

# Diversidade Genética Humana: Mutações e Polimorfismo

O estudo da variação genética e genômica é o pilar conceitual para a genética na medicina e para o campo mais amplo da genética humana. Durante o curso da evolução, o fluxo constante de nova variação de nucleotídeos tem assegurado um alto grau de diversidade genética e individualidade, e este tema se estende através de todos os campos da genética médica e humana. A diversidade genética pode manifestar-se como diferenças na organização do genoma, como alterações de nucleotídeos na sequência do genoma, como variações no número de cópias de grandes segmentos de DNA genômico, como alterações na estrutura ou na quantidade de proteínas encontradas em vários tecidos, ou como qualquer um destes no contexto de doenças clínicas.

Este capítulo é um dos vários em que exploraremos a natureza das diferenças geneticamente determinadas entre os indivíduos. A sequência de DNA nuclear é aproximadamente 99,5% idêntica entre dois seres humanos não aparentados. Ainda, é precisamente a diferença na pequena fração da sequência de DNA entre indivíduos que é responsável pela variabilidade geneticamente determinada, a qual é evidente tanto na existência diária quanto na medicina clínica. Muitas diferenças nas sequências de DNA têm pouco ou nenhum efeito na aparência externa, ao passo que outras diferenças são diretamente responsáveis por causar doença. Entre esses dois extremos está a variação responsável pela variabilidade geneticamente determinada na anatomia, na fisiologia, nas intolerâncias alimentares, na suscetibilidade à infecção, na predisposição ao câncer, nas respostas terapêuticas ou nas reações adversas a medicamentos, e talvez até mesmo a variabilidade em vários traços de personalidade, aptidão atlética e talento artístico.

Um dos conceitos importantes da genética humana e médica é que doenças com um componente claramente hereditário são apenas a manifestação mais óbvia e muitas vezes mais extrema de diferenças genéticas, uma das extremidades de um *continuum* de variação que se estende desde variantes deletérias raras que causam doença, através de variantes mais comuns que podem aumentar a suscetibilidade à doença, até a variação mais comum na população, a qual apresenta relevância incerta em relação à doença.

## A NATUREZA DA VARIAÇÃO GENÉTICA

Conforme descrito no Capítulo 2, um segmento de DNA que ocupa uma posição ou localização particular no cromossomo é um *locus* (plural, *loci*). O *locus* pode ser grande,

como um segmento de DNA que contém muitos genes, tal qual o *locus* do complexo principal de histocompatibilidade, envolvido na resposta do sistema imunológico a substâncias estranhas; pode ser um gene único, tal como o *locus* da  $\beta$ -globina que introduzimos no Capítulo 3; ou pode ainda ser apenas uma base única no genoma, como no caso de um único nucleotídeo variante (Fig. 2-6 e mais adiante neste capítulo). Versões alternativas da sequência de DNA no *locus* são chamadas de **alelos**. Para muitos genes, há um único alelo predominante, em geral presente em mais da metade dos indivíduos em uma população, que os geneticistas chamam de **tipo selvagem** ou alelo comum. (Em linguagem leiga, isso é algumas vezes referido como o alelo “normal”. No entanto, como a variação genética é por si só muito “normal”, a existência de alelos diferentes em indivíduos “normais” é trivial. Assim, deve-se evitar o uso “normal” para designar o alelo mais comum.) As outras versões do gene são alelos **variantes** (ou **mutantes**) que diferem do alelo selvagem, devido à presença de uma **mutação**, uma alteração permanente na sequência de nucleotídeos ou na disposição do DNA. Note que os termos *mutação* e *mutantes* referem-se ao DNA, mas não aos seres humanos portadores dos alelos mutantes. Os termos denotam uma alteração na sequência, mas por outro lado não transmitem qualquer conotação com respeito à função ou à capacidade dessa alteração.

A frequência de variantes diferentes pode variar amplamente em diferentes populações ao redor do mundo, como exploraremos detalhadamente no Capítulo 9. Se houver dois ou mais alelos relativamente comuns (definidos por convenção como tendo uma frequência alélica >1%) em um *locus* na população, diz-se que esse *locus* apresenta polimorfismo (literalmente “muitas formas”) nessa população. A maioria dos alelos variantes, no entanto, não é suficientemente frequente na população para ser considerada como polimorfismos; alguns são tão raros a ponto de serem encontrados em apenas uma única família e são conhecidos como alelos “particulares”.

## O Conceito de Mutações

Neste capítulo, começaremos a explorar a natureza da **mutação**, variando desde a alteração de um nucleotídeo único a alterações em um cromossomo inteiro. Para se reconhecer uma mudança deve-se comparar aquilo que a

variante mostra ser uma diferença com um “padrão-ouro”. Como vimos no Capítulo 2, não há um único indivíduo cuja sequência do genoma poderia servir como padrão para a espécie humana, e assim a sequência ou arranjo mais comum na população em qualquer posição no genoma foi arbitrariamente designada como a sequência de referência (Fig. 2-6). À medida que mais genomas de indivíduos ao redor do mundo são amostrados (e, portanto, mais variação é detectada entre os atualmente sete bilhões de genomas que compõem nossa espécie), este genoma de referência está sujeito a avaliação e mudanças constantes. De fato, uma série de colaborações internacionais compartilha e atualiza dados de ações sobre a natureza e a frequência de variação no DNA em diferentes populações no contexto da sequência de referência do genoma humano, e disponibiliza os dados através de bancos de dados públicos, que servem como recursos essenciais para cientistas, médicos e outros profissionais da saúde (Tabela 4-1).

As mutações são por vezes classificadas pelo tamanho da sequência de DNA alterada e, em outros momentos, pelo efeito funcional da mutação na expressão gênica. Embora a classificação por tamanho seja um pouco arbitrária, pode ser útil conceitualmente para distinguir as mutações em três níveis diferentes:

- Mutações que deixam cromossomos intactos, mas que alteram o número de cromossomos de uma célula (**mutações cromossômicas**).
- Mutações que mudam apenas uma parte do cromossomo e podem envolver uma alteração no número de cópias de um segmento subcromossômico ou um rearranjo estrutural que envolve partes de um ou mais cromossomos (**mutações regionais ou subcromossômicas**).
- Alterações na sequência de DNA que envolvem a substituição, deleção ou inserção de DNA, variando de um nucleotídeo único até um limite definido de modo arbitrário de aproximadamente 100 kb (**mutações gênicas ou de DNA**).

A base para e as consequências desse terceiro tipo de mutação são o principal foco deste capítulo, enquanto mutações cromossômicas e regionais serão apresentadas nos Capítulos 5 e 6.

Dependendo da localização precisa, da natureza e do tamanho da mutação no DNA, as suas consequências funcionais, mesmo aquelas que alteram um único par de bases, podem ir desde consequências completamente inócuas até causar sérias doenças. Por exemplo, uma mutação dentro de um éxon codificante de um gene pode não ter nenhum efeito sobre a forma como esse gene é, se a alteração não afetar a sequência primária de aminoácidos do produto polipeptídico; mesmo que isso aconteça, a modificação resultante na sequência de aminoácidos codificada pode não alterar as propriedades funcionais da proteína. Portanto, nem todas as mutações se manifestam em um indivíduo.

## O Conceito de Polimorfismo Genético

A sequência de DNA de uma determinada região do genoma é notavelmente semelhante entre os cromossomos transportados por muitos indivíduos diferentes em todo o mundo. De fato, qualquer segmento de DNA humano de aproximadamente 1.000 pb de comprimento, escolhido ao acaso, contém, em média, apenas um par de bases que é diferente entre os dois cromossomos homólogos herdados dos pais (assumindo que os pais não tenham parentesco). No entanto, em todas as populações humanas, têm sido identificados e catalogados muitas dezenas de milhões de diferenças de um único nucleotídeo e mais de um milhão de variantes mais complexas. Devido à amostragem limitada, esses números provavelmente subestimam a verdadeira extensão da diversidade genética em nossa espécie. Muitas populações ao redor do mundo ainda têm de ser estudadas, e mesmo naquelas que foram estudadas, o número

**TABELA 4-1 Bancos de Dados Úteis sobre Informações da Diversidade Genética Humana**

Descrição	URL
O Projeto Genoma Humano, concluído em 2003, foi uma colaboração internacional para sequenciar e mapear o genoma da nossa espécie. O rascunho da sequência do genoma foi divulgado em 2001, e a montagem do genoma de referência “essencialmente completo” foi publicada em 2004.	<a href="http://www.genome.gov/10001772">http://www.genome.gov/10001772</a> <a href="http://genome.ucsc.edu/cgi-bin/hgGateway">http://genome.ucsc.edu/cgi-bin/hgGateway</a> <a href="http://www.ensembl.org/Homo_sapiens/Info/Index">http://www.ensembl.org/Homo_sapiens/Info/Index</a>
O Single Nucleotide Polymorphism Database (dbSNP) e o Structural Variation Database (dbVar) são bancos de dados de variações em pequena e larga escala, incluindo variantes de nucleotídeo único, microssatélites, indels e CNVs.	<a href="http://www.ncbi.nlm.nih.gov/snp/">http://www.ncbi.nlm.nih.gov/snp/</a> <a href="http://www.ncbi.nlm.nih.gov/dbvar/">http://www.ncbi.nlm.nih.gov/dbvar/</a>
O 1.000 Genomes Project está sequenciando os genomas de um grande número de indivíduos para fornecer uma fonte abrangente sobre a variação genética em nossa espécie. Todos os dados estão disponíveis publicamente.	<a href="http://www.1000genomes.org">www.1000genomes.org</a>
O Human Gene Mutation Database é uma coleção abrangente de mutações germinativas associadas a ou causadoras de doenças hereditárias humanas (atualmente incluindo mais de 120.000 mutações em 4.400 genes).	<a href="http://www.hgmd.org">www.hgmd.org</a>
O Database of Genomic Variants é um catálogo de curadoria de variações estruturais no genoma humano. Desde 2012, o banco de dados contém mais de 400.000 entradas, incluindo mais de 200.000 CNVs, 1.000 inversões e 34.000 indels.	<a href="http://dgv.tcag.ca">http://dgv.tcag.ca</a>
O Japanese Single Nucleotide Polymorphisms Database (JSNP Database) relata SNPs descobertos como parte do Millennium Genome Project.	<a href="http://snp.ims.u-tokyo.ac.jp/">http://snp.ims.u-tokyo.ac.jp/</a>

CNV, variação no número de cópias; SNP, polimorfismo de nucleotídeo único.

Atualizado de Willard HF: The human genome: a window on human genetics, biology and medicine. In Ginsburg GS, Willard HF, editors: *Genomic and personalized medicine*, ed 2, New York, 2013, Elsevier.

de indivíduos examinados é demasiadamente pequeno para revelar a maioria das variantes com frequências alélicas menores, abaixo de 1% a 2%. Assim, à medida que mais pessoas são incluídas nos projetos de descoberta de variantes, variantes adicionais (e raras) serão certamente descobertas.

Se uma variante é considerada formalmente como um polimorfismo ou não, depende inteiramente de se sua frequência populacional é superior a 1% dos alelos na população, e não do tipo de mutação que o causou, de quão grande é o segmento do genoma envolvido, ou se ele tem um efeito aparente sobre o indivíduo. A localização de uma variante em um gene também não determina se a variante é um polimorfismo. Embora a maioria dos polimorfismos de sequência esteja localizada entre genes ou dentro de íntrons e seja irrelevante para o funcionamento de qualquer gene, outros podem estar localizados na sequência codificante dos próprios genes e resultar em proteínas variantes diferentes que podem, por sua vez, levar a diferenças distintivas em populações humanas. Outros ainda estão em regiões reguladoras e também podem ter efeitos importantes sobre a transcrição ou a estabilidade do RNA.

Pode-se esperar que mutações deletérias que causam doenças monogênicas raras provavelmente sejam muito raras para atingir a frequência necessária para serem consideradas um polimorfismo. Embora seja verdade que os alelos responsáveis pela maioria das condições clínicas claramente hereditárias sejam raros, alguns alelos que apresentam um efeito profundo sobre a saúde são relativamente comuns, tais como os alelos de genes que codificam enzimas que metabolizam medicamentos (p. ex., sensibilidade ao abacavir em alguns indivíduos infectados com o vírus da imunodeficiência humana [HIV]) (Caso 1), ou uma mutação falciforme nas populações africanas e afro-americanas (Cap. 11) (Caso 42). No entanto, estas são exceções, e à medida que mais variações genéticas são descobertas e catalogadas, fica evidente que a grande maioria das variantes no genoma, comuns ou raras, reflete diferenças na sequência de DNA que não têm nenhum significado conhecido para a saúde.

Os polimorfismos são elementos-chave para o estudo da genética humana e médica. A capacidade de distinguir diferentes formas de herança de um gene ou diferentes segmentos do genoma fornece ferramentas essenciais para uma vasta gama de aplicações, tanto na pesquisa quanto na prática clínica (Quadro).

## VARIAÇÃO HERDADA E POLIMORFISMO NO DNA

O Projeto Genoma Humano original e o estudo subsequente atual de milhares de indivíduos em todo o mundo têm proporcionado uma grande quantidade de informações sobre a sequência de DNA. Com essas informações em mãos, pode-se começar a caracterizar os tipos e as frequências de variações polimórficas encontrados no genoma humano e gerar catálogos da diversidade de sequência

## POLIMORFISMOS E VARIAÇÃO HERDADA EM GENÉTICA HUMANA E MÉDICA

As variantes alélicas podem ser utilizadas como “marcadores” para rastrear a herança do segmento genômico correspondente em famílias e em populações. Tais variantes podem ser usadas:

- Como ferramentas de pesquisa poderosas para mapear um gene em uma determinada região de um cromossomo por análise de ligação ou por associação alélica (Cap. 10)
- Para diagnóstico pré-natal de doenças genéticas e para detecção de portadores de alelos deletérios (Cap. 17), bem como em bancos de sangue e tipagem de tecidos para transfusões e transplantes de órgãos
- Em aplicações forenses, tais como testes de identificação para determinar a paternidade, identificar restos mortais de vítimas de crimes, ou para comparar o DNA do suspeito com o do agressor (neste capítulo)
- No esforço contínuo de fornecer medicina personalizada baseada em genômica (Cap. 18), na qual o cuidado médico individual é adaptado conforme o paciente carregue ou não variantes que aumentam ou diminuem o risco para transtornos comuns em adultos (como doenças cardíacas coronarianas, câncer e diabetes, Cap. 8) ou que influenciam a eficácia ou segurança de medicamentos específicos

do DNA humano ao redor do mundo. Os polimorfismos de DNA podem ser classificados de acordo com a forma como a sequência de DNA varia entre os diferentes alelos (Tabela 4-2 e Figs. 4-1 e 4-2).

## Polimorfismos de Nucleotídeo Único

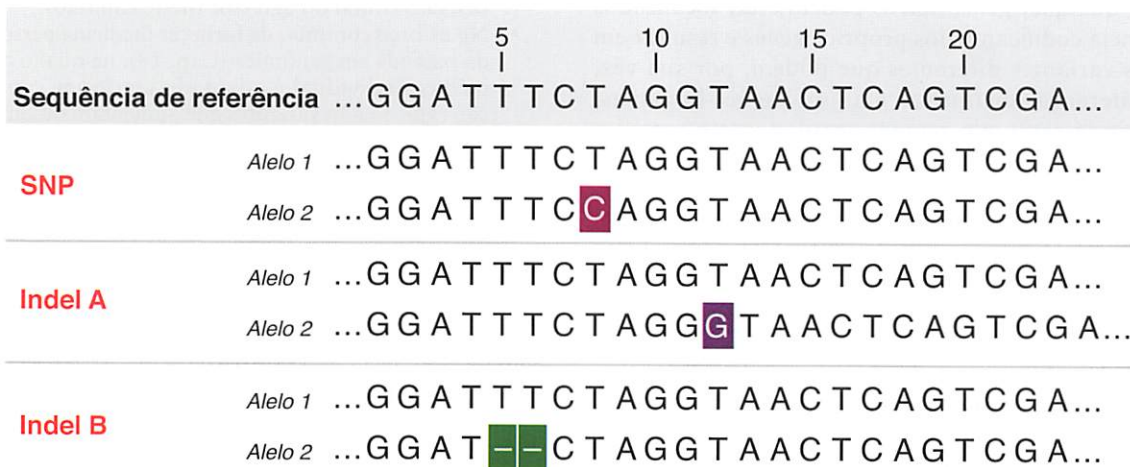
Os mais simples e comuns de todos os polimorfismos são os polimorfismos de nucleotídeo único (SNPs, do inglês, *single nucleotide polymorphisms*). Um *locus* caracterizado por um SNP geralmente tem apenas dois alelos, que correspondem a duas bases diferentes que ocupam uma localização específica no genoma (Fig. 4-1). Como mencionado anteriormente, os SNPs são comuns e são observados em média uma vez a cada 1.000 pb no genoma. Entretanto, a distribuição de SNPs é desigual em todo o genoma; muito mais SNPs são encontrados em regiões não codificantes do genoma, em íntrons e em sequências que estão a alguma distância de genes conhecidos. No entanto, há ainda um número significativo de SNPs que ocorrem em genes e em outros elementos funcionais conhecidos no genoma. Para o conjunto de genes codificantes de proteínas, mais de 100.000 SNPs exônicos foram documentados até o momento. Cerca de metade desses não alteram a sequência de aminoácidos prevista para a proteína codificada e são assim denominados de *sinônimos*, enquanto a outra metade altera a sequência de aminoácidos e compreende os chamados *não sinônimos*. Outros SNPs introduzem ou alteram um códon de parada (Tabela 3-1), e outros ainda alteram um sítio de *splicing* conhecido; esses SNPs são candidatos a apresentar consequências funcionais significativas.

A importância da grande maioria dos SNPs para a saúde é desconhecida e é objeto de pesquisas em andamento.

TABELA 4-2 Variação Comum no Genoma Humano

Tipo de Variação	Extensão do Tamanho (Aprox.)	Base para o Polimorfismo	Número de Alelos
Polimorfismos de nucleotídeo único	1 pb	Substituição de um ou outro par de bases em uma localização específica no genoma	Geralmente dois
Inserção/deleções (indels)	1 pb a > 100 pb	<i>Simples</i> : Presença ou ausência de um pequeno segmento de DNA de 100-1.000 pb de comprimento <i>Microsatélites</i> : Geralmente, uma unidade de 2, 3 ou 4 nucleotídeos repetida em <i>tandem</i> 5-25 vezes	<i>Simples</i> : 2 <i>Microsatélites</i> : tipicamente 5 ou mais
Variantes no número de cópias	10 kb a > 1 Mb	Tipicamente a presença ou ausência de segmentos de DNA de 200 pb a 1,5 Mb, embora a duplicação em <i>tandem</i> de 2, 3, 4 ou mais cópias também possa ocorrer	2 ou mais
Inversões	Poucos pb a > 1 Mb	Um segmento de DNA presente em qualquer uma das duas orientações com respeito ao DNA circundante	2

pb, par de bases; kb, par de quilobases; Mb, par de megabases



**Figura 4-1** Três polimorfismos no DNA genômico a partir de um segmento do conjunto de referência do genoma humano são demonstrados na parte superior (Fig. 2-6). O polimorfismo de nucleotídeo único (SNP) na posição 8 possui dois alelos, um com T (correspondente à sequência referência) e um com C. Existem duas indels nessa região. Na indel A, o alelo 2 apresenta uma inserção de um G entre as posições 11 e 12 na sequência de referência (alelo 1). Na indel B, o alelo 2 apresenta uma deleção de 2 pb nas posições 5 e 6 na sequência de referência.

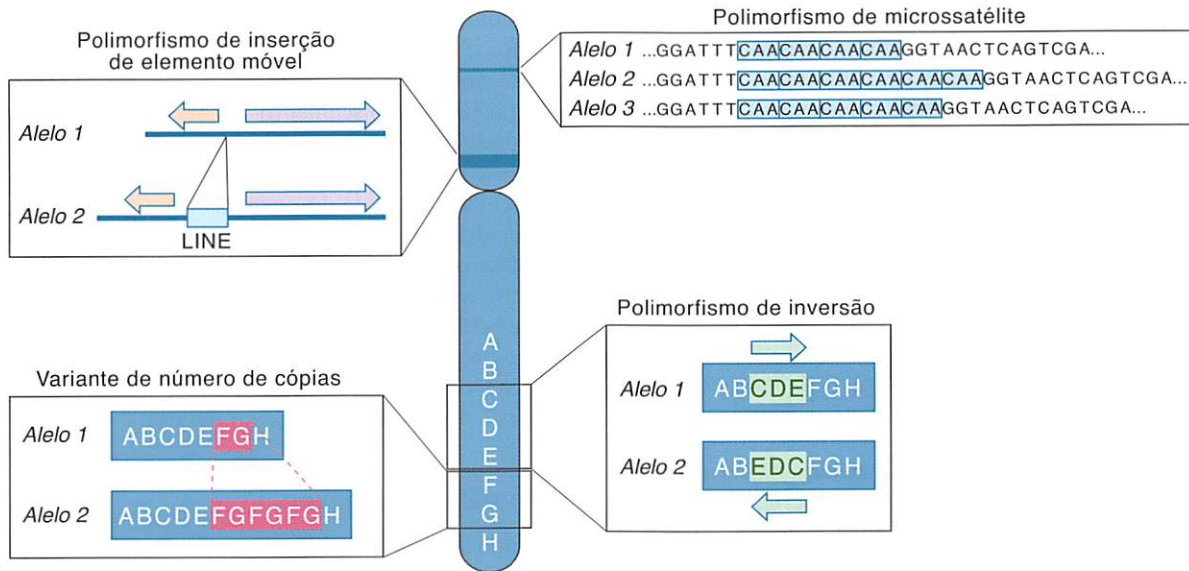
O fato de os SNPs serem comuns, não significa que eles não apresentem efeito na saúde e na longevidade. Isto quer dizer que qualquer efeito de SNPs comuns está mais provavelmente envolvido na alteração relativamente sutil de suscetibilidade a doenças do que na causa direta de doenças sérias.

### Polimorfismos de Inserção e Deleção

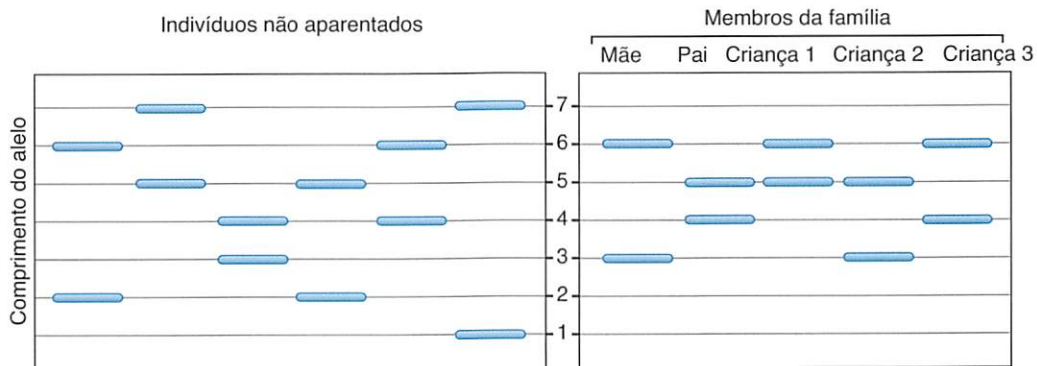
Uma segunda classe de polimorfismos resulta de variações causadas por **inserção** ou **deleção** (*in/dels* ou simplesmente **indels**) em qualquer parte, variando de um único par de bases até aproximadamente 1.000 pb, embora indels maiores também tenham sido bem documentadas. Mais de um milhão de indels têm sido descritas, na casa de centenas de milhares em qualquer genoma individual. Aproximadamente metade de todas as indels é referida como “simples”, porque elas apresentam apenas dois alelos — ou seja, presença ou ausência do segmento inserido ou deletado (Fig. 4-1).

### Polimorfismos de Microsatélites

Outras indels, entretanto, são multialélicas, devido a números variáveis de um segmento de DNA que é inserido em *tandem* em um determinado local, constituindo assim o que é chamado de **microsatélites**. Eles consistem em trechos de DNA compostos por unidades de dois, três ou quatro nucleotídeos, como TGTGTG, CAACAACAA ou AAATAAATAAAT, repetidos entre uma e algumas dúzias de vezes em um local específico no genoma (Fig. 4-2). Os diferentes alelos em um polimorfismo de microsatélite são resultado de diferentes números de unidades de nucleotídeos repetidas, contidas dentro de qualquer microsatélite, e são, portanto, por vezes também chamadas de **polimorfismos de repetições curtas em tandem** (STR, do inglês, *short tandem repeat*). Um *locus* de microsatélite frequentemente possui muitos alelos (tamanhos de repetição) que podem ser avaliados rapidamente por procedimentos laboratoriais padronizados para distinguir indivíduos diferentes e para inferir relações de parentesco (Fig. 4-3). Muitas dezenas de milhares de *loci*



**Figura 4-2** Exemplos de polimorfismos no genoma humano maiores que SNPs. No sentido horário da direita superior: O locus de microsatélite possui três alelos, com quatro, cinco ou seis cópias de uma repetição trinucleotídica CAA. O polimorfismo de inversão possui dois alelos correspondentes às duas orientações (indicados pelas setas) do segmento genômico mostrado em verde; tais inversões podem envolver regiões de até muitas megabases de DNA. As variantes de número de cópias envolvem deleção ou duplicação de centenas de pares de quilobases até mais de uma megabase de DNA genômico. No exemplo mostrado, o alelo 1 contém uma cópia única, enquanto o alelo 2 contém três cópias do segmento cromossômico que contém os genes F e G; outros alelos possíveis com zero, duas, quatro ou mais cópias de F e G não são mostrados. O polimorfismo de inserção por elemento móvel possui dois alelos, um com e outro sem inserção de um retroelemento repetido LINE de aproximadamente 6 kb; a inserção do elemento móvel altera o espaçamento entre os dois genes e pode alterar a expressão gênica na região.



**Figura 4-3** Esquema de um marcador de microsatélite hipotético no DNA humano. Os alelos de tamanho diferentes (numerados de 1 a 7) correspondem aos fragmentos de DNA genômico contendo diferentes números de cópias de uma repetição de microsatélites, e os seus tamanhos relativos são determinados separando-os por eletroforese em gel. O alelo mais curto (alelo 1) migra em direção à parte inferior do gel, enquanto o alelo mais longo (alelo 7) permanece mais próximo do topo. À esquerda, Para este microsatélite multialélico, cada um dos seis indivíduos não aparentados possui dois alelos diferentes. À direita, Dentro de uma família, a herança dos alelos pode ser seguida a partir de cada um dos pais para cada uma das três crianças.

de microsatélites polimórficos são conhecidas ao longo do genoma humano.

Os microsatélites são um grupo particularmente útil de indels. A determinação dos alelos nos múltiplos loci de microsatélites é atualmente o método de escolha para a impressão digital de DNA (*DNA fingerprinting*) utilizada para o teste de identificação. Por exemplo, o Federal

Bureau of Investigation (FBI) nos Estados Unidos utiliza atualmente uma coleção de alelos em 13 desses loci para o seu painel de impressão digital de DNA. É improvável que dois indivíduos (exceto gêmeos monozigóticos) tenham exatamente os mesmos alelos em todos os 13 loci para os quais o painel determinará em definitivo se duas amostras vieram de um mesmo indivíduo. A informação

é armazenada no *FBI's Combined DNA Index System* (CODIS), que vem crescendo desde dezembro de 2014 e inclui mais de 11.548.700 perfis de criminosos, 1.300.000 de perfis de presos, e 601.600 perfis forenses (material obtido em cenas de crimes). Muitos estados e o U.S. Department of Defense, assim como as unidades correspondentes em outros países, possuem bancos de dados similares de impressões digitais de DNA.

### Polimorfismos de Inserção de Elementos Móveis

Quase a metade do genoma humano é composta por famílias de elementos repetitivos que estão dispersos ao longo do genoma (Cap. 2). Embora a maioria das cópias dessas repetições seja fixa, algumas delas são móveis e contribuem para a diversidade genética humana através da **retrotransposição**, um processo que envolve a transcrição para um RNA, transcrição reversa em uma sequência de DNA e inserção (i.e., transposição) em outra região do genoma, como introduzimos no Capítulo 3, no contexto dos pseudogenes processados. As duas famílias mais comuns de elementos móveis são a família *Alu* e a família de repetições LINE, sendo que cerca de 10.000 polimorfismos de inserção de elementos móveis foram descritos em diferentes populações. Cada *locus* polimórfico consiste em dois alelos, um com e outro sem o elemento móvel inserido (Fig. 4-2). Os polimorfismos de elementos móveis são encontrados em todos os cromossomos humanos. Embora a maioria seja encontrada em regiões não gênicas do genoma, uma pequena parte deles é encontrada dentro dos genes. Pelo menos 5.000 desses *loci* polimórficos têm uma frequência de inserção superior a 10% em várias populações.

### Variação no Número de Cópias

Outro tipo importante de polimorfismo humano inclui a **variação no número de cópias** (CNVs, do inglês, *copy number variants*). As CNVs são conceitualmente relacionadas às indels e aos microssatélites, mas consistem em variações no número de cópias de segmentos grandes do genoma, que variam em tamanho de 1.000 pb a muitas centenas de pares de quilobases. Variações maiores que 500 kb são encontradas em 5% a 10% dos indivíduos na população em geral, ao passo que as variações abrangendo mais que 1 Mb são encontradas em 1% a 2%. As maiores CNVs são encontradas, às vezes, em regiões do genoma caracterizadas por blocos repetidos de sequências homólogas chamadas de **duplicações segmentares** (ou *segdups*). A sua importância em mediar a duplicação e a deleção dos segmentos correspondentes é discutida mais adiante no Capítulo 6, no contexto de várias síndromes cromossômicas.

As CNVs menores em particular podem apresentar apenas dois alelos (i.e., a presença ou ausência de um segmento), de modo semelhante às indels nesse contexto. As CNVs maiores tendem a apresentar alelos múltiplos, devido à presença de números diferentes de cópias em *tandem* de um segmento de DNA (Fig. 4-2). Em termos de diversidade genômica entre os indivíduos, a quantidade de DNA envolvida em CNVs excede amplamente aquele conteúdo que difere por

causa dos SNPs. *Nos loci com CNV, o conteúdo de quaisquer dos dois genomas humanos pode diferir em até 50 a 100 Mb por causa de diferenças no número de cópias.*

Notavelmente, o segmento variável em muitos *loci* com CNV pode incluir uma a várias dúzias de genes, e assim as CNVs são frequentemente associadas a características que envolvem alteração da dosagem gênica. Quando uma CNV é frequente o suficiente para ser polimórfica, ela representa um *background* de variação comum que deve ser compreendido caso as alterações no número de cópias observadas em pacientes forem interpretadas adequadamente. Assim como todos os polimorfismos de DNA, o significado de diferentes alelos na CNV sobre a saúde e sobre a suscetibilidade a doenças é objeto de investigação intensa.

### Polimorfismos de Inversão

Um grupo final de polimorfismos a ser discutido compreende as inversões, que diferem em tamanho — desde poucos pares de bases a grandes regiões do genoma (até vários pares de megabase) —, podendo estar presentes em qualquer uma das duas orientações nos genomas de indivíduos diferentes (Fig. 4-2). A maioria das inversões é caracterizada por regiões de homologia de sequência nas extremidades do segmento invertido, implicando um processo de recombinação homóloga na origem das inversões. Na sua forma balanceada, as inversões, independentemente da orientação, não envolvem ganho ou perda de DNA, e os polimorfismos de inversão (com dois alelos correspondentes às duas orientações) podem atingir frequências substanciais na população em geral. Entretanto, a recombinação anômala pode resultar na duplicação ou deleção de DNA localizado entre as regiões de homologia, associada a distúrbios clínicos que serão mais explorados nos Capítulos 5 e 6.

## A ORIGEM E A FREQUÊNCIA DE DIFERENTES TIPOS DE MUTAÇÕES

Ao longo do espectro da diversidade desde variantes raras até os polimorfismos mais comuns, diferentes tipos de mutações surgem no contexto de processos fundamentais da divisão celular, tais como replicação, reparo e recombinação de DNA, e a segregação cromossômica na mitose ou meiose. A **frequência de mutações por locus por divisão celular** é uma medida básica de quão propensos a erros estes processos estão, o que é de fundamental importância para a biologia e evolução do genoma. No entanto, de maior importância para os médicos geneticistas é a **frequência de mutações por locus da doença por geração**, em vez da taxa de mutação total em todo o genoma por divisão celular. Entretanto, quantificar as taxas de mutações causadoras de doenças pode ser difícil, porque muitas mutações causam letalidade embrionária precoce antes de a mutação ser reconhecida em um feto ou recém-nascido, ou porque algumas pessoas com uma mutação causadora de doença podem manifestar a condição tardiamente na vida ou nunca manifestar sinais da doença. Apesar dessas limitações, temos tido um ótimo progresso na determinação da frequência total — algumas

vezes referida como carga genética — de todas as mutações que afetam a espécie humana.

Os principais tipos de mutação apresentados de forma breve anteriormente ocorrem com frequências consideráveis em muitas células diferentes do corpo. Na prática da genética, estamos preocupados principalmente com a variação genômica herdada; no entanto, toda essa variação teria de se originar como uma alteração nova (*de novo*) nas células germinativas. Nesse ponto, tal variante seria bastante rara na população (ocorrendo apenas uma vez), e sua frequência final na população ao longo do tempo dependeria do acaso e dos princípios de herança e de genética de populações (Caps. 7 e 9). Embora a mutação original tenha ocorrido apenas no DNA das células da **linhagem germinativa**, qualquer pessoa que herdasse esta mutação a carregaria como uma mutação constitucional em todas as células do corpo.

Ao contrário, as **mutações somáticas** ocorrem em todo o corpo, mas não podem ser transmitidas à geração seguinte. Dada a taxa de mutação (veja mais adiante nesta seção), pode-se prever que, de fato, cada célula em um indivíduo tem uma versão ligeiramente diferente do seu genoma, dependendo do número de divisões celulares que ocorrem desde a concepção até o tempo de aquisição das amostras. Em tecidos altamente proliferativos, tais como os epiteliais intestinais ou células hematopoiéticas, tal heterogeneidade genômica é particularmente suscetível de estar evidente. No entanto, a maioria de tais mutações não é tipicamente detectada, porque em ensaios clínicos, o DNA é geralmente sequenciado a partir de coleções de muitos milhões de células; em tais coleções, a base mais prevalente em qualquer posição no genoma será a única presente no momento da análise, e mutações somáticas raras serão amplamente invisíveis e indeterminadas. Tais mutações podem ser de importância clínica, entretanto, em distúrbios provocados por mutação em apenas um subtipo de células em tecidos específicos, levando ao mosaicismo somático (Cap. 7).

A principal exceção à expectativa de que mutações somáticas sejam subdetectadas em qualquer amostra de DNA multicelular está no câncer. A base mutacional para as origens do câncer e a natureza clonal da evolução tumoral direcionam certas alterações somáticas a estar presentes essencialmente em todas as células de um tumor. De fato, de 1.000 a 10.000 mutações somáticas (e algumas vezes muito mais) são encontradas nos genomas da maioria dos cânceres de adultos, com frequências e padrões mutacionais específicos para diferentes tipos de câncer (Cap. 15).

## Mutações Cromossômicas

Mutações que produzem alteração no número de cromossomos devido a erros de segregação cromossômica estão entre as mutações mais observadas em humanos, com uma taxa de uma mutação por 25 a 50 divisões celulares meióticas. Essa estimativa é um valor mínimo, porque as consequências de muitos desses eventos no desenvolvimento são provavelmente tão graves que os fetos resultantes são abortados de modo espontâneo logo após a concepção sem serem detectados (Caps. 5 e 6).

## Mutações Regionais

As mutações que afetam a estrutura ou a organização regional dos cromossomos podem surgir por vários caminhos diferentes. As duplicações, deleções e inversões de um segmento de um único cromossomo são predominantemente o resultado da recombinação homóloga entre segmentos de DNA com alta homologia de sequência situados em mais de um local em uma região do cromossomo. No entanto, nem todas as mutações estruturais são resultado de recombinação homóloga. Algumas, como translocações cromossômicas e algumas inversões, podem ocorrer em locais de quebras espontâneas do DNA de dupla-fita. Uma vez que a quebra ocorra em dois locais no genoma, as duas extremidades quebradas podem ser unidas em conjunto, mesmo sem qualquer homologia óbvia na sequência entre as duas extremidades (um processo denominado *reparo por união de extremidades não homólogas*). Exemplos de tais mutações serão discutidos em profundidade no Capítulo 6.

## Mutações Gênicas

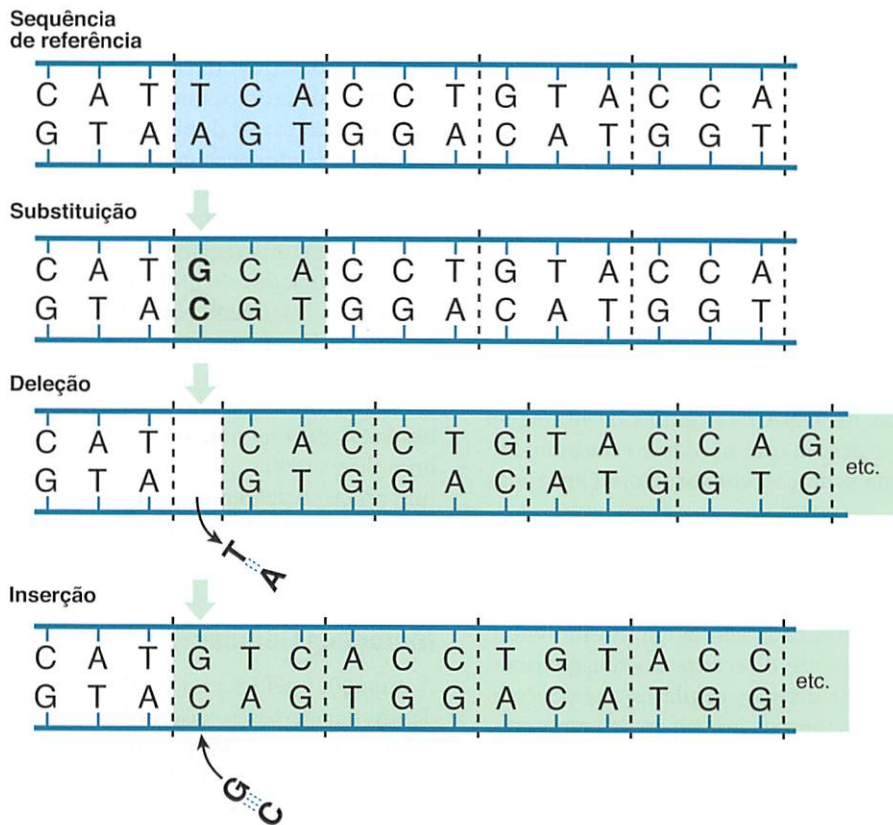
As mutações gênicas ou de DNA, incluindo a substituição de um par de bases, inserções e deleções (Fig. 4-4), podem originar-se por qualquer um de dois mecanismos básicos: erros introduzidos durante a replicação do DNA ou mutações decorrentes de uma falha no reparo correto do DNA após lesão. Muitas dessas mutações são espontâneas e surgem durante os processos normais (mas imperfeitos) de replicação e reparo do DNA, enquanto outras são induzidas por agentes físicos ou químicos, chamados de mutagênicos.

## Erros de Replicação do DNA

O processo de replicação do DNA (Fig. 2-4) é altamente preciso; a maioria dos erros de replicação (i.e., a inserção de uma base diferente da base complementar que restauraria o par de bases nessa posição da dupla-hélice) é rapidamente removida do DNA e corrigida por uma série de enzimas de reparo de DNA que primeiramente reconhecem qual fita na dupla-hélice recém-sintetizada contém a base incorreta e, em seguida, substituem-na com a base complementar adequada, um processo denominado **revisão do DNA** (*proofreading*). A replicação do DNA precisa ser um processo notavelmente exato; caso contrário, o ônus da mutação nos organismos e nas espécies seria intolerável. A enzima DNA polimerase duplica fielmente as duas fitas da dupla-hélice com base em regras rigorosas de pareamento de bases (A pareia com T, C pareia com G), mas introduz um erro a cada 10 milhões de pb. Uma revisão adicional, em seguida, corrige mais de 99,9% desses erros de replicação do DNA. Assim, a taxa de mutação total por base, como resultado de erros de replicação, é consideravelmente menor que  $1 \times 10^{-10}$  por divisão celular — *menor que uma mutação por genoma por divisão celular*.

## Reparo da Lesão do DNA

Em adição aos erros de replicação, estima-se que entre 10.000 e um milhão de nucleotídeos sejam danificados por célula humana por dia devido a processos químicos



**Figura 4-4** Exemplos de mutações em uma porção de um gene hipotético com cinco códons são demonstrados (delimitados pelas *linhas tracejadas*). O primeiro par de bases do segundo códon na sequência de referência (sombreados em *azul*) está mutado por uma substituição de base, deleção ou inserção. A substituição de base de um G para um T nesta posição leva a uma mudança de códon (sombreado em *verde*) e, supondo que a fita superior é a fita senso ou codificante, a uma alteração não sinônima predita de uma serina para uma alanina na proteína codificada (código genético na Tabela 3-1); todos os outros códons permanecem inalterados. Tanto a deleção quanto a inserção do par de bases único levam a uma mutação *frameshift*, na qual o quadro de leitura da tradução é alterado para todos os códons subsequentes (sombreados em *verde*), até que um códon de término seja alcançado.

espontâneos, tais como a depurinação, a desmetilação ou a desaminação; por reação com mutagênicos químicos (naturais ou não) no ambiente; e por exposição à radiação ultravioleta ou ionizante. Algumas dessas lesões, mas nem todas, são reparadas. Mesmo que a lesão seja reconhecida e destruída, a maquinaria de reparo pode criar mutações através da introdução de bases incorretas. Assim, em contraste com as alterações do DNA relacionadas à replicação, as quais são geralmente corrigidas por meio de mecanismos de revisão, as alterações de nucleotídeos introduzidos por lesão e reparo do DNA muitas vezes resultam em mutações permanentes.

Uma mutação espontânea particularmente comum é a substituição de T por C (ou A por G na outra fita). A explicação para essa observação vem da principal forma de modificação epigenética no genoma humano, a metilação do DNA, introduzida no Capítulo 3. A desaminação espontânea da 5-metilcitosina para timidina (compare as estruturas de citosina e timina na Fig. 2-2) no par CpG dá origem a mutações de C para T ou G para A (dependendo em qual fita a 5-metilcitosina é desaminada). Tais mutações espontâneas podem não ser reconhecidas pela maquinaria

de reparo do DNA e, assim, estabelecer-se no genoma após a rodada seguinte de replicação do DNA. Mais de 30% de todas as substituições de nucleotídeos únicos são deste tipo e ocorrem em uma taxa 25 vezes maior que quaisquer outras mutações de um único nucleotídeo. Assim, os dupletos de CpG representam um verdadeiro *hot spot* (“ponto quente”) de mutação no genoma humano.

### Taxa Total de Mutações de DNA

Embora a taxa de mutações de DNA em *loci* específicos tenha sido estimada utilizando-se uma variedade de abordagens ao longo dos últimos 50 anos, o impacto global de erros de replicação e reparo sobre a ocorrência de novas mutações ao longo do genoma atualmente pode ser determinado diretamente pelo **sequenciamento de genoma completo** de trios constituídos pela criança e seus pais, em busca de novas mutações na criança que não estão presentes na sequência genômica de nenhum dos pais. A taxa total média de novas mutações entre gametas maternos e paternos é de aproximadamente  $1,2 \times 10^{-8}$  mutações por par de bases por geração. Assim, cada pessoa provavelmente recebe cerca de 75 novas mutações em seu genoma de um ou do outro



*progenitor*. Essa taxa, no entanto, varia de gene para gene ao longo do genoma e, talvez, de população para população ou mesmo de indivíduo para indivíduo. De forma geral, essa taxa, combinada com considerações sobre o crescimento e a dinâmica populacional, prevê que deve haver um número enorme de mutações relativamente novas (e, portanto, muito raras) na atual população mundial atual de sete bilhões de indivíduos.

É previsto que a grande maioria dessas mutações seja de alterações de nucleotídeo único em porções não codificantes do genoma e provavelmente tenha pouco ou nenhum significado funcional. No entanto, em nível populacional, o impacto coletivo potencial dessas novas mutações em genes de importância médica não deve ser negligenciado. Nos Estados Unidos, por exemplo, com mais de quatro milhões de nascidos vivos por ano, aproximadamente seis milhões de novas mutações ocorrerão em sequências codificantes. Assim, mesmo para um gene codificante de uma proteína *única* de tamanho médio, podemos esperar várias centenas de recém-nascidos por ano com uma nova mutação na sequência codificante deste gene.

Conceitualmente, estudos semelhantes têm determinado a taxa de mutações em CNVs, nas quais a geração de uma nova variante de tamanho depende de recombinação, em vez de erros na síntese de DNA para gerar um novo par de bases. A quantificação da taxa de formação de novas CNVs ( $\approx 1,2 \times 10^{-2}$  por locus por geração) é de uma ordem de magnitude mais alta do que aquelas de substituições de bases.

### Taxa de Mutações Gênicas Causadoras de Doença

A maneira mais direta de estimar a taxa de mutações causadoras de doença por *locus* por geração é medindo a incidência de novos casos de doença genética que não está presente em nenhum dos progenitores e é causada por uma mutação única, que gera uma condição claramente reconhecível em todos os recém-nascidos que carregam essa mutação. A **acondroplasia**, uma condição de crescimento ósseo reduzido que leva a uma baixa estatura (Caso 2), é uma condição que atende a esses requisitos. Em um estudo, sete crianças acondroplásicas nasceram em uma série de 242.257 nascimentos consecutivos. Todas as sete nasceram de pais de estatura normal, e como a acondroplasia sempre se manifesta quando a mutação está presente, todas foram

consideradas por representarem novas mutações. A nova taxa de mutação nesse *locus* pode ser calculada como sendo de sete novas mutações em um total de  $2 \times 242.257$  cópias do gene relevante, ou seja, aproximadamente  $1,4 \times 10^{-5}$  mutações causadoras de doença por *locus* por geração. Essa taxa de mutação elevada é particularmente notável, porque foi verificado que virtualmente todos os casos de acondroplasia são devido a uma mutação *idêntica*, uma mutação de G para A, que altera um códon de glicina para uma arginina na proteína codificada.

A taxa de mutações gênicas causadoras de doença foi estimada por vários outros distúrbios, nos quais a ocorrência de uma nova mutação foi determinada pelo aparecimento de uma doença detectável (Tabela 4-3). A quantificação das taxas para esses e outros distúrbios varia dentro de uma faixa de 1.000 vezes, de  $10^{-4}$  a  $10^{-7}$  mutações por *locus* por geração. A base para essas diferenças pode estar relacionada com alguns ou todos os seguintes pontos: o tamanho de genes diferentes; a porção de todas as mutações naqueles genes que vão levar a doenças; a idade e o sexo do progenitor no qual a mutação ocorreu; o mecanismo mutacional; e a presença ou ausência de mutações nos *hot spots* no gene. Na verdade, a elevada taxa de mutação sítio-específica particular na acondroplasia pode ser parcialmente explicada pelo fato de que a mutação na outra fita é uma alteração de um C para T em uma posição sujeita à metilação de CpG, que é um *hot spot* para mutação por desaminação, como discutido anteriormente.

Mesmo com essa variedade de taxas entre os diferentes genes, a taxa de mutação gênica média é de aproximadamente  $1 \times 10^{-6}$ . Dado que há pelo menos 5.000 genes no genoma humano em que as mutações são atualmente conhecidas por causar uma doença ou qualquer característica discernível (Cap. 7), *é provável que aproximadamente uma em 200 pessoas receba uma nova mutação em um gene associado a uma doença conhecida a partir de um dos progenitores.*

### Diferenças Sexuais e Efeitos da Idade nas Taxas de Mutação

Como o DNA no esperma é submetido a muito mais ciclos de replicação do que o DNA nos óvulos (Cap. 2), há maior oportunidade de ocorrerem erros; pode-se prever, portanto, que muitas mutações sejam mais frequentemente de origem paterna que materna. Na verdade, quando estas foram

**TABELA 4-3** Estimativas de Taxas de Mutação para Genes de Doenças Humanas Selecionados

Doença	Locus (Proteína)	Taxa de Mutação*
Acondroplasia (Caso 2)	<i>FGFR3</i> (receptor do fator de crescimento de fibroblasto 3)	$1,4 \times 10^{-5}$
Aniridia	<i>PAX6</i> (Pax6)	$2,9-5 \times 10^{-6}$
Distrofia muscular de Duchenne (Caso 14)	<i>DMD</i> (distrofina)	$3,5-10,5 \times 10^{-5}$
Hemofilia A (Caso 21)	<i>F8</i> (fator VIII)	$3,2-5,7 \times 10^{-5}$
Hemofilia B (Caso 21)	<i>F9</i> (fator IX)	$2,3 \times 10^{-6}$
Neurofibromatose tipo 1 (Caso 34)	<i>NF1</i> (neurofibromina)	$4-10 \times 10^{-5}$
Doença renal policística, tipo 1 (Caso 37)	<i>PKD1</i> (policistina)	$6,5-12 \times 10^{-5}$
Retinoblastoma (Caso 39)	<i>RBI</i> (Rb1)	$5-12 \times 10^{-6}$

\*Expresso como mutações por *locus* por geração

Baseado nos dados de Vogel F, Motulsky AG: *Human genetics*, ed 3, Berlin, 1997, Springer-Verlag.

exploradas, as novas mutações responsáveis por determinadas condições (p. ex., acondroplasia, como nós acabamos de discutir) são geralmente mutações de sentido trocado (*missense*) que surgem quase sempre na linhagem paterna. Além disso, quanto mais velho o homem for, mais ciclos de replicação terão precedido as divisões meióticas, e, portanto, seria esperado que a frequência de novas mutações paternas aumentasse com a idade do pai. De fato, foram observadas correlações do aumento da idade do pai com a incidência de mutações gênicas em uma série de distúrbios (incluindo a acondroplasia) e com a incidência de mutações regionais envolvendo CNVs em transtornos do espectro autista (Caso 5). Em outras doenças, entretanto, a origem parental e os efeitos da idade no espectro mutacional podem, por razões desconhecidas, não ser tão surpreendentes.

## TIPOS DE MUTAÇÕES E SUAS CONSEQUÊNCIAS

Nesta seção, consideramos a natureza de diferentes mutações e seus efeitos sobre os genes envolvidos. Cada tipo de mutação discutido aqui é ilustrado por um ou mais exemplos de doenças. Notavelmente, a mutação específica encontrada em quase todos os casos de acondroplasia é uma exceção e não a regra, e as mutações que estão na base de uma única doença genética são mais tipicamente heterogêneas entre um grupo de indivíduos afetados. Diferentes casos de uma doença em particular, portanto, serão frequentemente causados por diferentes mutações subjacentes (Tabela 4-4). Nos Capítulos 11 e 12, vamos nos voltar para a maneira pelas quais mutações em genes específicos causam tais doenças.

### Substituições Nucleotídicas

#### Mutações de Sentido Trocado

Uma única substituição de nucleotídeo (ou **mutação pontual**) em uma sequência gênica, tal como a observada no exemplo de acondroplasia que acabamos de descrever, pode alterar o código em uma trinca de bases e causar a substituição não sinônima de um aminoácido por outro no produto gênico (veja código genético no Quadro 3-1 e o exemplo da Fig. 4-4). Tais mutações são denominadas **mutações de sentido trocado** (*missense*) porque alteram o “sentido” da codificação do gene ao especificar um aminoácido diferente. Embora nem todas as mutações de sentido trocado conduzam a uma alteração observável na função proteica, a proteína resultante pode não funcionar adequadamente, pode tornar-se instável e degradar-se rapidamente, ou pode falhar em localizar a sua posição intracelular correta. Em muitos distúrbios, tais como a  $\beta$ -talassemia (Caso 44), a maioria das mutações detectadas em diferentes pacientes compreende mutações de sentido trocado (Cap. 11).

#### Mutações sem Sentido

As mutações pontuais em uma sequência de DNA que causam a substituição de um códon normal para um aminoácido por um dos três códons de término (ou “parada”) são chamadas de **mutações sem sentido** (*nonsense*). Como

TABELA 4-4 Tipos de Mutação em Doenças Genéticas Humanas

Tipo de Mutação	Porcentagem de Mutações Causadoras de Doença
<b>Substituições de Nucleotídeos</b>	
• Mutações de sentido trocado ( <i>missense</i> ) (substituições de aminoácidos)	50%
• Mutações sem sentido ( <i>nonsense</i> ) (códon de término prematuros)	10%
• Mutações no processamento de RNA (destroem sítios de <i>splicing</i> consensuais, sítios de capeamento e sítios de poliadenilação ou criam sítios ocultos)	10%
• Mutações de sítios de <i>splicing</i> , levando a mutações de mudança de matriz de leitura ( <i>frameshift</i> ) e códons de término prematuros	10%
• Mutações reguladoras de longo alcance	Raras
<b>Deleções e Inserções</b>	
• Adição ou deleções de um pequeno número de bases	25%
• Deleções gênicas grandes, inversões, fusões e duplicações (podem ser mediadas pela homologia de sequência do DNA tanto dentro quanto entre as fitas de DNA)	5%
• Inserção do elemento LINE ou <i>Alu</i> (perturbação da transcrição ou interrupção da sequência codificante)	Rara
• Mutações dinâmicas (expansão de sequências de repetição de tri ou tetranucleotídeos)	Raras

a tradução do RNA mensageiro (RNAm) cessa quando o códon de término é atingido (Cap. 3), uma mutação que converte um éxon codificante em um códon de término promove a parada da tradução no meio da sequência codificante do RNAm. As consequências das mutações de término prematuras são duplas. Em primeiro lugar, o RNAm transportando uma mutação prematura é frequentemente alvo de rápida degradação (através de um processo celular conhecido como **decaimento do RNAm mediado por mutações sem sentido**), e a tradução não é possível. Em segundo, mesmo que o RNAm seja suficientemente estável para ser traduzido, a proteína truncada é tão instável que é rapidamente degradada dentro da célula (veja o Cap. 12 para exemplos). Enquanto algumas mutações pontuais criam um códon de término prematuro, outras podem destruir o códon de término normal e permitir, assim, que a tradução continue até que outro códon de término do RNAm seja alcançado a jusante. Tal mutação irá levar a um produto proteico anormal com aminoácidos adicionais em sua extremidade carboxiterminal, e poderá também perturbar as funções reguladoras normalmente exercidas pela região 3' não traduzida a jusante do códon de término normal.

### Mutações que Afetam a Transcrição, o Processamento e a Tradução do RNA

O mecanismo normal pelo qual os transcritos iniciais de RNA são feitos e depois convertidos em RNAs maduros (ou versões finais de RNAs não codificantes) requer uma

série de modificações, incluindo a ligação de fatores de transcrição, o capeamento 5', a poliadenilação e o *splicing* (Cap. 3). Todos esses passos de maturação do RNA dependem de sequências específicas dentro do RNA. No caso de *splicing*, duas classes gerais de mutações foram descritas. Para os íntrons serem excisados do RNA não processado e os éxons serem unidos para formar um RNA maduro são requeridas sequências particulares de nucleotídeos localizados dentro ou próximos das junções éxon-íntron (sítio doador 5') ou íntron-éxon (sítio aceptor 3') delas. As mutações que afetam essas bases necessárias, seja no sítio doador ou aceptor, interferem com (e em alguns casos anulam) o *splicing* normal de RNA naquele local. Uma segunda classe de mutações de *splicing* envolve substituições de bases que não afetam por si próprias as sequências do sítio doador ou aceptor. Em vez disso, criam sítios doadores ou aceptores alternativos que competem com os sítios normais durante o processamento do RNA. Assim, pelo menos uma proporção do RNAm maduro ou do RNA não codificante em tais casos pode conter sequências de íntron imprópriamente excisadas. Exemplos de ambos os tipos de mutação são apresentados no Capítulo 11.

Para genes codificantes de proteínas, mesmo se o RNAm for produzido e for estável, as mutações pontuais em regiões 5' e 3' não traduzidas também podem contribuir para doenças ao alterarem a estabilidade do RNAm ou a eficiência de tradução, reduzindo, assim, as quantidades de produtos proteicos produzidos.

### Deleções, Inserções e Rearranjos

As mutações também podem ser causadas por inserção, deleção ou rearranjo nas sequências de DNA. Algumas deleções e inserções envolvem apenas alguns nucleotídeos e são, em geral, mais facilmente detectadas pelo sequenciamento direto dessa parte do genoma. Em outros casos, um segmento substancial de um gene ou um gene inteiro é deletado, duplicado, invertido, ou translocado para criar uma nova organização de sequências gênicas. Dependendo da natureza exata da deleção, inserção ou rearranjo, uma variedade de diferentes abordagens laboratoriais pode ser usada para detectar a alteração genômica.

Algumas deleções afetam apenas um pequeno número de pares de bases. Quando tal mutação ocorre em uma sequência codificante e o número de bases envolvidas não é um múltiplo de três (i.e., não é um número completo de códons), o quadro de leitura será alterado começando no ponto de inserção ou deleção. O resultado das mutações é chamado de **mutações de mudança de matriz de leitura** (*frameshift*) (Fig. 4-4). A partir do ponto de inserção ou de deleção, uma sequência diferente de códons é, portanto, gerada, codificando aminoácidos incorretos seguidos por um códon de término na matriz alterada, o que levará a um produto proteico funcionalmente alterado. Em contraste, se o número de pares bases inserido ou deletado *for* um múltiplo de três, não ocorrerão mudanças na matriz de leitura e haverá uma simples inserção ou deleção de aminoácidos correspondentes no produto gênico normalmente traduzido. Inserções ou deleções maiores, que variam de cerca de

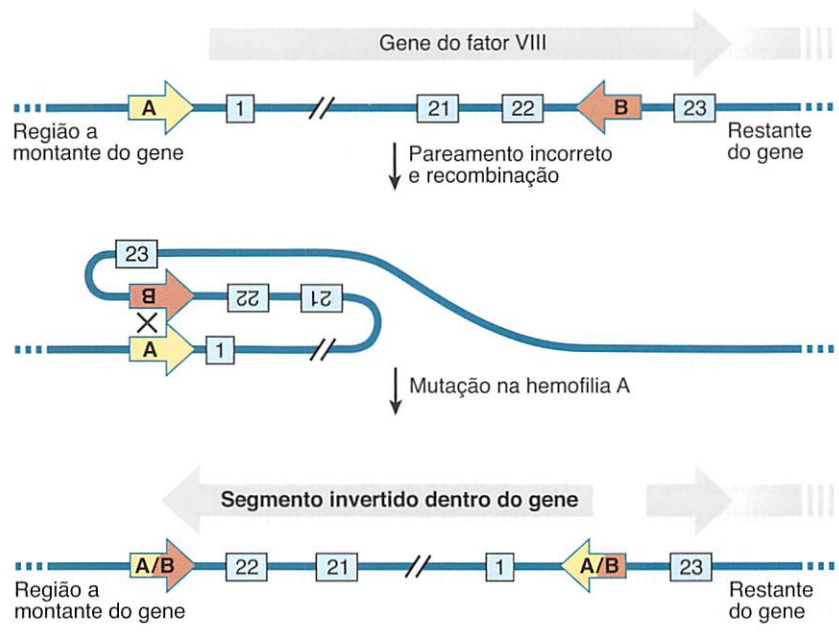
100 a mais de 1.000 pb, são tipicamente referidas como “indels”, como vimos no caso de polimorfismos anteriores. Elas podem afetar múltiplos éxons de um gene e causar distúrbios maiores na sequência codificante.

Um tipo de mutação de inserção envolve a inserção de um elemento móvel, como aqueles pertencentes à família de DNA repetitivo LINE. Estima-se que, em qualquer indivíduo, aproximadamente 100 cópias de uma subclasse particular da família LINE no genoma sejam capazes de se movimentar por **retrotransposição**, como abordado anteriormente. Tal movimento não só gera diversidade genética em nossa espécie (Fig. 4-2), como também pode causar doenças por mutagênese insercional. Por exemplo, em alguns pacientes com a hemorragia grave do tipo **hemofilia A** (Caso 21) são encontradas sequências LINE com vários pares de quilobases de tamanho inseridas em um éxon do gene do fator VIII, que interrompem a sequência de codificação e inativam o gene. Inserções LINE ao longo do genoma também são comuns em câncer de colo, refletindo a retrotransposição em células somáticas (Cap. 15).

Como discutido anteriormente no contexto de polimorfismos neste capítulo, duplicações, deleções e inversões de um segmento maior de um único cromossomo são predominantemente o resultado de recombinação homóloga entre segmentos de DNA com alta homologia de sequência (Fig. 4-5). Os distúrbios que surgem como resultado de tais trocas podem ser devido a outra forma de alteração na dosagem de produtos gênicos selvagens, quando os segmentos homólogos estão fora dos próprios genes (Cap. 6). Alternativamente, tais mutações podem, por si sós, levar a uma alteração da natureza da proteína codificada quando a recombinação ocorre entre genes diferentes dentro de uma família gênica (Cap. 11) ou entre genes em diferentes cromossomos (Cap. 15). O pareamento e recombinação anormais entre duas sequências similares em orientação oposta em uma única fita de DNA levam à inversão. Por exemplo, quase metade de todos os casos de hemofilia A se deve à recombinação que inverte uma série de éxons, interrompendo assim a estrutura gênica e tornando o gene incapaz de codificar um produto gênico normal (Fig. 4-5).

### Mutações Dinâmicas

As mutações em alguns distúrbios envolvem a amplificação de uma sequência de repetição de nucleotídeos simples. Por exemplo, repetições simples, tais como (CCG)<sub>n</sub>, (CAG)<sub>n</sub> ou (CCTG)<sub>n</sub> localizadas na porção codificante de um éxon, em uma região não traduzida de um éxon, ou mesmo em um íntron, podem expandir-se durante a gametogênese, o que é denominado **mutação dinâmica**, interferindo com a expressão gênica normal ou com a função proteica. Uma repetição expandida na região codificante irá gerar um produto proteico anormal, enquanto a expansão da repetição em regiões não traduzidas ou íntrons de um gene pode interferir com a transcrição, o processamento de RNA ou a tradução. Não se sabe plenamente como as mutações dinâmicas ocorrem; elas são conceitualmente semelhantes aos polimorfismos de microssatélites, mas expandem-se a



**Figura 4-5** Sequências homólogas invertidas, marcadas como A e B, localizadas a 500 kb uma da outra no cromossomo X, uma a montante do gene do fator VIII, e outra em um íntron entre os éxons 22 e 23 do gene. O pareamento intracromossômico incorreto e a recombinação resultam na inversão dos éxons 1 a 22 do gene, interrompendo desse modo o gene e causando hemofilia grave.

uma taxa muito maior que aquelas observadas para os *loci* de microssatélites.

O envolvimento de expansões de repetição de nucleotídeos simples nas doenças é discutido nos Capítulos 7 e 12. Em distúrbios causados por mutações dinâmicas, efeitos notáveis da origem parental são bem conhecidos e parecem ser característicos de doenças específicas e/ou da repetição de nucleotídeos simples específica envolvida (Cap. 12). Tais diferenças podem ser devido a diferenças biológicas fundamentais entre a ovogênese e a espermatogênese, mas também podem resultar da seleção contra gametas que carregam determinadas expansões de repetição.

### VARIAÇÃO EM GENOMAS INDIVIDUAIS

O inventário atual mais extenso da quantidade e do tipo de variação que se espera em um dado genoma vem da análise direta dos genomas humanos diploides individuais. A primeira dessas sequências genômicas, em um indivíduo do sexo masculino, foi relatada em 2007. Agora, dezenas de milhares de genomas individuais foram sequenciadas, algumas como parte de um grande consórcio internacional de pesquisa que exploram a diversidade genética humana na saúde e nas doenças, e outras no contexto sequenciamento clínico para determinar a base subjacente de um distúrbio em pacientes específicos.

Qual o grau de variação genômica detectado em tais estudos? Os genomas humanos individuais carregam tipicamente de cinco a 10 milhões de SNPs, dos quais — dependendo em parte da população — um quarto a um terço é novo (Quadro). Isso sugere que o número de SNPs descritos para nossa espécie ainda é incompleto, embora presumivelmente a fração desses novos SNPs diminua à medida que mais genomas de mais populações forem sequenciados.

Dentro dessa variação estão as variantes com impacto clínico conhecido, provável ou suspeito. Com base em estudos realizados até o momento, cada genoma carrega de 50 a 100 variantes previamente implicadas em doenças hereditárias conhecidas. Além disso, cada genoma carrega milhares de SNPs não sinônimos nos genes codificantes de proteína ao longo do genoma, alguns dos quais previstos para alterar a função proteica. Cada genoma também transporta aproximadamente 200 a 300 prováveis mutações de perda de função, algumas das quais estão presentes em ambos os alelos dos genes daquele indivíduo. No cenário clínico, essa percepção tem implicações importantes para a interpretação dos dados da sequência genômica dos pacientes, particularmente quando se tenta prever o impacto de mutações em genes cuja função atualmente é desconhecida (Cap. 16).

Um aspecto interessante e inesperado do sequenciamento genômico individual é que a montagem do genoma humano de referência ainda carece de conteúdos consideráveis de DNA não documentados e não anotados que são descobertos em literalmente todos os genomas individuais sequenciados. Essas “novas” sequências são reveladas apenas à medida que genomas adicionais são sequenciados. Estima-se que a coleção completa de todas as sequências genômicas humanas encontradas em nossa população atual de sete bilhões de indivíduos seja 20 a 40 Mb maior que o conjunto de referência existente, e ainda precisa ser totalmente elucidada.

É impressionante o inventário atual da diversidade genética humana, e é claro que ainda estamos em um modo de descoberta. Não há dúvida de que milhões de SNPs adicionais e outras variantes permanecem por serem descobertos, assim como o grau em que qualquer um deles pode afetar o estado clínico de um indivíduo no contexto dos cuidados de bem-estar e saúde.