

Uma alternativa à marcação com GFP para detectar a localização intracelular de uma proteína é modificar o gene de interesse fusionando-o a uma pequena sequência de DNA que codifica uma curta região de aminoácidos reconhecida por um anticorpo monoclonal conhecido. Esse peptídeo curto que pode ser ligado por um anticorpo é chamado de epítipo; assim, o método é conhecido como *marcação de epítipo*. Após a transfecção com um plasmídeo de expressão contendo o gene modificado, a proteína marcada com epítipo expressa pode ser detectada por marcação imunofluorescente das células com o anticorpo monoclonal específico para o epítipo. A escolha entre o uso de um epítipo 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 clonados 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 corregulados 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 clonado.
- Vetores de expressão eucarióticos podem ser usados para expressar genes clonados em leveduras ou células de mamíferos. Uma aplicação importante desses métodos é a marcação de proteínas com GFP ou um epítipo 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 comum. 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 comparação de sequências de um gene de doença e seu produto com aquelas de genes e proteínas cujas sequência e função 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^s* 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 enfim 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 exibem diferentes padrões de herança, dependendo da natureza e da localização cromossômica dos alelos 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 *autosômico 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
Autossômicas recessivas		
Anemia falciforme	Hemoglobina anormal causa deformação das hemácias, que podem ficar presas 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 (tirosina hidroxilase) resulta em excesso de fenilalanina, causando retardo 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 excesso nos lisossomos dos neurônios, prejudicando o desenvolvimento neural	1/1.000 judeus do leste europeu
Autossômicas recessivas		
Síndrome de Huntington	Proteína neural defeituosa (huntingina) 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	1/122 franco-canadenses
Recessivas ligadas ao X		
Distrofia muscular de Duchenne	Proteína citoesquelética distrofina defeituosa leva à função muscular prejudicada	1/3.500 homens
Hemofilia A	Elemento de coagulação sanguínea fator VII 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 (p. ex., 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 terceiro 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 e, portanto, fenotipicamente 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

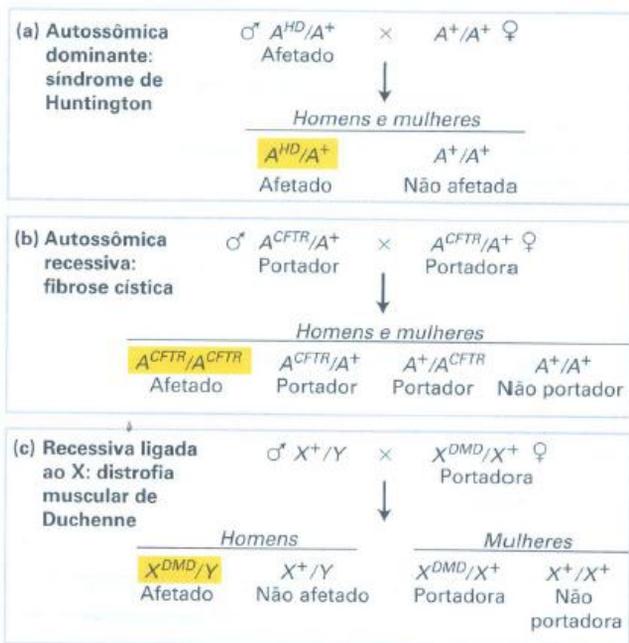


FIGURA 5-35 Três padrões de herança comuns para doenças genéticas humanas. Autossomos (A) e cromossomos sexuais (X e Y) selvagens são indicados por sinais de soma sobrescritos. (a) Em um distúrbio autossômico dominante como a doença de Huntington, apenas um alelo mutante é necessário para produzir a doença. Se um dos genitores for heterozigoto para o alelo mutante *HD*, 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 *CFTR* 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 exibe o típico padrão de segregação ligado ao sexo. Homens nascidos de mães heterozigotas para um alelo *DMD* mutante possuem 50% de chance de herdar 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 a 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 espalhados 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 associa-

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 diferenças nucleotídicas entre indivíduos não relacionados sejam detectadas em média a cada 10^3 nucleotídeos. Se essas variações na sequência de DNA, chamadas de *polimorfismos de DNA*, puderem ser acompanhadas de uma geração para a outra, poderão servir como marcadores genéticos para estudos de ligação. Atualmente, um painel com cerca de 10^4 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 (SSRs) ou *microsatélites*, são formados presumivelmente por recombinação ou por um mecanismo de deslizamento ou do molde ou das fitas recém-sintetizadas durante a replicação do DNA. Uma característica útil das SSRs é que indivíduos diferentes geralmente terão 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 heredograma e, portanto, têm 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

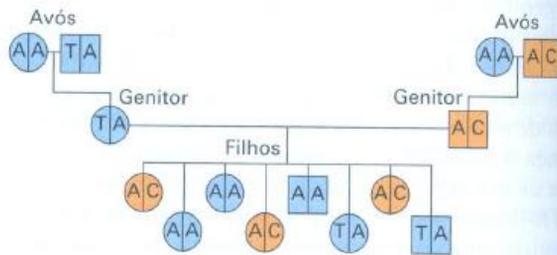


FIGURA EXPERIMENTAL 5-36 Polimorfismos de nucleotídeo único (SNPs) podem ser acompanhados como marcadores genéticos. Heredograma hipotético baseado na análise de SNPs do DNA de uma região de um cromossomo. Nesta família, o SNP existe como A, T ou C. Cada indivíduo possui dois alelos: alguns têm A em ambos os cromossomos, e outros são heterozigotos neste 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 segrega com um C no SNP.

mapeado. A primeira etapa é obter amostras de DNA de todos os membros de uma família contendo indivíduos que apresentam a doença. O DNA de cada indivíduo afetado e não afetado é analisado para se identificar um grande número de polimorfismos de DNA conhecidos (marcadores tipo SSR ou SNP podem ser usados). O padrão de segregação de cada polimorfismo de DNA em uma família é comparado com a segregação da doenças estudadas para se encontrar aqueles polimorfismos que tendem a segregar juntamente com a doença. Por fim, a análise computacional dos dados de segregação é utilizada para o cálculo da probabilidade de ligação entre cada polimorfismo de DNA e o alelo causador da doença.

Na prática, os dados de segregação são coletados de diferentes famílias que exibem a mesma doença e agrupados. Quanto mais famílias que exibem determinada doença puderem ser examinadas, maior será a significância estatística de evidência de ligação a ser obtida e maior a precisão com a qual a distância será medida entre um polimorfismo de DNA e um alelo de doença ligados. A maioria dos estudos familiares possui um máximo de 100 indivíduos nos quais a ligação entre um gene de doença e um painel de polimorfismos de DNA pode ser testada. Esse número de indivíduos define o limite superior prático da resolução do estudo de mapeamento em cerca de 1 centimorgan, ou uma distância física de aproximadamente $7,5 \times 10^5$ pares de bases.

Um fenômeno chamado de *desequilíbrio de ligação* é a base para uma estratégia alternativa, que geralmente permite um grau de resolução maior em estudos de mapeamento. A abordagem depende de circunstâncias particulares nas quais uma doença genética comumente encontrada em determinada população resulte de uma única mutação ocorrida há muitas gerações. Os polimorfismos de DNA carregados pelo cromossomo ancestral são coletivamente conhecidos como *haplótipo* daquele cromossomo. À medida que a doença é transmitida de uma geração para a outra, apenas os polimorfismos mais próximos ao gene da doença não serão separados dele por recombinação. Após muitas gerações, a região que contém o gene da doença ficará evidente porque será a única região do cromossomo que portará o haplótipo do cromossomo ancestral conservado por muitas gerações (Figura 5-37). Ao acessar a distribuição de marcadores específicos em todos os indivíduos afetados em uma população, os geneticistas podem identificar marcadores de DNA associados à doença, localizando o gene associado com a doença em uma região relativamente pequena. Em condições ideais, estudos de *desequilíbrio de ligação* melhoram a resolução para menos de 0,1 centimorgan. O poder de resolução do método provém da habilidade de se determinar se um polimorfismo e o alelo da doença foram alguma vez separados por um evento de recombinação meiótica em qualquer momento desde o surgimento do alelo da doença no cromossomo ancestral – em alguns casos, isso inclui a busca de marcadores que estão tão próximos ao gene da doença, que, mesmo após centenas de meioses, nunca tenham sido separados por recombinação.

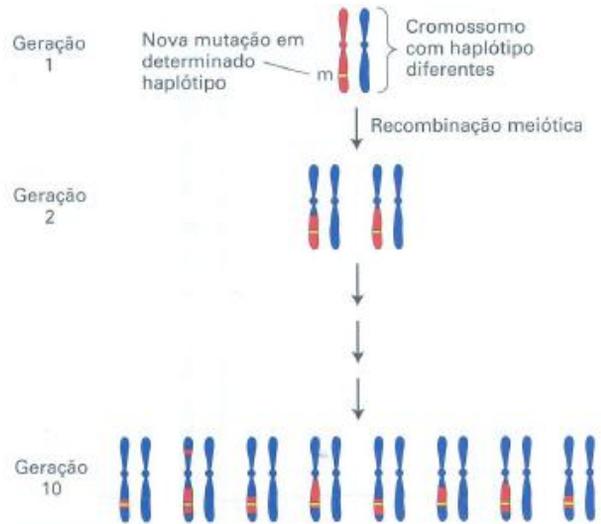


FIGURA 5-37 Estudos de desequilíbrio de ligação de populações humanas podem ser usados para mapear genes em alta resolução. Uma nova mutação de doença irá surgir no contexto de um cromossomo ancestral dentre um conjunto de polimorfismos conhecido como *haplótipo* (indicado em vermelho). Após várias gerações, os cromossomos portadores da mutação da 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 gerais derivados da população em geral, e não 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 triagem 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 10^5 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 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 determinada por *Northern blotting* ou hibridização *in situ* de DNA ou RNA marcado em seções do tecido. Experimentos de *Northern blots*, hibridização *in*

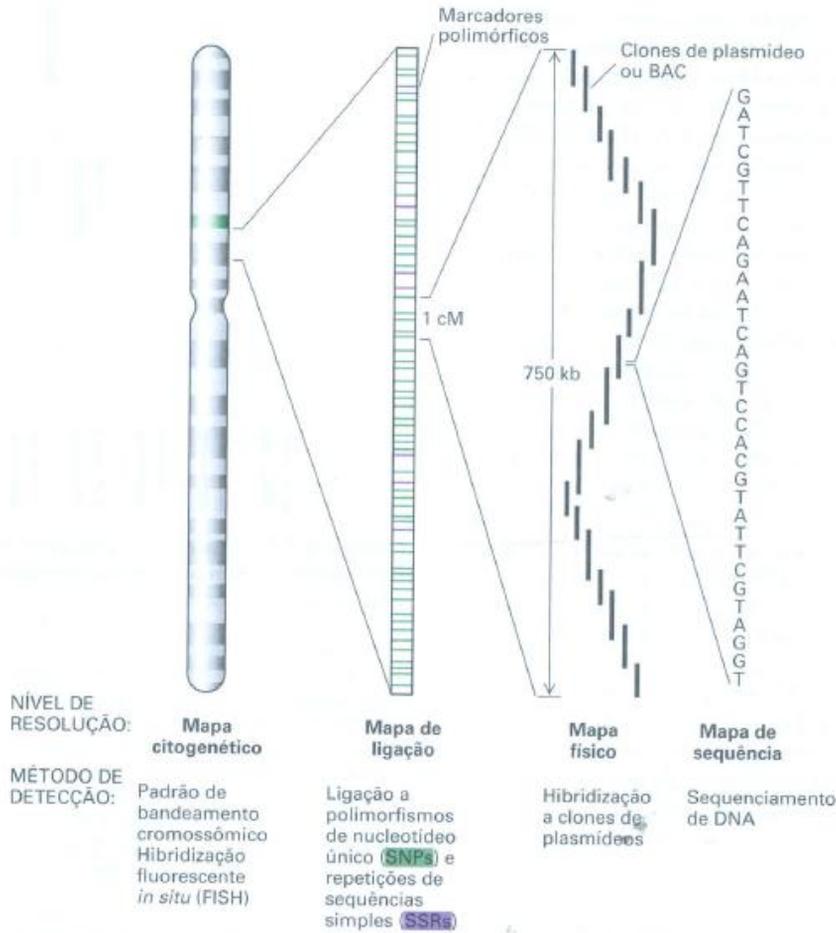


FIGURA 5-38 A relação entre os mapas genético e físico de um cromossomo humano. O diagrama representa um cromossomo analisado em diferentes níveis de resolução. O cromossomo como um todo pode ser visto sob a luz do microscópio quando se encontra em estado condensado que ocorre na metáfase, e a localização aproximada de sequências específicas pode ser determinada por hibridização 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 hibridização de *Southern* ou PCR. Finalmente, importantes diferenças genéticas podem ser definidas de forma mais precisa por diferenças na sequência de nucleotídeos do DNA cromossômico. (Adaptada de L. Hartwell et al., 2003, *Genetics: From Genes to Genomes*, 2. ed., McGraw-Hill.)

situ ou microarranjo permitem comparar ambos o nível de expressão e o tamanho dos mRNAs em tecidos mutantes e selvagens. Embora a sensibilidade da hibridização *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 de ponto. A estratégia global é buscar por sequências codificadoras que mostram de forma consistente alterações possivelmente deletérias no DNA de indivíduos que exibem a doença. Uma limitação da abordagem é que a região próxima ao gene afetado pode apresentar polimorfismos 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, quanto 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 compreendidas em nível molecular são doenças monogênicas; isto é, um distúrbio claramente discernível é produzido por um defeito em um único gene. Doenças monogênicas causadas por mutação em um gene específico exibem um dos padrões 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 subjacente muito 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 apresentada 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 dependia 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. Seguindo essa pista, os geneticistas concentraram sua atenção naqueles genes que são altamente expressos na retina quando triavam 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 dissecação 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 propriedade hereditária. Essas e muitas outras doenças são consideradas *poligênicas* no sentido de que alelos de múltiplos genes, atuando juntos em um indivíduo, contribuem para a ocorrência e gravidade da doença. Como mapear sistematicamente características poligênicas complexas em seres humanos 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 propensão 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 predispõe 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 candidatos, 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 análogos àqueles 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 contribuam 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 codificantes do genoma.
- Polimorfismos de DNA úteis no mapeamento de genes humanos incluem os polimorfismos de nucleotídeo único (SNPs) e repetições 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 similaridade 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 destas “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.

Três abordagens básicas estão na base dessas técnicas de inativação gênica: (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 compreende o marcador de seleção (p. ex., o gene *kanMX*, que assim como o *neo* confere 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 vias de segurança ou 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, este método fornece pouca informação acerca da verdadeira função da proteína codificada nas células. Para aprender mais a respeito da contribuição de um gene