Tal propriedade é vantajosa se um pesquisador estiver interessado em um gene que se transcreve em uma alta taxa em um determinado tipo celular. Nesse caso, uma biblioteca de cDNA preparada a partir de mRNAs expressos naquele tipo celular será enriquecida no cDNA de interesse, facilitando o isolamento de clones portadores deste cDNA da biblioteca. Entretanto, para ter uma chance razoável de incluir clones correspondentes a genes pouco transcritos, bibliotecas de cDNA de mamíferos devem ter de  $10^6$  a  $10^8$  clones recombinantes individuais.

**Bibliotecas de DNA podem ser triadas por hibridização a uma sonda de oligonucleotídeo**

Ambos os tipos de bibliotecas genômica e de cDNA de vários organismos, contêm centenas de milhares até mais de um milhão de clones individuais no caso de eucariotos superiores. Duas abordagens gerais estão disponíveis para triagem de bibliotecas para identificação de clones que portam um gene ou outra região do DNA de interesse: (1) detecção com sondas de oligonucleotídeos que se ligam ao clone de interesse e (2) detecção baseada na expressão da proteína codificada. Aqui será descrito o primeiro método; um exemplo do segundo método está presente na próxima seção.

A base para a triagem com sondas de oligonucleotídeo é a hibridização, a habilidade que moléculas complementares de DNA ou RNA de fita simples têm de se associar (hibridizar) especificamente umas com as outras por meio de pareamento de bases. Conforme discutido no Capítulo 4, o DNA de fita dupla (duplex) pode ser desnaturado em fitas simples por aquecimento em uma solução salina diluída. Se a temperatura for reduzida e a concentração de íons aumentada, as fitas simples complementares irão reassociar (hibridizar) em duplex. Em uma mistura de ácidos nucleicos, apenas fitas simples complementares (ou fitas contendo regiões complementares) irão reassociar, além disso, a extensão de sua reassociação praticamente não é afetada pela presença de fita de DNA ou RNA a partir de uma determinada sequência molecular pela hibridização de ácidos nucleicos e a base gênica.

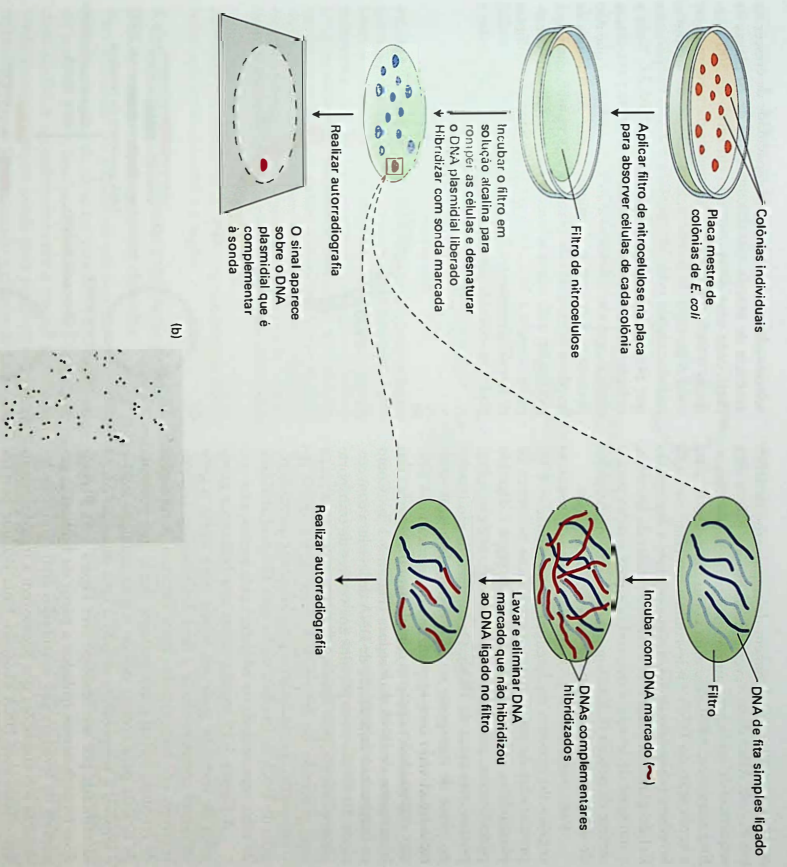
As etapas em olivadas na triagem de uma biblioteca de cDNA de plasmídeos de *E. coli* estão representadas na Figura 5-16. Primeiramente, o DNA a ser triado deve

ser ligado a um suporte sólido. Uma *replica* da placa de Petri contendo um grande número de clones individuais é reproduzida na superfície de uma membrana de nitrocelulose. O DNA na membrana é desnaturado e a membrana, então, incubada em uma solução contendo uma sonda específica para o DNA recombinante que possui o fragmento de interesse, que é marcada radioativamente ou fluorescentemente. Sob condições de hibridização (pH quase neutro, 40 a 65°C, 0,3 a 0,6 M NaCl), a sonda marcada hibridiza com qualquer fita de ácido nucleico complementar ligada à membrana. Todo excesso de sonda que não hibridiza é lavado, e os híbridos marcados são detectados por autoradiografia ou por imagem de fluorescência. Esta técnica pode ser usada para triar ambos os tipos de biblioteca genômica e de cDNA, mas é mais comum no isolamento de cDNAs específicos.

Explicitamente, a identificação de clones específicos pela técnica de hibridização em membrana depende da disponibilidade de sondas complementares radiativas. Para que seja útil como sonda, um oligonucleotídeo deve ser longo o suficiente para que sua sequência ocorra unicamente no clone de interesse, e não em quaisquer outros clones. Para a maioria dos propósitos, essa condição é satisfeita por oligonucleotídeos que contêm cerca de 20 nucleotídeos, porque uma sequência específica de 20 nucleotídeos ocorre uma vez a cada  $4^{20}$  (cerca de  $10^7$ ) nucleotídeos. Como todos os genomas são muito menores (cerca de  $3 \times 10^9$  nucleotídeos para seres humanos), uma sequência genômica específica de 20 nucleotídeos geralmente ocorre apenas uma vez. Com instrumentos automatizados disponíveis atualmente, pesquisadores podem programar a síntese química de oligonucleotídeos de sequência específica com até cerca de 100 nucleotídeos de comprimento. Sondas maiores podem ser preparadas pela reação em cadeia da polimerase (PCR), técnica bastante usada para amplificar sequências de DNA específicas descritas adiante neste capítulo.

Como um pesquisador poderia projetar uma sonda de oligonucleotídeo para identificar um clone que codifica uma determinada proteína? Ajuda se toda ou parte da sequência de aminoácidos da proteína for conhecida. Carças à disponibilidade de sequências genômicas completas para humanos e vários outros organismos tais como o camundongo, *Drosophila*, e o verme cilindrico *Caenorhabditis elegans*, um pesquisador pode usar um programa de computador adequado para buscar em bancos de dados de sequência genômica a sequência codificadora que corresponde a uma sequência de aminoácidos sob investigação. Se um pareamento for encontrado, cria-se uma única sonda de DNA baseada na sequência genômica conhecida hibridizará perfeitamente com o clone que codifica a proteína de interesse.

**Bibliotecas genômicas de leveduras podem ser construídas com vetores de transporte e triadas por complementação funcional**  
Em alguns casos, uma biblioteca de DNA pode ser triada quanto à habilidade para expressar uma proteína funcio-



**FIGURA EXPERIMENTAL 5-16 Bibliotecas de cDNA podem ser triadas com uma sonda radiativa para identificar um clone de interesse.** O surgimento de um ponto na autoradiografia indica a presença de um clone recombinante contendo o DNA complementar à sonda. A posição do ponto na autoradiografia é a imagem espelhada da posição daquele determinado clone na placa de Petri.

nal que complementa uma mutação recessiva. Essa estratégia de triagem seria uma maneira eficiente para isolar um gene clonado que corresponde a uma mutação recessiva interessante identificada em um organismo experimental. Para ilustrar o método, chamado de complementação funcional, será descrito como genes de leveduras clonados em plasmídeos especiais de *E. coli* podem ser introduzidos em células de leveduras mutantes para identificar o gene selvagem deficiente na linhagem mutante. Bibliotecas construídas com o propósito de rastrear sequências genômicas de leveduras geralmente são feitas a partir de DNA genômico em vez de cDNA. Como não

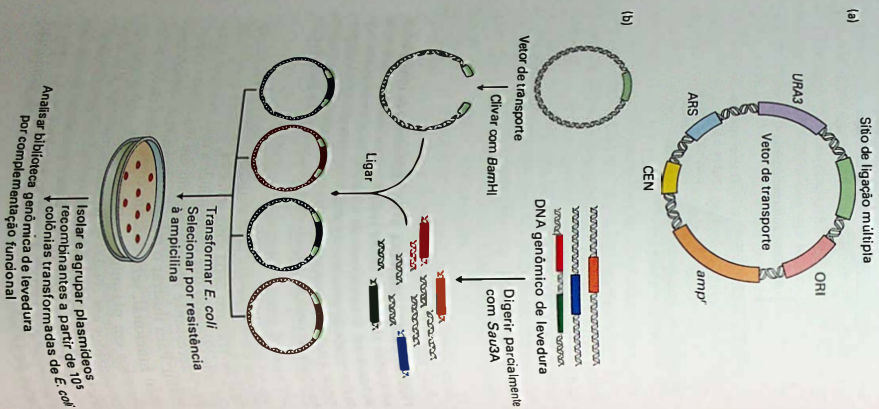
contêm introns múltiplos, os genes de *Saccharomyces* são suficientemente compactos para que a sequência inteira de um gene seja incluída em um fragmento de DNA genômico inserido em um vetor plasmidial. Para se construir um plasmídeo para biblioteca genômica a ser triada para a complementação funcional em células de levedura, o vetor plasmidial deve ser capaz de replicar em ambas as células, de *E. coli* e de levedura. Esse tipo de vetor, capaz de se propagar em dois tipos diferentes de hospedeiros, é chamado de vetor de transporte. A estrutura de um vetor de transporte de levedura típico é mostrada na Figura 5-17a. Este vetor contém os elementos básicos

cos que permitem a clonagem de fragmentos de DNA em *E. coli*. Além disso, o vetor de transporte possui uma sequência de replicação autônoma (*autonomous replicating sequence*, ARS), que atua como uma origem para a replicação do DNA em leveduras, um centromero de levedura (chamado de CEN), que permite a segregação fiel do plasmídeo durante a divisão celular da levedura, e um gene de levedura que codifica uma enzima para a síntese de uracila (URA3), que serve como um marcador de seleção em um mutante de levedura apropriado.

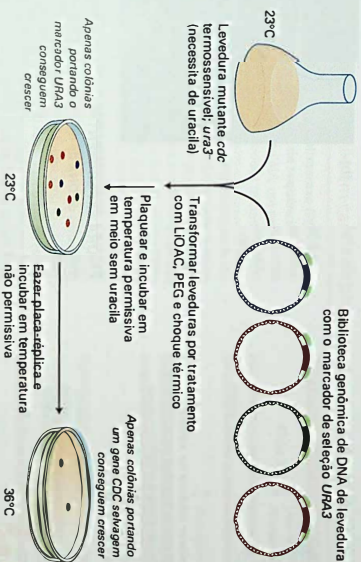
Para aumentar a probabilidade de que todas as regiões do genoma da levedura sejam clonadas e representadas de maneira bem-sucedida na biblioteca de plasmídeos, o DNA genômico via de regra é digerido parcialmente, para gerar fragmentos de restrição sobrepostos com cerca de 10 kb. Os fragmentos são ligados ao vetor de transporte no qual o sítio de ligação múltipla foi derivado com uma enzima de restrição que produz extremidades coesivas complementares àquelas dos fragmentos de DNA de levedura (figura 5-17b). Como os fragmentos de restrição de DNA de levedura com cerca de 10 kb são incorporados aos vetores de transporte aleatoriamente, pelo menos 10<sup>6</sup> colônias de *E. coli*, cada uma contendo um determinado vetor de transporte recombinante, são necessárias para assegurar que cada região do DNA de levedura tenha uma alta probabilidade de estar representada na biblioteca pelo menos uma vez.

A Figura 5-18 mostra como esta biblioteca genômica de levedura pode ser usada para isolar o gene selvagem correspondente a uma das mutações recessivas *cdc* mencionadas neste capítulo. A linhagem de levedura inicial é um mutante duplo que necessita de uracila para crescer devido a uma mutação *ura3* e é sensível à temperatura devido a uma mutação *cdc28* identificada por seu fenótipo (ver Figura 5-6). Plasmídeos recombinantes isolados a partir da biblioteca genômica de levedura são misturados com células de leveduras sob condições que promovem a transformação uma cópia do gene selvagem URA3 carregado pelo plasmídeo, as células de levedura transformadas podem ser

selecionadas com base em sua capacidade de crescer na ausência de uracila. Geralmente, cerca de 20 placas de Petri, cada uma delas contendo em torno de 500 leveduras transformantes, são suficientes para representar todo o genoma da levedura. Essa coleção de leveduras transformantes pode ser mantida a 23°C, temperatura que permite o crescimento do mutante *cdc28*. Toda a coleção nas 20 placas é, então, transferida para placas que permitem o crescimento do mutante *cdc28*. Toda a coleção nas 20 placas a 36°C, temperatura que não permite o crescimento de mutantes *cdc*. As colônias de levedura portadoras de plasmídeos recombinantes que expressam um cópia selvagem do gene *CDC28* conseguem crescer a 36°C. Uma vez identificadas as colônias de levedura resistentes à temperatura, o DNA plasmídico pode ser extraído das células de levedura cultivadas e



**FIGURA EXPERIMENTAL 5-17** Uma biblioteca genômica de levedura pode ser construída em um vetor de transporte plasmídico que pode replicar em levedura de *E. coli*. (a) Componentes de um vetor de transporte plasmídico típico para clonagem de genes de leveduras. A presença de uma origem de replicação de DNA de levedura (ARS) e de um centromero de levedura (CEN) permitem a criação e segregação estáveis em levedura. Indicado também está um marcador de seleção de levedura como URA3, que permite o crescimento de mutantes *ura3* em meio desprovido de uracila. *amp^r* é um marcador de seleção em *E. coli*. (b) Protocolo típico para construção de uma biblioteca genômica de levedura. Digestão parcial de DNA de levedura com *SmaI* e alíquotas para gerar fragmentos de tamanho médio de cerca de 10 kb. O vetor de transporte preparado para receber os fragmentos genômicos por digestão com *BamHI*, que produz as mesmas extremidades coesivas, que *SmaI*. Cada clone transformado de *E. coli* que cresce após seleção para resistência à ampicilina contém um único tipo de fragmento de DNA de levedura.



**FIGURA EXPERIMENTAL 5-18** A triagem de uma biblioteca genômica de levedura por complementação funcional pode identificar clones portadores da forma normal de um gene mutante de levedura. Neste exemplo, um gene *CDC selvagem* é isolado por complementação de um mutante de levedura *cdc*. A linhagem de *Saccharomyces* usada para analisar a biblioteca de levedura possui *ura3* e uma mutação termossensível *cdc*. Essa linhagem mutante é criada e mantida em uma temperatura permissiva (23°C). Plasmídeos recombinantes agrupados preparados de acordo com a figura

analisado por subclonagem e sequenciamento de DNA, temas que serão discutidos a seguir.

**A eletroforese em gel permite a separação entre o DNA do vetor e o de fragmentos clonados**

Para manipular ou sequenciar um fragmento de DNA clonado, às vezes ele deve ser primeiramente separado do vetor de DNA. Isso pode ser feito pela clivagem do clone de DNA recombinante com a mesma enzima de restrição originalmente usada para produzir os vetores recombinantes. O DNA clonado e o vetor de DNA são sublinhados à eletroforese em gel, um método eficaz para separar moléculas de DNA de tamanhos diferentes (Figura 5-19).

Em pH próximo ao neutro, moléculas de DNA possuem uma grande carga negativa e, portanto, movem-se em direção ao eletrodo positivo durante a eletroforese em gel. Como o material do gel restringe a difusão aleatória das moléculas, as que possuem o mesmo tamanho migram juntas como uma banda cuja largura é igual àquela do sulco onde as amostras foram originalmente aplicadas no início da eletroforese. Moléculas menores se movem pela matriz do gel com mais rapidez do que moléculas maiores, de forma que moléculas de tamanhos diferentes migram como bandas distintas. Moléculas de DNA menores, com cerca de 10 a 2.000 nucleotídeos, podem ser separadas eletroforéticamente em géis de *polyacrilamida*, e moléculas maiores, com cerca de 200 nucleotídeos a mais de 20 kb, em géis de *agarose*.

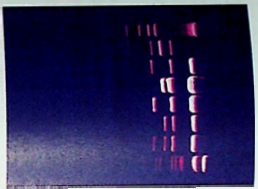
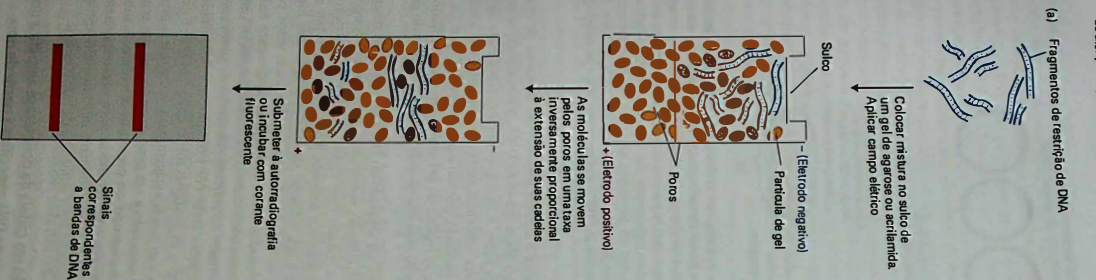
Um método comum para visualização de bandas de DNA separadas em um gel é incubá-lo em uma so-

lução com o corante fluorescente brometo de etídeo. A molécula plana liga-se ao DNA, intercalando-se entre os pares de bases. A ligação concentra o brometo de etídeo no DNA e também aumenta sua fluorescência intrínseca. Como resultado, quando o gel é iluminado por luz ultravioleta, as regiões do gel que contêm DNA fluorescem com muito mais intensidade do que as regiões do gel sem DNA.

Uma vez separado do DNA do vetor, um fragmento de DNA clonado, sobretudo um fragmento longo, é de costume tratado com várias enzimas de restrição para gerar fragmentos menores. Após a separação por eletroforese em gel, todos ou alguns dos fragmentos menores podem ser individualmente ligados a um vetor plasmídico e clonados em *E. coli* pelo procedimento habitual. O processo, conhecido como *subclonagem*, é uma etapa importante no rearranjo de partes de genes em novas configurações úteis. Por exemplo, um pesquisador que quer alterar as condições sob as quais um gene é expresso poderia usar a subclonagem para substituir o promotor normal associado a um gene clonado por um segmento de DNA contendo um promotor diferente. A subclonagem também pode ser usada para obter fragmentos de DNA clonados que possuam um tamanho apropriado para determinar sua sequência de nucleotídeos.

**A reação em cadeia da polimerase amplifica uma sequência de DNA específica a partir de uma mistura complexa**

Se as sequências de nucleotídeos das extremidades de uma determinada região de DNA forem conhecidas, o



**FIGURA EXPERIMENTAL 5-19** A eletroforese em gel separa moléculas de DNA de diferentes tamanhos. (a) Um gel é preparado vindo-se uma solução contendo agarose ou acrílica entre duas placas de vidro separadas por alguns milímetros. A medida que a agarose solidifica ou a acrílica polimeriza em poliacrílamida, forma-se uma matriz de gel (pilipes, laranja). Composta por emulsões de logras, corantes de polímeros. As dimensões dos canais comunitários, ou poros, dependem da concentração de agarose ou acrílica usada para fazer o gel. As bandas separadas podem ser visualizadas por autoradiografia (se os fragmentos estiverem marcados radioativamente) ou pela difusão de um corante fluorescente (p. ex., brometo de etídio) que se liga ao DNA. (b) Fotografia de um gel corado com brometo de etídio (EtBr). O EtBr liga-se ao DNA e fluoresce sob luz UV. As bandas nas caméias das extremidades esquerda e direita são conhecidas como marcadores de DNA – fragmentos de DNA de tamanho conhecido que servem de referência para determinar o tamanho dos fragmentos de DNA da amostra experimental. (Parte (b) Science Photo Library).

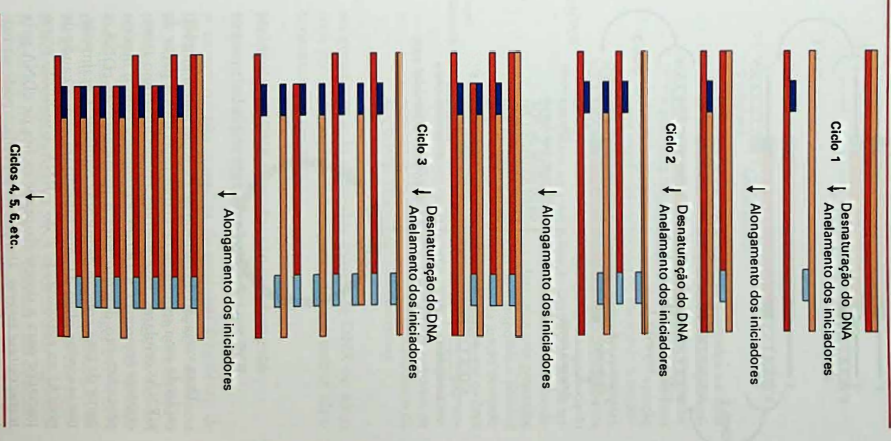
fragmento entre eles poderá ser amplificado diretamente pela reação em cadeia da polimerase (*polymerase chain reaction*, PCR). Serão descritos a técnica de PCR básica e três situações nas quais é usada.

A PCR depende da habilidade de desnaturar aleatoriamente moléculas de DNA de fita dupla e hibridizar fitas simples complementares de modo controlado. Conforme delineado na Figura 5-20, uma PCR típica começa com a desnaturação de uma amostra de DNA em fitas simples por calor, a 95°C. Depois, dois oligonucleotídeos sintéticos complementares à extremidade 3' do segmento de DNA-alvo são adicionados em grande excesso ao DNA desnaturado, e a temperatura é reduzida para 50 a 60°C. Os oligonucleotídeos específicos, que estão em concentração muito alta, hibridizam com suas sequências complementares na amostra de DNA, enquanto as longas fitas de amostra permanecem separadas por conta de sua baixa concentração. Os oligonucleotídeos hibridizados servem, então, como iniciadores para a síntese da cadeia de DNA na presença de desoxinucleotídeos (dNTPs) e uma DNA-polimerase termorresistente como aquela de *Thermus aquaticus* (uma bactéria que vive em águas termais). Essa enzima, chamada de *Taq-polimerase*, pode permanecer ativa mesmo depois de aquecida a 95°C e estender a síntese está completa, toda a mistura é aquecida a 95°C para desnaturar o DNA recém-formado. Depois que a temperatura é reduzida novamente, ocorre outro ciclo de

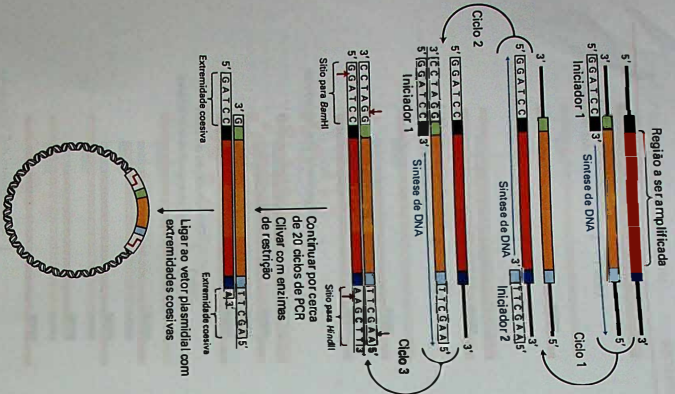
**ANIMAÇÃO DA TÉCNICA:** Reação em cadeia da polimerase

**FIGURA 5-20** A reação em cadeia da polimerase (PCR) é baseada usada para amplificar regiões do DNA de sequências conhecidas. Para amplificar uma região específica do DNA, um pesquisador irá sintetizar quimicamente dois oligonucleotídeos, iniciadores diferentes complementares a sequências de aproximadamente 18 bases flanqueando a região de interesse (designadas como barras azuis claras e azuis escuras). A reação completa é composta por uma mistura complexa de DNA de fita dupla (geralmente DNA genômico contendo a sequência-alvo de interesse), um excesso estequiométrico de ambos os iniciadores, os quatro desoxinucleotídeos trifosfato e uma DNA-polimerase termorresistente conhecida como *Taq-polimerase*. Durante cada ciclo de PCR, a mistura da reação é primeiramente aquecida para separar as fitas e resfriada para permitir a ligação dos iniciadores às sequências complementares flanqueando a região a ser amplificada. A *Taq-polimerase* estende cada iniciador a partir de sua extremidade 3', gerando fitas novas recém-sintetizadas que se estendem na direção 3' para a extremidade 5' da fita molde. Durante o terceiro ciclo, duas moléculas de DNA de fita dupla são geradas com tamanho igual à sequência da região a ser amplificada. Em cada ciclo sucessivo, o segmento-alvo, que irá ativar os iniciadores, é duplicado e eventualmente excede em número todos os outros segmentos de DNA na mistura da reação. Ciclos de PCR sucessivos podem ser automatizados pela criação da reação por intervalos cronometrados em alta temperatura para desnaturar o DNA e em determinadas temperaturas mais baixas para as etapas de anelamento e alongamento do ciclo. Uma reação com 20 ciclos irá amplificar a sequência-alvo específica 1 milhão de vezes.

síntese, pois o excesso de iniciadores ainda está presente. Reptidos ciclos de desnaturação (aquecimento) seguidos por hibridização e síntese (resfriamento) amplificar rapidamente a sequência de interesse. A cada ciclo, o número de cópias da sequência entre os sítios dos iniciadores é duplicado; portanto, a sequência de interesse aumenta exponencialmente – cerca de um milhão de vezes após 20 ciclos –, enquanto todas as outras sequências da amostra original de DNA permanecem não amplificadas. **Isolamento direto de um segmento específico de DNA genômico.** Para organismos nos quais todo o genoma, ou a maior parte dele, tenha sido sequenciado, a amplificação por PCR, conseguida com DNA genômico total, tende a ser a maneira mais fácil de se obter uma região específica do DNA de interesse para clonagem. Nessa aplicação, os dois oligonucleotídeos iniciadores são projetados para hibridizar em sequências que flanqueiam a região genômica de interesse e para incluir sequências que são reconhecidas por enzimas de restrição específicas (Figura 5-21). Após a amplificação da sequência-alvo desafiada por cerca de 20 ciclos de PCR, o DNA com as enzimas de restrição apropriadas produzidas em extremidades coesivas que permitem a ligação pela mesma enzima de restrição no sítio de ligação múltipla. Os plasmídeos recombinantes resultantes, todos carregando o fragmento de DNA genômico idêntico, podem, então, ser clonados em células de *E. coli*. Com determinados refinamentos da PCR, até mesmo fragmentos de DNA maiores do que 10 kb podem ser assim amplificados e clonados.



Esse método não envolve a clonagem de grandes números de fragmentos de restrição derivados de DNA genômico e sua subsequente triagem para identificar o fragmento de interesse específico. Na realidade, o método de PCR inverte a abordagem tradicional e evita seus aspectos mais entediantes. O método de PCR serve para isolar sequências genéticas a serem manipuladas em uma variedade de maneiras úteis descritas posteriormente. Além disso, o método de PCR pode ser usado para isolar sequências genéticas de organismos mutantes a fim de determinar como eles diferem do tipo selvagem.



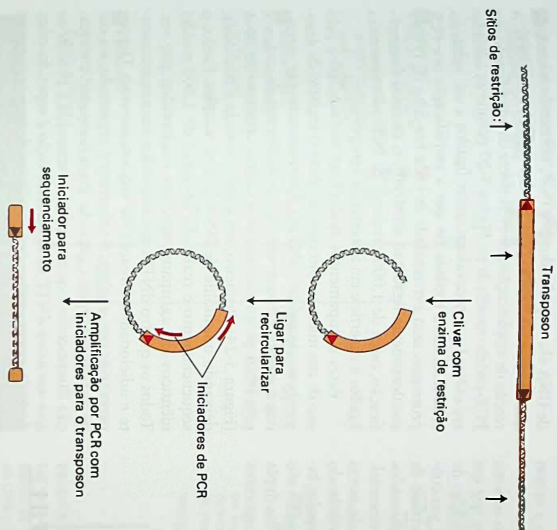
**FIGURA EXPERIMENTAL 5-21** Uma região-alvo específica do DNA genômico total pode ser amplificada por PCR para uso em clonagem. Cada iniciador para PCR é complementar a uma extremidade da sequência-alvo e inclui a sequência de reconhecimento para uma enzima de restrição que não possui um sítio dentro da sequência-alvo. Nesse exemplo, o iniciador 1 contém uma sequência de reconhecimento para a enzima *Hind*III, enquanto o iniciador 2 possui uma sequência para a enzima *Bam*HI, enquanto o iniciador 3 possui uma sequência para a enzima *Eco*RI. Após a amplificação, a região-alvo é tratada com as enzimas de restrição apropriadas, gerando fragmentos com extremidades coesivas. Estes podem ser incorporados em vetores plasmidiais complementares e clonados em *E. coli* por meio do procedimento habitual (ver Figura 5-13).

quantidade de mRNA prosequer. Pela extrapolação dessas quantidades, uma estimativa da quantidade inicial de sequências de mRNA pode ser obtida. A reação de RT-PCR quantitativa realizada em tecidos ou em organismos inteiros utilizando iniciadores direcionados a genes de interesse fornece uma das metodologias mais precisas para acompanhamento de alterações na expressão gênica.

**Preparação de sondas** Foi previamente mencionado que sondas de oligonucleotídeos para ensaios de hibridização podem ser quimicamente sintetizadas. A preparação das sondas por meio de amplificação por PCR requer a síntese química de apenas dois iniciadores relativamente pequenos, correspondendo às duas extremidades da sequência-alvo. A amostra inicial para a amplificação por PCR da sequência-alvo pode ser uma preparação de DNA genômico ou de cDNA sintetizado a partir do mRNA celular total. Para gerar um produto radioativamente marcado a partir de PCR, dNTPs marcados com <sup>32</sup>P são incluídos durante os vários últimos ciclos de amplificação ou um produto fluorescente pode ser obtido pelo uso de dNTPs marcados fluorescentemente durante os últimos ciclos de amplificação. Como as sondas preparadas por PCR são relativamente longas e possuem muitos nucleotídeos radiativos ou fluorescentes incorporados a elas, geralmente produzem um sinal mais forte e mais específico do que sondas sintetizadas quimicamente.

**Marcação de genes por mutações insercionais** Outra aplicação útil da PCR é a amplificação de um gene "marcado" a partir do DNA genômico de uma linhagem identificável. Essa abordagem consiste em um método para identificar genes associados com uma determinada característica, mas simples do que a triagem de uma biblioteca por complementação funcional (ver Figura 5-18).

A chave para esse uso da PCR é a habilidade para produzir mutações por inserção de uma sequência conhecida de DNA no genoma de um organismo experimental. As mutações insercionais podem ser geradas pelo uso de elementos móveis de DNA, que se movem (ou transgrem) de um sítio cromossômico para outro. Conforme discutido em mais detalhe no Capítulo 6, essas sequências de DNA ocorrem naturalmente nos genomas da maioria dos organismos e podem originar mutações de perda de função quando se transpõem para regiões codificadoras de proteínas.



**FIGURA 5-22** A sequência genômica no sítio de inserção de um transposon é revelada por amplificação por PCR e sequenciamento. Para obter a sequência no sítio de inserção de um transposon elemento P, é necessário amplificar por PCR a junção entre sequências conhecidas do transposon e sequências cromossômicas flanking desconhecidas. Um método para se fazer isso é clivar DNA genômico com uma enzima de restrição que cliva a sequência do transposon uma vez. A ligação dos fragmentos de restrição resultantes produzirá moléculas de DNA circulares. Utilizando iniciadores adequadamente projetados que parelham a sequência do transposon, é possível amplificar por PCR o fragmento da junção desejado. Finalmente, uma reação de sequenciamento de DNA (ver Figuras 5-23 e 5-24) é realizada usando o fragmento amplificado por PCR como molde e um oligonucleotídeo iniciador que parelha as sequências próximas da extremidade do transposon para obter a sequência da junção entre o transposon e o cromossomo.

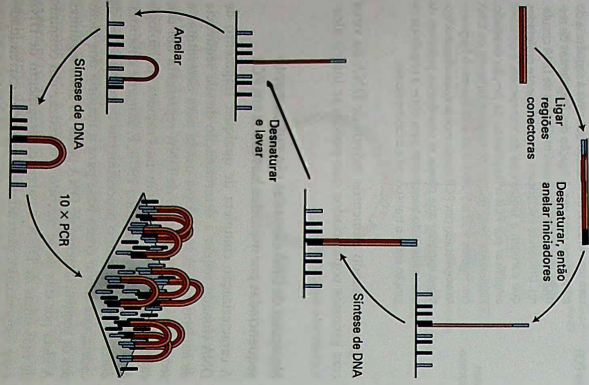
ser geradas usando elementos móveis de DNA ou vírus com genomas sequenciados que podem se inserir aleatoriamente no genoma.

**Moléculas de DNA clonadas são rapidamente sequenciadas por métodos baseados em PCR**

A caracterização completa de qualquer fragmento de DNA clonado requer a determinação de sua sequência de nucleotídeos. A tecnologia usada para determinar a sequência de um segmento de DNA representa um dos campos que se desenvolveu com mais rapidez em biologia molecular. No final dos anos 1970, F. Sanger e colaboradores desenvolveram o procedimento de terminação de cadeia, que serviu como base para a maioria dos métodos de sequenciamento de DNA pelos 30 anos seguintes. A ideia por trás desse método é sintetizar um conjunto de fitas-filhas de DNA a partir do fragmento de DNA a ser sequenciado, marcadas em uma das extremidades e terminadas em um dos quatro nucleotídeos. A separação das fitas-filhas terminadas por eletroforese em gel, que pode separar fitas que diferem em tamanho por um nucleotídeo, pode revelar o tamanho de todas as fitas, terminando em G, A, T ou C. A partir dessas coleções de fitas de diferentes tamanhos, a sequência de nucleotídeos do fragmento de DNA original pode ser estabelecida. O método de Sanger passou por vários aperfeiçoamentos e agora pode ser completamente automatizado, porém, como cada nova sequência de DNA requer uma reação de sequenciamento individual, a taxa global na qual novas sequências de DNA podem ser produzidas pelo método...

todo é limitada pelo número total de reações a serem realizadas ao mesmo tempo.

Um avanço na tecnologia de sequenciamento ocorreu quando foram concebidos métodos que permitem que um único equipamento de sequenciamento realize bilhões de reações de sequenciamento simultaneamente por sua concentração em pequenos agrupamentos na superfície de um substrato sólido. Desde 2007, quando os chamados sequenciadores de *high-throughput* se tornaram comercialmente disponíveis, a capacidade para produção de novas sequências aumentou bastante e desde então tem dobrado a cada ano. Em um método de sequenciamento popular, bilhões de fiéis de DNA diferentes a serem sequenciadas são preparadas pela ligação de regiões conectoras de fra dupla em suas extremidades (Figura 5-23). Depois, os fragmentos

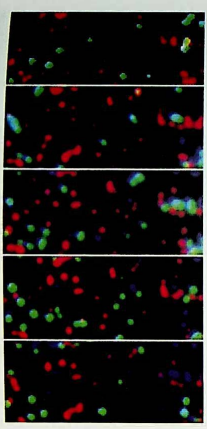
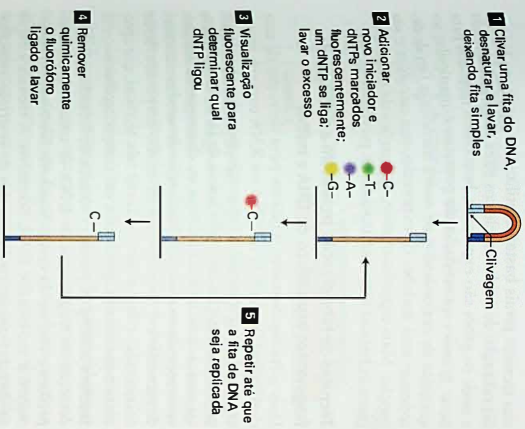


**FIGURA EXPERIMENTAL 5-23** Geração de agrupamentos de moléculas de DNA idênticas ligadas a um suporte sólido. Uma grande coleção de moléculas de DNA a ser sequenciada é ligada a regiões conectoras de fra dupla, que se unem a cada extremidade do componente as sequências das regiões conectoras que estão covalentemente ligadas a um substrato sólido. Dez ciclos de amplificação geram cerca de 1.000 cópias idênticas do fragmento de DNA localizado em um pequeno agrupamento, o qual está ligado por ambas as extremidades ao substrato sólido. Essas regiões são otimizadas para produzir cerca de  $3 \times 10^6$  agrupamentos distintos, não sobrepostos e prontos para serem sequenciados.

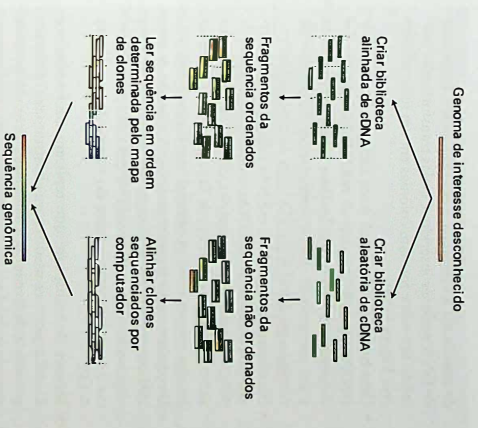
de DNA são amplificados por PCR, utilizando iniciadores que são pareados às sequências das regiões conectoras. A reação de amplificação por PCR difere da amplificação por PCR padrão mostrada na Figura 5-20 porque os iniciadores utilizados são covalentemente ligados a um substrato sólido. Assim, a medida em que a amplificação por PCR prossegue, uma extremidade de cada fra-filha de DNA é covalentemente ligada ao substrato, e, ao final da amplificação, cerca de 1.000 produtos de PCR idênticos estão ligados à superfície em um estreito agrupamento.

Esses agrupamentos podem ser sequenciados pelo uso de um microscópio especial para o registro de díxirbonucleotídeos (dNTPs) fluorescentemente marcadas à medida em que são incorporados um de cada vez pela DNA-polimerase em uma cadeia crescente de DNA (Figura 5-24). Primeiro, uma fra é escolhida e eliminada, deixando um molde de DNA de fra simples. Então, o sequenciamento é realizado em cerca de 1.000 moldes idênticos em agrupamentos, um nucleotídeo por vez. Todos os quatro dNTPs são marcados fluorescentemente e adicionados à reação de sequenciamento. Depois, permite-se que anelam o substrato e registrando, e a cor de cada agrupamento, gravada. Posteriormente, a marcação fluorescente é quimicamente removida e permite-se que um novo dNTP se ligue. Esse ciclo é repetido cerca de 100 vezes, resultando em bilhões de sequências com aproximadamente 100 nucleotídeos.

Para sequenciar uma longa região contínua de DNA genômico ou mesmo o genoma inteiro de um organismo, pesquisadores geralmente empregam uma das estratégias ilustradas na Figura 5-25. O primeiro método requer o isolamento de uma coleção de fragmentos de DNA clonados cujas sequências são sobrepostas. Uma vez determinada a sequência de um desses fragmentos, oligonucleotídeos baseados em um dos fragmentos podem ser quimicamente sintetizados para uso como iniciadores no sequenciamento de fragmentos sobrepostos adjacentes. Assim, a sequência de uma longa região de DNA é determinada de forma crescente pelo sequenciamento dos fragmentos de DNA clonados sobrepostos que a compõem. Um segundo método, chamado de sequenciamento *shotgun*, pula a etapa tra balhosa do isolamento de uma coleção ordenada de segmentos de DNA que abrangem todo o genoma. Este método envolve simplesmente o sequenciamento de clones aleatórios de uma biblioteca genômica. Um número total de clones é escolhido para sequenciamento de forma que em média cada segmento do genoma seja sequenciado cerca de 10 vezes. O grau de cobertura assegura que cada segmento do genoma seja sequenciado mais de uma vez. A sequência do genoma inteiro é, então, montada utilizando um algoritmo de computador que alinha todas as seqüências, usando suas regiões de sobreposição. O seqüenciamento *shotgun* é o método mais eficiente e prático para o seqüenciamento de longas regiões de DNA, e a maioria dos genomas, inclusive o genoma humano, foi sequenciada por tal método.



**FIGURA EXPERIMENTAL 5-24** Uso de trifosfato de desoxirribonucleotídeos marcados com fluorescência para determinar a seqüência. A reação começa pela clivagem de uma fra do DNA agrupado. Após a desnaturação, uma única fra do DNA permanece ligada à célula de fluxo. Um oligodesoxinucleotídeo sintético e utilizado como iniciador para a reação de polimerização que contém dNTPs, cada um deles marcado fluorescentemente com uma cor diferente. A marcação fluorescente é projetada para bloquear o grupo 3' OH do dNTP, o que previne a continuação do alongamento do DNA, que ele é incorporado. Como a DNA-polimerase já incorporar o mesmo dNTP fluorescente em cada uma das cerca de 1.000 cópias de DNA de um agrupamento, o agrupamento inteiro será marcado de maneira uniforme com a mesma cor, que poderá ser visualizada em um microscópio especial. Uma vez visualizados todos os agrupamentos, as marcações fluorescentes são removidas por uma reação química que deixa uma nova extremidade do iniciador disponível para o próximo ciclo de incorporação de dNTP marcado com fluorescência. Uma reação de sequenciamento típica pode realizar 100 ciclos de polimerização, possibilitando determinar 100 bases de seqüência para cada agrupamento. Assim, uma reação de seqüenciamento total desse tipo pode gerar informação sobre cerca de  $3 \times 10^6$  bases de seqüência em aproximadamente dois dias. (Usado de Ardecka Loefer em OpenStaxWare ([http://openstaxware.org/Wiki/User:Ardecka\\_Loefer11](http://openstaxware.org/Wiki/User:Ardecka_Loefer11))).



**FIGURA EXPERIMENTAL 5-25** Duas estratégias para montar seqüências genômicas inteiras. Um método (*à esquerda*) depende do isolamento e da formação de um conjunto de segmentos de DNA clonados que abrangem o genoma. Isso pode ser feito pelo pareamento de segmentos clonados por hibridização ou por alinhamento de mapas de sítio de restrição. A seqüência de DNA dos clones ordenado pode, então, ser montada em uma seqüência genômica completa. O método alternativo (*à direita*) depende da facilidade relativa do sequenciamento de DNA automatizado e ignora a etapa tra balhosa de ordenação da biblioteca. Sequenciando clones aleatórios da biblioteca em quantidade suficiente de forma que cada segmento do genoma seja representado de 3 a 10 vezes, é possível reconstruir a seqüência genômica por alinhamento computacional do grande número de fragmentos de seqüência.

**CONCEITOS-CHAVE da Seção 5.2**

- **Clonagem e caracterização do DNA**
- Na clonagem de DNA, moléculas recombinantes são formadas *in vitro* pela inserção de fragmentos de DNA em moléculas de vetores de DNA. As moléculas de DNA recombinantes são, então, introduzidas em células hospedeiras, nas quais replicam, produzindo grandes números de moléculas de DNA recombinantes.
- Enzimas de restrição (endonucleases) geralmente cortam o DNA em seqüências palindrômicas específicas de 4 a 8 pb, produzindo fragmentos definidos que geralmente têm caudas de fra simples complementares a si mesmas (extremidades coesivas).
- Dois fragmentos de restrição com extremidades complementares podem ser unidos com DNA-ligase para formar uma molécula de DNA recombinante (ver Figura 5-12).
- Vetores de clonagem de *E. coli* são pequenas moléculas circulares de DNA (plasmídeos) que incluem três regiões funcionais: uma origem de replicação, um gene

- Uma biblioteca de cDNA é um conjunto de clones de DNA preparado a partir de mRNAs isolados de um determinado tipo de tecido. Uma biblioteca genômica é um conjunto de clones portadores do genoma inteiro, restando produzidos por clivagem do genoma inteiro.
- Na clonagem de cDNA, mRNAs expressas são reversamente transcritas em DNAs complementares, ou cDNAs. Por uma série de reações, cDNAs de fita simples são convertidos em DNAs de fita dupla, que podem ser ligados em um vetor plasmídico (ver Figura 5-15).
- Um determinado fragmento de DNA clonado de uma biblioteca pode ser detectado por hibridização a um oligonucleotídeo radioativamente marcado cuja sequência é complementar à parte do fragmento (ver Figura 5-16).
- Vetores de transporte que replicam em leveduras e *E. coli* podem ser usados para construir uma biblioteca genômica de leveduras. Genes específicos são isolados por sua habilidade de complementar os genes mutantes correspondentes em células de leveduras (ver Figura 5-17).
- Longos fragmentos de DNA clonados podem ser clivados com enzimas de restrição, produzindo fragmentos menores que são então separados por eletroforese em gel e subclonados em vetores plasmídicos antes do sequenciamento ou manipulação experimental.
- A reação em cadeia da polimerase (PCR) permite a amplificação exponencial de um fragmento específico de DNA a partir de apenas uma molécula inicial de DNA-molde se a sequência flanqueadora da região do DNA a ser amplificada for conhecida (ver Figura 5-20).
- A PCR é um método muito versátil que pode ser empregado para amplificar uma sequência de DNA genômico específica, um cDNA, ou uma sequência da junção entre um elemento transponível e as sequências cromossômicas flanqueadoras.
- Fragmentos de DNA com até cerca de 100 nucleotídeos são sequenciados pela geração de aglomerados de moléculas idênticas por PCR e pelo registro de nucleotídeos precursoros marcados fluorescentemente incorporados pela DNA-polimerase (ver Figuras 5-23 e 5-24).
- Sequências do genoma inteiro podem ser montadas a partir de sequências de um grande número de clones sobrepostos de uma biblioteca genômica (ver Figura 5-25).

**5.3 Uso de fragmentos de DNA clonados para estudo da expressão gênica**

Na última seção, foram descritos as técnicas básicas de uso da tecnologia de DNA recombinante para isolamento de clones de DNA específicos, e marcadas pelas quais os clones podem ser mais bem caracterizados. Agora será considerado como o clone de DNA isolado pode ser usado para estudo da expressão gênica. Serão discutidas várias

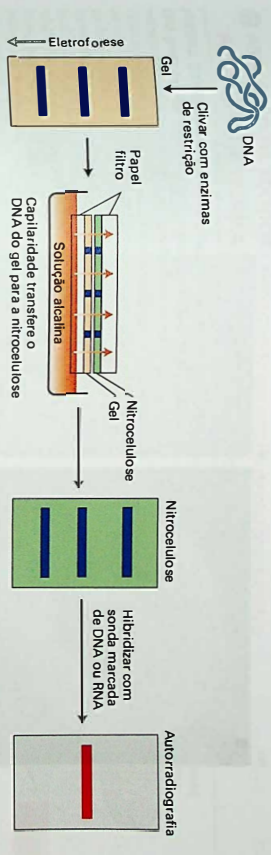
técnicas gerais bastante utilizadas que dependem da hibridização de ácidos nucleicos para esclarecer quando e onde os genes são expressos, bem como métodos para gerar grandes quantidades de proteína e manipular sequências de aminoácidos de outra maneira a fim de determinar seus padrões de expressão, sua estrutura e sua função. Mais especificamente, aplicações de todas essas técnicas básicas são examinadas nas próximas seções.

**Técnicas de hibridização permitem a detecção de fragmentos específicos de DNA e mRNA**

Dois métodos bastante sensíveis para detecção de uma determinada sequência de DNA ou RNA em uma mistura complexa combinam a separação por eletroforese em gel e a hibridização com uma sonda de DNA complementar marcada de forma radiativa ou fluorescente. Um terceiro método envolve a hibridização de sondas marcadas diretamente em uma amostra de tecido preparada. Serão encontradas referências a todas as três técnicas, que apresentam numerosas aplicações, em outros capítulos.

**Southern blotting** A primeira técnica de hibridização para detecção de fragmentos de DNA de uma sequência específica é conhecida como **Southern blotting**, em homenagem a seu criador E. M. Southern. A técnica é capaz de detectar um único fragmento de restrição específico em uma mistura altamente complexa de fragmentos produzidos por clivagem do genoma humano inteiro com uma enzima de restrição. Quando a mistura complexa é submetida à eletroforese em gel, tantos fragmentos de tamanho semelhante estão presentes, que não é possível distinguir nenhum fragmento de DNA como uma banda única no gel. Mesmo assim, é possível identificar um determinado fragmento migrando como uma banda no gel por sua capacidade de hibridização a uma sonda de DNA específica. Para realizar isso, os fragmentos de restrição presentes no gel são desnaturados com álcali e transferidos para um filtro de nitrocelulose ou uma membrana de nylon por transferência (*blotting*) (Figura 5-26). Esse processo preserva a distribuição dos fragmentos no gel, criando uma réplica do gel no filtro. O *blot* é usado porque as sondas não se difundem de imediato no gel original. O filtro é então incubado sob condições de hibridização com uma sonda de DNA específica marcada, geralmente produzida a partir de um fragmento de restrição clonado. O fragmento de restrição de DNA que for complementar à sonda hibridiza, e sua localização no filtro pode ser revelada por autoradiografia para uma sonda radioativamente marcada ou por registro de imagem fluorescente para uma sonda marcada fluorescentemente. Embora a PCR seja mais usada na detecção da presença de uma determinada sequência em uma mistura complexa, o *Southern blotting* é ainda útil na reconstrução da relação entre sequências genômicas que estão muito distantes para serem amplificadas por PCR em uma única reação.

**Northern blotting** Uma das formas mais básicas de se caracterizar um gene clonado é determinar quando e onde o gene é expresso em um organismo. A expressão de um



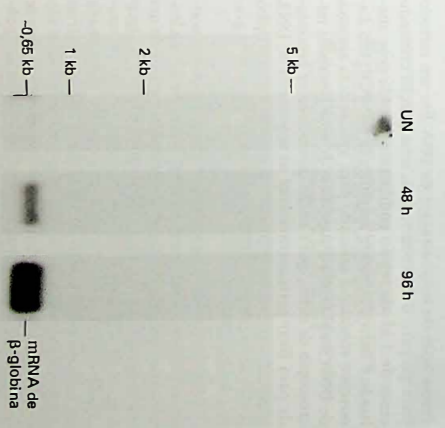
**FIGURA EXPERIMENTAL 5-26** A técnica de **Southern blotting** pode detectar um fragmento de DNA específico em uma mistura completa de fragmentos de restrição. O diagrama ilustra três fragmentos de restrição diferentes no gel, mas o procedimento pode ser aplicado em uma mistura de milhares de fragmentos de DNA. Apenas o determinado gene pode ser acompanhado pelo ensaio do mRNA correspondente por meio de **Northern blotting**, realizado, por um jogo de pláivas, em homenagem ao método relacionado de **Southern blotting**. Uma amostra de mRNA, geralmente o RNA celular total, é desnaturada pelo tratamento com um agente como o formaldéido, que rompe as ligações de hidrogênio entre os pares de bases, assegurando que todas as moléculas de RNA estejam em uma conformação linear. Os RNAs individuais são separados pelo tamanho por eletroforese em gel e transferidos para um filtro de nitrocelulose ao qual os RNAs escudidos e desnaturados são aderidos. Assim como no **Southern blotting**, o filtro é exposto a uma sonda de DNA marcada complementar a o gene de interesse; finalmente, o filtro marcado é submetido à autoradiografia. Como a quantidade de um RNA específico em uma amostra pode ser estimada por um **Northern blot**, o procedimento é bastante utilizado para comparar as quantidades de um determinado mRNA em células sob diferentes condições (Figura 5-27).

**Hibridização *in situ*** O **Northern blotting** necessita da extração de mRNA de uma célula ou de uma mistura de células, o que implica na remoção das células de seu local normal em um organismo ou tecido. Como resultado, a localização da célula e sua relação com suas vizinhas são perdidas. Para reter as informações posicionais em estudos de expressão gênica precisos, um tecido inteiro ou seccionado, ou até mesmo um embrião inteiro permeabilizado, pode ser submetido à hibridização *in situ* para detectar o mRNA codificado por um determinado gene. Esta técnica permite que a transferência genética seja monitorada no tempo e no espaço (Figura 5-28).

**Microrranjos de DNA podem ser utilizados para se avaliar a expressão de vários genes ao mesmo tempo**

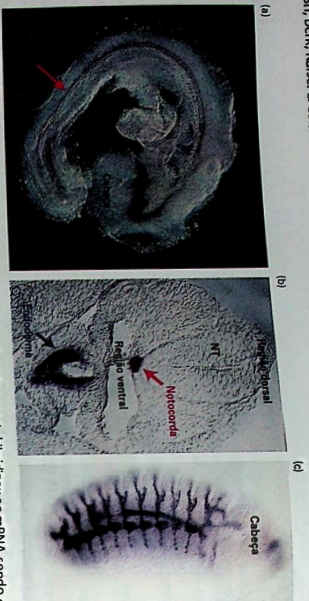
O monitoramento da expressão de milhares de genes simultaneamente é possível com a análise de microrranjos de DNA, outra técnica baseada no conceito de hibridização de ácidos nucleicos. Um microrranjo de DNA

consiste em um arranjo organizado de milhares de sequências genéticas específicas individuais agrupadas e ligadas à superfície de uma lâmina de microscópio de vidro. Acolando a análise de microrranjos com os resultados dos projetos de sequenciamento genômico, os pesquisadores podem analisar os padrões globais de expressão gênica de um organismo durante respostas fisiológicas específicas ou processos de desenvolvimento.



**FIGURA EXPERIMENTAL 5-27** Análise de **Northern blot** revela expressão aumentada de mRNA de  $\beta$ -globina em células de eritroleucemia diferenciadas. O mRNA total em extratos de células de eritroleucemia que foram cultivadas, mas não induzidas, e em células induzidas para parar o crescimento e diferenciadas por 48 ou 96 horas foi analisado por **Northern blotting** para o mRNA de  $\beta$ -globina. A densidade de uma banda é proporcional à quantidade de mRNA presente. O mRNA de  $\beta$ -globina é diferencialmente detectável em células não induzidas (caneta UV), mas aumenta em mais de 1.000 vezes após 96 horas de diferenciação induzida. (Cortesias de L. Kole.)





**FIGURA EXPERIMENTAL 5-28** A hibridização *in situ* pode detectar a atividade de genes específicos em embriões inteiros e seccionados. O espécime é permeabilizado por tratamento com detergente e uma proteína para expor o mRNA à sonda. Uma sonda de DNA ou RNA, específica para o mRNA de interesse, é feita com análogos de nucleotídeos contendo grupos químicos que podem ser reconhecidos por anticorpos. Depois que o espécime permeabilizado foi incubado com a sonda sob condições que promovem hibridização o excesso de sonda é removido por uma série de lavagens. O espécime é então incubado com uma solução contendo um anticorpo que se liga à sonda. Este anticorpo é covalentemente ligado a uma enzima reporter (p. ex., peroxidase ou fosfatase alcalina) que gera um produto detectável. Após a remoção do excesso de anticorpo, o substrato para a enzima reporter é adicionado. Um precipitado colorido é

formado onde a sonda hibridizou ao mRNA sendo detectado. (a) Um embrião inteiro de camundongo com cerca de 10 dias de desenvolvimento sondado para mRNA de *Sonic hedgehog*. O corante marca a notocorda (seu vermelho), um filamento de mesodermia que percorre a futura linha dorsal. (b) Uma seção de um embrião de camundongo semelhante a que da panela (a). O eixo dorsoventral do tubo neural (NT) pode ser visto, com a notocorda expressando *Sonic hedgehog* (NT) (pode ser visto, com a notocorda (seu azul) ainda bastante visível). (c) Um embrião inteiro de *Drosophila* sondado para um mRNA produzido durante o desenvolvimento da traqueia. O padrão de expressão do segmento corporal é visível. A região anterior (cabeça) está na porção superior, a região ventral, à esquerda. (Cortesia de L. Melnikovic e M. Scott)

**Preparação de microarranjos de DNA** Em um método para preparar microarranjos, uma porção contendo cerca de 1 kb da região codificante de cada gene analisado é individualmente amplificada por PCR. Um dispositivo robótico é utilizado para aplicar cada amostra de DNA amplificada à superfície de uma lâmina de microscópio de vidro, que é então quimicamente processada para fixar permanentemente as sequências de DNA à superfície do vidro e desnaturá-las. Um arranjo típico pode contar cerca de 6.000 pontos de DNA em uma grade com 2 x 2 cm.

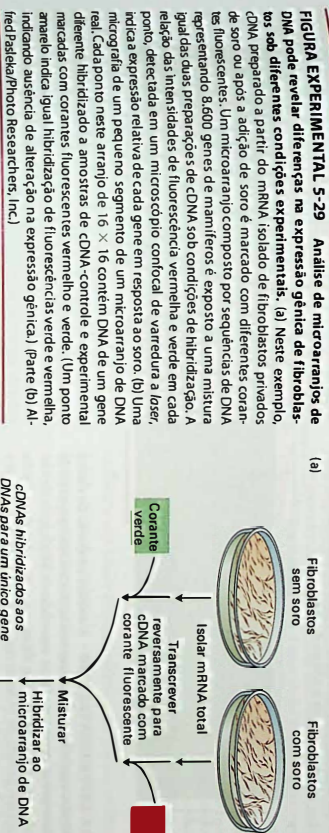
Em um método alternativo, múltiplos oligonucleotídeos de DNA, geralmente contendo cerca de 20 nucleotídeos de comprimento, são sintetizados a partir de um nucleotídeo inicial que é covalentemente ligado a superfície de uma lâmina de vidro. A síntese de um oligonucleotídeo de sequência específica pode ser programada em uma pequena região da superfície da lâmina. Várias sequências de oligonucleotídeos de um único gene são assim sintetizadas em regiões vizinhas da lâmina para se analisar a expressão daquele gene. Com tal método, polinucleotídeos que representam milhares de genes podem ser produzidos em uma única lâmina de vidro. Como os métodos para construção dos arranjos de oligonucleotídeos são diferentes foram adaptados de métodos de produção de circuitos integrados microscópicos utilizados em computadores, os tipos de microarranjos de oligonucleotídeos com frequência são chamados de *chips de DNA*.

**Uso de microarranjos para se comparar expressão gênica sob diferentes condições** A etapa inicial em um estudo

de expressão por microarranjo é o preparo de cDNAs marcados de forma fluorescente correspondendo aos mRNAs expressos pelas células estudadas. Quando a preparação de cDNAs é aplicada em um microarranjo, pontos representando genes expressos hibridizarão sob condições apropriadas a seus cDNAs complementares na mistura de sondas marcadas e poderão ser detectados em um microscópio de varredura a laser.

A Figura 5-29 mostra como esse método pode ser aplicado no exame das alterações na expressão gênica observadas depois que fibroblastos humanos privados de nutrientes são transferidos para um meio de crescimento rico conteúdo soro. Neste tipo de experimento, as preparações separadas de DNA de fibroblastos crescidos com privação de nutrientes e com soro são marcadas com corantes fluorescentes de cores diferentes. Um arranjo de DNA incluindo 8.600 genes de mamíferos é incubado com uma mistura contendo quantidades iguais das duas preparações de cDNA sob condições de hibridização. Após a lavagem e eliminação do cDNA não hibridizado, a intensidade das fluorescências verde e vermelha em cada ponto de DNA é medida utilizando-se um microscópio de fluorescência e armazenada em arquivos de computador sob o nome de cada gene de acordo com sua posição conhecida na lâmina. As intensidades relativas dos sinais de fluorescência verde e vermelha em cada ponto são uma medida do nível relativo de expressão daquele gene em resposta ao soro. Genes que não são transcritos sob essas condições de crescimento não produzem sinal detectável. Genes que são transcritos no mesmo nível sob ambas as condições hibridizarão igualmente

**ANIMAÇÃO DA TÉCNICA:** Sintetizando um arranjo de oligonucleotídeo ANIMAÇÃO DA TÉCNICA: Triagem para padrões de expressão gênica

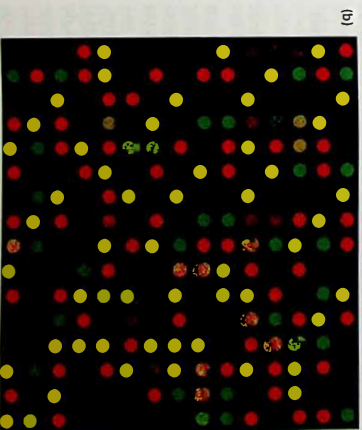
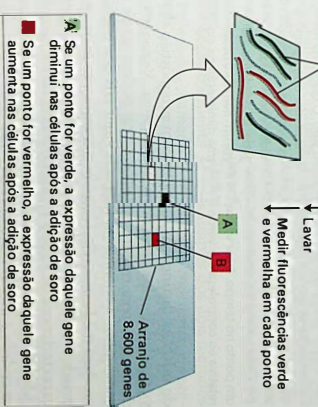


**FIGURA EXPERIMENTAL 5-29** Análise de microarranjos de DNA pode revelar diferenças na expressão gênica de fibroblastos sob diferentes condições experimentais. (a) Neste exemplo, cDNA preparado a partir do mRNA isolado de fibroblastos privados de soro ou após a adição de soro é marcado com diferentes corantes fluorescentes. Um microarranjo composto por sequências de DNAs representando 8.600 genes de mamíferos é exposto a uma mistura igual das duas preparações de fluorescência vermelha e verde em cada ponto, detectada em um microscópio confocal de varredura a laser, indica a expressão relativa de cada gene em resposta ao soro. (b) Uma micrografia de um pequeno segmento de um microarranjo de DNA real. Cada ponto neste arranjo de 16 x 16 contém DNA de um gene diferente hibridizado a amostras de cDNA controle e experimental marcadas com corantes fluorescentes vermelho e verde. (Um ponto arranja indica igual hibridização de fluorescências verde e vermelha, indicando ausência de alteração na expressão gênica.) (Parte (b) Altered by Photo Research, Inc.)

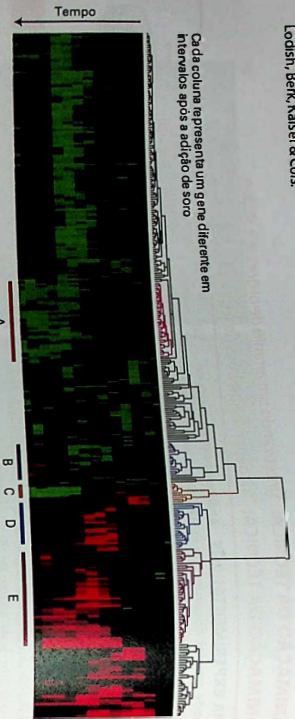
**Análise conjunta de múltiplos experimentos de expressão identifica genes correlacionados**

Raramente se obtêm conclusões definitivas a partir de um único experimento de microarranjo sobre genes que exibem alterações de expressão semelhantes serem correlacionados e, portanto, provavelmente relacionados funcionalmente. Por exemplo, muitas das diferenças observadas na expressão gênica recém-descritas em fibroblastos poderiam ser consequências indiretas de várias alterações na fisiologia celular que ocorrem quando as células são transportadas de um meio para o outro. Em outras palavras, genes que parecem correlacionados em um único experimento de expressão por microarranjo podem sofrer alterações de expressão por razões muito diferentes e a verdade ter funções biológicas muito diferentes. Uma solução para tal problema é combinar a informação de um conjunto de experimentos de expressão em um arranjo para encontrar genes regulados simultaneamente sob uma variedade de condições ou por um período de tempo.

O uso mais informativo de múltiplos experimentos de arranjos de expressão é ilustrado pelo exame de expressão relativa dos 8.600 genes em tempos diferentes após a adição de soro, gerando mais de 10<sup>7</sup> conjuntos de dados individuais. Um programa de computador, relacionado aquele utilizado para determinar a relação entre diferentes sequências de proteínas, pode organizar esses dados e agrupar genes que apresentaram expressão semelhante ao longo de um período de tempo após a adição de soro. Extraordinariamente, essa análise conjunta agrupou genes cujas proteínas codificadas participam de um processo celular comum, tais como biossíntese de colesterol ou ciclo celular (Figura 5-30).



No futuro, a análise de microarranjo será uma ferramenta de diagnóstico muito eficaz em medicina. Descobriu-se, por exemplo, que determinados conjuntos



**FIGURA EXPERIMENTAL 5-30 Análise conjunta de dados de múltiplos experimentos de expressão em microarranjos pode identificar genes corregulados.** A expressão de 8.600 genes de mamíferos foi detectada por análise de microarranjo em intervalos de tempo em um período de 24 horas depois que fibroblastos cultivados sem soro foram colocados em meio com soro. O diagrama mostrado fundamenta-se em um algoritmo de computador que agrupa genes que apresentam alterações de expressão semelhantes quando comparados um amostra-contrôle de exposição de soro ao longo do tempo. Cada coluna representa um único gene, e cada linha representa um ponto no tempo. A cor vermelha indica um aumento na expressão relativa ao controle; a cor verde indica uma diminuição na expressão.

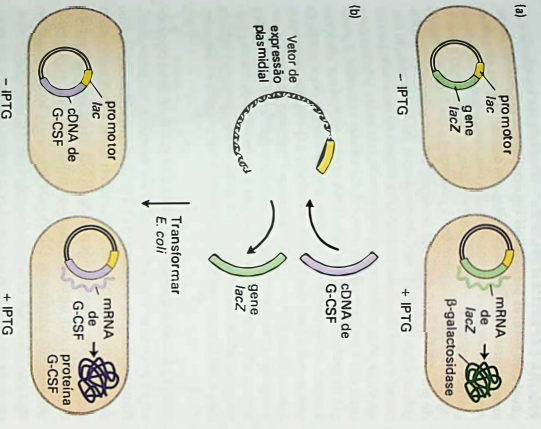
de mRNAs distinguem tumores com prognóstico desfavorável daqueles com bom prognóstico. Variações de doenças anteriormente imperceptíveis podem agora ser detectadas. A análise de biópsias tumorais dos mRNAs ajudará médicos a selecionarem o tratamento mais apropriado. A medida que mais padrões de expressão gênica característicos de vários tecidos doentes foram reconhecidos, o uso diagnóstico de microarranjos de DNA será estendido a outras condições.

**Sistemas de expressão em E. coli podem produzir grandes quantidades de proteínas a partir de genes clonados**

Diversos hormônios proteicos e outras proteínas sinalizadoras ou regulatórias são normalmente expressos em concentrações muito baixas, impedindo seu isolamento e purificação em grandes quantidades por técnicas bioquímicas padrão. O uso generalizado dessas proteínas, bem como a pesquisa básica de suas estruturas e funções, depende de procedimentos eficientes para sua produção em grandes quantidades a um custo razoável. Técnicas de DNA recombinante que transformam células de *E. coli* em fábricas para síntese de proteínas de baixa abundância são agora utilizadas para produzir comercialmente fator estimulador de colônias de granulócitos (G-CSF), insulina, hormônio de crescimento e outras proteínas humanas para uso terapêutico. Por exemplo, o G-CSF estimula a produção de granulócitos, leucócitos fagocíticos críticos para a defesa contra infecções bacterianas. A administração de G-CSF para pacientes com câncer ajuda a compensar a redução na produção de granulócitos causada por agentes quimioterápicos, protegendo os pacientes contra sérias infecções enquanto estão recebendo quimioterapia.

A primeira etapa na produção de grandes quantidades de uma proteína de baixa abundância é obter um clone de cDNA que codifica a proteína inteira por métodos descritos anteriormente. A segunda etapa é projetar vetores plasmidiais que expressem grandes quantidades da proteína codificada quando inseridos em células de *E. coli*. O ponto principal para o planejamento desses vetores de expressão é a inclusão de um promotor, uma sequência de DNA a partir da qual a transcrição do cDNA pode iniciar. Considere, por exemplo, o sistema relativamente simples para expressão de G-CSF apresentado na Figura 5-31. Nesse caso, o G-CSF é expresso em células de *E. coli* transformadas com vetores plasmidiais que contêm o promotor *lac* adjacente ao cDNA, codificando G-CSF clonado. A transcrição a partir do promotor *lac* ocorre em altas taxas apenas quando a lactose, ou um análogo da lactose, como o isopropilglicaricósido (IPTG), é adicionada ao meio de cultura. Quantidades ainda maiores de uma proteína desejada podem ser produzidas em sistemas de expressão em *E. coli* mais complicados.

Para auxiliar a purificação de uma proteína eucariótica produzida em um sistema de expressão de *E. coli*, pesquisadores geralmente modificam o cDNA codificando a proteína recombinante para facilitar sua separação de proteínas endôgenas de *E. coli*. Uma modificação bastante usada é a adição de uma pequena sequência nucleotídica à extremidade do cDNA, de maneira que a proteína expressa tenha seis resíduos de histidina na porção C-terminal. Proteínas assim modificadas ligam-se com alta afinidade



**FIGURA EXPERIMENTAL 5-31 Algumas proteínas eucarióticas podem ser produzidas em células de E. coli a partir de vetores plasmidiais contendo o promotor lac.** (a) O vetor de expressão plasmidial contém um fragmento do cromossomo de *E. coli* contendo o promotor *lac* e o gene vizinho *lacZ*. Na presença do análogo de lactose IPTG, a RNA-polimerase normalmente transcreve o gene *lacZ*, produzindo mRNA de *lacZ* que é traduzido na proteína codificada,  $\beta$ -galactosidase. (b) O gene *lacZ* pode ser removido do vetor de expressão por enzimas de restrição e substituído por um cDNA clonado, nesse caso o DNA que codifica o fator estimulador de colônia de granulócito (G-CSF). Quando o plasmídeo resultante é transformado em células de *E. coli*, a adição de IPTG e a subsequente transcrição a partir do promotor *lac* produzem mRNA de G-CSF, o qual é transcrito na proteína G-CSF.

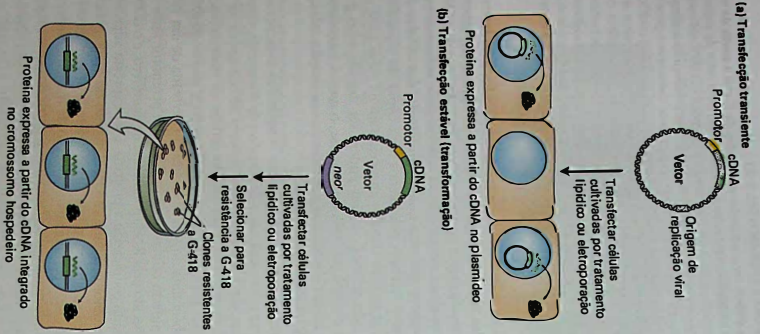
**Vetores plasmidiais de expressão podem ser projetados para uso em células animais**

Embora os sistemas de expressão bacterianos possam ser usados de maneira bem-sucedida para criar grandes quantidades de algumas proteínas, bactérias não podem ser utilizadas em todos os casos. Muitos experimentos que investigam a função de uma proteína em um contexto celular adequado requerem a expressão de uma proteína geneticamente modificada em células animais em cultivo. Os genes são clonados em vetores de expressão eucarióticos especializados e introduzidos em células animais cultivadas por um processo chamado de transfeção. Dois métodos comuns para transfectar células animais diferem quanto à integração ou não do vetor de DNA recombinante no DNA genômico da célula hospedeira. Em ambos os métodos, células animais em cultivo devem ser tratadas para facilitar sua captação inicial do vetor plasmidial recombinante. Isso pode ser feito pela exposição das células a uma preparação de lipídios que penetram na membrana plasmática, aumentando sua permeabilidade ao DNA. Alternativamente, submeter as

células a um breve choque elétrico de vários milhares de volts, técnica conhecida como *eletroporação*, torna-as transientemente permeáveis ao DNA. Em geral, o DNA plasmidial é adicionado em concentração suficiente para assegurar que uma grande proporção das células cultivadas recbam pelo menos uma cópia do plasmídeo. Pesquisadores também aproveitaram os vírus para uso em laboratório; os vírus podem ser modificados para que contenham DNA de interesse, o qual é introduzido em células hospedeiras por meio de sua simples infecção com o vírus recombinante.

**Transfeção transiente** O mais simples dos dois métodos de expressão, chamado de *transfeção transiente*, emprega um vetor semelhante aos vetores de transporte de levedura descritos. Para uso em células de mamíferos, os vetores plasmidiais são projetados para que contenham também uma origem de replicação derivada de um vírus que infecta células de mamífero, um promotor forte reconhecido pela RNA-polimerase de mamífero, e o cDNA clonado codificando a proteína a ser expressa adjacente ao promotor (Figura 5-32a). Quando tal vetor plasmidial entra em uma célula de mamífero, a origem de replicação viral permite sua replicação de maneira eficiente, gerando numerosos plasmídeos a partir dos quais a proteína é expressa. Entretanto, durante a divisão celular os plasmídeos não são fielmente segregados para ambas as células-filhas, e com o tempo uma fração substancial das células em cultivo não contém um plasmídeo, por isso o nome *transfeção transiente*.

**Transfeção estável (Transformação)** Se um vetor introduzido se integra ao genoma da célula hospedeira, o genoma é alterado de forma permanente, e a célula, transformada. É mais provável que a integração seja realizada por enzimas de mamíferos que normalmente atuam no reparo e na recombinção do DNA. Um marcador de seleção bastante usado é o gene da neomicina-fosforotransferase (designado como *neo<sup>r</sup>*), que confere resistência a um composto tóxico quimicamente relacionado à neomicina conhecido como G-418. O procedimento básico para se expressar um cDNA clonado pela *transfeção estável* está delineado na Figura 5-32b. Apenas aquelas células que integram o vetor de expressão no cromossomo hospedeiro irão sobreviver e originar um clone na presença de alta concentração de G-418. Como a integração ocorre em sítios aleatórios do genoma, clones transfor-



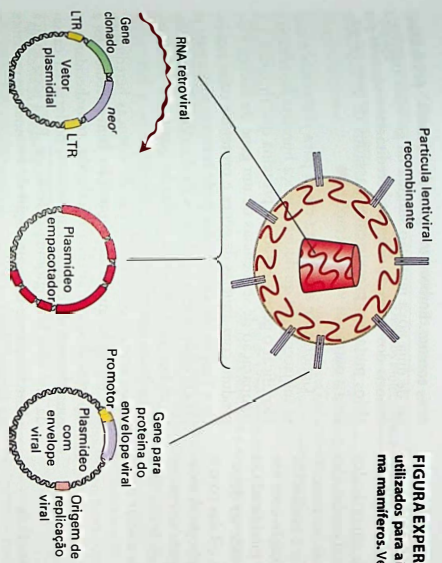
madros individuais resistentes a G-418 irão diferir em suas taxas de transcrição do cDNA inserido. Portanto, os transfectantes estáveis geralmente são triados para identificar aqueles que produzem a proteína de interesse em maiores quantidades.

**Sistemas de expressão retroviral** Pesquisadores exploraram o mecanismo básico utilizado por vírus para introdução de material genético em células animais e subsequente inserção no DNA cromossômico para aumentar a eficiência com a qual um gene modificado pode ser expresso de maneira estável em células animais. Um exemplo desse tipo de expressão viral deriva de uma classe de retrovírus conhecidos como *lentivírus*. Conforme mostrado na Figura 5-33, três plasmídeos diferentes, introduzidos em células por transfeção transiente, são usados para produzir partículas lentivirais recombinantes adequadas para introdução eficiente de um gene clonado em células-sílvio animais. O primeiro plasmídeo, conhecido

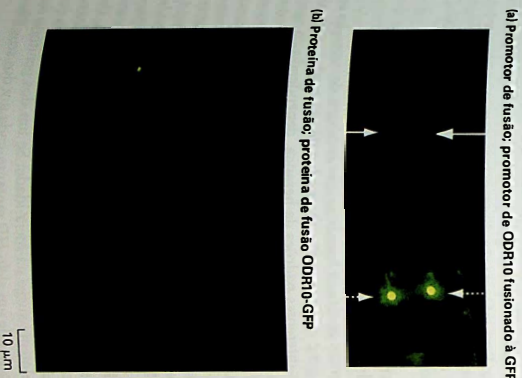
**FIGURA EXPERIMENTAL 5-32** Transfeções transiente e estável com vetores plasmidiais especialmente projetados permitem a expressão de genes clonados em células animais em cultivo. Ambos os métodos empregam vetores plasmidiais que contêm os elementos habituais - Ori, marcador de seleção (p. ex., *amp<sup>r</sup>*) e sítio de ligação múltipla - que permitem a propagação em *E. coli* e a inserção de um cDNA clonado com um promotor animal adjacente. Para simplificar, estes elementos não são mostrados. (a) Na transfeção transiente, o vetor plasmidial contém uma origem de replicação para o vírus, o que pode replicar em células animais cultivadas; a produção de proteína codificada pelo cDNA ocorre apenas por um tempo limitado. (b) Na transfeção estável, o vetor carrega um marcador de seleção como *neo<sup>r</sup>*, que confere resistência a G-418. As relativamente poucas células animais transfectadas que integram o DNA exógeno em seu genoma são selecionadas em meio contendo G-418. Como o vetor de integração contém uma origem de replicação para a proteína codificada pelo cDNA enquanto a cultura é mantida, Ver o texto para discussão.

como *plasmídeo vetor*, contém um gene de interesse clonado próximo a um marcador de seleção, tal como *neo<sup>r</sup>*, flanqueado por sequências LTR do lentivírus. Conforme descrito no Capítulo 6, sequências LTR virais controlam a síntese de uma molécula de RNA viral que, ao ser introduzida em uma célula-sílvio infectada por vírus, pode ser transformada em DNA por transcrição reversa e integrada ao DNA cromossômico. Um segundo plasmídeo, conhecido como empacotador, carrega todos os genes virais, exceto o principal proteína do envelope viral, necessária ao empacotamento de RNA viral contendo LTRs em partículas lentivirais funcionais. O plasmídeo final permite a expressão de uma proteína de envelope viral que quando incorporada em um lentivírus recombinante permitirá que as partículas virais híbridas resultantes infectem um tipo de célula-sílvio desejado. Uma proteína de envelope muito usada neste contexto é a glicoproteína do vírus da estomatite vesicular (proteína VSV-G), que pode substituir de imediato a proteína do envelope lentiviral normal na superfície de partículas virais completas e permitir às partículas virais resultantes que infectem uma grande variedade de tipos de células de mamíferos, inclusive células-tronco hematopoiéticas, neurônios e células musculares e hepáticas. Após a infecção celular, o gene clonado flanqueado por sequências LTR virais é retroativamente transcrito em DNA, que é transportado para o núcleo e então integrado no genoma hospedeiro. Se necessário, como no caso da transfeção estável, células com um gene clonado e um marcador *neo<sup>r</sup>* integrados de maneira estável podem ser selecionados por resistência a G-418. Muitas das técnicas para inativação da função de genes específicos (Ver Seção 5.5) necessitam que uma porção inteira de células em cultivo seja geneticamente modificada de forma simultânea. Lentivírus modificados são particularmente úteis em tais experimentos porque infectam células com alta eficiência de maneira que cada célula de uma população recebe pelo menos uma cópia do plasmídeo portador do lentivírus.

**Marcação de genes e proteínas** Vetores de expressão podem fornecer uma maneira de se estudar a expressão e a localização intracelular de proteínas eucarióticas. O



**FIGURA EXPERIMENTAL 5-33** Vetores retrovirais podem ser utilizados para a integração eficiente de genes clonados no genoma dos mamíferos. Ver o texto para discussão.



**FIGURA EXPERIMENTAL 5-34** A marcação de genes e proteínas facilita a localização celular de proteínas expressas a partir de genes clonados. Neste experimento, o gene codificando um receptor de odor químico, *ODR10*, de *C. elegans* foi fusionado à sequência de genes para a proteína fluorescente verde (*GFP*). (a) Um promotor de fusão foi gerado pela ligação da *GFP* ao promotor e os quatro primeiros códons de aminoácidos do *ODR10*. Esta proteína é expressa no citoplasma de neurônios sensoriais específicos na cabeça de *C. elegans*. O corpo celular (seta pontilhada) e os dendritos sensoriais (seta sólida) estão fluorescentemente marcados. (b) Uma proteína de fusão foi construída pela ligação da *GFP* à extremidade da sequência codificadora interna de *ODR10*. Neste caso, a proteína de fusão *ODR10-GFP* é direcionada à membrana da ponta dos neurônios sensoriais e fica aparentemente na extremidade distal dos cílios sensoriais. A distribuição observada pode ser inferida como reflexo da localização normal da proteína *ODR10* em neurônios específicos. Como o promotor de fusão mostrado em (a) é desprovido das sequências de localização da *ODR10*, a *GFP* expressa preenche o citoplasma todo, em vez de ser localizada apenas na ponta distal dos cílios sensoriais. (P. Sengupta et al., 1996, *Cell* 84:899 (derivado das Figuras 4 e 5))

Uma alternativa à marcação com GFP para detectar a localização intracelular de uma proteína e modificar o gene de interesse fixando-o a uma pequena sequência de DNA que codifica uma curta região de aminoácidos reconhecida por um anticorpo monoclonal conhecido como *marcação de epítopo*. Após a transcrição, o plasmídeo de expressão contendo o gene modificado, a proteína marcada com epítopo expressa pode ser detectada por marcação imunofluorescente das células com o anticorpo monoclonal específico para o epítopo. A escolha entre o uso de um epítopo ou a GFP para se marcar uma determinada proteína geralmente depende dos tipos de modificação que um gene clonado pode tolerar e ainda permanecer funcional.

**CONCEITOS-CHAVE da Seção 5.3**  
**Uso de fragmentos de DNA donados para estudo da expressão gênica**

- O *Southern blotting* pode detectar um único fragmento de DNA específico em uma mistura complexa pela combinação de eletroforese em gel, transferência (*blotting*) das bandas separadas para um filtro e hibridização com uma sonda de DNA complementar radiativamente marcada (ver Figura 5.26). A técnica semelhante de *Northern blotting* detecta um RNA específico em uma mistura.
- A presença e a distribuição de mRNAs específicos podem ser detectadas em células vivas por hibridização *in situ*.
- A análise de microarranjos de DNA detecta simultaneamente o nível relativo de expressão de milhares de genes em diferentes tipos de células sob diferentes condições (ver Figura 5.29).
- A análise conjunta de dados de múltiplos experimentos de microarranjos de expressão pode identificar genes regulados de maneira semelhante sob várias condições. Os genes correlacionados geralmente codificam proteínas que possuem funções biologicamente relacionadas.
- Vetores de expressão derivados de plasmídeos permitem a produção de quantidades abundantes de uma proteína a partir de um gene doado.
- Vetores de expressão encarióticos podem ser usados para expressar genes donados em leveduras ou células de mamíferos. Uma aplicação importante desses métodos é a marcação de proteínas com GFP ou um epítopo para detecção por anticorpo.

**5.4 Localização e Identificação de genes de doenças humanas**

Doenças hereditárias humanas são a consequência fenotípica de genes humanos defeituosos. A Tabela 5-2 lista várias das doenças hereditárias de ocorrência

mais comuns. Embora um gene de "doença" possa resultar de uma nova mutação surgida na geração anterior, a maioria dos casos de doenças hereditárias é causada por alelos mutantes preexistentes passados de uma geração para a outra por várias gerações. ■

Geralmente a primeira etapa para identificar a causa subjacente em qualquer doença humana hereditária é identificar o gene afetado e a proteína que ele codifica. A identificação de sequências de um gene de doença e seu comparativo de sequências de genes e proteínas cujas sequências são conhecidas pode fornecer pistas para as causas moleculares e celulares da doença. Historicamente, os pesquisadores têm utilizado quaisquer indicações fenotípicas que possam ser relevantes para fazer suposições sobre a base molecular de doenças hereditárias. Um exemplo inicial de suposição bem-sucedida foi a hipótese de que a anemia falciforme, reconhecidamente uma doença de células sanguíneas, poderia ser causada por uma hemoglobina defeituosa. Essa ideia levou à identificação de uma alteração de aminoácido específica na hemoglobina que causa polimerização das moléculas de hemoglobina defeituosas, causando a deformação em forma de foice das hemácias nos indivíduos que herdaram duas cópias do alelo *Hb<sup>S</sup>* da hemoglobina falcêmica.

O mais comum, no entanto, é que os genes responsáveis por doenças hereditárias precisem ser identificados sem nenhum conhecimento prévio ou hipótese razoável sobre a natureza do gene afetado ou seu produto. Nesta seção, será visto como geneticistas podem identificar o gene responsável por uma doença hereditária seguindo a segregação da doença nas famílias. A segregação da doença pode ser correlacionada com a segregação de vários outros marcadores genéticos, eventualmente levando à identificação da posição cromossômica do gene afetado. Esta informação, juntamente com o conhecimento da sequência do genoma humano, pode então permitir que o gene afetado e a mutação causadora da doença sejam identificados.

**Doenças monogênicas apresentam um dos três padrões de herança**

Doenças genéticas humanas que resultam de mutação em um gene específico são chamadas de *monogênicas* e existem diferentes padrões de herança, dependendo da natureza e da localização cromossômica dos alelos em envolvidos. Um padrão característico exibido por um alelo dominante em um autossomo (i.e., um dos 22 cromossomos humanos que não é um cromossomo sexual). Como um alelo *autossômico dominante* é expresso no heterozigoto, geralmente pelo menos um dos genitores de um indivíduo afetado também terá a doença. Muitas vezes as doenças causadas por alelos dominantes se manifestam tardiamente, após a idade reprodutiva. Se não fosse assim, a seleção natural teria eliminado o alelo durante a evolução humana. Um exemplo de doença autossômica dominante é a síndrome de Huntington, um distúrbio neurológico degenerativo que geralmente surge na meia-idade. Se um dos genitores for portador de um alelo *HD* mutante, cada um dos seus filhos

**TABELA 5-2** Doenças hereditárias humanas comuns

Doença	Defeitos moleculares e celulares	Incidência
<b>Autossômicas recessivas</b>		
Anemia falciforme	Hemoglobina anormal causa formação das hemácias, que podem ficar presa nos capilares; também confere resistência à malária	1/625 de origem subsaariana
Fibrose cística	Canal de cloro defeituoso (CFTR) em células epiteliais leva a muco excessivo nos pulmões	1/2.500 de origem europeia
Fenilcetonúria (PKU)	Enzima defeituosa do metabolismo de fenilalanina (fenilalanina hidroxilase) resulta em excesso de fenilalanina, causando retardamento mental; a menos que restrita pela dieta	1/10.000 de origem europeia
Doença de Tay-Sachs	Enzima hexoaminidase defeituosa leva ao acúmulo de esfingolípideos em excessos nos lisossomos dos neurônios, prejudicando o desenvolvimento neural	1/1.000 judeus do leste europeu
<b>Autossômicas recessivas</b>		
Síndrome de Huntington	Proteína neural defeituosa (huntingtina) pode formar agregados, danificando o tecido neural	1/10.000 de origem europeia
Hipercolesterolemia	Receptor de LDL defeituoso leva ao excesso de colesterol no sangue e ataques cardíacos precoces	1/225 franco-canadenses
<b>Recessivas ligadas ao X</b>		
Distrofia muscular de Duchenne	Proteína citosquelética distrofina defeituosa leva à função muscular prejudicada	1/3.500 homens
Hemofilia A	Elemento de coagulação sanguínea fator VIII defeituoso leva a hemorragias descontroladas	1-2/10.000 homens

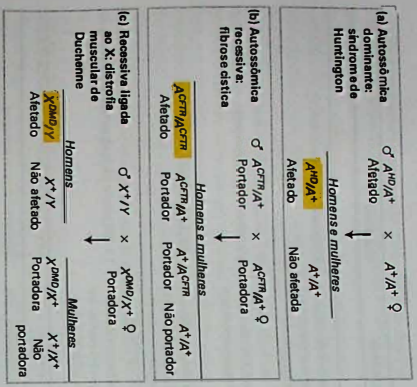
(independentemente de gênero) terá uma chance de 50% de herdar o alelo mutante e ser afetado (Figura 5-35a).

Um alelo recessivo em um autossomo exibe um padrão de segregação bastante diferente. No caso de alelo *autossômico recessivo*, ambos os genitores devem ser heterozigotos portadores do alelo para que seus filhos tenham risco de ser afetados pela doença. Cada filho de um genitor heterozigoto possui uma chance de 25% de receber ambos os alelos recessivos e assim ser afetado, uma chance de 50% de receber um alelo normal e um alelo mutante e assim ser um portador, e uma chance de 25% de receber dois alelos normais. Um exemplo claro de doença autossômica recessiva é a fibrose cística, que resulta de um gene para canal de cloro defeituoso chamado *CFTR* (Figura 5-35b). Indivíduos com grau de parentesco *fratres*, primos em primeiro ou segundo grau) apresentam probabilidade relativamente alta de serem portadores dos mesmos alelos recessivos. Assim, filhos de pais relacionados têm chance muito maior do que aqueles de pais não relacionados de serem homozigotos para uma doença autossômica recessiva e, portanto, afetados por ela. O receptor padrão comum de herança é aquele de um alelo *recessivo ligado ao X*. Um alelo recessivo no cromossomo X será com mais frequência expresso em

homens, que recebem apenas um cromossomo X de suas mães, mas não em mulheres, que recebem um cromossomo X de sua mãe e outro de seu pai. Isso leva a um padrão de segregação distinto ligado ao sexo no qual a doença é exibida com frequência muito maior em homens do que em mulheres. Por exemplo, a distrofia muscular de Duchenne (DMD), uma doença muscular degenerativa que afeta especificamente homens, é causada por um alelo recessivo no cromossomo X. A DMD apresenta o típico padrão de segregação ligado ao sexo no qual mães heterozigotas, portadoras, fentópicamente normais, podem atuar como portadoras, transmitindo o alelo da DMD e em consequência a doença para 50% de sua prole masculina (Figura 5-35c).

**Polimorfismos de DNA são utilizados como marcadores para o mapeamento de ligação de mutações humanas**

Uma vez determinado o modo de herança, a próxima etapa para se determinar a posição de um alelo de doença é mapear geneticamente sua posição em relação a marcadores genéticos conhecidos usando o princípio básico da ligação genética descrito na Seção 5.1. A presença



**FIGURA 5-35 Três padrões de herança comuns para doenças genéticas humanas.** Autosomos (A) e cromossomos sexuais (X e Y) selvagens são indicados por sinais de somo sobscritos. (a) Em um distúrbio autossômico dominante como a doença de Huntington, apenas um alelo mutante é necessário para produzir a doença. Se um dos genitores for heterozigoto para o alelo mutante  $H^D$ , seus filhos terão uma chance de 50% de herdar o alelo mutante e desenvolver a doença. (b) Em um distúrbio autossômico recessivo como a fibrose cística, dois alelos mutantes devem estar presentes para produzir a doença. Ambos os genitores precisam ser heterozigotos portadores do gene mutante  $Ctfr$  para que seus filhos corram risco de serem afetados ou portadores. (c) Uma doença recessiva ligada ao X como a distrofia muscular de Duchenne é causada por uma mutação recessiva no cromossomo X e exige o tipo padrão de segregação ligado ao sexo. Homens não doentes não herdam alelos para um alelo DMD mutante, pois possuem 50% de chance de receber um alelo mutante e ser afetado. Mulheres nascidas de mães heterozigotas possuem 50% de chance de serem portadoras.

de vários traços, ou marcadores, genéticos diferentes já mapeados, distribuídos ao longo de um cromossomo, facilita o mapeamento de uma nova mutação; determina-se sua possível ligação com esses genes marcadores em cruzamentos apropriados. Quanto mais marcadores estiverem disponíveis, mais preciso será o mapeamento de uma mutação. A densidade de marcadores genéticos necessária para um mapa genético humano de alta resolução é de cerca de um marcador para cada 5 centimorgans (cM) (conforme discutido, uma unidade de mapa genético, ou centimorgan, é definida como a distância entre duas posições ao longo de um cromossomo que resulta em um indivíduo recombinante a cada 100 descendentes). Portanto, um mapa genético de alta resolução necessita de 25 ou mais marcadores genéticos em posições conhecidas espalhadas ao longo de cada cromossomo humano.

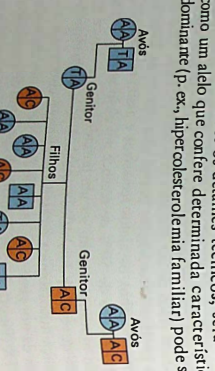
Nos organismos experimentais comumente utilizados em estudos genéticos, numerosos marcadores com fenótipos facilmente identificáveis estão disponíveis para o mapeamento genético de mutações. Isso não acontece no mapeamento de genes cujos alelos mutantes estão associ-

dos a doenças hereditárias em seres humanos. Contudo, a tecnologia de DNA recombinante disponibilizou uma variedade de marcadores moleculares úteis baseados em DNA. Como a maior parte do genoma humano não codifica proteínas, existe uma grande quantidade de variação de sequência entre os indivíduos. De fato, estima-se que diferentes nucleotídeos entre indivíduos não relacionados sejam detectadas na sequência de DNA, chamadas de *polimorfismos de DNA*, podem ser acompanhadas de uma geração para a outra, podendo servir como marcadores genéticos para estudos de ligação. Atualmente, um painel com cerca de 10<sup>5</sup> polimorfismos diferentes conhecidos, cuja localização foi mapeada no genoma humano, é utilizado para estudos de ligação genética em seres humanos.

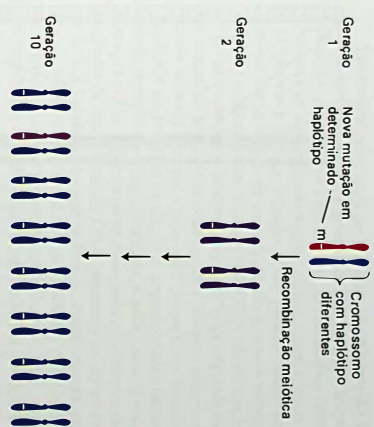
*Polimorfismos de nucleotídeo único (single nucleotide polymorphisms, SNPs)* constituem o tipo mais abundante e, assim, útil para construir mapas genéticos de resolução máxima (figura 5-36). Outro tipo útil de polimorfismos de DNA consiste em uma sequência com número variável de repetições de uma, duas ou três bases. Esses polimorfismos, conhecidos como *repetições de sequência simples (STRs)* ou *microsatélites*, são formados por deslizamento do molde ou das fitas recém-sintetizadas durante a replicação do DNA. Uma característica útil das STRs é que indivíduos diferentes geralmente têm diferentes números de repetições. A existência de múltiplas versões de uma SSR torna mais provável a produção de um padrão de segregação informativo em um dado pedigree, e, portanto, tem uso mais geral no mapeamento das posições de genes de doenças. Os polimorfismos podem ser detectados por meio de amplificação por PCR e sequenciamento de DNA.

**Estudos de ligação podem mapear genes relacionados a doenças com resolução de aproximadamente 1 centimorgan**

Sem considerar todos os detalhes técnicos, será visto como um alelo que confere determinada característica dominante (p. ex., hipercolesterolemia familiar) pode ser relacionado a genes de doenças com resolução de aproximadamente 1 centimorgan. A existência de múltiplas versões de uma SSR torna mais provável a produção de um padrão de segregação informativo em um dado pedigree, e, portanto, tem uso mais geral no mapeamento das posições de genes de doenças. Os polimorfismos podem ser detectados por meio de amplificação por PCR e sequenciamento de DNA.



**FIGURA EXPERIMENTAL 5-36 Polimorfismos de nucleotídeo único (SNPs) podem ser acompanhados como marcadores genéticos.** Heterodígrafa hipotética baseada no análise de SNPs do DNA A1 ou C. Cada indivíduo possui dois alelos, A1 ou S em ambos os cromossomos, e o outro sexo heterozigoto nesse sítio. Os círculos indicam mulheres, quadrados indicam homens; azul indica indivíduos não afetados; laranja indica indivíduos com a característica. A análise revela que a característica segregava com um C no SNP.

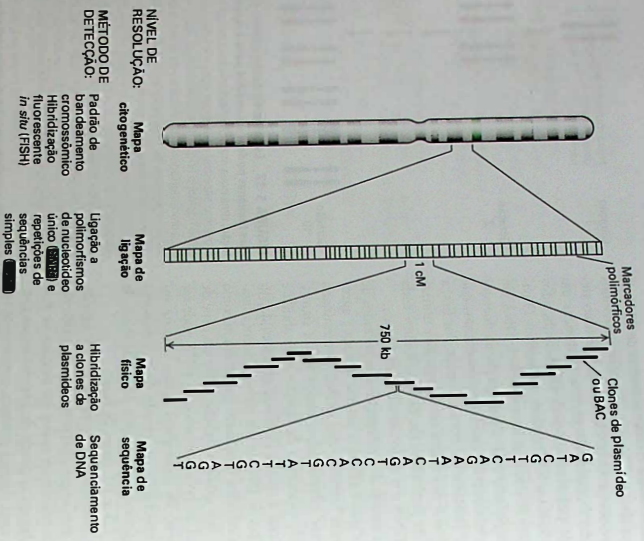


**FIGURA 5-37 Estudos de desequilíbrio de ligação de populações humanas podem ser usados para mapear genes em alta resolução.** Uma nova mutação de doença irá surgir no contexto de um cromossomo ancestral dentro um conjunto de polimorfismos conhecido como *haplótipo* (indicado em vermelho). Após várias gerações, os cromossomos portadores da mutação de doença também serão portadores de segmentos do haplótipo ancestral que não foram separados da mutação da doença por recombinação. Os segmentos em azul destes cromossomos representam haplótipos genais derivados da população em geral, e o alelo do haplótipo ancestral no qual a mutação originalmente surgiu. Este fenômeno é conhecido como *desequilíbrio de ligação*. A posição da mutação da doença pode ser localizada pela ligação de cromossomos contendo a mutação da doença para polimorfismos bem conservados correspondendo ao haplótipo ancestral.

**Análises adicionais são necessárias para se localizar um gene de doença em um DNA clonado**

Embora o mapeamento de ligação geralmente localize um gene de doença humano em uma região contendo cerca de 10<sup>5</sup> pares de bases, mais de 10 genes diferentes podem estar localizados na mesma região. O objetivo final de um estudo de mapeamento é localizar o gene dentro de um segmento clonado de DNA e, então, determinar a sequência nucleotídica deste segmento. As escalas relativas de um mapa genético cromossomal e de mapas físicos que correspondem a conjuntos ordenados de clones de plasmídeos e a sequência nucleotídica são apresentados na Figura 5-38.

Uma estratégia para localização adicional de um gene de doença em um genoma é identificar o mRNA codificado pelo DNA na região do gene em estudo. A comparação de da expressão gênica em tecidos de indivíduos normais e afetados pode sugerir tecidos nos quais um determinado gene de doença é normalmente expresso. Por exemplo, uma mutação que afeta fenotipicamente o músculo, mas nenhum outro tecido, pode estar em um gene expresso apenas no tecido muscular. A expressão de mRNA em ambos os indivíduos normal e afetado é geralmente determinado por *Northern blotting* ou hibridização *in situ* de DNA ou RNA marcada em seções do tecido. Experimentos de *Northern blots*, hibridização *in*



**FIGURA 5-38 A relação entre os mapas genético e físico de um cromossomo humano.** O diagrama representa um cromossomo analisado em diferentes níveis de resolução. O cromossomo como um todo pode ser visto sob a luz do microscópio quando se encontra em estado condensado que ocorre na metáfase, e a localização aproximada de sequências específicas pode ser determinada por hibridação fluorescente *in situ* (FISH). No próximo nível de resolução, características podem ser mapeadas em relação a marcadores de DNA. Segmentos locais do cromossomo podem ser analisados no nível de sequências de DNA identificadas por hibridação de *Southern* ou PCR. Finalmente, importantes diferenças genéticas podem ser definidas de forma mais precisa por diferenças na sequência de nucleotídeos do DNA. cromossomo, (adaptada de L. Hartwell et al., 2003. *Genetics from Genes to Genomes*, 2. ed., McGraw-Hill).

*in situ* ou microarranjo permitem comparar ambos o nível de expressão e o tamanho dos mRNAs em tecidos miúdos e seixos. Embora a sensibilidade da hibridação *in situ* seja mais baixa do que aquela das análises de *Northern blot*, ela pode ser muito útil na identificação de um mRNA que é expresso em baixos níveis em um dado tecido, mas em níveis muito altos em uma subclasse de células dentro daquele tecido. Um mRNA alterado ou ausente em vários indivíduos afetados por uma doença, quando comparado com indivíduos com o fenótipo selvagem, seria um excelente candidato à codificação da proteína cuja função defeituosa causa a doença.

Em muitos casos, mutações de ponto que originam alelos causadores de doenças podem resultar em alteração indetectável no nível de expressão ou na mobilidade eletroforética dos mRNAs. Assim, se a comparação dos mRNAs expressos em indivíduos normais ou afetados não revelar diferenças detectáveis nos mRNAs candida-

tos, uma busca por mutações de ponto nas regiões do DNA que codificam mRNAs é realizada. Agora que métodos altamente eficientes para o sequenciamento de DNA estão disponíveis, os pesquisadores com frequência determinam a sequência de regiões candidatas do DNA isolado de indivíduos afetados para identificar mutações codificadoras que mostram de forma consistente alterações possíveis de deletar no DNA de indivíduos que exibem a doença. Uma limitação do abordagem é que a região próxima ao gene afetado pode apresentar por limorfismos de ocorrência natural não relacionados ao gene de interesse. Estes polimorfismos, não relacionados funcionalmente à doença, podem levar à identificação errônea do fragmento de DNA portador do gene de interesse. Por tal razão, quando mais alelos mutantes disponíveis para análise, maior é a chance de se identificar corretamente um gene.

**Muitas doenças hereditárias são o resultado de múltiplos defeitos genéticos**

A maioria das doenças hereditárias humanas hoje conhecidas em nível molecular são doenças monogênicas; porém, um distúrbio claramente discernível é produzido por um defeito em um único gene. Doenças monogênicas causadas por mutação em um gene específico exibem um padrão de herança característicos mostrados na Figura 5-35. Os genes associados com a maioria das doenças monogênicas comuns já foram mapeados usando-se marcadores baseados em DNA, conforme descrito.

Várias outras doenças, porém, apresentam padrões de herança mais complicados, tornando a identificação da causa genética substancialmente mais difícil. Um tipo de complexidade adicional com frequência encontrada é a *heterogeneidade genética*. Nesses casos, mutações em qualquer um entre múltiplos genes diferentes podem causar a mesma doença. Por exemplo, a retinite pigmentosa, caracterizada pela degeneração da retina geralmente levando à cegueira, pode ser causada por mutações em qualquer um de mais de 60 genes diferentes. Em estudos de ligação com seres humanos, dados de várias famílias precisam ser combinados para determinar se uma ligação estatisticamente significativa existe entre um gene de doença e marcadores moleculares conhecidos. A heterogeneidade genética tal como aquela apreendida pela retinite pigmentosa pode confundir a abordagem, pois qualquer tendência estatística nos dados de mapeamento de uma família tende a ser anulada por dados obtidos de outra família com um gene causador não relacionado.

Os pesquisadores em genética humana usam duas abordagens diferentes para identificar vários genes associados com a retinite pigmentosa. A primeira abordagem dependeu de estudos de mapeamento em famílias únicas excepcionalmente grandes que continham um número suficiente de indivíduos afetados para fornecer evidência de ligação estatisticamente significativa entre polimorfismos de DNA conhecidos e um único gene causador. Os genes identificados por estes estudos mostraram que várias das mutações que causam retinite pigmentosa estão em genes que codificam proteínas abundantes na retina. Segundo essa pista, os geneticistas concentraram sua atenção naqueles genes que são altamente expressos na retina quando estavam outros indivíduos com retinite pigmentosa. A abordagem usando informações adicionais para focar os esforços de triagem em um conjunto de genes candidatos levou à identificação de mutações causais raras adicionais em vários genes diferentes que codificam proteínas da retina.

Uma complicação adicional na descrição genética das doenças humanas é representada por diabetes, doença cardíaca, obesidade, predisposição ao câncer e uma variedade de doenças mentais que apresentam pelo menos alguma complexidade hereditária. Essas e muitas outras doenças são consideradas *poligênicas* no sentido de que alelos de múltiplos genes, atuando juntos em um indivíduo, contribuem sistematicamente e gravemente para a doença. Como mapa genético humano é um dos problemas mais importantes e desafiadores em genética humana atualmente.

Um dos métodos mais promissores para o estudo de doenças que exibem heterogeneidade genética ou são poligênicas é a busca por uma correlação estatística entre a herança de determinada região de um cromossomo e a predisposição à doença, usando-se um procedimento conhecido como estudo de associação do genoma inteiro (*genome-wide association study*, GWAS). A identificação de genes causadores de doenças por GWAS depende do fenômeno de desequilíbrio de ligação descrito anteriormente. Se um alelo que causa uma doença, ou mesmo predisposição um indivíduo a ela, tiver origem relativamente recente durante a evolução humana, o alelo causador da doença tenderá a permanecer associado a determinado conjunto de marcadores de DNA presentes na vizinhança de sua localização cromossômica. Examinando um grande número de marcadores de DNA em populações de indivíduos com uma determinada doença, bem como em populações de indivíduos-controle sem a doença, regiões cromossômicas que tendem a estar correlacionadas com a ocorrência da doença podem ser identificadas por GWAS. O sequenciamento genômico e outros métodos são, então, usados para identificar possíveis mutações causadoras de doença nas regiões. Embora o GWAS seja uma ferramenta poderosa para identificar genes de doenças candidatas, é necessário muito trabalho adicional para determinar como um indivíduo portador de uma determinada mutação pode estar predisposto à doença.

Modelos de doenças humanas em organismos experimentais também contribuem para desvendar o componente genético de características complexas, como obesidade ou diabetes. Por exemplo, experimentos controlados de acasalamentos de camundongos em grande escala identificam genes murinos associados com doenças análogas a aquelas em seres humanos. Os ortólogos humanos dos genes murinos identificados nos estudos seriam candidatos prováveis para o envolvimento na doença humana correspondente. O DNA de populações humanas poderia ser examinado para definir se determinados alelos de genes candidatos apresentam tendência a estar presentes em indivíduos afetados com a doença, mas ausentes em indivíduos não afetados. Essa abordagem de "gene candidato" está sendo amplamente usada na busca por genes que contribuíam para as principais doenças poligênicas em seres humanos.

**CONCEITOS-CHAVE DA Seção 5.4**

- Localização e identificação de genes de doenças humanas**
- Doenças hereditárias e outras características em seres humanos apresentam três padrões de herança principais: autossômico dominante, autossômico recessivo e recessivo ligado ao X. (ver Figura 5-35).
  - Genes para doenças humanas e outras características podem ser mapeados pela determinação de sua segregação conjunta durante a meiose com marcadores cuja localização no genoma é conhecida. Quanto mais próximo um gene estiver de um determinado marcador, maior sua chance de segregação conjunta.

- O mapeamento de genes humanos com grande precisão requer milhares de marcadores moleculares distribuídos ao longo dos cromossomos. Os marcadores mais úteis são diferenças na sequência de DNA (polimorfismos) entre indivíduos em regiões não codificadas do genoma.
- Polimorfismos de DNA úteis no mapeamento de genes humanos incluem os polimorfismos de nucleotídeo único (SNPs) e espectros de sequência simples (SSRs).
- Pelo mapeamento de ligação, consegue-se, geralmente, localizar um gene de doença humana em uma região cromossômica que inclui 10 ou mais genes. Identificar o gene de interesse dentro da região candidata requer geralmente a análise de expressão e a comparação de sequências de DNA entre indivíduos afetados e não afetados pela doença.
- Algumas doenças hereditárias são o resultado de mutações em genes diferentes em diferentes indivíduos (heterogeneidade genética). A ocorrência e a gravidade de outras doenças dependem da presença de alelos mutantes de múltiplos genes no mesmo indivíduo (características poligênicas). O mapeamento dos genes associados com as doenças pode ser feito pela busca de uma correlação estatística entre a doença e uma determinada região cromossômica em um estudo de associação do genoma inteiro.

5.5 Inativação da função de genes específicos em eucariotos

A elucidação das sequências de DNA e de proteínas nos últimos anos levou à identificação de vários genes, usando-se padrões de sequências do DNA genômico e similitude das proteínas codificadas com proteínas de função conhecida. Conforme discutido no Capítulo 6, as funções gerais das proteínas identificadas por buscas de sequências podem ser previstas por analogia com proteínas conhecidas. No entanto, os papéis *in vivo* precisos dessas "novas" proteínas talvez não fiquem claros na ausência de formas mutantes dos genes correspondentes. Nesta seção, serão descritas várias maneiras de se interromper a função normal de um gene específico no genoma de um organismo. A análise do fenótipo mutante resultante geralmente ajuda a revelar a função *in vivo* do gene normal e de sua proteína codificada.

Tês abordagens básicas estão na base dessas técnicas de inativação genica: (1) substituir um gene normal por outras sequências; (2) introduzir um alelo cuja proteína codificada iniba o funcionamento da proteína normal expressa; (3) promover a destruição do mRNA expresso a partir de um gene. O gene normal endógeno é modificado por técnicas baseadas na primeira abordagem, mas não é modificado nas outras abordagens.

Genes normais de levedura podem ser substituídos por alelos mutantes por recombinação homóloga

A modificação do genoma da levedura *S. cerevisiae* é particularmente fácil por duas razões: as células de levedura

captam de imediato DNA exógeno sob certas condições, e o DNA introduzido é de maneira eficiente trocado pelo sítio cromossômico homólogo da célula recipiente. Essa recombinação específica direcionada de duas regiões idênticas de DNA permite que qualquer gene dos cromossomos de levedura seja substituído por um alelo mutante. (Como discutido na Seção 5.1, a recombinação entre cromossomos homólogos também ocorre naturalmente durante a meiose.)

Em um método popular para interromper genes de leveduras com tal método, a PCR é usada para gerar um construto interrompido contendo um marcador de seleção que após é transfectado em células de levedura. Conforme mostrado na Figura 5-39a, iniciadores para amplificação por PCR do marcador de seleção são projetados para incluir cerca de 20 nucleotídeos idênticos às sequências flanqueadoras do gene de levedura a ser substituído. O construto amplificado resultante e compreendido o marcador de seleção (p. ex., o gene *kanMX*, que assim como o *ura0* contém resistência a G-418) flanqueado por cerca de 20 pares de bases que pareiam com as extremidades do gene-alvo de levedura. Células diploides de levedura transformadas nas quais uma das duas cópias do gene-alvo endógeno foi substituído pelo construto interrompido são identificadas por sua resistência a G-418 ou por outro fenótipo de seleção. As células diploides heterozigotas de levedura crescem de modo normal, independentemente da função do gene-alvo, mas metade dos esporos haploides derivados dessas células carregará apenas o alelo interrompido (Figura 5-39b). Se o gene for essencial para a viabilidade, então os esporos portadores do alelo interrompido não sobreviverão.

A interrupção de genes de levedura por esse método tem provado ser particularmente útil na avaliação do papel de proteínas identificadas pela análise de toda a sequência de DNA genômica (ver Capítulo 6). Um grande consórcio de cientistas substituiu cada um dos 6.000 genes identificados por essa análise pelo construto interrompido com *kanMX* e determinou quais interrupções gênicas levam a esporos haploides inviáveis. As análises mostraram que cerca de 4.500 dos 6.000 genes de leveduras não são necessários para viabilidade, um número inesperadamente grande de genes que não parecem essenciais. Em alguns casos, a interrupção de um determinado gene pode originar defeitos sutis que não comprometem a viabilidade das células de levedura crescendo em condições laboratoriais. Alternativamente, células portadoras de um gene interrompido podem ser viáveis devido às suas sequências de vias compensatórias. Para investigar tal possibilidade, geneticistas de levedura estão buscando mutações letais sintéticas que revelem genes não essenciais com funções redundantes (ver Figura 5-9c).

A transcrição de genes ligados a um promotor regulado pode ser experimentalmente controlada

Embora a interrupção de um gene essencial necessário para o crescimento celular leve à geração de esporos inviáveis, esse método fornece pouca informação acerca da verdadeira função da proteína codificada nas células. Para aprender mais a respeito da contribuição de um gene

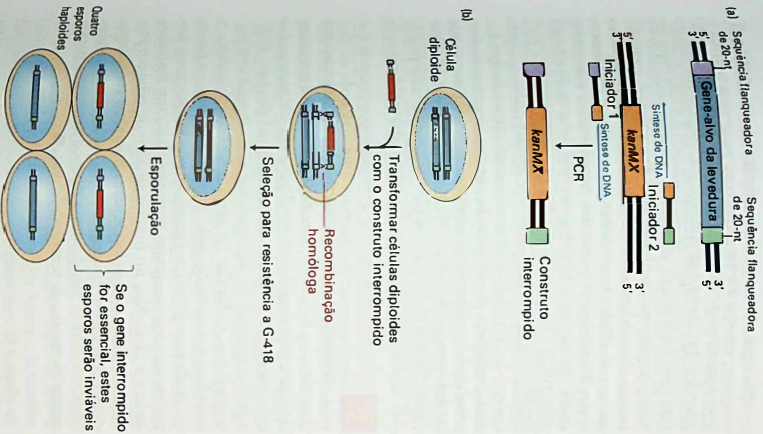


FIGURA EXPERIMENTAL 5-39 Recombinação homóloga com esporos de interrupção transfectados pode inativar genes-alvo em levedura. (a) Um construto adequado para interromper um gene-alvo pode ser preparado por PCR. Cada um dos dois iniciadores (ni) homóloga a uma sequência do gene-alvo de levedura, bem como sequências necessárias à amplificação de um segmento de DNA portador de um gene de marcador de seleção como *kanMX*, que confere resistência a G-418. (b) Quando células de *Saccharomyces glabrata* recombinações são transformadas com o construto de interrupção e as sequências correspondentes do cromossomo irá integrar o gene *kanMX* ao cromossomo, substituindo a sequência do gene-alvo. Assim, quando as células não transformadas não crescerão. Se o gene-alvo for essencial para a viabilidade, metade dos esporos haploides formados após a esporulação de células diploides recombinantes será inviável.

no crescimento e na viabilidade celulares, os investigadores devem ser capazes de inativar seletivamente o gene em uma população de células em crescimento. Um método para se fazer isso em prega um promotor regulado para inativar seletivamente a transcrição de um gene essencial.

O promotor *GAL1* de levedura é útil a este propósito e está ativo em células que crescem na presença de galactose, mas completamente inativo em células que crescem na presença de glicose. Nessa abordagem, a sequência codificadora de um gene essencial (X) ligada ao promotor *GAL1* é inserida em um vetor de transporte de levedura (ver Figura 5-17a). O vetor recombinante é então introduzido em células haploides de levedura nas quais o gene X foi interrompido. Células haploides transformadas crescem em meio com galactose, já que a cópia normal do gene X no vetor é expressa na presença de galactose. Quando as células são transferidas para um meio com glicose, o gene X não é mais transcrito; à medida que as células se dividem, a quantidade de proteína X codificada diminui, alcançando eventualmente um estado de depleção que mimetiza uma mutação de perda de função completa. As alterações observadas no fenótipo dessas células após a mudança para o meio com glicose podem sugerir quais processos celulares dependem da proteína codificada pelo gene essencial X.

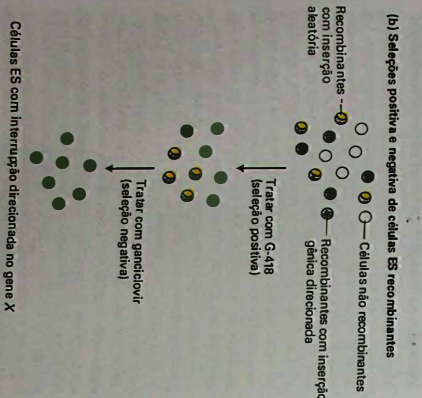
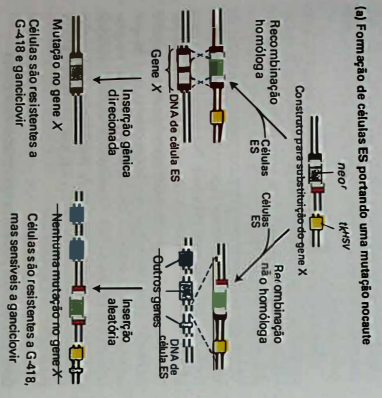
Em uma aplicação inicial do método, pesquisadores exploraram a função dos genes citosólicos *Hsc70* em levedura. Células haploides com uma interrupção em todos os quatro genes *Hsc70* redundantes eram inviáveis a menos que portassem um vetor contendo uma cópia do gene *Hsc70* que pudesse ser expresso a partir do promotor *GAL1* em meio com galactose. Na transferência para glicose, as células portadoras do vetor eventualmente pararam de crescer em decorrência da atividade insuficiente de *Hsc70*. O exame minucioso dessas células revelou que suas proteínas de secreção não conseguiram entrar no retículo endoplasmático (RE). O estudo forneceu a primeira evidência do papel inesperado de *Hsc70* na translocação de proteínas de secreção para o RE, processo examinado em detalhe no Capítulo 13.

Genes específicos podem ser permanentemente inativados na linhagem germinativa de camundongos

Muitos dos métodos de interrupção de genes em leveduras podem ser aplicados a genes de eucariotos superiores. Os genes alterados podem ser introduzidos na linhagem germinativa via recombinação homóloga para produzir animais com um nocaute gênico, ou simplesmente "nocaute". Camundongos-nocaute nos quais um gene específico foi interrompido são um poderoso sistema experimental para estudo do desenvolvimento, do comportamento e da fisiologia de mamíferos. São úteis também ao estudo das bases moleculares de certas doenças genéticas humanas.

Camundongos com genes-alvo nocauteados (*gene-targeted knockout mice*) são gerados por um procedimento de duas etapas. Na primeira etapa, um construto de DNA contendo um alelo interrompido de um determinado gene-alvo é introduzido em células tronco embrionárias (ES, do inglês *embryonic stem*). Essas células, derivadas do blastocisto, podem ser criadas em cultura por várias gerações (ver Figura 21-7). Em uma pequena fração de células transferidas, o DNA introduzido sofre recombinação homóloga com o gene-alvo, embora a recombinação em sítios cromossômicos não homólogos ocorra com bem mais

frequência. Para selecionar as células nas quais a inserção homologa direcionada ao gene-*alvo* ocorreu, o construto de DNA recombinante introduzido nas células ES precisa incluir dois genes de marcadores de seleção (Figura 5-40). Um destes genes (*neo<sup>r</sup>*), que confere resistência a G-418, é inserido no gene-*alvo* (X), interrompendo-o desse modo. O outro gene de seleção, o da timidina-cinase do vírus herpes simplex (*tk<sup>cat</sup>*), é inserido no construto fora da sequência do gene-*alvo*. As células ES que sofreram recombinação entre o DNA do construto e o sítio homologo do construto não irão conter o *neo<sup>r</sup>*, mas não irão incorporar o *tk<sup>cat</sup>*. Como o *tk<sup>cat</sup>* confere sensibilidade ao análogo tóxico de nucleotídeo ganciclovir, as células ES recombinantes de *tk<sup>cat</sup>* podem ser selecionadas por sua habilidade de sobreviver na presença de G-418 e ganciclovir. Nestas células, um alelo do gene X estará interrompido.



Na segunda etapa da produção de camundongos-nocutivos, células ES heterozigotas para uma mutação-nociva no gene X são injetadas em um blastocisto de camundongo selvagem recipiente, que é então transferido para uma fêmea substituta pseudogravídica (Figura 5-41). A prole resultante será de quimeras, contendo tecidos derivados tanto das células ES transplantadas quanto de células hospedeiras. Se as células ES também forem homozigotas para uma característica marcadora visível (p. ex., cor da pelagem), então os descendentes quiméricos nos quais as células ES sobreviveram e proliferaram poderão ser facilmente identificados. Camundongos quiméricos são cruzados com camundongos homozigotos para outro alelo da característica marcadora a fim de determinar se a mutação-nociva é incorporada na linhagem germinativa. Finalmente, o cruzamento de camundongos heterozigotos para o alelo nocivo produzirá descendentes homozigotos para a mutação-nociva.

O desenvolvimento de camundongos-nocutivos que mimetizam certas doenças humanas pode ser ilustrado no exemplo da fibrose cística. Por meio de métodos descritos na Seção 5.4, a mutação recessiva que causa a doença foi eventualmente localizada em um gene conhecido como *CFTR*, que codifica um canal de cloro. Utilizando o gene *CFTR* selvagem humano clonado, os pesquisadores isolaram o gene homólogo murino e introduziram mutações nele. A técnica de nocutivos genéticos usada para produzir camundongos homozigotos mutantes, que apresentam sintomas (i.e., um fenotipo), inclusive distúrbios no funcionamento de células ciliadas, semelhantes àqueles de seres humanos com fibrose cística. Os camundongos-nocutivos estão sendo usados como modelo de estudo dessa doença genética e para o desenvolvimento de terapias efetivas.

**FIGURA EXPERIMENTAL 5-40 O isolamento de células ES de camundongo com uma interrupção gênica direcionada é o primeiro estágio no produção de camundongos-nocutivos.** (a) Quando DNA exógeno é introduzido em células-tronco embrionárias (ES), a inserção aleatória via recombinação não homologa ocorre com muita frequência maior do que a inserção direcionada ao gene por meio da recombinação homologa. Células recombinantes nas quais um alelo do gene X (*kan<sup>r</sup>* e *neo<sup>r</sup>*) foi interrompido podem ser obtidas pelo uso de um vetor recombinante que carrega o gene X interrompido com o gene *neo<sup>r</sup>* (verde) que confere resistência a G-418, e fora do sítio de homologia, *tk<sup>cat</sup>*, o gene da timidina-cinase do vírus herpes simplex. A timidina-cinase viral, ao contrário da enzima endógena murina, pode converter o análogo de nucleotídeo ganciclovir em sua forma tóxicos. As células recombinantes sobreviventes são transferidas para a aplicação de este e análise modificada para sua forma trifásica que inibe a replicação do DNA, cultura em células ES. Assim, o ganciclovir é tóxico para células ES recombinantes portadoras do gene *tk<sup>cat</sup>*. Além disso, apenas células com inserção não homologa são sensíveis a ganciclovir. (b) Células recombinantes são selecionadas por tetragenon com G-418, já que células que não cataram DNA ou não o integram em seu genoma são sensíveis a este como prototóxico. As células recombinantes sobreviventes são transferidas para ganciclovir. Apenas as células com uma interrupção direcionada do gene X e, portanto, desprovidas do gene *tk<sup>cat</sup>* e sua toxicidade sobreviverão. (Ver S. L. Mansour et al., 1998, *Nature* 393:348).

**VIDEO: Microinjeção de células-tronco embrionárias em blastocistos**

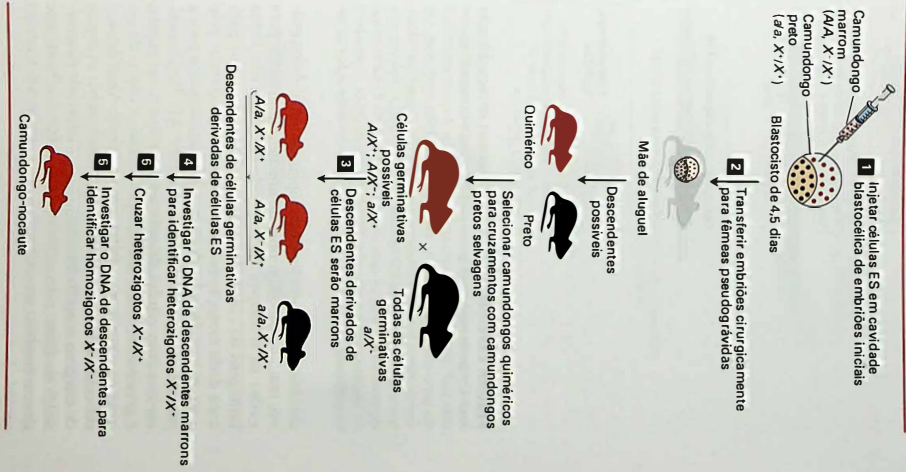
**FIGURA EXPERIMENTAL 5-41 Células ES heterozigotas para um gene-*alvo* nocivo usado para produzir camundongos com gene-*alvo* nocivo.** Etapa 1: células-tronco embrionárias (ES) heterozigotas para um alelo nocivo em um gene de interesse (M) e homozigotas para um alelo dominante de um gene marcador (D) são transplantadas em uma cavidade blastocística de embriões com 4,5 dias que são homozigotos para um alelo recessivo do marcador (cor da pelagem preta, *al*). Etapa 2: os embriões incluídos são implantados em uma fêmea pseudogravídica. Aberturas de descendentes portadores de células derivadas de ES são quimeras, indicadas por sua pelagem misturada preta e marrom. Etapa 3: camundongos quiméricos são então retrocruzados com camundongos pretos descendentes marroms deste cruzamento possuem células germinativas de ES em sua linhagem germinativa. Etapas 4 e 5: análise de DNA isolado de um pequeno pedaço de tecido da cauda pode identificar camundongos marroms heterozigotos para o alelo nocivo. Intercrianismo dos camundongos produz alguns indivíduos homozigotos para o alelo interrompido, ou seja, camundongos-nocutivos. (Adaptado de M. R. Capecchi, 1989, *Trends Genet.* 5:70).

**A recombinação celular somática pode inativar genes em tecidos específicos**

Pesquisadores estão geralmente interessados em examinar os efeitos de mutações-nocutivas em um determinado tecido do camundongo, em um estágio específico do desenvolvimento, ou em ambos. Entretanto, camundongos portadores de um nocutiva na linhagem germinativa podem ter defeitos em numerosos tecidos ou morrer antes do estágio do desenvolvimento de interesse. Para resolver tal problema, geneticistas murinos desenvolveram uma técnica, inteligente de inativação de genes-*alvo* em tipos específicos de células somáticas ou durante determinados estágios do desenvolvimento.

A técnica emprega sítios de recombinação de DNA sítio-específicos (chamados de *sítios loxP*) e a enzima Cre que catalisa a recombinação entre eles. O sistema de recombinação loxP-Cre é derivado do bacteriófago P1, mas o sistema de recombinação sítio-específico também funciona, quando aplicada em células murinas. Uma característica essencial da técnica é o controle da expressão de Cre por um promotor celular-específico. Em camundongos loxP-Cre gerados pelo procedimento representado na Figura 5-42, a inativação do gene de interesse ocorre apenas em células nas quais o promotor que controla o gene Cre está ativo.

Uma aplicação alternativa da técnica forneceu forte evidência de que um determinado receptor de neurotransmissor é importante para a aprendizagem e memória. Estudos farmacológicos e fisiológicos prévios haviam indicado que a aprendizagem normal requer a classe de receptores de glutamato NMDA no hipocampo, uma região do cérebro. No entanto, camundongos nos quais o gene codifican- do o receptor NMDA foi nocutivado não apresentaram aprendizagem na fase neonatal, impossibilitando a análise do papel do receptor no desenvolvimento. Seguindo o protocolo da Figura 5-42, pesquisadores geraram camundongos nos quais o gene da subunidade do receptor estava inativado no hipocampo, mas expresso em outros tecidos. Os

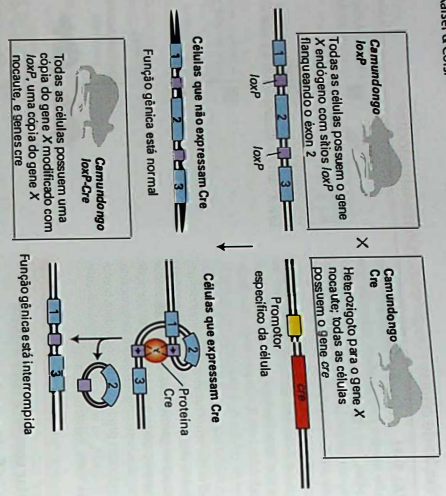


camundongos sobreviveram à vida adulta e apresentaram defeitos de aprendizagem e de memória, confirmando um papel para estes receptores na habilidade dos camundongos de codificar suas experiências em memória.

**Alelos dominantes negativos podem inibir funcionalmente alguns genes**

Em organismos diploides, como observado na Seção 5.1, o efeito fenotípico de um alelo recessivo é expresso apenas em indivíduos homozigotos, enquanto alelos dominantes são expressos em heterozigotos. Assim o in-





**FIGURA EXPERIMENTAL 5-42 O sistema de recombinção loxP-Cre pode nocautear genes em tipos celulares específicos.** Um sítio loxP (lilas) é inserido em cada extremidade do éon essencial 2 do gene *loxP* (X azul) por recombinção homóloga, produzindo um camundongo *loxP*. Uma vez que estão em ítrons, os sítios *loxP* não interagem à função de X. O camundongo Cre porta um báculo nocaute do gene X e um gene cre (laranja) introduzido a partir do bacteriófago P1 ligado a um promotor específico (amarelo) da célula. O gene cre é incorporado ao genoma do camundongo por recombinção não ho-

divíduo precisa portar duas cópias de um alelo recessivo, mas apenas uma cópia de um alelo dominante para exibir os fenótipos correspondentes. Foi visto como linhagens de camundongos que são homozigotas para uma dada mutação-nocaute podem ser produzidas pelo cruzamento de indivíduos que são heterozigotos para a mesma mutação nocaute (ver Figura 5-41). Para experimentos com células de animais em cultivo, entretanto, é geralmente difícil interromper ambas as cópias de um gene para produzir um fenótipo mutante. Além disso, a dificuldade em produzir linhagens com ambas as cópias de um gene mudado com frequência é agravada pela presença de genes relacionados de função semelhante que devem também ser inativados para revelar um fenótipo visível.

Para certos genes, as dificuldades em produzir mutantes-nocaute homozigotos podem ser evitadas pelo uso de um alelo portador de uma mutação dominante recessiva. Estes alelos são geneticamente dominantes; isto é, produzem um fenótipo mutante mesmo em células portadoras de uma cópia do gene selvagem. Contudo, ao contrário de outros tipos de alelos dominantes, alelos dominantes negativos produzem um fenótipo equivalente a quele da inativação de função.

Alelos dominantes negativos úteis foram identificados para uma variedade de genes e podem ser introduzidos em células cultivadas por transfecção ou na linhagem germinativa de camundongos ou outros orga-

**ANIMAÇÃO TÉCNICA:** Criando um camundongo transgênico

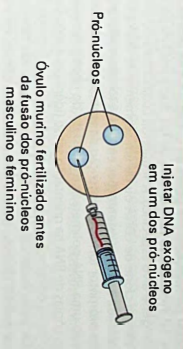
**FIGURA EXPERIMENTAL 5-43 Camundongos transgênicos são produzidos pela integração aleatória de um gene exógeno na linhagem germinativa murina.** DNA exógeno injetado em um dos dois pró-núcleos (os núcleos haploides masculino e feminino) com os quais contribuem os cromossomos do zigoto diploide. Como é totalmente integrado ao cromossomo do zigoto diploide, um transgene não interrompe genes endógenos. (Ver R. L. Brinster et al., 1981, *Cell* 27:223).

do ativado ligado a GTR, impedindo-as, assim, de realizar suas funções de inibidores (Figura 5-44).

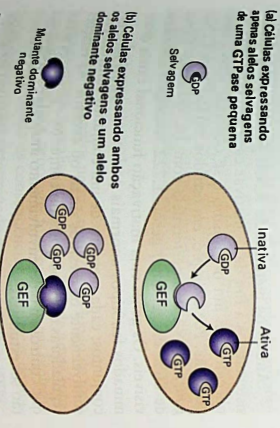
**ORNA de interferência causa a inativação gênica pela destruição do mRNA correspondente**

O fenômeno conhecido como RNA de interferência (RNAi) talvez seja o método mais direto de inibir a função de genes específicos. Essa abordagem é tecnicamente mais simples do que os métodos já descritos para inibição de genes. Inicialmente observado no verme *Caenorhabditis elegans*, o RNAi refere-se à capacidade que o RNA de fita dupla tem de bloquear a expressão de seu mRNA de fita simples correspondente, mas não aquela de mRNAs com uma sequência diferente.

Conforme descrito no Capítulo 8, o fenômeno do RNAi reside na capacidade geral de as células eucarióticas tem de clivar segmentos de RNA de fita dupla em segmentos curtos (23 nt) conhecidos como RNA inibidores pequenos (siRNA). A endonuclease de RNA que catalisa essa reação, conhecida como Dicer, é encontrada em



**FIGURA 5-43 Criando um camundongo transgênico.** DNA exógeno injetado em um dos pró-núcleos de um óvulo murino fertilizado antes da fusão dos pró-núcleos masculino e feminino. Transferir óvulos injetados em mães de aluguel. Cerca de 10 a 30% dos descendentes conterão DNA exógeno em cromossomos de todos os seus tetraploides e na linhagem germinativa. Cruzar camundongos expressando DNA exógeno para propagar o DNA na linhagem germinativa.



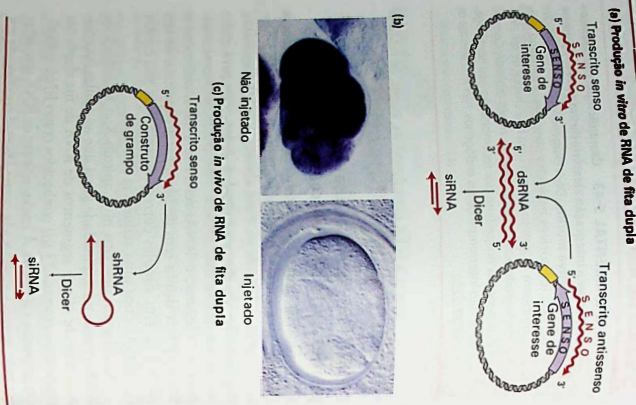
**FIGURA 5-44 Inativação da função de um GTPase selvagem por um alelo mutante dominante negativo.** (a) GTPases (lilas) pequenas (monoméricas) são ativadas por sua interação com GEF (verde) de fita de nucleotídeo guanina (GDP) que catalisa a troca de GDP por GTP. (b) Introdução de um alelo dominante negativo (GTPase mutante) em células cultivadas ou animais transgênicos leva à expressão de uma GTPase mutante que se liga a GEF inativando-o. Como resultado, cópias selvagens endógenas da única GTPase pequena ficam presas no estado inativo ligadas a GDP. Em um único alelo dominante negativo causa um fenótipo de perda de função nos heterozigotos semelhante a quele observado em homozigotos portadores de cópias de perda de função recessivos.

todos os metazoários, mas não em eucariotos simples, como a levedura. As moléculas de siRNA, por sua vez, podem causar a clivagem de moléculas de mRNA de sequência correspondente, em uma reação catalisada por um complexo proteico conhecido como RISC. O RISC medeia o reconhecimento e a hibridização entre uma fita do siRNA e sua sequência complementar no mRNA-alvo; depois, nucleases específicas no complexo RISC clivam o híbrido mRNA-siRNA. Esse modelo representa a especificidade do RNAi. Esse modelo representa a especificidade do RNAi, já que ele depende do pareamento de bases, e o seu potencial para silenciar a função gênica, já que o mRNA complementar é permanentemente destruído por degradação nucleotídica. A função normal de ambos Dicer e RISC é permitir a regulação gênica por pequenas moléculas de RNA endógenas conhecidas como microRNAs, ou miRNAs.

Pesquisadores exploram a via do microRNA para silenciar intencionalmente um gene de interesse usando um dos dois métodos gerais para produção de siRNAs de sequência definida. No primeiro método, um RNA de fita dupla correspondente à sequência do gene-alvo é produzido pela transcrição *in vitro* de ambas as cópias senso e antissenso da sequência (Figura 5-45a). O

**RECURSO DE MÍDIA: RNA de Interferência**

**FIGURA EXPERIMENTAL 5-45 O RNA de interferência (RNAi) pode inativar funcionalmente genes de *C. elegans* de outros organismos.** (a) Produção *in vitro* de RNA de fita dupla (dsRNA) para o RNAi de um gene-alvo específico. A sequência codificadora do gene, derivada de um clone de cDNA ou de um segmento de DNA genômico, é colocada em duas orientações em um vetor plasmidial adequado a um promotor forte. A transcrição de ambos os construtos *in vitro*, usando uma RNA-polimerase e fitosilido de ribonucleosídeos genos, gera cópias de RNA na orientação senso (identicas à sequência do mRNA) ou na orientação antissenso complementar. São condições apropriadas, estas moléculas de RNA complementares irão hibridizar para formar dsRNA. Quando o dsRNA é injetado nas células, ele é clivado por Dicer em siRNAs. (b) Inibição da expressão do RNA *mez1* em embriões do verme por RNAi. (ver o texto para o mecanismo). (A esquerda) A expressão do RNA *mez1* em embriões (fotomontagem por hibridização *in situ* com uma sonda específica para esse RNA, *hiz1*). (A direita) O embrião derivado de um verme injetado com RNA *mez1* de fita dupla autêntica de cor. Cada embrião no estágio de quatro células tem cerca de 50 µm de comprimento. (c) A produção de RNA de fita dupla *in vivo* ocorre por meio de um plasmídeo projetado especificamente para as células. O construto genético sintético é um arranjo sequencial de ambas as sequências senso e antissenso do gene-alvo. Quando transcrita, formam-se pequenos gempos de RNA (siRNA) de fita dupla. O siRNA é clivado por Dicer para formar siRNA. (Parte b) reproduzida de A. Fire et al., 1998, *Nature* 391:806).



RNA de fita dupla (dsRNA) é injetado na gônada de um verme adulto, onde será convertido a siRNA pela Dicer nos embriões em desenvolvimento. Juntamente com o complexo RISC, as moléculas de siRNA causam a destruição rápida de moléculas de mRNA correspondentes. Os vermes resultantes exibem um fenótipo semelhante àquele que resultaria da interrupção do gene correspondente propriamente dito. Em alguns casos, a entrada de apenas algumas moléculas de um determinado dsRNA em uma célula é suficiente para inativar várias cópias do mRNA correspondente. A Figura 5-45b ilustra a capacidade que o dsRNA injetado tem de interferir na produção do mRNA endógeno correspondente em embriões de *C. elegans*. No experimento, os níveis de mRNA dos embriões foram determinados por hibridização *in situ*, como previamente descrito, usando-se uma sonda marcada fluorescentemente.

O segundo método consiste na produção de um RNA de fita dupla específico *in vivo*. Uma maneira eficiente para se fazer isso é expressar um gene sintético projetado para que contenha segmentos adjacentes de ambas as sequências senso e antissenso correspondentes ao gene-alvo (Figura 5-45c). Quando o gene é transcrito, um "gampo" de RNA de fita dupla se forma, sendo chamado de "pequeno gampo de RNA", ou siRNA (de *small hairpin RNA*). O siRNA será clivado pela Dicer para formar moléculas de siRNA. Vetores de expressão lentivirais são particularmente úteis para introduzir genes sintéticos para expressão de construtos de siRNA em células animais.

**CONCEITOS-CHAVE da Seção 5.5**

- Uma vez que um gene é clonado, indicadores importantes acerca de sua função normal *in vivo* podem ser obtidos a partir dos efeitos fenotípicos observados quando ele é inativado.
- Genes podem ser interrompidos em leveduras pela injeção de um gene de marcador de seleção em um alelo de um gene selvagem por meio da recombinação homóloga, produzindo um mutante heterozigoto. Quando um heterozigoto assim é esporulado, a interrupção de um gene essencial produzirá dois esporos haploides inviáveis (ver Figura 5-39).
- Um gene de levedura pode ser inativado de maneira controlada pelo uso do promotor *GAL1* para inativar a transcrição de um gene quando as células são transferidas para um meio com glicose.
- Em amniondógos, genes modificados podem ser incorporados na linhagem germinativa em sua localização genômica original por recombinação homóloga, produzindo nocautes (ver Figuras 5-40 e 5-41). Camundongos-nocaute fornecem modelos para doenças genéticas humanas com o fitofose crítica.
- O sistema de recombinação *loxP-Cre* permite a produção de camundongos nos quais um gene é nocauteado em um tecido específico.
- Na produção de células ou organismos transgênicos, DNA exógeno é integrado no genoma do hospedeiro por recombinação não homóloga (ver Figura 5-43). A introdução de um alelo dominante negativo por esse modo pode inativar funcionalmente um gene sem alterar sua sequência.
- Em muitos organismos, incluindo o verme cilindrico *C. elegans*, moléculas de RNA de fita dupla provocam a destruição de todas as moléculas de mRNA de mesma sequência (ver Figura 5-45). Esse fenômeno, conhecido como RNAi (*RNA de interferência*), fornece um método específico e potente de inativação funcional de genes sem que se alterem suas estruturas.

• Ambos os métodos de RNAi servem para estudos sistêmicos a fim de se inativar cada um dos genes conhecidos em um organismo e se observar suas consequências. Por exemplo, em estudos iniciais com *C. elegans*, o RNA de interferência com 16.700 genes (cerca de 86% do genoma) gerou 1.722 fenótipos anormais visíveis. Os genes cuja inativação funcional causava determinados fenótipos anormais podem ser agrupados em conjuntos, cada membro de um conjunto controla presumivelmente os mesmos sinais ou eventos. As relações reguladoras entre os genes do conjunto – p. ex., os genes que controlam o desenvolvimento muscular – podem então ser estudadas.

Outros organismos nos quais a inativação gênica mediada por RNAi foi bem-sucedida incluem *Drosophila*, vários tipos de plantas, peixe-zebra, o sapo *Xenopus* e camundongos, agora os sujeitos de triagens de RNAi de larga escala. Por exemplo, vetores lentivirais foram projetados para inativar por RNAi mais de 10.000 genes diferentes para expressos em células de mamífero em cultura. A função dos genes inativados pode ser inferida a partir de defeitos no crescimento ou na morfologia dos clones de células transfectados com vetores lentivirais.

rosas bioquímicas fundamentais, a disponibilidade de informação sobre sequências genômicas completas sobre a maioria dos organismos experimentais comuns mudou fundamentalmente a maneira como os experimentos genéticos são conduzidos. Utilizando vários métodos computacionais, cientistas identificaram as sequências de genes codificadores de proteínas da maioria dos organismos experimentais, inclusive *E. coli*, levedura, *C. elegans*, *Drosophila*, *Arabidopsis*, camundongo e seres humanos. As sequências genéticas, por sua vez, revelaram as sequências primárias de aminoácidos dos produtos proteicos codificados, fornecendo-nos uma lista quase completa das proteínas encontradas em cada um dos principais organismos experimentais.

A abordagem adotada pela maioria dos pesquisadores passou da descoberta de novos genes e proteínas para a descoberta das funções de genes e proteínas cujas sequências já são conhecidas. Uma vez identificado um gene interessante, a informação da sequência genômica acerca bastante a manipulação subsequente do gene, inclusive sua inativação projetada, de forma que mais possa ser aprendido sobre sua função.

Conjuntos de vetores para a inativação por RNAi da maioria dos genes definidos no nematode *C. elegans* já permitem que triagens genéticas eficientes sejam realizadas neste organismo multicelular. Os métodos estão agora sendo aplicados em grandes coleções de genes de células de mamífero em cultura, e no futuro próximo, ou RNAi ou o método de nocaute terão sido empregados para inativar cada um dos genes de camundongo.

No passado, um cientista poderia passar muitos anos estudando apenas um único gene, mas hoje cientistas geralmente estudam conjuntos completos de genes ao mesmo tempo. Por exemplo, os microrranços de DNA o nível de expressão de todos os genes de um organismo pode ser medido tão facilmente quanto a expressão de um único gene. Um dos grandes desafios dos geneticistas no século XXI será explorar a grande quantidade de dados disponíveis sobre a função e regulação de genes individuais para entender como grupos de genes se organizam a fim de formar em rotas bioquímicas complexas e redes reguladoras.

**Termos-chave**

- heterozigoto 173
- hibridização 188
- hibridização *in situ* 199
- homodímero 173
- micorranço de DNA 199
- mutação de ponto 173
- mutação de sensibilidade à temperatura 176
- mutagênica 172
- enzimas de restrição 183
- nocaute gênico 213
- genômica 195
- plasmídeos 184

**Perspectivas**

Conforme os exemplos neste capítulo e ao longo de todo o livro mostram, a análise genética é a base de nossa compreensão de muitos processos fundamentais em biologia celular. Examinando as consequências fenotípicas de mutações que inativam determinado gene, geneticistas conseguem conectar os conhecimentos acerca de sequência, estrutura e atividade bioquímica da proteína codificada com sua função no contexto de uma célula viva ou organismo multicelular. A abordagem clássica para o estabelecimento dessas conexões em seres humanos e organismos mais complexos experimentalmente acessíveis tem sido identificar genes mutações de interesse com base em seus fenótipos e, embora o gene a ser estudado e seu produto proteico, *in vitro*, seja conhecido, para dissociar processos celulares e

relação em cadeia da polimerase (PCR) 192	selvagem 172
recombinação 180	sondas 188
RNA de interferência (RNAi) 217	<i>Southern blotting</i> 198
segr. ligação 175	transfecção 203
	transformação 184
	transgenes 216
	vetor 182

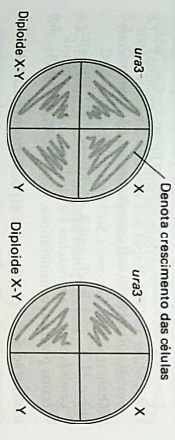
**Revisão dos conceitos**

- Mutações genéticas podem fornecer ideias sobre os mecanismos de complexos processos celulares ou de desenvolvimento. Como sua análise de uma mutação genética poderia ser diferente, conforme determinada mutação seja recessiva ou dominante; minúscula mutação seja recessiva ou dominante?
- O que é uma mutação termosensível? Por que as mutações termosensíveis são úteis para revelar a função de um gene?
- Descreva como a análise de complementação pode ser utilizada para determinar se duas mutações estão no mesmo gene ou em genes diferentes. Explique por que a análise de complementação não funcionará para mutações dominantes.
- Jane isolou uma linhagem mutante de levedura que forma colônias vermelhas em vez de brancas quando crescida em uma placa. Para determinar o gene mutante, ela decidiu usar a complementação funcional com uma biblioteca de DNA que contém um marcador de seleção para listria. Além da mutação genética descrita, as leveduras são desprovidas do gene necessário à síntese dos aminoácidos leucina e lisina. Que meio Jane usará no cultivo de suas leveduras para assegurar que capture os plasmídeos da biblioteca? Como ela saberá quando um plasmídeo da biblioteca terá complementado a mutação de suas leveduras?
- Enzimas de restrição e DNA-ligase desempenham papéis fundamentais na clonagem de DNA. Como uma bactéria que produz uma enzima de restrição não cliva seu próprio DNA? Descreva algumas características gerais dos sítios para enzimas de restrição. Quais são os três tipos de extremidades de DNA que podem ser gerados após sua clivagem com enzimas de restrição? Que reação é catalisada pela DNA-ligase?
- Plasmídeos bacterianos com frequência são usados como vetores de clonagem. Descreva a característica essencial de um vetor plasmídico. Quais são as vantagens e aplicações dos plasmídeos como vetores de clonagem?
- Uma biblioteca de DNA é uma coleção de clones, cada um deles contendo um fragmento diferente de DNA, inserido em um vetor de clonagem. Qual a diferença entre uma biblioteca de cDNA e uma de DNA genômico? Você gostaria de doar o gene X, um gene expresso apenas em neurônios, em um vetor usando uma biblioteca como fonte para o inserto. Se você tivesse as seguintes bibliotecas à disposição (biblioteca genômica de células da pele, biblioteca de cDNA de células da pele, biblioteca genômica de neurônios, biblioteca de cDNA de neurônios), qual delas gostaria usar e por quê?
- Em 1953, Kaye Mullis ganhou o Prêmio Nobel em Química por sua invenção do processo da PCR. Descreva os três passos em cada ciclo de uma reação de PCR. Por que a descoberta de uma DNA-polimerase termossensível (p. ex., *Taq*-polimerase) foi tão importante para o desenvolvimento da PCR?
- Southern* e *Northern blotting* são ferramentas poderosas em biologia molecular baseadas na hibridização de ácidos nucleicos. Como essas técnicas se assemelham? Como se diferenciam? Forneça algumas aplicações específicas para cada técnica de *blotting*.
- Um número de proteínas exógenas foi expressa em células bacterianas e mamíferas. Descreva as características essenciais de um plasmídeo recombinante necessárias à expressão de um gene exógeno. Como você poderia modificar a proteína exógena para facilitar sua purificação? Qual é a vantagem de expressar uma proteína em células de mamíferos em vez de bactérias?
- Northern blotting*, RT-PCR e microrranjos podem ser usados na análise da expressão gênica. Um laboratório estuda células de levedura, comparando seu crescimento na presença de dois açúcares diferentes, glicose e galactose. Um aluno está comparando a expressão do gene *HMG2* sob diferentes condições. Quais técnicas (ele poderia usar e por quê? Outro aluno quer comparar a expressão de todos os genes do cromossomo 4, os quais são cerca de 800. Qual técnica(s) ele poderia usar e por quê?
- Na determinação da identidade de uma proteína que corresponde a um gene recém-descoberto, geralmente ajuda saber o padrão de expressão tecidual daquele gene. Por exemplo, pesquisadores descobriram que um gene chamado *SERPINA6* é expresso no fígado, no rim e no pâncreas, mas não em outros tecidos. Quais técnicas os pesquisadores poderiam usar para descobrir que tecidos expressam um determinado gene?
- Polimorfismos de DNA podem ser usados como marcadores de DNA. Descreva as diferenças entre polimorfismos SNP e SSR. Como esses marcadores podem ser usados para estudos de mapeamento do DNA?
- Como o mapeamento de desequilíbrio de ligação pode às vezes fornecer a localização de um gene com resolução muito maior do que o clássico mapeamento de ligação?
- Estudos de ligação genética podem localizar apenas grosseiramente a posição cromossômica de um gene de "doença". Como a análise de expressão e a análise de seqüências de DNA ajudam a localizar um gene de doença dentro da região identificada por mapeamento de ligação?

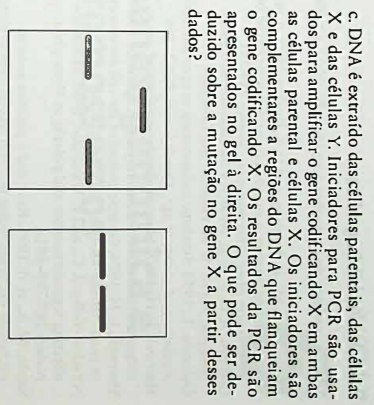
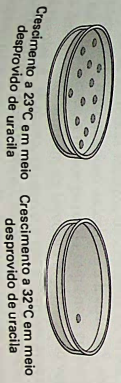
- A capacidade de modificar seletivamente o genoma do camundongo revolucionou a genética murina. Faça um esquema demonstrando o procedimento para se gerar um camundongo-nocaut de um locus genético específico. Como o sistema *LoxP-Cre* pode ser usado para nocautar um gene de maneira condicional? Que aplicação médica importante têm os camundongos-nocaut?
- Dois métodos para inativar funcionalmente um gene sem alterar sua seqüência incluem mutações dominantes negativas e RNA de interferência (RNAi). Descreva como cada método pode inibir a expressão de um gene.

**Análise dos dados**

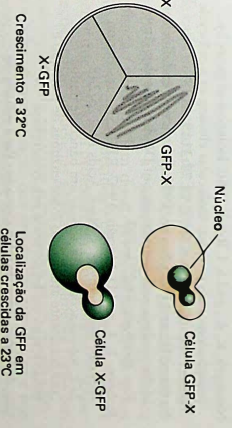
- Uma cultura de leveduras que requer uracila para crescer (*ura3*<sup>-</sup>) foi mutagenizada, e duas colônias mutantes, X e Y, foram isoladas. Células de acasalamento do tipo a da colônia mutante X são cruzadas com células de acasalamento do tipo α da colônia Y para gerar células diploides. As células parentais (*ura3*<sup>-</sup>, X e Y) e diploides são semeadas em placas com ágar contendo uracila e incubadas a 23°C ou 32°C. O crescimento celular foi monitorado pela formação de colônias nas placas de cultura conforme demonstrado na figura abaixo.



- O que pode ser deduzido sobre os mutantes X e Y a partir dos dados fornecidos?
- Uma biblioteca de cDNA de leveduras selvagens, preparada em um plasmídeo que contém o gene selvagem *URA3*<sup>+</sup>, é usada para transformar células, as quais são então cultivadas como indicado. Cada ponto preto abaixo representa um único clone crescendo em uma placa de Petri. Quais são as diferenças moleculares entre os clones crescendo nas duas placas? Como os resultados podem ser usados para identificar o plasmídeo que contém uma cópia do gene X selvagem? Baseando-se nesses resultados, como pode a identidade do gene X ser descoberta?



- Um construído do gene selvagem X é projetado para codificar uma proteína de fusão na qual a proteína fluorescente verde (GFP) está presente na porção N-terminal (GFP-X) ou C-terminal (X-GFP) da proteína X. Ambos os construídos, presentes em um plasmídeo *URA3*<sup>+</sup>, são usados para transformar células X crescidas na ausência de uracila. As transformantes são, então, monitoradas durante o crescimento a 32°C, mostrado abaixo à esquerda. A direita estão imagens fluorescentes típicas de células X-GFP e GFP-X crescidas a 23°C nas quais a cor verde denota a presença da proteína fluorescente verde. Que explicação sensata pode haver para o crescimento de células GFP-X, mas não de X-GFP, a 32°C?



- Descendentes haploides das células diploides da parte (a) são geradas. Mutantes duplos XY constroem 1/4 destes descendentes. Células X e Y haploides, e células XY em cultivo líquido são sincronizadas em um estágio logo antes do brotamento e então passadas a 23°C para 32°C. O exame das células 24 horas depois revela que as células X pararam de proliferar com pequenos brotamentos, as células Y pararam de proliferar com grandes brotamentos, e as células XY pararam de proliferar com pequenos brotamentos. Qual é a relação entre X e Y?

**Referências bibliográficas**

Análise genética de mutações para identificação de genes de genes

Adams, A. E. M., D. Bostein, and D. B. Drubin. 1989. A yeast actin-binding protein is encoded by *sec6*, a gene found by suppression of an actin mutation. *Science* 243:231.

Griffiths, A. G. F., et al. 2000. *Artificially Directed Genetic Analysis*. 7th ed. W. H. Freeman and Company.

Guarente, L. 1993. Synthetic enhancement in gene interaction: a genetic tool comes of age. *Trends Genet.* 9:362-366.

Harwell, L. H. 1967. Macromolecular synthesis of temperate sensitive mutants of yeast. *J. Bacteriol.* 93:1662.

Harwell, L. H. 1974. Genetic control of the cell division cycle in yeast. *Science* 183:46.

Nybakken-Volhard, C., and E. Wieschaus. 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287:795-801.

Simon, M. A., et al. 1991. Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the yeast protein tyrosine kinase. *Cell* 67:701-716.

Tong, A. H., et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294:2364-2368.

**Clonagem e caracterização do DNA**

Anshel, F. M., et al. 2002. *Current Protocols in Molecular Biology*. Wiley.

Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 252:53-589.

Han, J. H., C. Stratawa, and W. J. Rutter. 1987. Isolation of full-length putative rat tyrosine phosphatase cDNA using improved methods for mRNA isolation and cDNA cloning. *Biochem.* 26:6167-1632.

Itakura, K., J. I. Roski, and R. B. Wallace. 1984. Synthesis and use of synthetic oligonucleotides. *Ann. Rev. Biochem.* 53:323-356.

Mamiani, T., et al. 1978. The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 15:887-901.

Nasmyth, K. A., and S. I. Reed. 1980. Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. *Proc. Natl Acad. Sci. USA* 77:2119-2123.

Nathans, D., and H. O. Smith. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Ann. Rev. Biochem.* 44:273-293.

Roberts, R. J., and D. Macelis. 1997. REBASE - restriction enzymes and methylases. *Nucl. Acids Res.* 25:248-262.

262. Informação para acessar um banco de dados continuamente atualizado sobre enzimas de restrição e modificação em <http://www.neb.com/rtbase>.

**Uso de fragmentos de DNA clonados para estudo de expressão gênica**

Andrews, A. T. 1986. *Electrophoresis*, 2d ed. Oxford University Press.

Efthymiou, H., ed. 1992. *PCR Technology: Principles and Applications for DNA Amplification*. W. H. Freeman and Company.

Reilcke, A. M., W. G. A. Axel, and S. Silverstein. 1978. The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. *Cell* 41:133-141.

Saito, R. A., et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:948-951.

Sanger, F. 1981. Determination of nucleotide sequences in DNA. *Science* 214:1205-1210.

Souza, L. M., et al. 1986. Recombinant human granulocyte colony stimulating factor: effects on normal and granulocyte myeloid cells. *Science* 232:61-65.

Walsh, G. M., J. L. Meinkoth, and A. R. Kimmel. 1987. Northern and Southern Blots. *Methods Enzymol.* 152:572-581.

Wallace, R. B., et al. 1981. The use of synthetic oligonucleotides as hybridization probes. II: Hybridization of oligonucleotides of mixed sequence to riblet  $\beta$ -globin DNA. *Nucl. Acids Res.* 9:879-887.

**Localização e identificação de genes de doenças humanas**

Bostein, D., et al. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Genet.* 32:314-331.

Donis-Keller, H., et al. 1987. A genetic linkage map of the human genome. *Cell* 51:319-337.

Harwell, L., et al. 2006. *Genetics: From Genes to Genomes*. McGraw-Hill.

Harsbacka, T., et al. 1994. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 78:1073.

Ohta, M., et al. 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5:874.

Tabor, H. K., N. J. Risch, and R. M. Myers. 2002. Opinion: candidate gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.* 3:391-397.

**Inativação da função de genes específicos em eucariotos**

Capechi, M. R. 1989. Altering the genome by homologous recombination. *Science* 244:1288-1292.

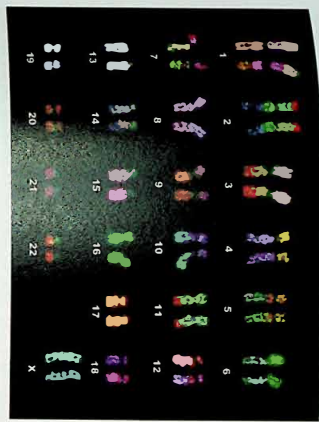
DeSilva, R. J., et al. 1988. A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature* 332:800-805.

Fitz, A., et al. 1998. Roten and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.

Gou, H., et al. 1994. Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265:103-106.

Zamore, P. D., et al. 2000. RNAi double-stranded RNA directs the ATP-dependent cleavage of mRNA in 23 nucleotide intervals. *Cell* 101:25-33.

Zimmer, A. 1992. Manipulating the genome by homologous recombination in embryonic stem cells. *Ann. Rev. Neurosci.* 15:113.



Os cromossomos marcados com a técnica de R-banding são ao mesmo tempo brancos e úteis para a visualização de anormalias e para a comparação de cariótipos de diferentes espécies. (© Departamento de Genética Clínica, Hospital Adenauer/Phibro Researches, Inc.)

**SUMÁRIO**

6.1 Estrutura gênica dos eucariotos	225	6.5 Genômica: análise da estrutura e expressão de genes em genomas eucarióticos	252
6.2 Organização cromossômica dos genes e do DNA não codificante	231	6.6 Organização estrutural dos cromossomos eucarióticos	256
6.3 Elementos móveis de DNA transportáveis	234	6.7 Morfologia e elementos funcionais dos cromossomos eucarióticos	266
6.4 DNA de organelas	245		

**Genes, genômica e cromossomos**

**CAPÍTULO 6**

**E**m capítulos anteriores, foi discutido como a estrutura e a composição das proteínas permitem que elas desempenhem uma grande variedade de funções celulares. Também foi visto outro componente vital das células, os ácidos nucleicos, e o processo pelo qual a informação codificada na sequência de DNA é traduzida em proteína. Neste capítulo, nosso foco novamente será as características dos genomas eucarióticos e organelas de DNA que compõem o genoma e o modo como as proteínas estruturam e organizam este DNA no interior da célula.

No início do século XXI, foi completado o sequenciamento de genomas inteiros de centenas de vírus, de bactérias e de um eucarioto unicelular, a levedura *S. cerevisiae*. Atualmente, a maior parte da sequência genômica foi determinada também na levedura *S. pombe*, no vegetal simples *A. thaliana*, no arroz e em múltiplos animais multicelulares (metazoários), inclusive no nematódeo *C. elegans*, na mosca-da-fruta *D. melanogaster*, em camundongos, em humanos e pelo menos em um representante de cada um dos cerca de 35 filos de metazoários. A análise detalhada dos dados obtidos pelo sequenciamento revelou aspectos sobre a evolução, organização genômica e função gênica. Além disso, permitiu identificar genes previamente desconhecidos e estimar o número total de genes que codificam proteínas em cada genoma. Comparações entre sequências gênicas normalmente fornecem indicações para possíveis funções dos genes recém-identificados. A comparação das sequências genômicas e sua organização entre espécies também auxilia a compreensão da evolução dos organismos.

Surpreendentemente, o sequenciamento de DNA revelou que uma grande proporção do genoma de eucariotos superiores não codifica mRNAs ou qualquer outro RNA necessário ao organismo. É notável que este DNA não codificante constitui cerca de 98,5% do DNA cromossômico humano! O DNA não codificante presente nos organismos multicelulares contém diversas regiões semelhantes, mas não idênticas. As variações nas porções desses DNAs repetitivos entre indivíduos são tão impressionantes que cada pessoa pode ser distinguida por uma "impressão digital" de DNA baseada nas variações dessas