

Variabilidade Genética Humana e suas Consequências

13

CONCEITOS PRINCIPAIS

- Variantes humanas de DNA podem ser classificadas em larga escala *versus* pequena escala, comuns *versus* raras e patogênicas *versus* não patogênicas.
- Polimorfismos de nucleotídeo único (SNPs, do inglês *single nucleotide polymorphisms*) são as variantes mais numerosas.
- Polimorfismos de repetições curtas em *tandem* (microssatélites ou STRs, do inglês *short tandem repeat polymorphisms*) são muito comuns. As unidades de repetição comumente apresentam 1,2 ou 4 pb de comprimento e o conjunto de repetições em *tandem* normalmente varia de uma pessoa para outra, como resultado de deslizamentos da polimerase durante a replicação do DNA.
- Minissatélites são repetições em *tandem* com unidades de repetição mais longas, entre 10 a 50 pb. Novamente, o número de unidades de repetição em um bloco varia entre as pessoas, mas isso ocorre com mais frequência devido a recombinação desigual. Minissatélites são mais comuns nos telômeros dos cromossomos.
- Muitas sequências maiores (entre 1 kb e 1 Mb) apresentam variação no número de cópias entre pessoas saudáveis. Cerca de 5% de todo o genoma podem variar desta forma. Uma causa comum de toda esta variação é a recombinação malpareada entre sequências repetidas.
- O DNA genômico é submetido a processos constantes de dano e reparo. A maioria dos danos passa despercebida, pois é eficientemente reparada. Falhas no reparo de dano ou na correção de erros de replicação são as principais causas de variação de sequência.
- Variantes mais comuns são não patogênicas, embora elas possam, em combinação com outras variantes, aumentar ou reduzir a suscetibilidade a doenças multifatoriais. Elas são úteis como marcadores (variantes que podem ser utilizadas para identificar um cromossomo ou uma pessoa) em análises de ascendência, forenses e estudos sobre as origens e as relações entre as populações.
- Efeitos patogênicos podem ser mediados por perda ou ganho de função de um produto gênico. Muitas diferenças distintas em um gene podem causar perda de função. Geralmente apenas uma ou algumas mudanças específicas podem causar ganho de função.
- Mudanças com perda de função normalmente levam a fenótipos recessivos, enquanto mudanças com ganho de função levam a fenótipos dominantes. Fenótipos dominantes também podem resultar da perda de função se uma dose de 50% do produto gênico normal não for suficiente para produzir um fenótipo normal (haploinsuficiência), ou se a proteína produzida pelo alelo mutante interferir na função do produto normal (efeito dominante negativo).
- A patologia molecular procura explicar porque algumas mudanças genéticas causam um determinado fenótipo clínico. No entanto, correlações bem definidas entre genótipo e fenótipo são raras em humanos, uma vez que humanos diferem muito entre si quanto a sua constituição genética e seu ambiente.

Este capítulo analisa as diferenças que existem entre diferentes humanos quanto à sequência de DNA de seus genomas (diferenças entre humanos e outras espécies foram abordadas no Capítulo 10). Agora que se têm sequências genômicas completas de vários indivíduos conhecidos, pode-se notar que eles diferem de um para o outro em inúmeros modos. A maior parte das diferenças entre genomas humanos individuais parece não ter qualquer efeito. Outras diferenças afetam o fenótipo, produzindo o espectro normal de determinantes genéticos das variações na constituição do corpo, na pigmentação, no metabolismo, no temperamento e assim por diante, tornando-nos únicos. Todas estas são variantes normais. Algumas são patogênicas – ou seja, causam doenças ou tornam o seu portador suscetível a doenças que podem ou não ser desenvolvidas, dependendo de seus outros genes, seu estilo de vida, seu ambiente ou mero acaso. Neste capítulo, considera-se primeiramente as variantes normais e depois aquelas que são patogênicas. Existe outro nível de variação genética que não será abordado aqui: variação epigenética (variações na metilação do DNA e na conformação da cromatina) entre indivíduos. Isto foi descrito no Capítulo 11, mas ainda está sob debate a questão de quanto do fenótipo é conferido pela genética e quanto se deve a variações epigenéticas entre nós.

13.1 TIPOS DE VARIAÇÃO ENTRE GENOMAS HUMANOS

A variação genética humana inclui desde mudanças nucleotídicas pontuais até o ganho (ou perda) de cromossomos inteiros. Variantes podem ser classificadas como de larga ou de pequena escala, com base no fato de poderem ou não ser detectadas pelo sequenciamento de um produto de PCR convencional, com poucas centenas de nucleotídeos. Variantes de pequena escala normalmente causam efeito maior, caso tenham algum efeito, sobre um único gene, enquanto variantes de larga escala geralmente afetam muitos genes. Na realidade, evidentemente, existe um contínuo entre estes limites.

Polimorfismos de nucleotídeo único são numericamente o tipo de variação genética mais abundante

Quando genomas de diferentes humanos são comparados, a grande maioria dos nucleotídeos é quase sempre a mesma em todo mundo. Esta é a razão pela qual é possível falar de forma generalizada sobre o genoma humano. Variantes ocasionais podem ser vistas na posição de qualquer nucleotídeo, mas são quase sempre raras. No entanto, cerca de um nucleotídeo a cada 300 é **polimórfico** – isto é, mais de uma forma é comum na população (**Quadro 13.1**). Estes **polimorfismos de nucleotídeo único (SNPs, do inglês *single nucleotide polymorphisms*)** são catalogados no banco de dados público dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP>) e designados por números iniciados por rs (do inglês, *reference SNP*; **Quadro 13.2**).

Na maioria dos casos, um SNP possui duas formas alternativas (alelos) – por exemplo, A ou G em uma certa posição (**Figura 13.1A**). A frequência do alelo mais raro pode

QUADRO 13.1 Polimorfismo e mutação: palavras com vários significados

A palavra **polimorfismo** é utilizada por geneticistas humanos para se referir a várias questões diferentes, em momentos diferentes. Parece desnecessário dizer, mas isso pode causar alguma confusão.

- Geneticistas moleculares geralmente descrevem uma variante como um polimorfismo se sua frequência na população for superior a algum valor arbitrário, frequentemente 0,01. Por exemplo, o SNP rs1447295 (ver Quadro 13.2 para uma explicação sobre a notação), localizado em 8q24, é um polimorfismo com dois alelos, C e A. Suas frequências são 0,93 e 0,07, respectivamente, na população europeia do HapMap. Este é o sentido que se utilizará neste livro, a menos que especificado diferentemente. Variantes com uma frequência abaixo do ponto de corte arbitrário podem ser descritas como variantes raras.
- Em genética de populações, os pesquisadores definem um polimorfismo como a coexistência estável na população de mais de um genótipo em frequências tais que o tipo raro não poderia ser mantido apenas por mutações recorrentes. Nesta definição, algumas mutações patogênicas também seriam consideradas polimorfismos. Por exemplo, a

mutação mais comum causadora de fibrose cística nas populações do norte europeu (classificação p.F508del; ver Quadro 13.2 para uma explicação sobre a notação) apresenta uma frequência entre 0,01-0,02 em populações do norte europeu. Conforme apresentado no Capítulo 3, isto não poderia ser mantido por mutações recorrentes em função da pressão seletiva contra pessoas com fibrose cística.

- Geneticistas clínicos muitas vezes utilizam polimorfismo para se referir a variantes não patogênicas, independentemente de sua frequência. Variantes patogênicas, sejam comuns ou raras, são descritas como mutações.

A palavra mutação também pode ser utilizada com sentidos distintos, referindo-se tanto ao processo como ao produto:

- um evento que altera a sequência de DNA – a radiação UV produziu uma mutação no DNA.
- uma mudança na sequência de DNA que pode ter acontecido há muito tempo – ela herdou uma mutação de seu pai.

QUADRO 13.2 Nomenclatura para descrever variantes de DNA e de aminoácidos

Cada polimorfismo de nucleotídeo único (SNP, na sigla em inglês) presente no banco de dados público dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP>) pode ser referido por seu identificador único, como o rs212570, em que rs se refere a “SNP referencia” (na sigla em inglês) e 212570 é um número serial único. Polimorfismos de repetições curtas em *tandem* (STRPs, na sigla em inglês) possuem identificadores tais como D6S282, em que D se refere a segmento de DNA, 6 é o número do cromossomo no qual o marcador está localizado, S significa mutação de cópia única (do inglês *single copy sequence*) e 282 é um número serial único.

Qualquer mudança na sequência pode ser descrita utilizando-se as convenções definidas no endereço eletrônico da *Human Genome Variation Society*, <http://www.hgvs.org/mutnomen/>. Os casos mais comuns são descritos a seguir.

Todas as variantes apresentam o prefixo g. (genoma), c. (cDNA), r. (RNA) ou p. (proteína).

Substituições de nucleotídeos

Para mudanças em um gene, o A do códon iniciador ATG é numerado como +1; a base imediatamente anterior a ele é -1. Não existe zero. O número do nucleotídeo alterado é seguido da mudança.

- **g.1162G>A** – no DNA genômico, substituição de uma guanina na posição 1162 por uma adenina.

Para mudanças nos íntrons, quando apenas a sequência do cDNA é conhecida por completo, especifica-se o número do íntron com IVS (do inglês *intervening sequence*) ou o número do éxon mais próximo.

- **g.621+1G>T** ou **IVS4+1G>T** – substituição de G por T na primeira base do íntron 4 (o nucleotídeo 621 é a última base do éxon 4).

Substituições de aminoácidos

Utiliza tanto códigos de uma letra (X indica um códon de parada) como códigos de três letras. Proteínas são numeradas com o iniciador metionina como códon 1.

- **p.R117H** ou **Arg117His** – substitui a arginina 117 por uma histidina.
- **p.G542X** ou **Gly542Stop** – substitui o códon da glicina 542 por um códon de parada.

Deleções e inserções

Utilize *del* para deleções e *ins* para inserções, precedido pela posição do nucleotídeo ou do intervalo (para mudanças no DNA) ou o símbolo do aminoácido (código de uma letra; para trocas de aminoácidos).

- **p.F508del** – em uma proteína (p), deleção da fenilalanina (F) 508.
- **c.6232_6236del** ou **c.6232_6236delATAAG** – deleção de cinco nucleotídeos começando pelo nucleotídeo 6232 do cDNA. A identidade dos nucleotídeos deletados pode ser especificada.
- **g.409_410insC** – insere C entre os nucleotídeos 409 e 410 do DNA genômico.

O programa, Mutalyzer, foi desenvolvido para garantir o nome correto para qualquer variante de sequência que o usuário inserir; ver em <http://www.lovd.nl/mutalyzer/>.

ser qualquer uma abaixo de 0,5. Não seria de se esperar que alelos de SNP com alta frequência tenham grande efeito fenotípico, pois a seleção natural deve ter garantido sua eliminação – caso seja danoso, ou sua fixação (presente em toda a população) – caso seja benéfico. No entanto, estes processos evolutivos podem ser incompletos no presente para um dado alelo. Adicionalmente, a vantagem do heterozigoto é um mecanismo que pode produzir polimorfismos estáveis mesmo quando um dos alelos é danoso em homozigose. Isto acontece quando portadores assintomáticos de uma condição deletéria recessiva apresentam alguma vantagem seletiva sobre os homozigotos normais, como ocorre com a anemia falciforme (ver Capítulo 3, p. 64). O argumento da seleção natural é menos poderoso com SNPs raros, mas mesmo assim, de modo geral, também não seria de se esperar que estes tivessem um efeito fenotípico, pois a maioria dos SNPs não está localizada em sequências codificantes ou regulatórias.

Alguns sítios polimórficos possuem variantes mais complexas. Um SNP pode possuir três alelos, A, G ou C, ou dois SNPs podem ser adjacentes um ao outro. A grande maioria das variantes mais complexas no banco de dados dbSNP (cerca de 2 dos 12 milhões de entradas) são polimorfismos de inserção/deleção (DIPs, do inglês *deletion/insertion polymorphisms*, ou *indels*). Geralmente um ou dois nucleotídeos não repetidos são inseridos ou deletados (Figura 13.1B). Inserções e deleções de unidades repetidas possuem causas e dinâmica distintas, sendo discutidas após a seção sobre repetições curtas em *tandem* (p. 408).

Algumas vezes um SNP afetar um nucleotídeo que é parte do sítio de reconhecimento de uma ou de outra enzima de restrição. Embora muitas centenas de enzimas de restrição sejam conhecidas, muitas delas reconhecem uma sequência **palindrômica** como GAATTC. A sequência é palindrômica porque a fita complementar, lida na direção 5'→3', também é GAATTC:



Apenas cerca de 10% de todos os nucleotídeos se localizam em sequências palindrômicas, de modo que a maioria dos SNPs não afeta nenhum sítio de restrição. Quando ocorre de um SNP se localizar em um sítio de restrição, apenas um alelo irá reter a sequência necessária para o reconhecimento. Assim sendo, o polimorfismo irá criar ou abolir um sítio de restrição. SNPs deste tipo (Figura 13.1C), conhecidos como **polimorfismos**

(A)
GCCTGTTTTATATTAC/TGATCCAATTTTTTCA
(B)
GAGACAGAGTTTCGC(T)TCTTGTGCCAGGCT
(C)
CCAAGCCTGGAGCTA/GGCCGTGGCCAGGCAAG

Figura 13.1 Tipos de polimorfismos de nucleotídeo único (SNP). (A) Um SNP simples. O rs212570 é uma variante C/T no cromossomo 20. (B) Um polimorfismo de deleção/inserção. O rs36126541 é uma inserção ou deleção de um único nucleotídeo T na sequência do cromossomo 12. (C) Um polimorfismo de fragmento de restrição (RFLP, do inglês *restriction fragment length polymorphism*). O SNP rs36078338 cria ou elimina a sequência GCTAGC, a qual é o sítio de reconhecimento para a enzima de restrição *NheI*.

de fragmento de restrição (RFLP, do inglês *restriction fragment length polymorphisms*) ou polimorfismos de sítio de restrição (RSPs, do inglês *restriction site polymorphisms*), foram os primeiros marcadores de DNA amplamente utilizados e empregados na criação do primeiro mapa genético de ligação do genoma humano (ver Capítulo 8).

Por que alguns nucleotídeos devem ser polimórficos enquanto os nucleotídeos vizinhos raramente apresentam variantes? Em geral não é porque algo neste nucleotídeo o torna especialmente suscetível a sofrer mutação. Provavelmente, as variantes alternativas distinguem segmentos cromossômicos ancestrais alternativos que são comuns na população atual. Em comparação com repetições curtas em *tandem*, SNPs são estáveis ao longo do tempo evolutivo. Poucos SNPs são específicos a uma raça humana ou a um grupo étnico, embora as frequências alélicas possam variar entre os grupos – portanto, a maioria deles tem origem desde o passado remoto da evolução humana. A habilidade de definir segmentos cromossômicos ancestrais por meio de seu conteúdo de SNPs é extremamente importante na pesquisa genética e será uma importante preocupação no Capítulo 15.

Tanto sequências intercaladas como as repetidas em *tandem* podem apresentar variação polimórfica

Conforme observado no Capítulo 9, cerca de 50% do genoma humano são constituídos de sequências repetidas. Muitas são repetições intercaladas – isto é, sequências repetidas que estão dispersas por todo o genoma (ou por parte do genoma), em vez de estarem agrupadas. A maioria das sequências intercaladas no genoma humano é derivada de transposons (genes “saltadores” que podem se espalhar pelo genoma como uma infecção intracelular – ver Capítulo 9). Várias famílias de repetições derivadas de transposons estão presentes em número elevado – existe talvez 1,5 milhão de cópias do elemento SINE (do inglês *short interspersed nuclear elements*), com 100 a 300 pb, e 850 mil elementos LINE (do inglês *long interspersed nuclear elements*), com 6 a 8 kb dispersos pelo genoma. Algumas inserções específicas são polimórficas no genoma humano, estando presentes em alguns cromossomos, mas ausentes em outras cópias dos mesmos cromossomos.

Repetições em *tandem* também são comuns. Unidades de repetição podem ser desde nucleotídeos únicos (p. ex., uma sequência de nucleotídeos A) até 100 ou mais nucleotídeos. Repetições com unidades maiores são chamadas **satélites**. O nome vem de estudos iniciais nos quais o DNA genômico era submetido à sedimentação por gradiente de densidade em ultracentrífugas. O DNA repetido tinha uma densidade de sedimentação diferente do restante do DNA, formando uma pequena banda satélite. Um exemplo conhecido é o DNA α -satélite, encontrado nos centrômeros dos cromossomos, o qual consiste em repetições em *tandem* de uma unidade com 171 pb. Repetições com unidades de 10-50 nucleotídeos são chamadas de **minissatélites** e aquelas com unidades menores são chamadas de **microssatélites**. A maioria dos microssatélites possui unidades de repetição com 1, 2 ou 4 nucleotídeos (**Figura 13.2**).

Polimorfismos de repetições curtas em *tandem*: a mão de obra para estudos forenses e de parentesco

Caso genomas de diferentes indivíduos sejam comparados, é perceptível uma variação no número de unidades de muitas das repetições em *tandem*. Diferentemente das variações de SNPs, variações nas repetições em *tandem* são o resultado de eventos relativamente recentes. São de dois tipos:

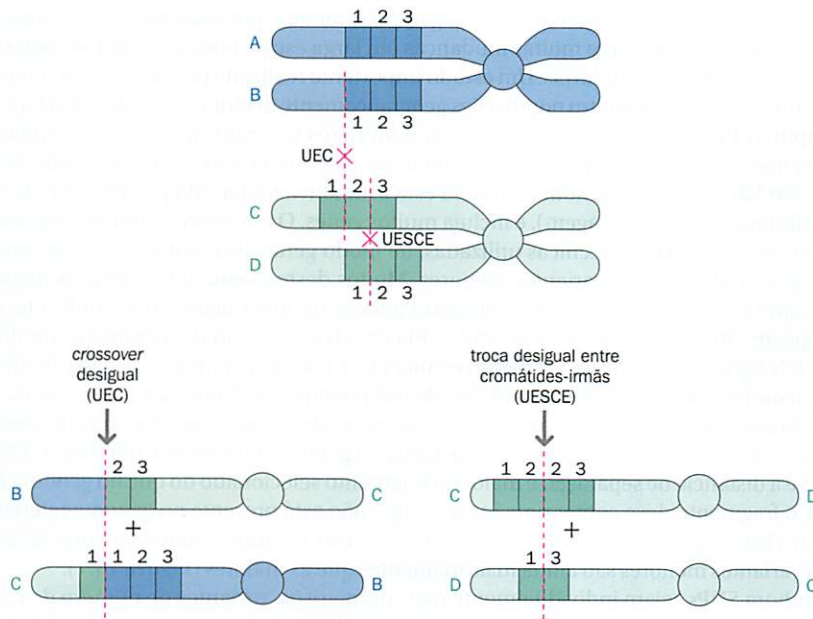
- Recombinação meiótica entre repetições malpareadas produz mudanças graduais no número de unidades (**Figura 13.3**). Acredita-se que esta seja o principal mecanismo gerador de diversidade nos minissatélites.
- Deslizes da polimerase durante a replicação do DNA podem alterar o comprimento das repetições por uma ou talvez duas unidades. Estes são o principal mecanismo gerador de polimorfismos de microssatélite (**Figura 13.4**).

Figura 13.2 Microssatélites. (A) D6S282 é uma repetição $(CA)_n$ no cromossomo 6. (B) D12S391 é uma repetição de quatro nucleotídeos, $(AGAT)_n$, no cromossomo 12. (C) D21S11 é uma mutação complexa de quatro nucleotídeos no cromossomo 21. Sequências sublinhadas variam no número de repetições; a figura mostra exemplo de alelos típicos.

(A)
ACAGAGATAGACACACACACACACACACACACAAACAAGCATGCTC

(B)
ATCAATGGATGCATACGT (AGAT)₁₅GAGAGGGGATTTATTAGAGGAATTAGC

(C)
TGAATTGCCT (TCTA)₄ (TCTG)₆ (TCTA)₃ TA (TCTA)₃ TCA (TCTA)₁₀ TCGTCTATC



Microsatélites polimórficos (**polimorfismos de repetições curtas em tandem**, ou **STRPs**, do inglês *short tandem repeat polymorphisms*) têm sido os marcadores genéticos de escolha para estudos forenses e de parentesco desde o início dos anos 1990. Eles são mais informativos que os SNPs para a distinção entre indivíduos ou para seguir um segmento cromossômico específico por meio de uma linhagem, pois deve haver um número maior de alelos na população (p. ex., um microsatélite pode ter entre 5 e 20 unidades de repetição em diferentes indivíduos). Os primeiros anos do Projeto Genoma Humano foram amplamente dedicados à definição e ao mapeamento de STRPs suficientes para construir um mapa com alta definição utilizando marcadores ao longo de todo o genoma humano; cerca de 150 mil STRPs foram identificados. Embora os STRPs permaneçam sendo a ferramenta de escolha para trabalhos forenses, trabalhos de ligação passaram a utilizar SNPs, uma vez que eles podem ser genotipados em larga escala utilizando microarranjos.

Variações em larga escala no número de cópias são surpreendentemente frequentes em genomas humanos

Variantes suficientemente grandes para serem vistas por um citogeneticista em um microscópio são quase sempre o resultado de acidentes patogênicos isolados, não sendo parte da variação humana normal. Citogeneticistas reconhecem apenas três tipos de variantes normais relativamente comuns:

- O tamanho das regiões heterocromáticas nos centrômeros dos cromossomos 1, 9 ou 16, assim como o braço longo do cromossomo Y, podem variar.
- Os braços curtos dos cromossomos acrocêntricos (cromossomos 13, 14, 15, 21 e 22, os quais possuem o centrômero próximo a uma extremidade - ver Figura 2.15) variam consideravelmente em tamanho e morfologia. Frequentemente a parte mais distal aparece como um satélite conectado ao corpo principal do cromossomo por uma haste curta e estreita. Note que este uso da palavra *satélite* não apresenta relação com seu uso na descrição de repetições de DNA em *tandem*.
- Uma variedade de **sítios frágeis** - segmentos de cromatina desenrolada - pode ser vista quando células são cultivadas sob condições que dificultam a replicação do DNA - por exemplo, privando a célula de timidina ou adicionando afidicolina, um inibidor da DNA polimerase (Figura 13.5). A maioria dos sítios frágeis é variante normal, embora os sítios *FRAXA* e *FRAXE*, discutidos a seguir (ver p. 424), sejam patogênicos.

Todas estas mudanças refletem variações no número de cópias de seqüências repetidas em *tandem*. Até recentemente se assumia que qualquer deleção ou inserção de um trecho longo de DNA não repetitivo seria patogênica. O advento de técnicas como a hibridização genômica de arranjo comparativo (na qual o DNA teste e o controle competem para hibridizar em um microarranjo; ver Capítulo 2) e arranjo de SNP do genoma completo tornaram

Figura 13.3 Crossover desigual e trocas desiguais entre cromátides irmãs causam inserções e deleções. A figura mostra o resultado da recombinação entre repetições mal-alinhadas em um arranjo de repetições em *tandem*. As cromátides mal-alinhadas podem estar em cromossomos homólogos, em cada caso o resultado é um crossover desigual (UEC, do inglês *unequal crossover*), ou podem estar em cromátides-irmãs, produzindo uma troca desigual entre cromátides-irmãs (UESCE, do inglês *unequal sister chromatid exchange*). Em ambos os casos, o resultado irá produzir duas cromátides, uma contendo algumas repetições extras, a outra com falta das respectivas unidades de repetição. Por uma questão de simplificação, as quebras são indicadas entre as unidades de repetição, mas elas poderiam igualmente ocorrer em posições dentro das unidades.

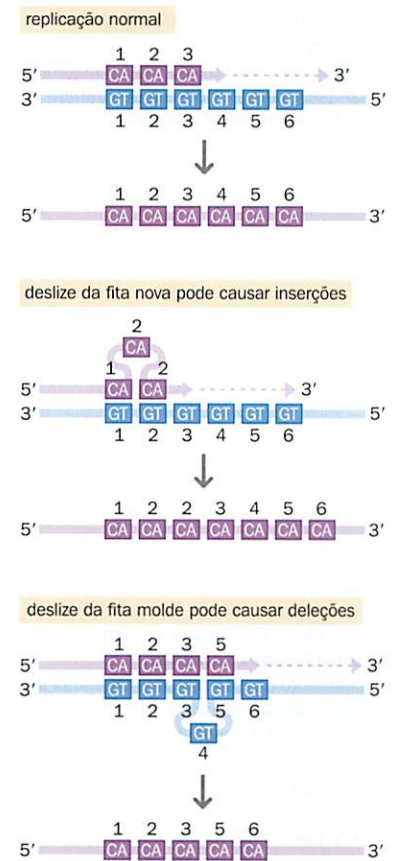


Figura 13.4 Mau pareamento causado por deslizamento de fita durante a replicação do DNA. Uma nova fita de DNA (rosa) está sendo sintetizada, utilizando a fita azul como molde. Durante a replicação normal do DNA, a fita nascente se dissocia parcialmente do molde e depois se associa novamente. Quando existe uma seqüência repetida em *tandem*, a fita nascente pode ficar malpareada com o molde durante a reassociação. Isso pode resultar na perda ou no ganho de unidades de repetição na nova fita, com relação à fita molde.

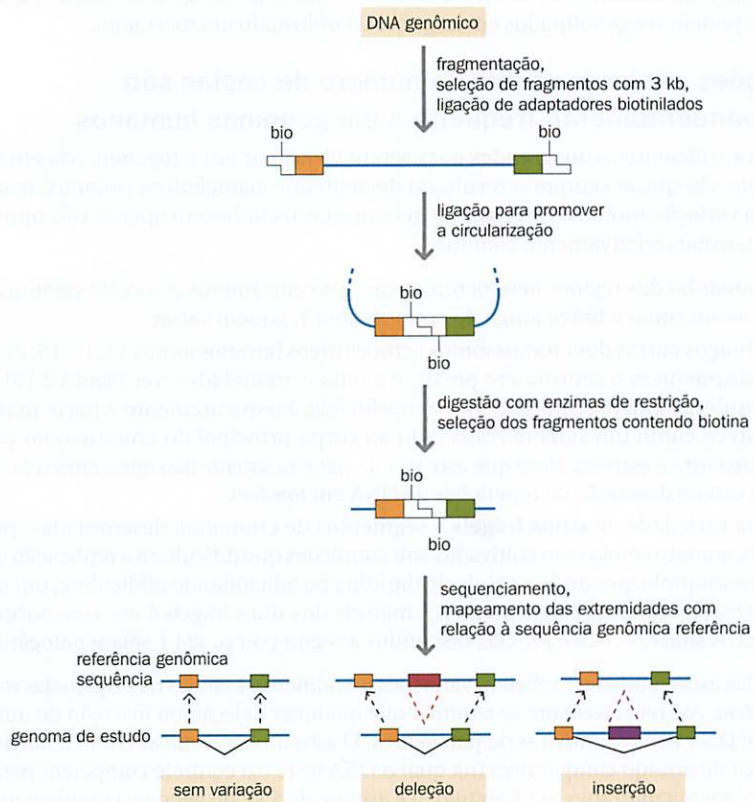


Figura 13.5 Um sítio frágil no cromossomo. Uma porção de cromatina relativamente não condensada em um cromossomo metafásico altamente condensado. Existem cerca de 120 localizações no genoma humano onde o cromossomo de um indivíduo normal possui uma tendência a apresentar um sítio frágil quando células são cultivadas sob condições especiais. A imagem mostra um cromossomo X com o sítio frágil patogênico *FRAXA* (seta cinza). (Cortesia de Graham Fewes, University of Birmingham.)

relativamente fácil, pela primeira vez, triar genomas inteiros por material extra ou ausente. Logo se tornou aparente que muitas mudanças em larga escala podem ser encontradas nos genomas de indivíduos saudáveis. Um estudo importante realizado por Redon et al., com 269 indivíduos saudáveis de quatro populações geograficamente distintas (amostras do HapMap; ver Capítulo 15), identificou 1.447 regiões com números variáveis de cópias de seqüências de pelo menos 1 kb de comprimento. No total, foi relatado que estas regiões variáveis cobriam 360 Mb, ou 12% do genoma (embora esta estimativa tenha sido posteriormente revista, diminuindo a porcentagem), e incluía muitos genes. O tamanho médio dos segmentos variáveis foi de 250 kb. As técnicas utilizadas, de modo geral, apresentaram pouca sensibilidade para a detecção de variantes menores. Muitos destes segmentos variáveis menores teriam sido perdidos. Essas variantes menores podem ser identificadas utilizando a técnica de mapeamento de extremidades pareadas. Ela envolve a seleção de fragmentos randômicos de DNA genômico, com um tamanho conhecido, e o sequenciamento de algumas dúzias de nucleotídeos de ambas as extremidades de cada fragmento. É possível então comparar a distância entre as seqüências terminais no fragmento selecionado (cujo tamanho é conhecido) com a distância entre as mesmas seqüências no genoma humano de referência (**Figura 13.6**). Se a distância de separação é maior no fragmento selecionado do que no genoma referência, o fragmento deve conter uma inserção que não está presente no genoma referência; por outro lado, uma distância menor indica uma deleção. Estudos recentes têm demonstrado que as variantes menores são ainda mais frequentes que as maiores (**Figura 13.7**).

Embora SNPs sejam individualmente mais numerosos, variantes de número de cópias (CNVs, do inglês *copy-number variants*) são responsáveis, de longe, pelo maior número de nucleotídeos que difere entre dois genomas. Assim sendo, o genoma humano é consideravelmente mais variável entre pessoas normais (saudáveis) do que havia se admitido inicialmente. Os estudos mencionados não revelaram se, nos casos em que havia ganho no número de cópias, as várias cópias estavam distribuídas ao longo de vários cromossomos ou se estavam agrupadas em *tandem*. Outros estudos indicam que eles comumente formam agrupamentos em *tandem*. O *Database of Genomic Variants* (TCAG; <http://projects.tcag.ca/variation/>) é um banco de dados de variantes observadas em pessoas aparentemente saudáveis, enquanto o banco de dados *Decipher* (<http://decipher.sanger.ac.uk/>) armazena variantes observadas em pessoas com anormalidades fenotípicas (embora raramente se saiba com certeza se quaisquer variantes específicas são a causa da anormalidade fenotípica).

Figura 13.6 O princípio do mapeamento de extremidades pareadas. No mapeamento de extremidades pareadas são utilizados pequenos fragmentos (tamanho definido) de DNA genômico compartilhado de clones BAC ou PAC, os quais são marcados pela ligação com adaptadores biotinilados. Em ambos os casos, o objetivo é identificar uma seqüência representando as duas extremidades do DNA original, reunidas pela circularização. Na técnica de salto sobre o cromossomo (*chromosome jumping*), isto é utilizado para encontrar seqüências (verde) localizadas a 80 a 130 kb de uma seqüência conhecida (azul), com o objetivo de acelerar o processo de construção de um clone *contig* por meio de uma grande região genômica. No mapeamento de extremidades pareadas o objetivo é identificar variantes estruturais por meio da detecção de seqüências que estão separadas a uma distância diferente no genoma teste, quando comparadas a seqüência do genoma humano de referência.



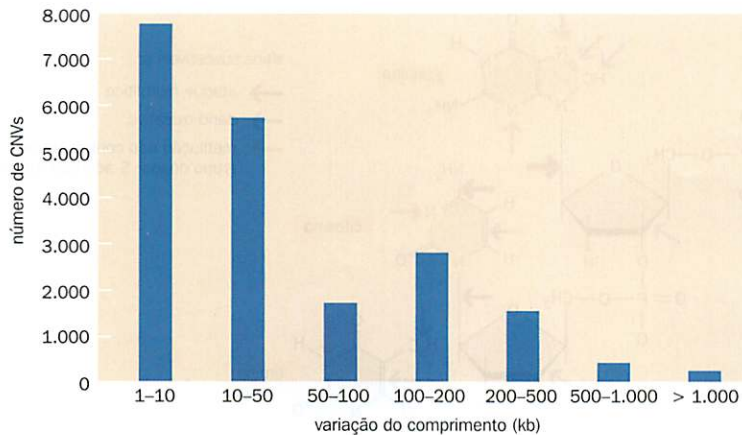


Figura 13.7 Distribuição de tamanho das variantes de número de cópias no genoma humano. Apenas variantes maiores que 1 kb são apresentadas. O pico na distribuição de tamanho entre 100 a 200 kb é parcialmente um artefato da tecnologia de arranjo em BAC, utilizado em muitos estudos; muitas das variantes relatadas nessa variação de comprimento são na verdade menores. [Dados reproduzidos com a permissão do *Database of Genomic Variants*, <http://projects.tcag.ca/variation>, acessado em janeiro de 2009.]

As sequências genômicas individuais que estão agora se tornando disponíveis fornecem uma imagem impressionante da variabilidade geral presente em indivíduos saudáveis.

O cientista pioneiro Craig Venter foi o primeiro indivíduo cuja sequência completa do genoma diploide foi determinada. A sequência referência do genoma humano foi composta por doadores anônimos. Comparada com esta referência, as seguintes variantes foram observadas no genoma de Venter:

- 3,2 milhões de SNPs
- 290 mil variantes de inserção/deleção em heterozigose (variando de 1 a 571 pb de comprimento)
- 559 mil variantes de inserção/deleção em homozigose (variando de 1 a 82.711 pb de comprimento)
- 90 inversões grandes
- 62 variantes com alto número de cópias

Um total de 12.290.978 nucleotídeos diferiram da sequência referência. Embora a maior parte desta variação tenha sido no DNA não codificante, 44% dos genes de Venter apresentavam uma variação de sequência com relação à sequência referência do genoma humano; 17% codificavam proteínas alteradas, incluindo 317 genes, nos quais algumas variantes foram identificadas como patogênicas.

O genoma diploide de James Watson também foi publicado e apresenta uma quantidade semelhante de variação, com 3,3 milhões de SNPs, dos quais 10.654 causam trocas de aminoácido em proteínas.

13.2 DANO AO DNA E MECANISMOS DE REPARO

Algumas variantes de DNA surgem de erros na replicação ou na recombinação, mas a principal fonte são falhas no reparo a danos no DNA. A visão do DNA como um arquivo genético estável, seguramente protegido no núcleo celular, esconde a dificuldade da manutenção dessa estabilidade. Ataques químicos por agentes endógenos ou exógenos e erros gerados durante sua função normal representam ameaças constantes à integridade do genoma (Figura 13.8).

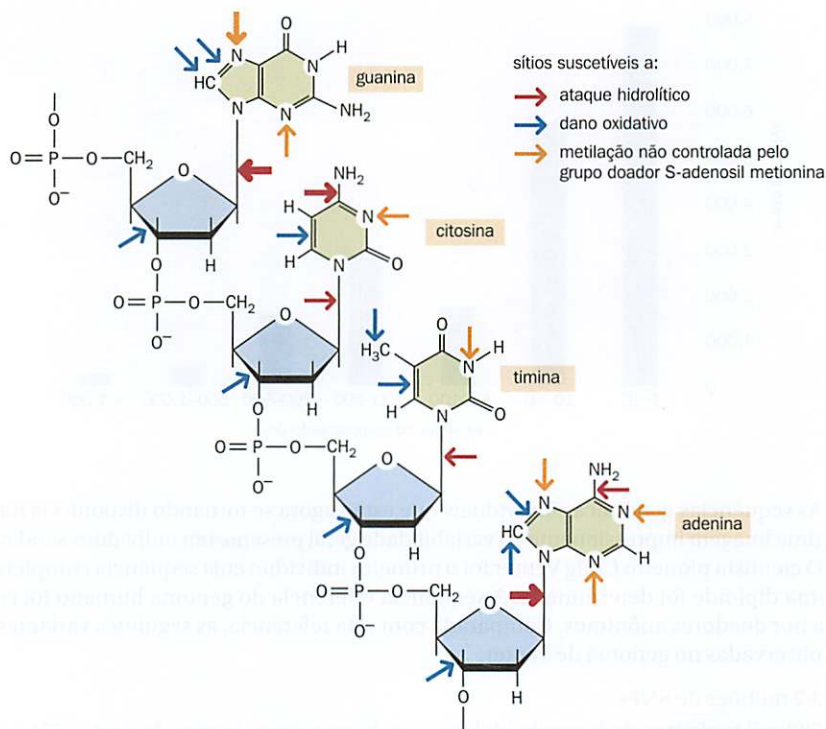
O DNA celular requer constante manutenção para reparar danos e corrigir erros

Os agentes que causam dano ao DNA podem ser tanto externos à célula como provenientes de efeitos indesejados da química intracelular.

Existem três agentes externos principais, com grande probabilidade de causar dano ao DNA:

- **Radiação ionizante** - raios gama e raios X podem causar quebras unifilamentares ou bifilamentares na cadeia principal de açúcar-fosfato.
- **Radiação ultravioleta** - raios UV-C (com um comprimento de onda de aproximadamente 260 nm) são especialmente danosos, mas a maior fonte de dano em humanos são os raios UV-B (280-315 nm) na luz solar que penetra a camada de ozônio. Radiação

Figura 13.8 Ameaças à integridade do DNA genômico. A figura indica os sítios em cada nucleotídeo que são reconhecidamente alterados pelo ataque hidrolítico (setas vermelhas), dano oxidativo espontâneo (setas azuis) e metilação não controlada pelo grupo doador S-adenosil metionina (setas amarelas). A largura de cada seta indica a frequência relativa de cada evento. Radiações iônica e ultravioleta representam ameaças adicionais não indicadas nesta figura. [De Alberts B, Johnson A, Lewis J et al. (2007) *Molecular Biology of the Cell*, 5th ed. Garland Science/Taylor & Francis LLC, após Lindahl T (1993) *Nature* 362, 709-715. Com a permissão de Macmillan Publishers Ltd.]



UV causa ligação cruzada entre pirimidinas adjacentes em uma fita de DNA, formando dímeros de ciclobutano pirimidina (Figura 13.9D) e outros fotoprodutos anormais.

- **Produtos químicos ambientais** - estes incluem hidrocarbonetos (p. ex., na fumaça do cigarro), algumas plantas e produtos microbiológicos, tais como as aflatoxinas encontradas em amendoim mofo, além de agentes químicos utilizados no tratamento quimioterápico contra o câncer. Agentes alquilantes podem transferir um grupo metil ou outro grupo alquila para as bases de DNA e podem causar ligação cruzada dentro de uma fita ou entre diferentes fitas de DNA.

Embora agentes externos sejam mais visíveis, as principais ameaças para a estabilidade do DNA celular vêm de eventos químicos internos, tais como os ilustrados nas Figuras 13.8 e 13.9:

- **Depurinação** - cerca de 5 mil bases de adenina e guanina são perdidas todos os dias de cada célula nucleada pela hidrólise espontânea da ligação base-açúcar (ver Figuras 13.8 e 13.9A).
- **Deaminação** - diariamente, em cada célula nucleada humana, pelo menos 100 citosinas são espontaneamente deaminadas para produzir uracila (ver Figuras 13.8 e 13.9B). Com menor frequência, a deaminação espontânea da adenina produz hipoxantina.
- **Ataque por espécies reativas de oxigênio** - ânions superóxido (O_2^-) altamente reativos e moléculas relacionadas são gerados como um subproduto do metabolismo oxidativo na mitocôndria. Eles também podem ser produzidos pelo impacto da radiação ionizante em constituintes celulares. Estas espécies reativas de oxigênio atacam anéis de purina e de pirimidina (ver Figura 13.8).
- **Metilação não enzimática** - metilação accidental do DNA (não enzimática) pela S-adenosil metionina produz cerca de 300 moléculas da base citotóxica 3-metiladenina, por célula por dia, além de uma quantidade da molécula menos danosa 7-metilguanina (ver Figura 13.8). Esse é um processo muito distinto da metilação enzimática da citosina para produzir 5-metilcitosina, o qual as células utilizam como um importante método de controle da expressão gênica (ver Capítulo 11). As bases metiladas de adenina e guanina distorcem a dupla-hélice e interferem com interações vitais entre DNA e proteínas.

Além do dano causado por esses diversos processos, erros podem ocorrer durante o metabolismo normal do DNA. Certa taxa de erro (incorporação do nucleotídeo errado) é inevitável durante a replicação do DNA. Mecanismos de revisão corrigem a grande maio-

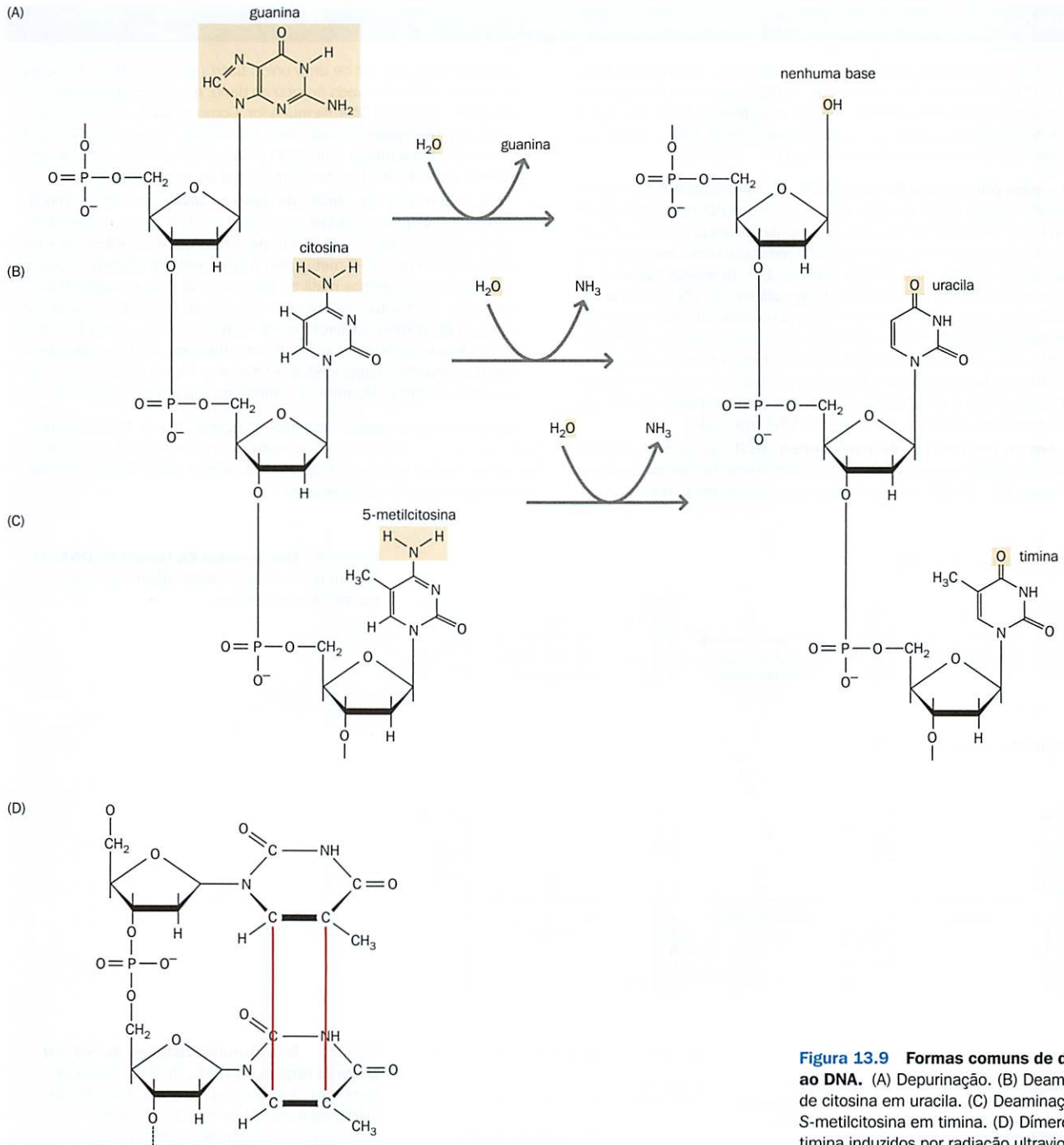


Figura 13.9 Formas comuns de dano ao DNA. (A) Depurinação. (B) Deaminação de citosina em uracila. (C) Deaminação de 5-metilcitosina em timina. (D) Dímeros de timina induzidos por radiação ultravioleta.

ria dos maus pareamentos resultantes, mas algumas podem persistir, produzindo variações na sequência. Falhas nos mecanismos de revisão nas células somáticas constituem uma das causas de câncer (ver Capítulo 17). Além disso, erros ocasionais na replicação ou na recombinação deixam quebras na fita de DNA, as quais precisam ser reparadas para que a célula sobreviva.

Os efeitos do dano ao DNA

O dano ao DNA acarreta dois possíveis efeitos sobre a célula. Primariamente ele é citotóxico. Quando o DNA é replicado, muitos tipos de dano causarão a parada da forquilha de replicação. Mesmo fora da fase S do ciclo celular, a RNA-polimerase irá parar em lesões no DNA durante a transcrição do RNA, prevenindo a expressão do gene danificado. Estes problemas são potencialmente letais à célula. Complexos multiproteicos especiais patrulham constantemente o DNA celular, detectando e respondendo ao dano. A progressão no ciclo

QUADRO 13.3 Mecanismos de reparo do DNA em células humanas

Células humanas utilizam pelo menos seis diferentes mecanismos de reparo de DNA. Três deles são usados para corrigir bases anormais quando estas estão presentes em apenas uma das duas fitas do DNA – ou seja, a base modificada está pareada com uma base normal. A base danificada pode ser reparada ou removida e substituída.

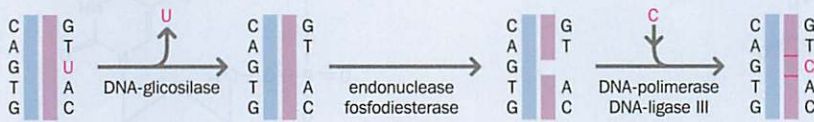
- O **reparo por excisão de base (BER, do inglês *base excision repair*)** corrige a maior parte do dano mais comum ao DNA (na ordem de 20 mil bases alteradas por dia em cada célula nucleada do corpo humano). Glicosilases envolvidas com o BER removem bases anormais pela quebra da ligação açúcar-base (Figura 1A). Humanos apresentam pelo menos oito genes que codificam diferentes DNA-glicosilases, cada uma responsável por identificar e remover um tipo específico de base danificada. Após a remoção da base danificada, uma endonuclease e uma fosfodiesterase cortam a cadeia principal de açúcar-fosfato na posição da base faltante e removem o resíduo de açúcar-fosfato. O *gap* é preenchido pela síntese com uma DNA-polimerase, e o corte remanescente é selado pela DNA-ligase III.
- O **reparo por excisão de nucleotídeo (NER, do inglês *nucleotide excision repair*)** remove dímeros de timina e grandes adutos químicos (Figura 1B). O NER remove e ressintetiza uma ampla região no en-

torno do dano, em vez de uma única base (como no BER). A cadeia de açúcar-fosfato é clivada no sítio do dano, e exonucleases removem um amplo trecho do DNA flanqueador. Como no BER, o *gap* é preenchido pela ressíntese e selado pela DNA-ligase. O reparo por excisão de nucleotídeo também é utilizado para corrigir quebras simples (em apenas uma das fitas) na cadeia principal de açúcar-fosfato.

- O **reparo por reversão direta do dano ao DNA** é um mecanismo de reparo de DNA pouco utilizado em humanos. Três genes humanos foram implicados neste mecanismo, dentre os quais o melhor caracterizado codifica uma O-6-metilguanina-DNA-metiltransferase, a qual é capaz de remover grupos metil de guaninas que foram incorretamente metiladas. Em muitos organismos, dímeros de timina produzidos por radiação UV podem ser diretamente resolvidos pela enzima fotoliase utilizando a energia da luz visível (fotorreativação). Embora mamíferos possuam enzimas relacionadas à fotoliase, eles a utilizam para um propósito bastante diferente, o controle de seu relógio circadiano.

Nos processos de reparo descritos, a segunda fita de DNA (não danificada) serve como molde para a acurada reconstrução da fita danificada. Danos que afetam ambas as fitas do DNA exigem mecanismos distintos. Existem dois mecanismos principais.

(A) reparo por excisão de base



(B) reparo por excisão de nucleotídeo

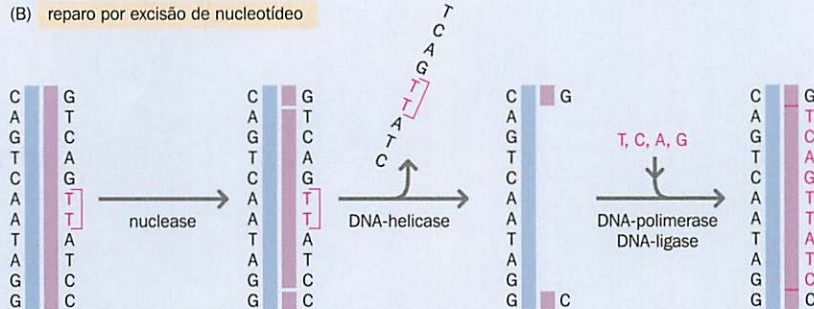
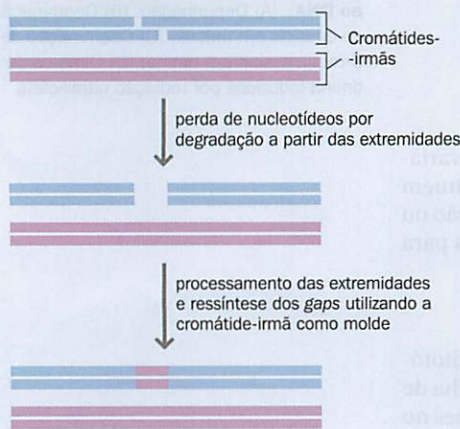


Figura 1 Mecanismos de reparo do DNA. (A) Reparo por excisão de base. (B) Reparo por excisão de nucleotídeo.

(A) recombinação homóloga



(B) reunião de extremidades não homólogas

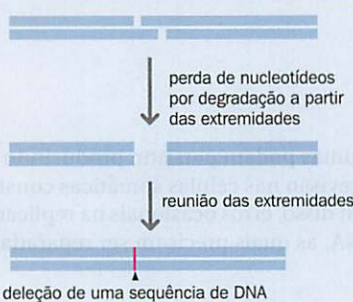


Figura 2 Duas formas distintas de reparar quebras duplas no DNA. (A) Recombinação homóloga. A sequência correta é reconstruída utilizando a cromátide-irmã intacta como um molde (detalhes moleculares não são apresentados). Isso produz um reparo acurado, mas nem sempre é possível. (B) A reunião de extremidades não homólogas é sempre possível, mas a fita reparada apresentará nucleotídeos para mais ou para menos na região da junção. [De Alberts B, Johnson A, Lewis J et al. (2007) *Molecular Biology of the Cell*, 5th ed. Garland Science/Taylor & Francis LLC.]

- Na **recombinação homóloga**, uma fita simples de um cromossomo homólogo invade o DNA danificado e atua como um molde para o reparo (**Figura 2A**). Este tipo de reparo normalmente ocorre após a replicação do DNA, mas antes da divisão celular, envolvendo a cromátide-irmã. A maquinaria eucariótica para o reparo por recombinação não é tão bem definida quanto os sistemas de reparo por excisão. Genes humanos envolvidos nesta via incluem *NBS* (*Nijmegen breakage syndrome*; OMIM 602667), *BLM* (*Bloom syndrome*; OMIM 604610) e os genes de suscetibilidade ao câncer de mama *BRCA2* (OMIM 600185) e *BRCA1* (OMIM 113705).
- Na **reunião de extremidades não homólogas**, complexos multiproteicos são montados nas extremidades rompidas da molécula de DNA, e DNAs-ligasas reúnem as extremidades independentemente de sua sequência (**Figura 2B**). Existe sempre alguma perda da sequência

de DNA nas extremidades reunidas. Esta é uma medida desesperada que provavelmente causará mutações ou rearranjos cromossômicos, mas é melhor do que deixar as extremidades sem reparo. Uma das razões pelas quais os telômeros precisam de uma estrutura especial é para proteger as extremidades normais dos cromossomos da ação deste tipo de resposta a quebras duplas (envolvendo ambas as fitas).

Um último mecanismo de reparo é relacionado à correção de incompatibilidades causadas por erros da replicação (*mismatches*). Células deficientes no reparo de *mismatches* apresentam taxas de mutação de 100 a 1.000 vezes superior ao normal, com uma tendência particular ao deslizamento da replicação em sequências homopoliméricas (ver Figura 13.4). Em humanos, o mecanismo envolve pelo menos cinco proteínas, e defeitos causam a síndrome de Lynch (OMIM 120435 e OMIM 609310). Detalhes do mecanismo serão fornecidos no Capítulo 17.

celular é adiada até que o dano tenha sido reparado, e danos irreparáveis desencadeiam a **apoptose** (morte celular programada). Disfunções do sistema para a detecção do DNA danificado e coordenação da resposta celular ao dano possuem um papel central no desenvolvimento do câncer, conforme será descrito no Capítulo 17.

Mesmo que uma célula seja capaz de sobreviver com DNA danificado, o DNA modificado e não reparado será provavelmente mutagênico. Caso mutações ocorram nas células germinativas, estas podem originar novas variantes na população, fornecendo a matéria-prima para a evolução, bem como as variantes intraespecíficas consideradas neste capítulo. Uma importante fonte de mutação é a síntese de DNA translesão suscetível a erro. As polimerases suscetíveis a erro ζ (zeta) e ι (iota) são capazes de replicar DNA danificado, ultrapassando forquilhas de replicação que pararam em função de lesões – mas à custa de uma elevada taxa de erro. Adicionalmente, bases modificadas podem não parear durante a replicação do DNA, introduzindo mudanças permanentes na sequência. Por exemplo, a uracila produzida pela deaminação da citosina pareia com a adenina, assim como a 8-hidroxiguanina, que é um produto do ataque oxidativo ao DNA.

Um caso especial é a 5-metilcitosina. Quando a citosina é deaminada, o produto é a uracila. Esta é uma base não natural no DNA, a qual é, de maneira eficiente, reconhecida e corrigida. No entanto, 5-metilcitosina é deaminada em timina, uma base natural no DNA (**Figura 13.9C**). Se o não pareamento G-T resultante não for corrigido antes da próxima etapa de replicação, uma célula filha terá uma mutação C→T permanente. Evidências oriundas tanto de estudos evolutivos como de doenças humanas demonstram a importância da deaminação da 5-metilcitosina como uma fonte de mudanças de sequência. Citosinas metiladas são quase sempre encontradas em sequências CpG (i.e., o nucleotídeo 3' da citosina é uma guanina), as quais são regiões com elevada taxa de mutação (*mutational hotspots*) (ver também Capítulos 8 e 11).

Em resposta a essas várias formas de dano, as células utilizam uma ampla gama de mecanismos de reparo. Diferentes mecanismos corrigem diferentes tipos de lesão (**Quadro 13.3**). A importância de mecanismos eficientes de reparo ao DNA é ressaltada pelos aproximadamente 130 genes humanos envolvidos em reparo de DNA, bem como pelas doenças graves que afetam pessoas com sistemas de reparo deficientes.

Replicação do DNA, transcrição, recombinação e reparo utilizam complexos multiproteicos que compartilham componentes

Os diferentes sistemas de reparo, com exceção do reparo direto, requerem exonucleases, endonucleases, helicases, polimerases e ligases (ver Quadro 13.3). Muitas das mesmas funções também são requeridas para a replicação do DNA, para a transcrição e para a recombinação. Em cada caso, grandes complexos multiproteicos são montados no sítio de ação, e a cromatina é remodelada. Embora alguns componentes sejam especializados para funções distintas (como as várias polimerases de DNA), muitos atuam em vários complexos diferentes. Estes incluem TFIIA (fator de transcrição geral IIA), PCNA (do inglês *proliferating cell nuclear antigen*) e RPA (do inglês *replication protein A*). O fator de transcrição TFIIH é um complexo multiproteico que existe em duas formas, uma voltada à transcrição geral e outra voltada ao reparo, provavelmente de forma específica para o reparo de DNA transcricionalmente ativo.

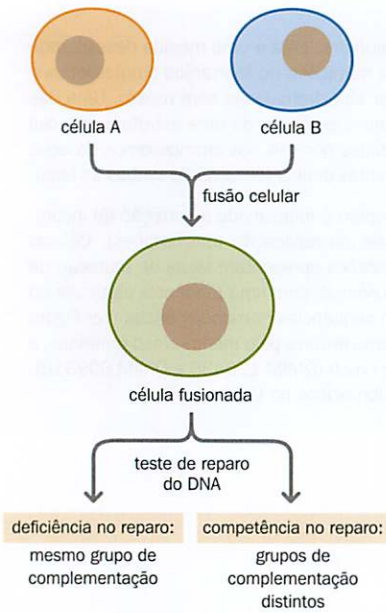


Figura 13.10 Designando genes defeituosos de reparo de DNA para grupos de complementação. Células com deficiência nos mecanismos de reparo e pertencentes a dois pacientes distintos (A e B) são fundidas em cultivo celular, de modo que a célula fusionada possa ser posteriormente testada quanto à sua habilidade de reparar o DNA. Caso as células fusionadas sejam competentes no reparo de DNA, então as células A e B pertencem a grupos de complementação distintos e provavelmente contêm defeitos em genes distintos envolvidos com o reparo de DNA. Caso a célula fusionada ainda não seja capaz de reparar o DNA, então A e B pertencem ao mesmo grupo de complementação e devem ter defeitos no mesmo gene de reparo do DNA.

Defeitos no reparo de DNA são a causa de diversas doenças humanas

Triagens de mutações em *E. coli* e leveduras identificaram componentes de vias de reparo por meio da busca por mutantes hipersensíveis aos efeitos danosos da radiação ou a agentes químicos. Em humanos, investigações similares utilizaram linhagens celulares derivadas de pacientes. Uma confusa variedade de doenças genéticas humanas apresenta fenótipos parcialmente sobrepostos, os quais sugerem defeitos em alguns aspectos do reparo de DNA. Essas incluem o *xeroderma pigmentosum* (OMIM 278700), síndrome da cocaína (OMIM 216400), tricotodistrofia (OMIM 601675), anemia de Fanconi (OMIM 227650), ataxia-telangiectasia (OMIM 208900), síndrome de quebra de Nijmegen (OMIM 251260) e a síndrome de Bloom (OMIM 210900). Clinicamente, pacientes podem apresentar sintomas que incluem hipersensibilidade a luz solar, problemas neurológicos e/ou esqueléticos, anemia e uma alta incidência de vários tipos de câncer. No laboratório, células derivadas de pacientes com essas desordens apresentam hipersensibilidade a vários agentes que causam dano ao DNA.

Grupos de complementação

Muitas dessas condições podem ser divididas em vários grupos de complementação por meio de testes de fusão celular (Figura 13.10). Se duas células, A e B, apresentam perda da função de genes de reparo distintos, então a fusão dessas células produzirá um híbrido contendo cópias funcionais de ambos os genes. A célula A, com defeito no gene A, fornecerá uma cópia funcional do gene B, enquanto a célula B, na qual o gene B é defeituoso, fornecerá uma cópia funcional do gene A. Assim sendo, o híbrido deve recuperar o fenótipo selvagem de resistência ao dano ao DNA. Utilizando esta abordagem, células de pacientes com *xeroderma pigmentosum*, causado por defeitos no reparo por excisão de nucleotídeo (ver Quadro 13.3), foram divididas em sete grupos diferentes. Células de um grupo irão complementar o defeito de células de qualquer outro grupo. A anemia de Fanconi, causada por uma resposta celular defectiva ao dano ao DNA, foi dividida em pelo menos 12 grupos. Em geral, um gene diferente está mutado em cada grupo de complementação, embora os fenótipos clínicos se sobreponham. Detalhes dessas doenças genéticas e seus grupos de complementação podem ser encontrados no banco de dados OMIM (<http://www.ncbi.nlm.nih.gov/omim>).

Estudos moleculares desses vários grupos definiram um grande número de genes envolvidos no reparo de DNA em humanos. A identificação das vias individuais foi amplamente auxiliada pela grande conservação dos mecanismos de reparo por meio de todo o espectro da vida. Não apenas os mecanismos de reação, mas também a estrutura das proteínas e a sequência dos genes são frequentemente conservadas desde *E. coli* até humanos.

Geralmente, eucariotos apresentam múltiplos sistemas que correspondem a cada sistema independente em *E. coli*. Por exemplo, o mecanismo de reparo por excisão de nucleotídeo requer seis proteínas em *E. coli* mas pelo menos 30 em mamíferos. Um contraponto à conservação é a confusa nomenclatura gênica, referindo-se algumas vezes a doenças humanas (p. ex., *xeroderma pigmentosum* tipo D, ou XPD), às vezes a mutantes de leveduras (genes *RAD*) e por vezes a sistemas de complementação de células de mamíferos (ERCC, do inglês *excision repair cross-complementing*). Então, por exemplo, XPD, ERCC2 e RAD3 são os mesmos genes em humanos, camundongos e leveduras.

Nem todas as doenças que envolvem hipersensibilidade a agentes que causam dano ao DNA são provocadas por defeitos nos sistemas de reparo ao DNA. Em alguns casos é a resposta celular ao dano, de modo mais abrangente, que está defectiva. Células normais reagem ao dano ao DNA atrasando o progresso por meio do ciclo celular – permanecendo em um determinado ponto de verificação (*checkpoint*) até que o dano tenha sido reparado – ou disparando a apoptose, caso o dano seja irreversível. Pacientes com ataxia-telangiectasia e anemia de Fanconi possuem mecanismos de reparo intactos, mas são deficientes na percepção do dano ou nos mecanismos de resposta. Defeitos no controle do ciclo celular e na reposta apoptótica são centrais no desenvolvimento do câncer, sendo discutidos posteriormente no Capítulo 17.

13.3 VARIANTES PATOGENICAS DE DNA

Conforme mencionado, os genomas de indivíduos saudáveis apresentam um grande número de variações na sequência. Em sua grande maioria, estas variantes são completamente inofensivas e não apresentam efeito conhecido sobre o fenótipo. Mesmo entre aquelas que afetam o fenótipo, a maioria faz parte da variação normal que nos torna únicos. Uma atenção especial, no entanto, é direcionada àquelas variantes que são patogênicas – ou seja, causam ou nos tornam suscetíveis a uma doença.

Decidir se uma mudança na sequência de DNA é patogênica pode ser difícil

Nem todas as variantes de sequência vistas em uma pessoa afetada serão patogênicas. Do mesmo modo que pessoas perfeitamente saudáveis carregam inúmeras variações na sequência, o mesmo será verdade para uma pessoa com uma doença genética. Como se pode decidir se uma mudança de sequência identificada em uma pessoa é a causa de sua doença ou apenas uma variante inofensiva? Apenas um teste funcional pode fornecer uma resposta definitiva – mas testes funcionais são muitas vezes difíceis de integrar o trabalho de um laboratório de diagnóstico. De qualquer modo, para muitos produtos gênicos não existe um teste laboratorial disponível, capaz de testar todos os aspectos da função gênica *in vivo*. Algumas variantes podem ser patogênicas apenas em períodos de estresse ambiental e outras podem apresentar efeitos sutis que se manifestam suscetivelmente a uma doença, mas apenas quando combinadas com algumas outras variantes genéticas.

Na ausência de um teste funcional definitivo, a natureza da mudança na sequência com frequência fornece uma pista. Primeiro pode-se questionar se a variante afeta uma sequência que é sabidamente funcional. Tais sequências incluiriam as sequências codificantes de genes, as sequências franqueadoras de junções éxon-íntron (sítios de *splicing*), a sequência promotora imediatamente a montante de um gene, assim como qualquer outra sequência regulatória conhecida. A grande maioria das variantes patogênicas afeta sequências que são sabidamente funcionais, as quais representam uma parcela pequena de DNA total. No entanto, existe sempre a possibilidade de que uma variante localizada fora de qualquer sequência funcional conhecida possa estar presente em um elemento funcional ainda não identificado. Conforme visto no Capítulo 11, o projeto ENCODE está revelando muitos elementos funcionais previamente ignorados no genoma humano. Suspeita-se que tais elementos sejam sítios de variantes que alteram meramente a suscetibilidade a uma determinada doença, em vez de a causar diretamente.

Se uma variante afeta uma sequência conhecida, deve-se tentar prever o seu efeito. Uma tabela do código genético (ver Figura 1.25) pode ser utilizada para identificar os efeitos de uma variação na sequência codificadora sobre o produto gênico. Conforme descrito a seguir, mutações sem sentido (*nonsense*), trocas de fase de leitura, assim como muitas deleções, podem garantidamente prever a destruição da proteína. De modo similar, alterações nas sequências invariantes GT...AG nos sítios de *splicing* apresentam elevada probabilidade de ser patogênicas. Mudanças que simplesmente substituem um aminoácido por outro (**mutação não sinônima ou missense**) são mais difíceis de ser interpretadas.

Outra abordagem envolve a busca por precedentes. Talvez uma variante já esteja documentada no dbSNP, o banco de dados de mutações pontuais (ver anteriormente). Alternativamente, ela pode estar documentada em um dos bancos de dados de mutações patogênicas listadas nas Leituras adicionais. Um tipo diferente de precedente pode ser buscado pela verificação de sequências normais de genes relacionados, os quais podem estar em humanos (parálogos) ou em outras espécies (ortólogos). Caso as variantes estejam presentes em sequências normais, do tipo selvagem, é improvável que sejam patogênicas.

Mutações pontuais e outras mudanças de escala menor são tipos comuns de mudança patogênica

Mudanças patogênicas são causadas por pequenas alterações de sequência na região codificante ou na região regulatória de um dado gene.

Mutações não sinônimas (*missense*)

Substituições de um único nucleotídeo na sequência codificadora de um gene podem ou não alterar a sequência da proteína codificada. O código genético é degenerado: os 64 códons codificam apenas 20 aminoácidos (mais três códons de parada). Portanto, algumas mudanças de códon não alteram o aminoácido – elas são silenciosas ou sinônimas. Quando a mudança no códon resulta em troca de aminoácido (uma alteração não sinônima), o efeito depende parcialmente das diferenças químicas entre os aminoácidos antigos e os novos. Conforme explicado no Capítulo 1, os 20 aminoácidos podem ser classificados em ácidos, básicos, polares não carregados e apolares não carregados. A substituição de um aminoácido por outro da mesma classe – uma substituição conservada possui menos efeito sobre a estrutura da proteína do que uma substituição não conservada. A adição ou a remoção de uma cisteína altera o potencial para a formação de pontes dissulfeto, podendo causar grandes alterações estruturais. Matrizes de similaridade foram construídas para fornecer um escore quantitativo para o provável efeito disruptivo de cada substituição (ver Leituras adicionais).

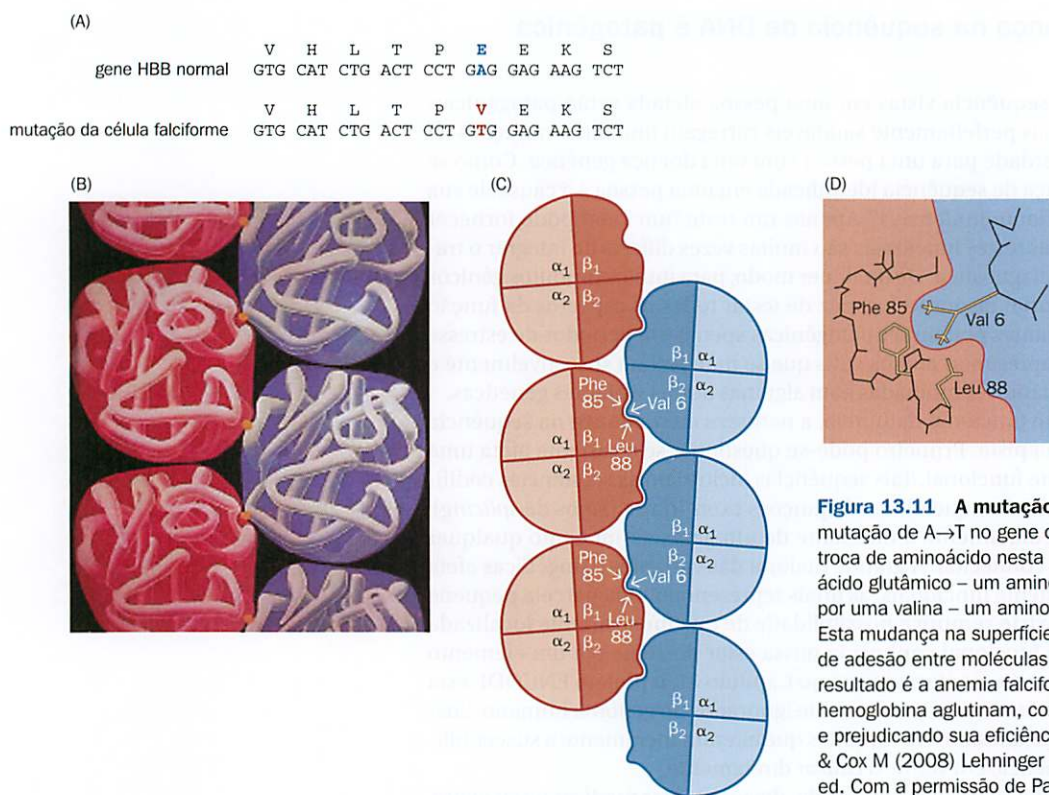


Figura 13.11 A mutação da célula falciforme. (A) Uma mutação de A→T no gene da β-globina (*HBB*) causa uma troca de aminoácido nesta proteína. A mutação substitui um ácido glutâmico – um aminoácido hidrofílico e carregado, por uma valina – um aminoácido hidrofóbico e não polar. Esta mudança na superfície da proteína favorece interações de adesão entre moléculas de hemoglobina. (B, C e D) O resultado é a anemia falciforme, na qual as moléculas de hemoglobina aglutinam, contorcendo os glóbulos vermelhos e prejudicando sua eficiência. [Partes B, C e D de Nelson DL & Cox M (2008) *Lehninger Principles of Biochemistry*, 4th ed. Com a permissão de Palgrave Macmillan.]

Alguns aminoácidos são cruciais para o funcionamento de uma proteína específica – por exemplo, aquelas no sítio ativo de uma enzima. Outras podem ser importantes para a manutenção da estrutura da proteína. Proteínas globulares tendem a apresentar aminoácidos não carregados no interior e aminoácidos carregados no exterior; qualquer substituição que altere essa distribuição pode comprometer a estrutura tridimensional da proteína. A mutação da célula falciforme é patogênica porque substitui um aminoácido polar por um aminoácido não polar na parte externa da molécula de globina (Figura 13.11). Isso resulta na tendência de as moléculas permanecerem unidas. A agregação proteica, que é o resultado da existência de proteínas anormais que apresentam áreas externas unidas, demonstrou-se um mecanismo patogênico comum a uma variedade de doenças, com destaque para as condições neurodegenerativas progressivas, sendo posteriormente discutidas na p. 425.

Raramente é possível prever esses efeitos com elevada confiança. Ajuda se a estrutura tridimensional da proteína tiver sido resolvida, de modo que se torne possível modelar o provável efeito estrutural de uma substituição. Caso sequências de aminoácidos de proteínas relacionadas (de humanos e outros organismos) sejam conhecidas, pode-se identificar quais aminoácidos são invariáveis e quais parecem livres para variar amplamente entre as espécies. A maioria das substituições de aminoácidos provavelmente não apresenta efeito no funcionamento da proteína.

Mutações sem sentido (*nonsense*)

Três dos 64 códons do código genético são códons de parada (*stop codons*), de modo que é bem comum que uma substituição nucleotídica converta o códon correspondente a um aminoácido no interior de uma proteína em um códon de parada. Quando os ribossomos encontram um códon de parada eles dissociam-se do mRNA, liberando a cadeia polipeptídica nascente (Figura 13.12A). No entanto, genes que contêm códons de terminação prematuros raramente produzem a proteína truncada que poderia ser predita. As células possuem um mecanismo, *nonsense-mediated decay* (NMD), que detecta e degrada mRNAs contendo códons de terminação prematuros. Assim sendo, mutações sem sentido normalmente acabam prevenindo qualquer expressão do gene.

O NMD funciona porque o mRNA que sofreu *splicing* conserva uma memória da posição dos íntrons. O mecanismo de *splicing* deixa proteínas do complexo de junção de éxons (EJC, do inglês *exon junction complex*) ligadas ao sítio de *splicing*. Durante a primeira rodada de tradução, à medida que o ribossomo ultrapassa cada sítio de *splicing*, ele libera as proteínas

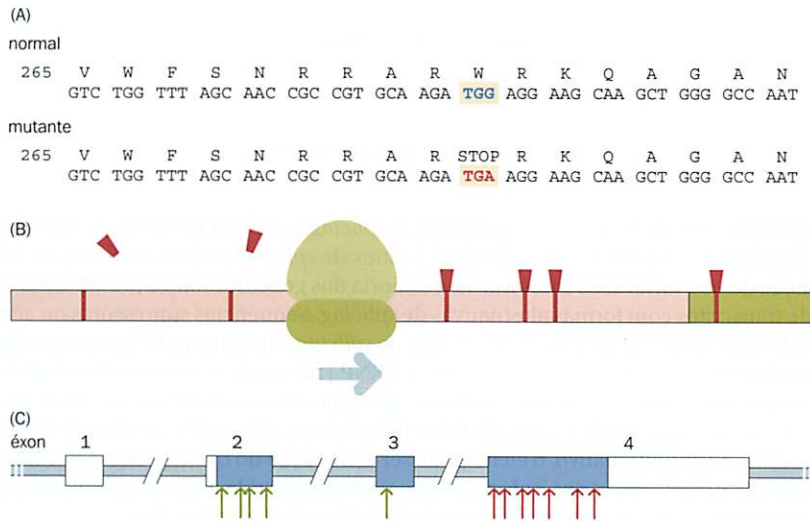


Figura 13.12 Mutações sem sentido e o nonsense-mediated decay. (A) Uma troca G→A no éxon 6 do gene *PAX3* substitui o códon TGG do triptofano 274 por um códon de parada. (B) Nonsense-mediated decay (NMD). Este mRNA maduro foi transcrito de um gene que possui sete éxons. Junções de *splicing* (barras vermelhas) mantêm proteínas do complexo de junção de éxons (EJC, triângulos vermelhos). À medida que o primeiro ribossomo se desloca ao longo do mRNA, ele desloca as proteínas do EJC. Caso ele encontre um códon de parada prematuro e se desprenda antes de remover todos os EJCs, o mRNA é marcado para degradação. Códon de parada no último éxon ou a menos de 50 nucleotídeos a montante da última junção de *splicing* (zona verde) não desencadeiam a NMD. (C) Dependendo se um códon de parada desencadeia ou não a NMD, as consequências de uma mutação sem sentido (*nonsense*) podem variar bastante. Mutações no gene *SOX10* que desencadeiam o NMD (setas verdes), resultam na Síndrome de Waardenburg tipo 4 (perda auditiva, anormalidades pigmentárias; doença de Hirschsprung; OMIM 277580). Mutações sem sentido na região 3' do mRNA que escapam da NMD (setas vermelhas) causam um fenótipo neurológico muito mais grave. Áreas sombreadas indicam sequências codificadoras, áreas claras indicam as regiões não traduzidas 5' e 3' do gene.

EJC que estavam ligadas ao sítio. Se existe um códon de terminação prematuro, o ribossomo não terá atravessado todos os sítios de *splicing* antes de se desprender. Algumas proteínas EJC irão permanecer ligadas ao mRNA, e isto marca o mRNA para a degradação (Figura 13.12B).

Nem sempre o *nonsense-mediated decay* é completamente eficiente. Ele não se aplica a códon de parada prematuros que se encontram no último éxon do gene ou a menos de 50 nucleotídeos a montante da última junção de *splicing*. Em alguns casos, certa quantidade de proteína truncada é produzida mesmo quando o códon de parada não se encontra nessa região protegida. Proteínas truncadas são potencialmente mais patogênicas do que a simples ausência de proteína (Figura 13.12C), uma vez que elas podem interferir na função do produto normal. Tal efeito dominante negativo será discutido posteriormente neste capítulo (ver p. 431). Acredita-se que a NMD tenha se desenvolvido como uma proteção contra esse problema.

Mudanças que afetam o *splicing* do transcrito primário

A posição dos sítios de *splicing* é marcada pela sequência (praticamente) invariável GT...AG, imersa em uma sequência de reconhecimento de sítio de *splicing* que apresenta um consenso menos estrito (ver Capítulo 1). Mutações que alteram os nucleotídeos canônicos GT ou AG irão sempre impedir o reconhecimento desse sítio pelo spliceossomo e desse modo desfazer o *splicing* nesse sítio (Figura 13.13A), mas uma variedade de outras alterações de

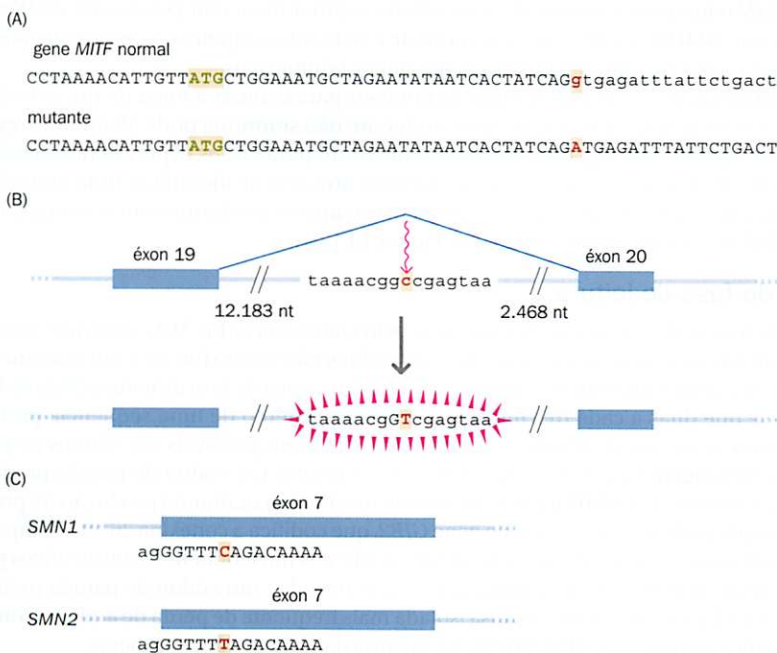


Figura 13.13 Mutações que afetam o *splicing*. (A) No gene *MITF*, uma troca G→A na sequência canônica GT, que marca a posição do primeiro íntron, irá sempre impedir o *splicing*. A sequência do éxon é representada em letras maiúsculas e a do íntron, em minúsculas. O códon de início da tradução está indicado em verde. (B) Uma mudança única de nucleotídeo no íntron 19 do regulador da condutância transmembrana da fibrose cística, codificado pelo gene *CFTR* (referida como 3849+10 kb C→T, embora o nucleotídeo modificado não esteja na verdade a 10 kb, mas a 12.191 nucleotídeos da extremidade 3' do éxon 19), ativa um sítio oculto de *splicing*. (C) Uma mudança aparentemente silenciosa que afeta o *splicing*. O gene *SMN1* é altamente expresso, enquanto o *SMN2* quase não produz proteína. A diferença é dada por uma mudança TTT→TTC no éxon 7. Embora ambos os códon codifiquem uma fenilalanina, a mudança inativa um acentuador de *splicing* e evita o *splicing* correto na junção entre o íntron 6 e o éxon 7 do *SMN2*.

Figura 13.14 A fase de leitura. Esta sequência contínua de letras (A) utiliza o sinal de início da tradução AUG (verde) para estabelecer a fase de leitura correta (B). A inserção (ou deleção) de uma letra (vermelho) destrói o significado (C).

(A)
 ACAUUGUUAUGNOWYOUCANSEEHOWTHERNACANGETHIT
 (B)
 ACAUUGUUAUG NOW YOU CAN SEE HOW THE RNA CAN GET HIT
 (C)
 ACAUUGUUAUG NOW YOU CAN TSE EHO WTH ERN ACA NGE THI T

sequência também pode afetar esse processo. O *splicing* não é um processo do tipo tudo ou nada. Conforme mencionado no Capítulo 11, sítios de *splicing* podem ser fortes ou fracos. Em função do uso variável de sítios fracos, a maioria dos genes humanos produz uma variedade de transcritos com formas alternativas de *splicing*. Sequências supressoras ou acentuadoras de *splicing* modulam a força de um sítio de *splicing* adjacente pela ligação a proteínas das famílias SR (ricas em serina e arginina) e hRNP (ribonucleoproteínas heterogêneas), o que possivelmente ocorre de um modo estágio-específico ou tecido-específico. Se uma dessas sequências moduladoras é mutada em um gene que naturalmente produz várias isoformas com *splicing* alternativo, o efeito pode ser a alteração do balanço das isoformas. Dependendo da função das várias isoformas, isso pode abolir toda a função gênica ou causar alterações mais sutis, como afetar o padrão de isoformas tecido-específicas.

A inativação de sítio de *splicing* irá normalmente abolir a função do gene, ou pelo menos de todas as isoformas que utilizam aquele sítio, mas é difícil prever com precisão as consequências moleculares. Por vezes um éxon é pulado; outras vezes uma sequência intrônica é mantida no mRNA maduro; com frequência um sítio de *splicing* oculto adjacente é utilizado. **Sítios ocultos de *splicing*** são sequências em um transcrito primário (em éxons ou íntrons) que lembram sítios verdadeiros de *splicing*, mas não suficientemente semelhantes a ponto de serem assim reconhecidos pela célula. (Figura 13.13B). Uma substituição de nucleotídeo em um sítio críptico pode aumentar essa semelhança o suficiente para convertê-lo em um sítio funcional, o que irá afetar o correto processamento do transcrito. Alternativamente, uma mudança de sequência pode reduzir a força de um sítio de *splicing* verdadeiro, de modo que um sítio oculto adjacente passe a ser utilizado preferencialmente.

É difícil prever a partir da sequência de DNA se uma mudança irá ou não afetar o *splicing*. Mutações aparentemente não sinônimas ou silenciosas podem, na verdade, ser patogênicas por afetarem o *splicing*. A diferença entre os dois genes *SMN* humanos ilustra esse efeito (Figura 13.13C). Pelo menos duas cópias duplicadas mas sutilmente divergentes do gene *SMN* são encontradas no braço longo do cromossomo 5 (5q13); a cópia mais próxima do centrômero é altamente expressa, enquanto a cópia (ou cópias) próxima ao telômero praticamente não produz proteína. A diferença é dada por uma troca TTTTTC no éxon 7. Embora essa troca seja aparentemente silenciosa (ambos os códons codificam uma fenilalanina), a mudança inativa um acentuador de *splicing* e evita o *splicing* correto na junção entre o íntron 6 e o éxon 7. Indivíduos homocigotos para a perda da função do gene *SMN* teloméricos sofrem de uma atrofia espinal muscular (síndrome de Werdnig-Hoffmann; OMIM 253300), mas a severidade é reduzida naqueles pacientes que possuem múltiplas cópias fracamente expressas dos genes teloméricos.

Programas de computador estão disponíveis para estimar a força de um sítio de *splicing*, ou verificar se uma mutação aparentemente não sinônima pode afetar um supressor ou acentuador de *splicing*, mas não existe substituto para dados experimentais reais obtidos por RT-PCR. Esse também é o único método provável de identificar uma mutação que ative um sítio oculto de *splicing* escondido em um íntron, conforme ocorre com a mutação 3849+10 kb C→T da fibrose cística (ver Figura 13.13B).

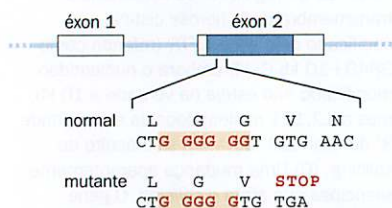


Figura 13.15 Uma simples troca de fase. Uma deleção com troca de fase de um nucleotídeo no gene *GJB2* surge pelo deslizamento da replicação em uma sequência com seis nucleotídeos G. A mutação cria um códon de parada prematuro no éxon 2.

Troca de fase de leitura

A fase de leitura da tradução é determinada pelo códon iniciador AUG; qualquer troca a jstante que adicione ou remova um número de códons não inteiro (ou seja, um número de nucleotídeos que não seja múltiplo de três) causará uma troca de fase de leitura (Figura 13.14). Espera-se que duas a cada três mudanças no comprimento de uma sequência produzam uma alteração na fase de leitura. Como três dos 64 códons possíveis são códons de parada, ler uma mensagem fora de fase irá normalmente levar a um códon de parada prematuro. Consequentemente, o NMD irá provavelmente resultar em nenhuma produção de proteína. Um exemplo pode ser encontrado no gene *GJB2*, que codifica a conexina 26, um componente das junções *gap* entre as células. Uma sequência de seis nucleotídeos G consecutivos predispõe ao deslizamento durante a replicação, o que introduz um códon de parada prematuro (Figura 13.15). Esta mutação é a causa isolada mais frequente de perda de audição congênita autossômica recessiva (OMIM 220290) na maioria das populações europeias.

íntron/éxon	tamanho (pb)	mudança na fase de leitura	deleções que causam DMD ou BMD
íntron 41	31.823		
éxon 42	195	0	
íntron 42	22.380		
éxon 43	173	-1	
íntron 43	70.465		
éxon 44	148	+1	
íntron 44	248.401		
éxon 45	176	-1	
íntron 45	36.111		
éxon 46	148	+1	
íntron 46	2.334		
éxon 47	150	0	
íntron 47	54.222		
éxon 48	186	0	
íntron 48	38.368		
éxon 49	102	0	
íntron 49	16.634		

Figura 13.16 Efeitos das deleções no gene da distrofina. O gene consiste em 79 éxons pequenos rodeados por íntrons maiores. Cerca de 65% de todas as mutações neste gene são deleções intragênicas. Como os íntrons são muito maiores que os éxons, os pontos de quebra quase sempre se localizam nos íntrons. O efeito é a deleção de um ou mais éxons inteiros. Quando isso resulta em uma troca de fase (barras vermelhas), ocasiona a distrofia muscular de Duchenne grave (DMD; OMIM 310200). Se os éxons deletados não alteram a fase de leitura (barras verdes) o resultado é a distrofia muscular de Becker (mais branda) (BMD; OMIM 300376), mesmo que a deleção seja maior do que aquela causadora da DMD.

Muitos tipos de eventos distintos podem produzir uma mudança na fase de leitura. Assim como pequenas inserções e deleções em um único éxon, *splicing* incorreto, deleções de éxons inteiros e duplicações também causam mudanças na fase de leitura. Muitos éxons não são de fase neutra (ou seja, o número de nucleotídeos no éxon não é um múltiplo de três), portanto a exclusão de um ou mais éxons em função de uma deleção ou de um erro no *splicing*, na maioria das vezes, causará um mudança de fase. A maioria dos íntrons é muito maior que os éxons adjacentes, e, assim sendo, os pontos de quebra de deleções intragênicas ou duplicações irão normalmente se localizar nos íntrons. O resultado será a deleção ou a duplicação de um ou mais éxons inteiros (Figura 13.16).

Mudanças que afetam o nível da expressão gênica

Uma variante de uma sequência controladora poderá afetar o nível de transcrição de um gene, de modo que, apesar de o produto gênico ser completamente normal, muito pouco (ou uma quantidade muito maior) desse produto é gerado. Isso ocorrerá de maneira mais óbvia se a variante alterar a sequência promotora. Na verdade, poucas dessas variantes foram descritas. Isso ocorre parcialmente porque o sequenciamento de promotores não faz parte da rotina dos laboratórios de diagnóstico. Mesmo que eles encontrassem uma alteração, raramente saberiam como interpretar o impacto de tal alteração. Conforme descrito no Capítulo 1, promotores podem incluir sítios consenso de ligação a uma variedade de fatores de transcrição, e o efeito de uma mudança de sequência na maioria dos casos só poderá ser determinado experimentalmente.

Pacientes com α - ou β -talassemia têm sido minuciosamente investigados na busca por mutações que afetam a transcrição. Suas doenças são formas de anemia causadas por uma deficiência significativa de α - ou β -globina, respectivamente, tendo sido, portanto, os primeiros candidatos para mutações desse tipo. No entanto, α -talassemia é mais comumente causada por números reduzidos de genes ativos de α -globina (ver a seguir). Algumas mutações no promotor do gene da β -globina foram identificadas (ver OMIM 141900, mutações 370 a 381 e também a Figura 13.17A), mas a grande maioria das mutações na β -talassemia atua na produção de um mRNA instável, ou uma proteína globina instável em vez de afetar diretamente a transcrição. Três tipos comuns de eventos foram observados:

- Muitas das mutações causadoras de talassemia são erros de *splicing* ou mutações sem sentido que produzem um códon de parada prematura, resultando na NMD do mRNA.

Figura 13.17 Duas mutações no gene da globina que causam talassemia. (A)

Substituições de nucleotídeos localizados entre as posições -31 a -28 (vermelho) são associadas com β-talasseмии. Estes nucleotídeos constituem o chamado TATA box, o qual é uma parte essencial do promotor. A sequência dos éxons está representada em letras maiúsculas e o códon de início da tradução está em verde. (B) A variante *Hb Constant Spring* do gene *HBA2*, principal gene de α-globina, possui uma substituição TAA→CAA no códon de parada normal. A proteína resultante possui 31 aminoácidos extras (apresentados em vermelho), o que a torna instável e causa a α-talassemia. A sequência do éxon 3 do gene também é indicada.



- Mudanças na região 3' não traduzida podem levar à instabilidade do mRNA. Uma mudança nessa região pode criar ou eliminar um sítio crítico para a ligação de um microRNA ou de uma proteína que regule a tradução (ver Capítulo 11).
- As células são muito eficientes na detecção e na degradação de proteínas anormais, com dobramento incorreto. Por exemplo, uma forma de α-talassemia é causada pela modificação do códon de parada normal TAA, do gene da globina, para CAA. Esse códon codifica uma glutamina e a tradução do mRNA continua por mais 31 aminoácidos até que outro códon de parada seja encontrado (Figura 13.17B). Esta hemoglobina modificada (*Hb Constant Spring*) é instável e o efeito clínico é uma α-talassemia resultante da deficiência quantitativa de cadeias de α-globina.

Mutações sinônimas (silenciosas) patogênicas

Conforme mencionado anteriormente, mudanças sinônimas são substituições de nucleotídeos que convertem um códon de um dado aminoácido em outro códon que codifica o mesmo aminoácido. Não é esperado que tais mudanças tenham qualquer efeito fenotípico. Algumas vezes, no entanto, o nucleotídeo alterado é parte de um acentuador ou supressor de *splicing*, de modo que sua mudança altera o *splicing*. Conforme mencionado, na atrofia espinal muscular uma troca silenciosa TTC→TTT marca a diferença entre um gene ativo, com *splicing* correto, e um gene inativo, com *splicing* incorreto (ver Figura 13.13C). Mesmo mutações silenciosas que não afetam o *splicing* podem não ser completamente neutras. O gene *MDR1* codifica a glicoproteína P, a qual é importante no transporte de muitos fármacos; sua superexpressão é um fator determinante da resistência a múltiplos fármacos na quimioterapia. Uma mudança silenciosa, c.3435C→T, no éxon 26 do gene, resulta na produção de uma proteína que possui a mesma sequência de aminoácidos (ocorre a troca entre dois códons de isoleucina 1145). No entanto, ela apresenta uma sutil diferença no dobramento e possui um perfil de atividade distinto. Acredita-se que isso ocorra em função de o códon alternativo requisitar um tRNA alternativo que seja mais escasso, o que consequentemente altera a cinética da tradução. Todavia, a frequência de efeitos como esses não é conhecida.

Variações em curtas repetições em tandem são ocasionalmente patogênicas

Ao passo que mutações pontuais podem afetar qualquer sequência codificadora, a maioria das repetições curtas em *tandem* é localizada em DNA não codificante. Contudo, variantes de repetições em *tandem* localizadas próximas a promotores ou sítios de *splicing* de genes podem algumas vezes afetar a expressão de genes. Por exemplo, alelos distintos de um minissatélite com 14 pb localizado na proximidade do promotor do gene da insulina (11p15) estão associados com o risco diferencial ao diabetes tipo 2 (ver OMIM 176730). Outro exemplo ocorre entre o gene regulador da condutância transmembrana da fibrose cística (*CFTR*), em que uma sequência de nucleotídeos T próximos a extremidade 3' do íntron 8 afeta a eficiência do sítio de *splicing* adjacente. Alelos com cinco, sete ou nove nucleotídeos são comuns. Enquanto o *splicing* de alelos 7-T ou 9-T é normal, alelos 5-T geralmente sofrem *splicing* incorreto, de modo que o éxon 9 é perdido. Alelos 5-T não reduzem por conta própria a produção de mRNAs com *splicing* correto com uma intensidade que os leve a ser



Figura 13.18 Variações em um íntron do gene *CFTR* afetam a eficiência do *splicing*. Próximo à extremidade 3' do íntron 8 (letras minúsculas) existem sequências $(TG)_n$ (verde) e T_n (vermelho) com comprimento variável. Alelos com as variantes 5T e $(TG)_{13}$ apresentam *splicing* incorreto de uma grande quantidade de transcritos, levando a uma produção muito reduzida da proteína *CFTR* funcional. Estas mudanças não são, por si só, patogênicas. No entanto, se esta cópia gênica também possui uma variante codificante, como a p.R117H, que abaixa a atividade da proteína intacta, o efeito combinado de baixos níveis de proteína e a função reduzida da proteína o tornam um alelo CF patogênico. Note que na literatura sobre fibrose cística, por razões históricas, este par íntron/éxon é normalmente referido como íntron 8/éxon 9, mas nos bancos de dados genômicos eles estão anotados como íntron 9/éxon 10.

patogênicos. No entanto, em conjunto com outras variantes de baixa funcionalidade, eles podem estar entre as causas da fibrose cística. Seu efeito é acentuado se uma repetição $(TG)_n$ próxima ao íntron apresenta mais do que 11 repetições (Figura 13.18).

Repetições em *tandem* localizadas em sequências codificantes normalmente não são polimórficas, mas podem predispor a mutações patogênicas em função de deslizamentos da polimerase. Mutações somáticas desse tipo são a maior causa de doença em pessoas com defeitos no sistema de reparo por mau pareamento pós-replicativo (ver Capítulo 17). Sequências longas de polialaninas em certas proteínas são responsáveis por várias doenças hereditárias. Os exemplos incluem a proteína *PHOX2B* nas pessoas com a síndrome de hipoventilação central congênita (OMIM 209880) e a proteína *HOXD13* nos pacientes com simpolidactilia 1 (OMIM 186000). Essas variantes foram provavelmente originadas pelo deslizamento da polimerase, mas dentro de cada família elas são estávelmente transmitidas, do mesmo modo que qualquer outro alelo de STRP. Em pelo menos alguns casos, a sequência longa de alaninas interfere na localização correta da proteína na célula.

Mutações dinâmicas: uma classe especial de variantes patogênicas de microssatélites

As **mutações dinâmicas** são STRPs que, acima de determinado tamanho, tornam-se intensamente instáveis. As causas moleculares não são bem compreendidas, mas elas podem ser uma consequência do modo como uma das fitas de DNA (fita retardada) é sintetizada por meio de uma série de fragmentos descontínuos – os fragmentos de Okazaki (ver Capítulo 1). Uma endonuclease especial, *FEN1*, corta as extremidades sobressalentes dos fragmentos de Okazaki. Um mecanismo proposto para a expansão de repetições se baseia na falha da *FEN1* em realizar esses cortes, de modo que fragmentos sobrepostos acabam sendo reunidos por suas extremidades. Repetições até determinado comprimento são estáveis, e pode ser significativo que, na maioria dos casos, o ponto de corte da instabilidade ocorra quando a sequência repetida alcançar o tamanho típico de um fragmento de Okazaki. Nem todas as mutações dinâmicas são patogênicas, mas muitas são (Tabela 13.1). Outras são responsáveis pelos sítios frágeis não patogênicos vistos por citogeneticistas quando células de algumas pessoas são sujeitas a estresse replicativo (ver Figura 13.5). Por exemplo, o sítio frágil *FRA16A* no cromossomo 16 é devido a uma repetição $(CCG)_n$, enquanto o sítio *FRA16B*, no mesmo cromossomo, é causado pela expansão de um minissatélite com 33 pb.

As doenças na Tabela 13.1 são heterogêneas em vários aspectos. Existem diferentes tamanhos da unidade de repetição, diferentes graus de expansão, diferentes localizações com relação ao gene afetado, além de diferentes mecanismos. Entre essas, as doenças de poliglutamina formam um grupo bem definido, no qual a doença de Huntington (HD; OMIM 143100) é o fenótipo. Nessas condições, expansões modestas de repetições $(CAG)_n$ na sequência codificadora do gene levam a uma sequência expandida de poliglutamina na proteína codificada (Figura 13.19C). Isto, por sua vez, predispõe a proteína a formar agregados intracelulares que são tóxicos à célula, especialmente neurônios (Quadro 13.4). O resultado é uma doença neurodegenerativa progressiva de início tardio.

Outras mutações dinâmicas afetam a sequência do gene fora das regiões codificantes, podendo estar envolvidas em expansões muito maiores. As expansões de repetições em *tandem* podem estar no promotor, na região 5' ou 3' não traduzida ou em um íntron. Normalmente o efeito é a não expressão do gene (Figura 13.19A) e a patologia se desenvolve pela ausência da função do gene. Quando isso ocorre, a mesma doença pode algumas vezes ser causada por outras mutações de perda de função no mesmo gene. Na distrofia miotônica tipo 1 (DM1; OMIM 160900) o mecanismo é bastante diferente. Existe um ganho de função envolvendo um mRNA tóxico. Uma sequência $(CTG)_n$ expandida massivamente na região 3' não traduzida do gene *DMPK* produz um mRNA que sequestra proteínas de ligação a CUG, no núcleo. Essas proteínas são necessárias para o *splicing* correto do transcrito primário de vários genes não relacionados, os quais, conseqüentemente, deixam de funcionar corretamente. O resultado é uma doença multissistêmica cujas características não têm qualquer relação com o produto do gene *DMPK*, uma proteína quinase, a qual ainda é produzida em quantidades normais. Nenhuma outra mutação no gene *DMPK* produz distrofia miotônica,

TABELA 13.1 Doenças causadas por expansões instáveis de repetições nucleotídicas

Doença	Nº OMIM	Modo de Herança ^a	Nome e localização do gene	Localização das repetições no gene	Sequência repetida	Nº estável de repetições	Nº instável de repetições
1. GRANDES EXPANSÕES DE REPETIÇÕES FORA DA REGIÃO CODIFICANTE							
Sítio A do X frágil (FRAXA)	309550	X	<i>FMR1</i> Xq27.3	região 5' não traduzida	(CGG) _n	6-54	200-1.000+
Sítio E do X frágil (FRAXE)	309548	X	<i>FMR2</i> Xq28	região 5' não traduzida	(CCG) _n	4-39	200-900
Ataxia de Friedreich (FRDA)	229300	AR	<i>FXN</i> 9q13	íntron 1	(GAA) _n	6-32	200-1.700
Distrofia miotônica tipo 1 (DM1)	160900	AD	<i>DMPK</i> 19q13	região 3' não traduzida	(CTG) _n	5-37	50-10.000
Distrofia miotônica tipo 2 (DM2)	602668	AD	<i>ZNF9</i> 3q21.3	íntron 1	(CCTG) _n	10-26	75-11.000
Ataxia espinocerebelar tipo 10 (SCA10)	603516	AD	<i>ATXN10</i> 22q13.31	íntron 9	(ATTCT) _n	10-20	500-4.500
Epilepsia miotônica de Unverricht e Lundborg	254800	AR	<i>CSTB</i> 21q22.3	promotor	(CCCCGCCCGCG) _n	2-3	40-80
2. EXPANSÕES INTERMEDIÁRIAS DE REPETIÇÕES CAG EM SEQUÊNCIAS CODIFICANTES							
Doença de Huntington (HD)	143100	AD	<i>HD</i> 4p16.3	região codificante	(CAG) _n	6-34	36-100+
Doença de Kennedy (SBMA)	313200	X	<i>AR</i> Xq12	região codificante	(CAG) _n	9-35	38-62
Ataxia espinocerebelar tipo 1 (SCA1)	164400	AD	<i>ATXN1</i> 6p23	região codificante	(CAG) _n	6-38	39-82
Ataxia espinocerebelar tipo 2 (SCA2)	183090	AD	<i>ATXN2</i> 12q24	região codificante	(CAG) _n	15-24	32-200
Doença de Machado-Joseph (SCA3)	109150	AD	<i>ATXN3</i> 14q32.1	região codificante	(CAG) _n	13-36	61-84
Ataxia espinocerebelar tipo 6 (SCA6)	183086	AD	<i>CACNA1A</i> 19p13	região codificante	(CAG) _n	4-17	21-33
Ataxia espinocerebelar tipo 7 (SCA7)	164500	AD	<i>ATXN7</i> 3p14.1	região codificante	(CAG) _n	4-35	37-306
Ataxia espinocerebelar tipo 17 (SCA17)	607136	AD	<i>TBP</i> 6q27	região codificante	(CAG) _n	25-42	47-63
Atrofia dentatorrubro-palidolusiana (DRPLA)	125370	AD	<i>DRPLA</i> 12p13.31	região codificante	(CAG) _n	7-34	49-88
3. OUTRAS EXPANSÕES DINÂMICAS							
Ataxia espinocerebelar tipo 8 (SCA8)	608768	AD	? 13q21	RNA não traduzido	(CTG) _n	16-34	74+
Ataxia espinocerebelar tipo 12 (SCA12)	604326	AD	<i>PPP2R2B</i> 5q32	promotor	(CAG) _n	7-45	55-78
Doença Huntington-like tipo 2 (HDL2)	606438	AD	<i>JPH3</i> 16q24.3	éxon com variação de <i>splicing</i>	(CTG) _n	7-28	66-78

^a Modos de herança: X, ligada ao X; AD, autossômica dominante; AR, autossômica recessiva.

mas uma doença muito semelhante (DM2; OMIM 116955) pode ser causada pela expansão massiva de uma sequência (CCTG)_n em um gene completamente não relacionado (*ZNF9*).

Uma característica das doenças causadas por repetições dinâmicas é a *antecipação* – em gerações sucessivas a idade de início é mais baixa e/ou a severidade é maior, em função de expansões sucessivas das repetições. Veja no Capítulo 3 uma nota de advertência sobre o viés no qual o modo como os membros da família são verificados pode produzir uma falsa aparência de antecipação. Em alguns casos, o ponto de corte para a instabilidade é menor do que o ponto de corte para os efeitos patogênicos. Nesses casos, é visto que **alelos de pré-mutação** de tamanho intermediário e não patogênicos são prontamente expandidos em alelos de mutação completa quando transmitidos a prole (p. ex., repetições *FRAXA* com 50-200 unidades na síndrome do X frágil; OMIM 300624). Em outros ca-

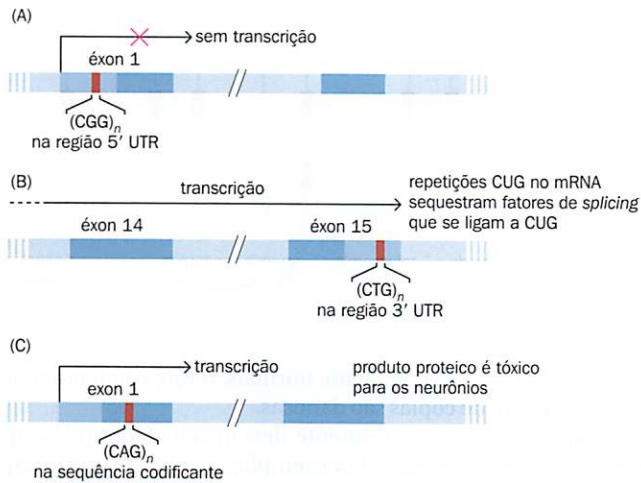


Figura 13.19 Três mecanismos pelos quais mutações dinâmicas podem ser patogênicas. (A) Na síndrome do X frágil, a repetição expandida na região 5' não traduzida (UTR) do gene desencadeia a metilação do promotor e previne a transcrição. (B) Na distrofia miotônica, a repetição expandida na região 3' não traduzida provoca o mRNA transcrito a sequestrar fatores de *splicing* no núcleo da célula, impedindo o *splicing* correto de diversos genes não relacionados. (C) Na doença de Huntington, o gene contendo a repetição expandida é transcrito e traduzido normalmente, mas o produto proteico possui um trecho expandido de poliglutamina que o torna tóxico. Repetições expandidas são apresentadas em vermelho e regiões codificantes em azul-escuro.

so, alelos abaixo do ponto de corte patogênico expandem apenas ocasionalmente (p. ex., alelos HD com 29 a 35 repetições). Em alguns casos existe um efeito sexual, de modo que grandes expansões são vistas principalmente em alelos herdados de parentes de um determinado sexo (o pai na HD, a mãe na distrofia miotônica). Isso reflete uma sobrevivência diferencial de gametas que carregam grandes expansões e não uma tendência hereditária para expandir preferencialmente em um dos sexos.

Comparações entre pais e filhos quanto aos comprimentos das repetições são difíceis de ser interpretadas mecanicamente, pois podem refletir a instabilidade mitótica ou meiótica da repetição na linhagem germinativa parental ou, alternativamente, a habilidade dos gametas em transmitir repetições extensas. Além disso, comprimentos de sequência são muitas vezes estudados em DNA extraído de linfócitos de sangue periférico e, se existe uma instabilidade mitótica, os linfócitos podem ter um comprimento de repetições muito diferente daquele presente nos espermatozoides, nos óvulos ou nos tecidos envolvidos com a patologia da doença. Algumas repetições são somaticamente instáveis, produzindo uma banda borrada quando o DNA sanguíneo é analisado por eletroforese. Outras são instáveis entre pais e filhos, mas estáveis na meiose. Nesse caso o DNA sanguíneo apresenta bandas bem definidas, mas com tamanhos diferentes entre pais e filhos.

Variantes que afetam a dose de um ou mais genes podem ser patogênicas

O potencial patogênico de doses anormais de um gene é conhecido há muito tempo em função dos fenótipos severos produzidos por trissomias cromossômicas. No entanto, apenas uma minoria dos genes é sensível a efeitos de dose. Se uma condição é recessiva, significa que os heterozigotos são fenotipicamente normais apesar de terem apenas uma cópia funcional do gene em questão. Para tais genes, a dosagem evidentemente não interessa. As descobertas recentes de variações abundantes e de larga escala, quanto ao número de

especialmente para os neurônios. Algumas vezes parece que a mudança conformacional pode propagar-se por meio da população de proteínas, convertendo as de sua conformação nativa (estável) para uma conformação nova, com propriedades distintas, em um processo que talvez seja análogo à cristalização.

O comportamento das proteínas priônicas, em doenças como a Creutzfeldt-Jakob, é o exemplo mais evidente. O *misfolding* pode começar por um mau dobramento ao acaso, de uma molécula recém-sintetizada e estruturalmente normal (casos esporádicos), por uma sequência mutante com uma maior propensão ao mau dobramento (uma doença genética) ou ainda por uma molécula maldobrada que seja de algum modo adquirida a partir do meio ambiente (doença infecciosa). Assim sendo, esta última rota comum reúne um conjunto de doenças com origens muito diversas. Esse tópico é revisado por Hardy & Orr (ver Leituras adicionais).

QUADRO 13.4 Doenças causadas pela agregação de proteínas anormais

A formação de agregados proteicos é atualmente reconhecida como uma característica comum a várias doenças neurológicas com início na fase adulta. Estas incluem as doenças de poliglutamina, das quais a doença de Huntington é o fenótipo, doença de Alzheimer, doença de Parkinson, doença de Creutzfeldt-Jakob e o grupo heterogêneo das amiloidoses.

Moléculas de proteínas globulares assemelham-se a gotas de óleo, com resíduos hidrofóbicos no interior e grupos polares no exterior. O dobramento correto é um processo crítico e muito específico, de modo que as proteínas normais são selecionadas dentre todas as possíveis sequências de polipeptídeos, parcialmente em função de sua habilidade de sofrer o dobramento correto. Proteínas mutadas podem ser mais suscetíveis ao mau dobramento (*misfolding*). Moléculas maldobradas com grupos hidrofóbicos expostos podem se agregar umas às outras, ou com outras proteínas, o que é de algum modo tóxico para as células,

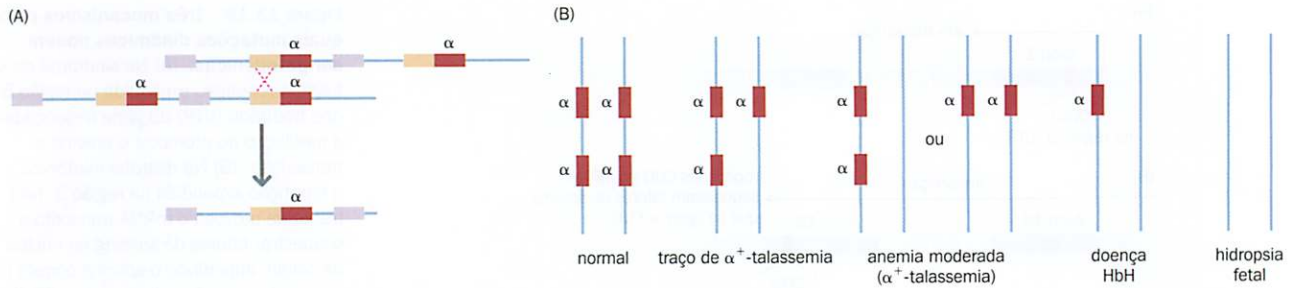


Figura 13.20 Deleções de genes de α -globina na α -talassemia. (A) Cópias normais do cromossomo 16, cada uma carregando dois genes de α -globina (vermelho) arranjados em *tandem*. Blocos repetidos flanqueando os genes (apresentados em cinza e amarelo) podem desalinhar, permitindo *crossover* desigual. A figura apresenta uma dentre várias destas recombinações desalinhadas que podem produzir um cromossomo que carrega apenas um gene α ativo. *Crossovers* desiguais entre outras repetições (não apresentado) podem produzir cromossomos que não carregam um gene α funcional. Portanto, o número de genes de α -globina pode variar entre indivíduos, desde nenhum, até quatro ou mais. (B) As consequências tornam-se mais graves à medida que o número de genes α diminui. Ver Weatherall et al., (nas Leituras adicionais) para mais detalhes.

cópias de genes entre pessoas aparentemente normais, reforça a mensagem de que nem todas as variantes de número de cópias são danosas.

Trissomias cromossômicas provavelmente devem seus fenótipos a apenas alguns poucos genes sensíveis a efeitos de dose. Por exemplo, acredita-se que os aspectos característicos da síndrome de Down sejam em grande parte ocasionados por efeitos de dose de apenas dois genes, *DSCR1* e *DYRK1A*. É esperado que mais genes produzam efeitos fenotípicos quando expressos em metade da dose do que quando expressos 1,5 vezes acima da dose normal. Assim sendo, grandes deleções ou mesmo a monossomia de um cromossomo inteiro são menos toleradas durante o desenvolvimento humano do que a duplicação gênica ou a trissomia cromossômica.

Um mecanismo comum capaz de gerar mudanças na dose gênica é a **recombinação homóloga não alélica** (NAHR, do inglês *non-allelic homologous recombination*). Duplicações de segmentos (frequentemente definidas como sequências com 1 kb ou mais e com identidade acima de 95%) podem desalinhar quando cromossomos homólogos formam pares na meiose. A NAHR produz então deleções ou duplicações. As repetições mal-alinhadas possuem a mesma sequência, mas não possuem a mesma localização cromossômica, de modo que a recombinação é homóloga, mas as sequências não são alélicas. Muitas (mas não todas) das variantes comuns de número de cópias não patogênicas encontradas em pessoas normais são geradas por esse mecanismo. A α -talassemia fornece um bom exemplo de NAHR produzindo uma variação patogênica na dose gênica. A maioria das pessoas possui quatro cópias do gene da α -globina ($\alpha\alpha/\alpha\alpha$) como resultado de uma duplicação ancestral em *tandem*. Conforme apresentado na **Figura 13.20**, a NAHR entre sequências repetidas com baixo número de cópias flanqueando os genes de α -globina podem produzir cromossomos que carregam mais ou menos genes de α -globina. O número de cópias reduzido de genes da α -globina produz efeitos sucessivamente mais graves. Pessoas com três cópias ($\alpha\alpha/\alpha-$) são saudáveis; aquelas com duas (seja na fase $\alpha-/alpha-$ ou $\alpha\alpha/--$) sofrem de uma α -talassemia moderada; aquelas com apenas um gene ($\alpha/--$) apresentam doença grave; enquanto a ausência de todos os genes α ($--/--$) causa a hidropsia fetal (acumulação fatal de fluídos no feto).

A monossomia e a trissomia do cromossomo X são particularmente interessantes porque a inativação do X (inativação de todos os - exceto um - cromossomos X na célula; ver Capítulo 3) deve torná-las assintomática nos tecidos somáticos. Todavia, conforme observado no Capítulo 11, um número surpreendentemente grande de genes do cromossomo X escapa da inativação. Alguns destes possuem um homólogo no cromossomo Y, mas a maioria não possui. Para aqueles genes que escapam da inativação do X mas não possuem um homólogo no Y, fêmeas normais teriam duas cópias funcionais e homens apenas uma. Fêmeas Turner (45,X) teriam a mesma cópia simples ativa, como nos machos - mas talvez no contexto de desenvolvimento da fêmea uma cópia única não seja suficiente. As anormalidades esqueléticas da síndrome de Turner são causadas por uma haploinsuficiência de *SHOX* (50% do produto gênico normal não são suficientes para produzir um fenótipo normal). Este é um gene *homeobox* localizado na região pseudoautosômica Xp/Yp, estando, portanto, presente em duas cópias tanto em homens como em mulheres normais. Abaixo do nível convencional de resolução citogenética mas acima do nível de genes únicos, variações patogênicas no número de cópias são classificadas como microdeleções ou microduplicações (**Tabela 13.2**). Entre estas, três patologias moleculares distintas podem ser distinguidas:

- **Síndromes de gene único**, nas quais todos os efeitos fenotípicos são causados pela deleção (ou às vezes duplicação) de um gene único. Por exemplo, a síndrome de Alagille (OMIM 118450) é vista em pacientes com microdeleção em 20p11. No entanto, 93% dos pacientes com essa síndrome não apresentam deleção, sendo na verdade

TABELA 13.2 Exemplos de síndromes associadas com microdeleções cromossômicas

Síndrome	Nº OMIM	Localização	Tipo de anomalia genética
Wolf-Hirschhorn	194190	4pter	aneuploidia segmentar
<i>Cri-du-chat</i>	123450	5pter	aneuploidia segmentar
Williams-Beuren	194050	7q11.23	aneuploidia segmentar
Langer-Giedon	150230	8q24	genes contíguos (<i>TRPS1</i> , <i>EXT1</i>)
WAGR	194072	11p13	genes contíguos (<i>PAX6</i> , <i>WT1</i>)
Prader-Willi	176270	15q11-q13	aneuploidia segmentar (ausência da cópia paterna)
Angelman	105830	15q11-q13	gene único (ausência do <i>UBE3A</i> materno)
Rubinstein-Taybi	180849	16p13.3	gene único (dosagem do <i>CBP</i>)
Miller-Dieker (Lisencefalia)	247200	17pter	genes contíguos (<i>LIS1</i> , etc.)
Smith-Magenis	182290	17p11.2	aneuploidia segmentar
Alagille do tipo 1	118450	20p12	gene único (dosagem do <i>JAG1</i>)
DiGeorge/VCFs	188400/192430	22q11.21	aneuploidia segmentar

Aquelas síndromes causadas pela presença de apenas uma cópia funcional de um único gene também são vistas em pacientes que, em vez de uma microdeleção, possuem uma mutação de ponto no gene. WAGR, Tumor de Wilms, anirídia, anormalidades genitais, retardo mental; VCFS, síndrome velocardiofacial.

heterozigotos para mutações de ponto no gene *JAG1*, localizado em 20p12. A causa da síndrome, em todos os casos, é uma dosagem da metade do produto do gene *JAG1*.

- **Síndromes de genes contíguos** são vistas primariamente em machos com deleções no cromossomo X (Figura 13.21A). O caso clássico foi um garoto BB que possuía distrofia muscular de Duchenne (DMD; OMIM 310200), doença granulomatosa crônica (CGD; OMIM 306400) e retinite pigmentosa (OMIM 312600), em conjunto com retardo mental. Ele teve uma deleção cromossômica em Xp21, a qual removeu um conjunto contíguo de genes e, incidentalmente, forneceu aos pesquisadores um meio de clonar os genes cuja ausência causava duas de suas doenças (DMD e CGD). Deleções da extremidade do braço curto do cromossomo X (Xp) são vistas em um outro conjunto de síndromes de genes contíguos. Deleções sucessivamente maiores removem mais genes e adicionam mais doenças à síndrome. Microdeleções são relativamente frequentes em algumas partes do cromossomo X (tais como Xq21 e a porção proximal de Xp) mas são raras ou desconhecidas em outras (como Xp22.1-22.2 e Xq28). Sem dúvida a deleção de alguns genes individuais e as deleções visíveis em regiões ricas

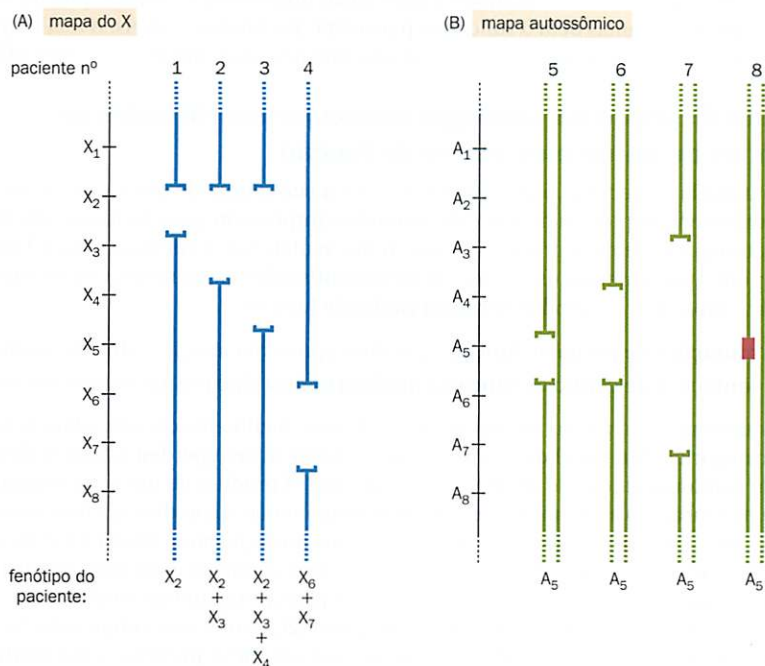


Figura 13.21 Síndromes com microdeleção autossômica e ligada ao X. (A) No cromossomo X, pacientes masculinos 1-3 apresentam uma série combinada de síndromes de genes contíguos, enquanto o paciente 4 apresenta uma síndrome de genes contíguos distinta. Deleção do gene *X₁* ou *X₅* é letal nos machos. (B) Nos autossomos, síndromes de genes contíguos são raras em função do equilíbrio proporcionado pelo segundo cromossomo. Neste exemplo, apenas o gene *A₅* é sensível a efeitos de dose. Os pacientes 5-7 apresentam deleções de diferentes tamanhos, todas incluindo o gene *A₅*. Todos apresentam o mesmo fenótipo do paciente 8, o qual é heterozigoto para uma mutação pontual de perda de função, localizada no gene *A₅*.

em genes seriam letais. Síndromes de genes contíguos, semelhantes às descritas para o X, são muito menos comuns com autossomos em função do equilíbrio fornecido pelo cromossomo normal (**Figura 13.21B**). A síndrome de Langer-Giedon (síndrome tricorriofalangeana, tipo II; OMIM 150230) é um exemplo raro.

- **Síndromes de aneuploidia segmentar** são um tipo especial de síndrome de genes contíguos, a qual regularmente ocorre com um fenótipo bem conhecido. Os exemplos incluem as síndromes Williams-Beuren (OMIM 194050), Prader-Willi (OMIM 176270), Angelman (OMIM 105830), Smith-Magenis (OMIM 182290) e DiGeorge/velocardiofacial (OMIM 188400/192430) (ver Tabela 13.2). Todas essas síndromes possuem deleções produzidas por NAHR entre repetições com baixo número de cópias que flanqueiam a região em questão. A NAHR também irá produzir duplicações dessas regiões, embora essas possam não ser patogênicas. O exemplo das síndromes de Prader-Willi e Angelman (ver Figura 11.20, p. 367) envolve uma região de *imprinting*, o que complica o fenótipo, mas o mesmo mecanismo produz as outras síndromes mencionadas. Assim como em outras síndromes de genes contíguos, o fenótipo depende de efeitos de dose de mais de um gene e não é visto nas pessoas com mutação de ponto em apenas um dos genes. A síndrome de Williams-Beuren é um exemplo típico. Os pacientes são heterozigotos para a deleção de 1,5 Mb no cromossomo 7q11.23 que remove cerca de 20 genes. Foram descritos casos apresentando deleções menores, mas não foi encontrado nenhum caso típico com apenas um gene deletado ou mutado.

Algumas outras síndromes recorrentes são produzidas por deleções terminais aleatórias em cromossomos nos quais um gene sensível à dose se localiza próximo aos telômeros. São exemplos as síndromes de Wolf-Hirschhorn (OMIM 194190) e *Cri-du-chat* (OMIM 123450). Na síndrome de Miller-Dieker (Lisencefalia; OMIM 247200), deleções terminais aleatórias no braço curto do cromossomo 17p podem remover um ou mais genes sensíveis à dose, produzindo uma síndrome de deleção de genes contíguos.

13.4 PATOLOGIA MOLECULAR: ENTENDENDO O EFEITO DAS VARIANTES

Para o aconselhamento genético e para análise de *pedigree*, os alelos são simplesmente designados como *A* e *a*, com a letra maiúscula indicando o alelo cujo efeito é dominante. Este é o nível apropriado de descrição para tais propósitos. No entanto, para patologia molecular, precisa-se olhar mais atentamente. Quando se descreve o genótipo de um portador de fibrose cística como *Aa*, representa-se como *a* qualquer sequência do gene *CFTR* que seja mutada de modo que não produza um canal de cloro funcional. Mais de 1.500 alelos distintos foram descritos com esse perfil. De modo similar, *A* significa qualquer sequência que funcione suficientemente bem de modo a não causar uma doença – ela não precisa funcionar perfeitamente, apenas bem o suficiente para evitar problemas visíveis. A real sequência de DNA dos alelos *A* em pessoas não relacionadas não necessariamente será 100% idêntica.

A grande distinção na patologia molecular ocorre entre as mudanças de perda e de ganho de função

Para a patologia molecular, o que importa não é a sequência de um mutante, mas seu efeito. Conhecer a sequência de um alelo mutante é importante para testes genéticos, mas para a patologia molecular precisa-se saber o que o alelo faz ou deixa de fazer. Um gene mutado pode apresentar todos os tipos de efeitos sutis sobre o organismo, mas uma questão inicial valiosa é se ele produz perda ou ganho de função:

- Nas **mutações de perda de função** o produto apresenta função reduzida ou ausente.
- Nas **mutações de ganho de função** o produto realiza algo positivamente anormal.

Inevitavelmente, algumas mutações não podem ser facilmente classificadas em nenhuma das categorias. Um canal de íon permanentemente aberto perdeu a função de fechamento ou ganhou a função de abertura inapropriada? O produto de um gene mutante que interfere na função do alelo normal em um heterozigoto não só perdeu sua função normal, como também ganhou uma nova função danosa. Uma mutação pode alterar o balanço entre diversas funções ou produtos de um gene. No entanto, a distinção entre perda ou ganho de função é a primeira ferramenta essencial quando se pensa em patologia molecular.

O modo de herança fornece uma pista da patologia molecular subjacente. Se existe um ganho de função, a presença de um gene normal não deve prevenir o comportamen-

to anormal do alelo mutante. A função anormal seria vista em uma pessoa heterozigota e seria esperado, portanto, que o fenótipo fosse dominante. Por exemplo, tanto a doença de Huntington como a distrofia miotônica são dominantes. Para fenótipos de perda de função, o quadro é menos evidente. Para a maioria dos produtos gênicos a quantidade exata não é crucial, de modo que se pode manter a função com apenas metade da quantidade normal.

Portanto, a maioria das mutações de perda de função produz fenótipos recessivos. Por exemplo, a maior parte dos erros inatos do metabolismo é recessiva. No entanto, condições de perda de função podem ser dominantes se houver uma haploinsuficiência ou um efeito dominante negativo, conforme descrito a seguir.

Heterogeneidade alélica é uma característica comum em fenótipos de perda de função

Existem várias maneiras de reduzir ou eliminar a função de um produto gênico. Quando um fenótipo clínico resulta da perda de função de um gene, seria esperada *qualquer* mudança que inativasse o produto gênico, produzindo o mesmo resultado clínico. O ganho de função, por outro lado, é um fenômeno bastante específico. Provavelmente apenas uma mudança muito específica em um gene pode causar o ganho de função. Portanto, a queda da heterogeneidade alélica é outro forte indicativo da patologia molecular subjacente, ainda que esteja longe de ser infalível. Por exemplo, entre as doenças causadas por repetições instáveis de trinucleotídeos (ver Tabela 13.1), a síndrome do X frágil (OMIM 300624) e a ataxia de Friedreich (OMIM 229300) são ocasionalmente causadas por outros tipos de mutações nos seus respectivos genes, apontando para a perda de função, enquanto a doença de Huntington e a distrofia miotônica nunca são vistas com outro tipo de mutação, sugerindo o ganho de função.

Quando um fenótipo clínico resulta da perda de função de um gene, deveria-se ser capazes de encontrar mutações de ponto que possuam o mesmo efeito como deleção ou ruptura do gene. Por exemplo, o gene mutado na ataxia-telangiectasia (*ATM*), no cromossomo 11q23, codifica uma proteína com 3.056 aminoácidos de comprimento, a qual é envolvida na detecção de DNA danificado. Quando o dano é detectado, a *ATM* fosforila proteínas a jusante, o que aciona o processo de reparo. Pacientes com ataxia-telangiectasia são deficientes na reparação de dano ao DNA. Pacientes diferentes apresentam uma ampla variedade de mutações do gene *ATM*. Na maioria dos casos, são pequenas inserções/deleções com troca de fase de leitura, mas também ocorrem mutações sem sentido, mutações de sítio de *splicing*, algumas mutações não sinônimas e ocasionais deleções de grande porte (Figura 13.22). Claramente, a causa da ataxia-telangiectasia é a perda da função de ambas as cópias do gene *ATM*.

Mutações de perda de função produzem fenótipos dominantes quando há haploinsuficiência

Conforme mencionado, para a maioria dos produtos gênicos a quantidade precisa não é crucial. A maioria das mutações de perda de função produzem fenótipos recessivos, o que significa que uma pessoa que possui apenas uma cópia funcional do gene será fenotipicamente normal. Em alguns casos, no entanto, 50% dos níveis normais não são suficientes para a função normal. Pessoas heterozigotas para uma mutação de perda de função apresentam um fenótipo herdado como um traço dominante. Isso é chamado **haploinsuficiência**. A síndrome de Waardenburg do tipo I (OMIM 193500: perda de audição e anomalias de pigmentação) fornece um exemplo. Conforme a Figura 13.23 apresenta, mutações

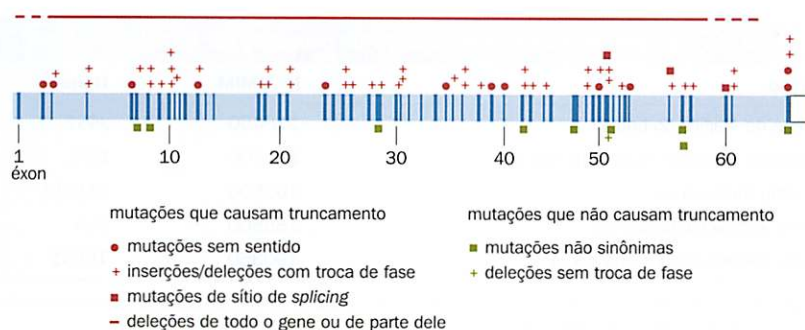
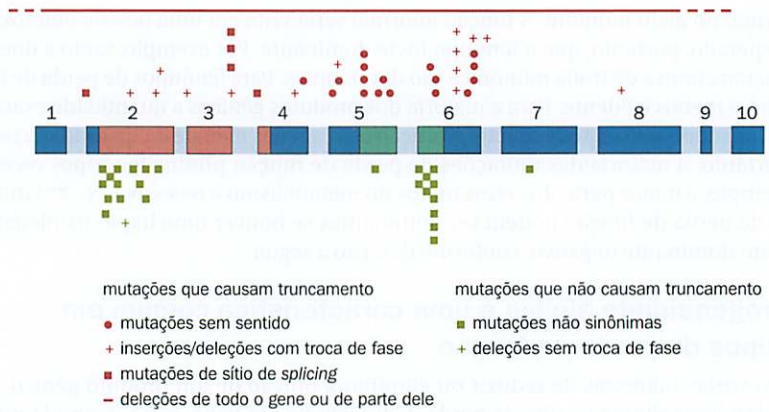


Figura 13.22 O espectro de mutações no gene *ATM* em pacientes com ataxia-telangiectasia. A grande variedade de mutações, cuja maioria produziria um produto gênico truncado, mostra que esta doença recessiva é causada pela perda de função do gene *ATM*. [Dados relatados em Li A & Swift M (2000) *Am. J. Med. Genet.* 92, 170-177. Com a permissão de Wiley-Liss, Inc, uma subsidiária de John Wiley & Sons, Inc.; e de Esembl e OMIM.]

Figura 13.23 Mutações de perda de função no gene *PAX3* em pacientes com a síndrome de Waardenburg do tipo I.

Um amplo espectro de mutações que causam e que não causam truncamento é identificado, de modo similar aos dados apresentados para o gene *ATM* na Figura 13.21. Trocas não sinônimas são concentradas nas duas áreas sombreadas que codificam importantes domínios funcionais de ligação ao DNA (laranja, domínio pareado; verde, homodomínio) da proteína *PAX3*. Embora a doença seja dominante, os dados na figura indicam que a causa deve ser a perda de função do gene *PAX3*. Isto é, portanto, um exemplo de haploinsuficiência.



causais no gene *PAX3* incluem substituições de aminoácidos, trocas de fase, mutações de *splicing* e, em alguns pacientes, deleção completa do gene. Como todos esses eventos produzem o mesmo resultado clínico, sua causa deve ser uma perda de função do gene *PAX3*. No entanto, a síndrome de Waardenburg é dominante, ou seja, é um exemplo de haploinsuficiência. A Tabela 13.3 enumera alguns outros exemplos.

Seria razoável perguntar por que deveria haver haploinsuficiência para *qualquer* produto gênico. Por que a seleção natural não gerenciou isso de uma maneira melhor? Se um gene é expresso de modo que duas cópias fazem apenas a quantidade necessária do produto, a seleção por variantes com níveis mais elevados de expressão deveria levar à evolução de um organismo mais robusto, sem nenhum custo evidente a ser pago por isso. A resposta é que na maioria dos casos isso de fato aconteceu – o que explica por que relativamente poucos genes são sensíveis a efeitos de dose. Existem alguns casos nos quais a célula com apenas uma cópia funcional do gene não pode suprir a demanda de um produto gênico que é necessário em grandes quantidades. Um exemplo pode ser a elastina. Em pessoas heterozigotas para uma deleção ou mutações de perda de função do gene da elastina, tecidos que necessitam apenas moderadas quantidades de elastina (p. ex., pele e pulmões) não são afetados, mas a aorta, onde muito mais elastina é requerida, geralmente apresenta uma anormalidade, a estenose aórtica supravascular (OMIM 185500). Contudo, certas funções gênicas são inerentemente sensíveis a efeitos de dose. Estas incluem:

- produtos gênicos que são parte de um sistema de sinalização quantitativo cuja função depende da ocupação de um receptor ou de um sítio de ligação ao DNA, por exemplo;
- produtos gênicos que competem entre si para determinar uma mudança no desenvolvimento ou no metabolismo;
- produtos gênicos que cooperam em interações com estequiometria fixa (tal como as α - e β -globinas ou muitas proteínas estruturais).

Em cada caso, o produto gênico é titulado contra algo mais na célula. O que importa não é o nível absoluto do produto gênico, mas o correto nível relativo dos produtos que interagem. Os efeitos são sensíveis a mudanças em todos os parceiros envolvidos; portanto, estas condições dominantes frequentemente apresentam expressão altamente variável. Genes cujos produtos atuam essencialmente sozinhos, tal como muitas enzimas solúveis do metabolismo, raramente apresentam efeitos de dose.

TABELA 13.3 Exemplos de fenótipos provavelmente causados por haploinsuficiência

Condição	Nº OMIM	Gene ^a
Síndrome de Alagille do tipo 1	118450	<i>JAG1</i>
Síndrome da exostose múltipla do tipo 1	133700	<i>EXT1</i>
Neuropatia tomaculosa	162500	<i>PMP22</i>
Estenose aórtica supravascular	185500	<i>ELN</i>
Síndrome tricorriñofalangiana tipo 1	190350	<i>TRPS1</i>

^a Indivíduos com um alelo de perda de função são afetados, pois o produto de uma única cópia funcional não é suficiente para o funcionamento ou desenvolvimento normais.

O efeito dominante negativo ocorre quando um produto gênico mutado interfere na função do produto normal

Efeitos dominantes negativos fornecem outra razão pela qual algumas mutações de perda de função produzem fenótipos dominantes. Efeitos dominantes negativos são vistos apenas em heterozigotos, nos quais causam efeitos mais graves que simples alelos nulos do mesmo gene. Uma pessoa heterozigota para todas as mutações nulas de um gene ainda terá 50% do nível normal de função, pois ainda possui um alelo normal. Contudo, se o produto do alelo mutado, além de ser não funcional, também interferir com a função do alelo normal, existirá uma atividade residual inferior a 50%. O mecanismo NMD (*nonsense-mediated decay*; ver p. 418) provavelmente evoluiu como uma proteção contra possíveis efeitos dominantes negativos de proteínas truncadas anormais; pode ser melhor não possuir o produto de um gene mutado do que possuir um produto anormal. Proteínas que constituem estruturas multiméricas são particularmente vulneráveis a efeitos dominantes negativos. O colágeno fornece um exemplo clássico.

Colágenos fibrilares, as principais proteínas estruturais do tecido conectivo, são formados por uma tripla-hélice de cadeias polipeptídicas, podendo ser homotrímeros ou heterotrímeros, as quais são montadas em arranjos cruzados fortemente empacotados, formando fibras rígidas. Nas cadeias polipeptídicas recém-sintetizadas (pré-pró-colágeno), pró-peptídeos N- e C-terminais flanqueiam uma sequência regular repetitiva (Gly-X-Y)_n, em que X e Y são aminoácidos variáveis, sendo que pelo menos um deles frequentemente é uma prolina. Três cadeias de pré-pró-colágeno associam-se para formar uma tripla-hélice sob controle do pró-peptídeo C-terminal. Após a formação da tripla-hélice, os pró-peptídeos N- e C-terminais são removidos. Mutações que substituem a glicina por qualquer outro aminoácido normalmente apresentam forte efeito dominante negativo por desfazerem o empacotamento da tripla-hélice. As mutações não sinônimas (*missense*) no colágeno tipo I são responsáveis pelas formas mais graves de osteogênese imperfeita (*brittle bone disease* ou *osteogenesis imperfecta* tipo IIA; OMIM 166210), em função destes efeitos dominantes negativos. Nos heterozigotos, os polipeptídeos mutados de colágeno associam-se a cadeias normais, mas acabam impedindo a formação da tripla-hélice. Isso pode reduzir para menos de 50% a disponibilidade de colágeno funcional. Seria de se esperar que mutações nulas no mesmo gene causassem efeitos mais graves, mas na verdade causam uma doença mais branda. A simples ausência de um colágeno é menos danosa que a presença de cadeias anormais (Figura 13.24).

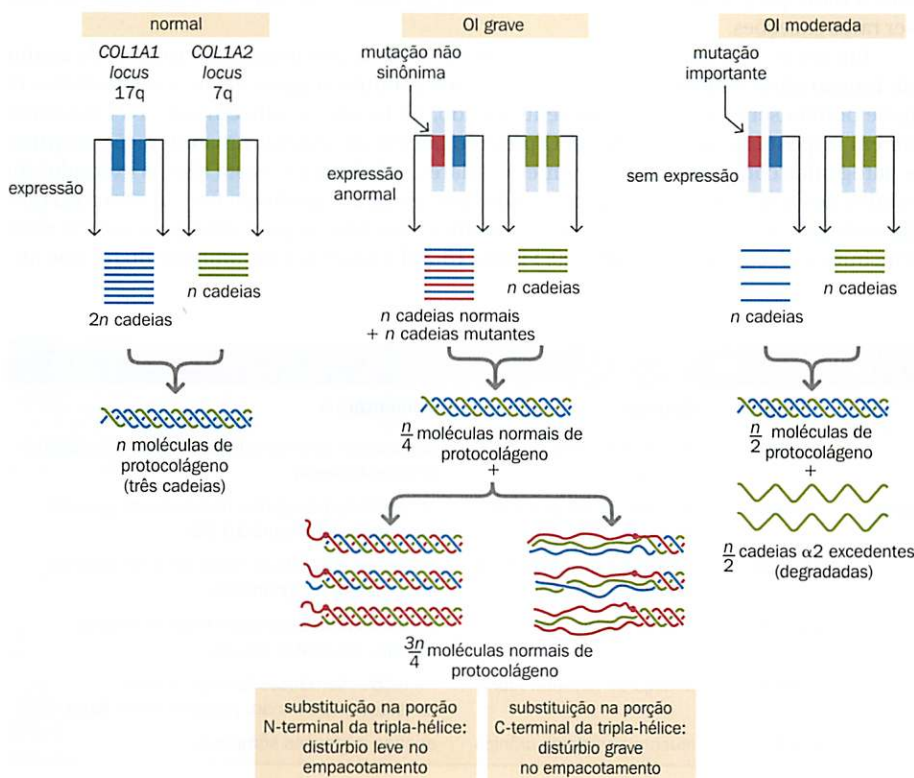


Figura 13.24 Efeito dominante negativo de mutações no gene do colágeno. Fibras de colágeno são construídas por arranjos cruzados de unidades da tripla-hélice de protocólágeno. O protocólágeno do tipo I compreende duas cadeias codificadas pelo gene *COL1A1*, e uma codificada pelo gene *COL1A2*. Na tripla-hélice, cada cadeia polipeptídica consiste em unidades repetidas (Gly-X-Y)_n. Na *osteogenesis imperfecta* do tipo IIA (OI; OMIM 166210), mutações que substituem a glicina por qualquer outro aminoácido normalmente apresentam forte efeito dominante negativo, uma vez que elas atrapalham o empacotamento. A hélice é montada iniciando pela extremidade C-terminal, e substituições de glicinas próximas a esta extremidade apresentam um efeito mais grave do que substituições de glicinas próximas à extremidade N-terminal. Mutações nulas em qualquer dos genes resultam na redução da formação de hélices, embora elas sejam normais (simples haploinsuficiência) e o fenótipo clínico seja menos grave.

Proteínas não estruturais que dimerizam ou oligomerizam também apresentam efeito dominante negativo. Os fatores de transcrição da família bHLH-Zip, por exemplo, ligam-se ao DNA na forma de dímeros. Mutantes que não formam dímeros em geral causam fenótipos recessivos, mas mutantes capazes de converter moléculas funcionais em moléculas que não formam dímeros apresentam fenótipo dominante. Os canais de íon na membrana celular fornecem outro exemplo de estruturas multiméricas que são sensíveis a efeito dominante negativo. A conexina 26 fornece um exemplo. Seis moléculas da proteína conexina 26 associam-se para formar um *connexon*, a metade de uma junção *gap* que permite que pequenos íons se movam entre as células. A **Figura 13.15** apresentou um exemplo de uma mutação nula no gene que codifica a conexina 26. Pessoas homocigotas para essa mutação não podem produzir conexina 26; elas não possuem junções *gap* em sua orelha interna, os íons de potássio não conseguem recircular como deveriam, e os pacientes são surdos. Heterocigotos são fenotipicamente normais (isto é um problema se você deseja identificar casais com risco de terem filhos surdos). No entanto, certas mutações não sinônimas produzem moléculas estruturalmente anormais de conexina 26. Estas interferem na função dos *connexons*, mesmo que algumas das seis moléculas de conexina sejam normais; os heterocigotos têm perda auditiva, e o fenótipo é dominante.

Mutações de ganho de função afetam a forma pela qual um gene ou seu produto reagem a sinais regulatórios

É improvável que uma mutação forneça uma função radicalmente nova para um produto gênico. O único mecanismo que garantidamente gera novas funções é quando um rearranjo cromossômico cria um novo gene quimérico reunindo éxons de dois genes distintos. Embaralhamentos de éxons deste tipo têm sido importantes para a evolução, mas não em doenças hereditárias. As translocações ou inversões cromossômicas que produzem estes genes quiméricos podem propagar-se através de divisões mitóticas sucessivas, quando cada cromossomo se comporta de maneira independente. No entanto, cromossomos homólogos pareiam e recombinam durante a meiose. Conforme visto no Capítulo 2, em portadores de rearranjos cromossômicos, o pareamento envolve conformações cromossômicas complicadas, as quais podem resultar em segregação incorreta; em alguns casos, a replicação irá produzir duplicações e deleções. Como resultado, genes quiméricos são frequentemente vistos em tumores, os quais dependem unicamente da mitose – muitos exemplos são descritos no Capítulo 17 – mas não em condições hereditárias, nas quais eles teriam que passar por meiose. Os exemplos evolutivamente bem-sucedidos devem ser raras exceções.

Em vez de produzir uma função radicalmente nova, a maioria das mutações de ganho de função afeta o modo como um gene ou seu produto reagem a sinais regulatórios. O gene poderá ser expresso no momento errado, no tecido errado, em um nível incorreto ou em resposta ao sinal incorreto. O produto poderá apresentar uma interação anormal e patogênica com outros componentes celulares. A **Tabela 13.4** lista alguns exemplos de modos nos quais um produto gênico pode apresentar um ganho de função. Um caso raro de mutação de ponto herdada conferindo uma nova função para uma proteína é o alelo Pittsburg no locus *PI* (OMIM 107400; **Figura 13.25**). Uma troca de aminoácido no sítio ati-

TABELA 13.4 Mecanismos das mutações de ganho de função

Defeito	Gene	Nº OMIM	Doença	Comentários
Superexpressão	NROB1	300018	reversão de sexo de macho para fêmea	duplicações gênicas causam a superexpressão e a reversão sexual
Adquire um novo substrato	<i>PI</i> (alelo de Pittsburgh)	107400	distúrbio de sangramento letal	um alelo raro de ganho de função do gene da α_1 -antitripsina (Figura 13.25)
Canal de íon aberto de maneira imprópria	SCN4A	168300	paramiotonia congênita de von Eulenburg	mutações específicas neste gene de canal de sódio atrasa o fechamento
Agregação proteica	HD	143100	doença de Huntington	proteínas com expansão de poliglutamina formam agregados tóxicos
Receptor permanentemente ativado	GNAS1	174800	doença de McCune-Albright	mutações somáticas apenas; formas constitucionais seriam provavelmente letais
Gene quimérico	BCR-ABL	151410	leucemia mieloide crônica	apenas mutações somáticas

vo converte um inibidor de elastase em um novo inibidor de trombina constitutivamente ativo, resultando em uma coagulopatia letal.

Desordens causadas pelo ganho de função de receptores de hormônio associados à proteína G

Os receptores de hormônio associados à proteína G fornecem bons exemplos de mutações ativadoras. Muitos hormônios exercem seus efeitos sobre as células-alvo por meio da ligação a domínios extracelulares de receptores transmembrana. A associação de um ligante ao receptor motiva a cauda citoplasmática do receptor a catalisar a conversão de uma proteína G inativa (ligada a GDP) em sua forma ativa (ligada a GTP), o que leva o sinal adiante por meio da estimulação de canais de íons ou enzimas como a adenilil ciclase. Alguns receptores mutados são constitutivamente ativos: eles disparam (sinalizam para efetores subjacentes) mesmo na ausência do ligante:

- A puberdade precoce masculina familiar (OMIM 176410: início da puberdade aos 4 anos em meninos afetados) é encontrada com um receptor de hormônio luteinizante ativo constitutivamente.
- A hiperplasia autossômica dominante da tireoide pode ser causada por uma mutação ativadora no receptor do hormônio estimulador da tireoide (ver OMIM 275200).
- O distúrbio ósseo conhecido como condrodissplasia metafisária tipo Jansen (OMIM 156400) pode ser causado por um receptor hormonal da paratireoide, constitutivamente ativo.

A homogeneidade alélica não é sempre dada pelo ganho de função

Foi apresentado anteriormente que a heterogeneidade alélica é normalmente uma marca característica dos fenótipos de perda de função. No entanto, não é seguro assumir que qualquer condição que apresente homogeneidade alélica seja causada por ganho de função. Existem outras explicações possíveis:

- Em alguns casos, o fenótipo é diretamente associado ao próprio produto gênico, em vez de ser uma consequência mais remota da mudança genética. A doença pode então ser definida em termos de uma variante particular do produto, como na anemia falciforme.
- Alguns mecanismos moleculares específicos podem ocasionar uma certa mudança na sequência de um gene com uma probabilidade maior do que qualquer outra mudança – por exemplo, a expansão (CGG)_n na síndrome do X frágil.
- Pode haver um **efeito fundador**. Por exemplo, algumas mutações patogênicas são comuns entre judeus asquenazi. Os judeus asquenazi da atualidade descendem de um número bastante reduzido de fundadores. Se um dos fundadores carregava um alelo recessivo, é possível que este alelo seja encontrado em alta frequência na atual população asquenazi.
- A seleção em favor do heterozigoto acentua o efeito fundador e geralmente resulta em uma, ou algumas poucas, mutações específicas que são comuns em uma população.

Mutações de ganho e de perda de função em um mesmo gene irão causar fenótipos distintos

Inicia-se esta seção sobre patologia molecular enfatizando a distinção entre mutações de perda e de ganho de função. Contudo, é possível ver ambos os tipos de mutação no mesmo gene. Quando isso acontece, os fenótipos resultantes são frequentemente muito distintos.

Conforme descrito, mutações de perda de função no gene *PAX3* causam o defeito de desenvolvimento conhecido como síndrome de Waardenburg do tipo I (ver Figura 13.23). Um fenótipo totalmente diferente é visto quando uma translocação cromossômica cria um novo gene quimérico, fusionando o gene *PAX3* com outro gene de fator de transcrição, o *FKHR*, em células somáticas. O ganho de função deste fator de transcrição híbrido causa o desenvolvimento de um tumor, o rabiomiossarcoma (OMIM 268220). Outro exemplo marcante diz respeito ao gene *RET*. Este gene codifica um receptor transmembrana do tipo tirosina quinase. A associação do ligante (GDNF, do inglês *glial cell line-derived neurotrophic factor*) ao domínio extracelular induz a dimerização dos receptores. Isso ativa módulos de tirosina quinase no domínio citoplasmático, os quais transmitem o sinal para a célula (ver Figura 4.15, p. 110).



Figura 13.25 Uma mutação que causa ganho de função na molécula de α_1 -antitripsina. A α_1 -antitripsina é um inativador de proteases específicas. A protease cliva a ligação peptídica entre os resíduos 358 e 359 na molécula de α_1 -antitripsina. Como resultado, os dois resíduos (apresentados como esferas em verde) estão afastados cerca de 65 Å, conforme indicado pelas setas. A mudança conformacional inativa a protease. No alelo selvagem, os resíduos 358 e 359 são uma metionina e uma serina, respectivamente. Isto cria um substrato para a elastase; portanto, a α_1 -antitripsina tipo selvagem atua como um agente antielastase. Na variante Pittsburgh, uma mutação não sinônima (p.M358R) substitui a metionina por uma arginina. A estrutura e o efeito de clivagem da ligação peptídica entre os resíduos 358 e 359 não são alterados, mas a especificidade agora é pela trombina e não mais pela elastase. Portanto, a variante Pittsburgh não possui efeito antielastase, tendo se convertido em um poderoso agente antitrombina. O resultado é uma coagulopatia letal. Azul e amarelo indicam as partes a C-terminal e a N-terminal do sítio de clivagem da molécula de α_1 -antitripsina; resíduos de carboidrato são representados em cinza. [Imagem S3D00427 do University of Geneva ExPASy molecular biology Web server, <http://www.expasy.ch/>]

Uma variedade de mutações de perda de função no gene *RET* corresponde a uma das causas da doença de Hirschsprung (OMIM 142623; ausência de gânglios entéricos no intestino). Estas mutações incluem trocas de fase, mutações sem sentido e substituições de aminoácidos que interferem na maturação pós-traducional da proteína RET. Algumas mutações não sinônimas muito específicas no gene *RET* são vistas em um conjunto totalmente diferente de doenças, como no carcinoma medular de tireoide familiar (OMIM 155240) e a patologia relacionada (mais extensa) conhecida como neoplasia endócrina múltipla tipo II (OMIM 162300 e 171400). Essas são mutações de ganho de função, produzindo moléculas de receptores que reagem extensivamente ao ligante ou que são constitutivamente ativas e dimerizam mesmo na ausência do ligante. Curiosamente, algumas pessoas com mutações não sinônimas que afetam as cisteínas 618 ou 620, importantes para a dimerização do receptor, sofrem tanto de câncer da tireoide como da doença de Hirschsprung – ganho e perda de função simultâneos. Isso nos lembra de que perda e ganho de função nem sempre são simples quantidades escalares: as mutações podem ter efeitos distintos nos diferentes tipos celulares nos quais o gene é expresso.

A **Tabela 13.5** lista vários casos nos quais mutações de perda e ganho de função afetando um único gene resultam em doenças distintas. Normalmente, o mutante de ganho de função produz uma proteína qualitativamente anormal, mas simples efeitos de dose podem ocasionalmente ser patogênicos em ambas as direções: doses aumentadas e reduzidas da expressão do gene são patogênicas, mas causam doenças distintas. O gene da proteína da mielina periférica (*PMP22*) é um exemplo. *Crossover* desigual entre sequências repetidas no cromossomo 17p11 criam duplicações ou deleções de uma região de 1,5 Mb que contém o gene *PMP22* (**Figura 13.26**). Portadores heterozigotos da deleção ou da duplicação possuem uma ou três cópias do gene, respectivamente. Pessoas com apenas uma cópia do gene sofrem de neuropatia hereditária com suscetibilidade a paralisia por pressão (HNPP) ou neuropatia tomaculosa (OMIM 162500), enquanto pessoas com três cópias apresentam uma neuropatia clinicamente diferente, a doença de Charcot-Marie-Tooth 1A (CMT1A; OMIM 118220). Essas duas doenças também podem ser causadas por mutações de ponto no gene *PMP22*. Apesar dos eventos moleculares, a causa da patologia é a mudança da atividade desse gene.

TABELA 13.5 Exemplos de genes nos quais mutações de perda e de ganho de função causam doenças distintas

Gene	Localização	Perda (–) ou ganho (+)	Doenças	Símbolo	Nº OMIM
PAX3	2q35	–	Síndrome de Waardenburg do tipo 1	WS1	193500
		+	Rabdomiossarcoma alveolar	RMS2	268220
RET	10q11.2	+	Neoplasia endócrina múltipla tipo IIA	MEN2A	171400
		+	Neoplasia endócrina múltipla tipo IIB	MEN2B	162300
		+	Carcinoma medular de tireoide familiar	FMTc	155240
PMP22	17p11.2	–	Doença de Hirschsprung	HSCR	142623
		–	Neuropatia de Charcot-Marie-Tooth tipo 1A	CMT14	118220
GNAS1	20q13.2	+	Neuropatia tomaculosa	HNPP	162500
		–	Osteodistrofia hereditária de Albright	PHP1A	103580
AR	Xq12	+	Síndrome McCune-Albright	MAS	174800
		–	Síndrome de feminilização testicular	TFM	300068
		+	Atrofia muscular espinobulbar	SBMA	313200

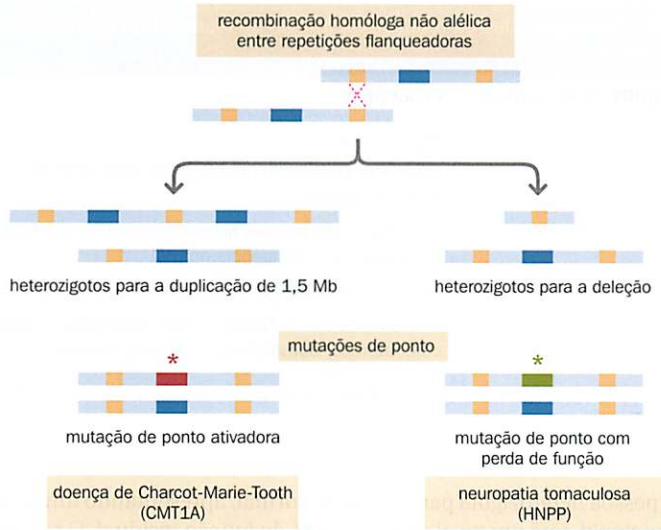


Figura 13.26 Tanto o aumento como a redução de dose do gene *PMP22* são patogênicos, mas causam patologias distintas. O crossover desigual entre repetições flanqueadoras (quadros amarelos) gera duplicações ou deleções de 1,5 Mb da região contendo o gene *PMP22* (quadro azul). A duplicação causa a doença de Charcot-Marie-Tooth; a deleção causa a neuropatia tomaculosa. As mesmas duas doenças podem resultar de mutações de ponto que aumentam ou reduzem a atividade do gene *PMP22* ou que o inativam.

13.5 A BUSCA POR CORRELAÇÕES ENTRE GENÓTIPO E FENÓTIPO

Dada a complexidade das interações genéticas, não é surpreendente que a patologia molecular seja uma ciência bastante imperfeita. O maior sucesso até agora se deu na compreensão do câncer e das hemoglobinopatias: no caso do câncer, o fenótipo para ser explicado – proliferação celular descontrolada – é relativamente simples, enquanto as hemoglobinopatias são um resultado direto de anormalidades em uma proteína abundante e de fácil acesso. Para a maioria das doenças genéticas, os aspectos clínicos são os resultados finais de uma longa cadeia causal, e o santo graal da patologia molecular – a correlação entre genótipo e fenótipo será sempre elusiva. Na realidade, mesmo simples doenças mendelianas não são nem um pouco simples. As revisões neste tópico realizadas por Scriver & Waters e por Weatherall são fortemente recomendadas (ver Leituras adicionais).

O efeito fenotípico das mutações de perda de função depende do nível residual da função gênica

Mudanças na sequência de DNA podem causar graus variáveis de perda de função. Muitas substituições de aminoácidos apresentam efeito reduzido ou mesmo inexistente, enquanto algumas mutações eliminam completamente a função. Uma mutação pode estar presente em uma ou nas duas cópias de um gene. Pessoas com condições autossômicas recessivas geralmente são **heterozigotos compostos** – ou seja, eles têm mutações diferentes em cada alelo. Se ambas as mutações causam uma perda de função, mas com graus distintos, o alelo menos severo irá determinar o nível residual da função. A **Figura 13.27** representa quatro possíveis relações entre o nível residual da função gênica e o fenótipo

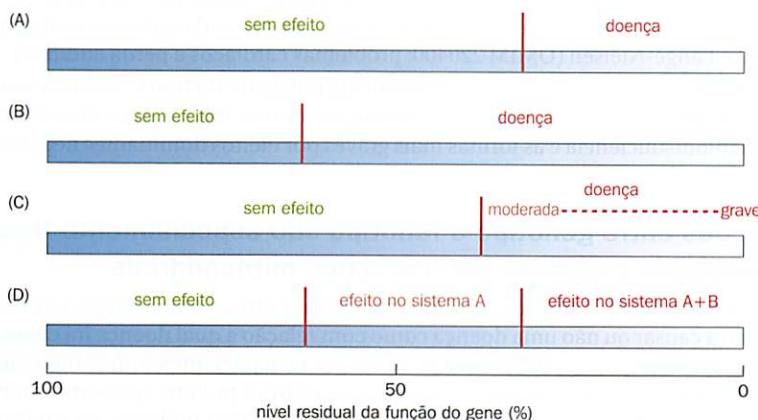


Figura 13.27 Quatro possíveis relações entre a perda da função gênica e o fenótipo clínico. As barras indicam o nível geral de função produzido pelo efeito combinado de dois alelos do gene. (A) Essa condição, com 50% da função residual do gene sem causar efeitos, será um recessivo simples. (B) Essa condição será dominante em função da haploinsuficiência; a perda de 50% da função do gene causa a doença. (C) Essa condição é recessiva, mas a gravidade depende do nível da função residual, de modo que existe uma correlação entre genótipo e fenótipo. (D) Se as consequências clínicas são muito diferentes, dependendo do grau da função residual, os resultados podem ser descritos como síndromes distintas (A e A + B), e elas podem ter modos distintos de herança, como aqui. Exemplos específicos são discutidos no texto.

TABELA 13.6 Consequências da redução na função da hipoxantina guanina fosforribosil transferase (HPRT)

Atividade da HPRT (% do normal)	Fenótipo
> 60	sem doença
8-60	gota; sem problemas neurológicos (síndrome de Kelley-Seegmiller)
1,6-8	gota associada a sinais neurológicos variáveis (falta de jeito, coreoatetose); inteligência normal
1,4-1,6	síndrome de Lesch-Nyhan, mas com inteligência normal
< 1,4	síndrome de Lesch-Nyhan (OMIM 300322): espasticidade, coreoatetose, automutilação, retardo mental

A HPRT é codificada por um gene ligado ao X (OMIM 308000); o fenótipo nos homens reflete diretamente a atividade de seu único gene *HPRT*.

clínico. Uma pessoa heterozigota para um alelo normal, apresentando um alelo completamente não funcional, terá um nível global de 50% de função residual. O resultado poderia ser tanto uma condição recessiva (**Figura 13.27A**) como dominante (**Figura 13.27B**), dependendo se 50% da atividade forem suficientes para a função normal.

Mutações no gene da distrofina (ligado ao X) fornecem um bom exemplo da situação apresentada na **Figura 13.27C**, na qual a severidade da doença se correlaciona bem com a amplitude da atividade residual. Perda completa da função gênica em machos afetados causa a distrofia muscular de Duchenne grave (DMD; OMIM 310200), enquanto um grau menor de perda causa uma condição mais branda, a distrofia muscular de Becker (BMD, OMIM 300376). Apesar de seus nomes distintos, ambas as condições possuem a mesma causa e a mesma patologia; elas diferem apenas quanto à gravidade. A razão para esta diferença foi apresentada na **Figura 13.16** Mutações no gene da hipoxantina guanina fosforribosil transferase (*HPRT*), também ligado ao X, apresenta uma série similar, mas clinicamente mais evidente (**Tabela 13.6**). Portanto, nestes casos, existe uma clara correlação genótipo-fenótipo.

Às vezes, quantidades distintas de função gênica residual podem originar condições clínicas diferentes, as quais podem inclusive apresentar modos diversos de herança (**Figura 13.27D**), por exemplo:

- Mutações no gene do transportador de sulfato (*DTDST*), que provocam diferentes graus de perda de função, causam displasias esqueléticas autossômicas recessivas, as quais receberam nomes distintos de acordo com a severidade: displasia diastrófica (OMIM 226600), displasia epifisária múltipla do tipo 4 (OMIM 226900), atelosteogênese do tipo II (OMIM 256050) e acondrogênese tipo 1B (OMIM 600972). A matriz extracelular é rica em proteoglicanos sulfatados, como o heparan sulfato e o condroitina sulfato. Defeitos no transporte de sulfato interferem no desenvolvimento esquelético.
- Mutações simples de perda de função no canal de potássio (K^+) *KVLQT1* não têm efeito em heterozigotos. Todavia, uma mutação dominante negativa no mesmo gene reduz a função geral para cerca de 20% do normal, produzindo uma condição de herança dominante conhecida como síndrome de Romano-Ward (OMIM 192500: arritmia cardíaca). A perda total de função em pessoas heterozigotas para mutações de perda de função causam uma condição recessiva mais grave, conhecida como síndrome de Jervell e Lange-Nielsen (OMIM 220400: problemas cardíacos e perda auditiva).
- Nos genes *COL1A1* ou *COL1A2*, que codificam colágeno do tipo I, mutações no mesmo gene podem produzir outras duas condições dominantes, a mais branda por simples haploinsuficiência e as formas mais graves por efeitos dominantes negativos (ver **Figura 13.24**).

Correlações entre genótipo e fenótipo são especialmente fracas para condições causadas por mutações mitocondriais

Mutações no DNA mitocondrial apresentam efeitos praticamente imprevisíveis, tanto com relação a causar ou não uma doença como com relação a qual doença irá causar. Uma dada mudança de sequência de mtDNA pode ser vista em pacientes com várias patologias distintas, e pacientes com a mesma doença mitocondrial podem apresentar mutações distintas em seu mtDNA. Além disso, como as células contêm milhares de moléculas de

mtDNA, elas podem ser *homoplásmicas* (todas as moléculas de mtDNA são idênticas) ou *heteroplásmicas* (uma população mista de mtDNA normal e mutante). Diferentemente do mosaïcismo, a heteroplasmia pode ser transmitida da mãe para o filho por meio de um óvulo heteroplásmico. O óvulo contém mais de 10 mil mitocôndrias, de modo que todos os filhos de uma mulher infectada irão herdar pelo menos algumas das mitocôndrias mutantes, mesmo se a mãe for heteroplásmica. No entanto, doenças mitocondriais frequentemente apresentam baixa penetrância (ver Figura 3.10, p. 68).

A neuropatia óptica de Leber (OMIM 535000: perda de visão súbita e irreversível) ilustra alguns dos problemas. Dezoito diferentes mutações de ponto no mtDNA foram associadas com essa condição; cinco delas apresentam efeitos suficientemente graves para, sozinhas, causarem a doença. Outras treze apresentam efeitos que contribuem para o quadro. Cinquenta por cento das pessoas afetadas apresentam a substituição g.11778GA. A maioria destes pacientes é homoplásmico, mas cerca de 14%, apresentando o mesmo grau de severidade, são heteroplásmicos. Mesmo nas famílias homoplásmicas a condição é bastante variável; a penetrância geral é de 33-60%, e 82% dos indivíduos afetados são homens.

Existem diversas razões possíveis para essa fraca correlação:

- O DNA mitocondrial é muito mais variável que o DNA nuclear e algumas síndromes podem depender da combinação da mutação reportada com outras variantes ainda não identificadas.
- Algumas doenças mitocondriais parecem ser de uma natureza quantitativa: pequenas mudanças mutacionais acumulam-se, reduzindo a capacidade mitocondrial de gerar energia e, em um dado limite, surgem os sintomas clínicos.
- A heteroplasmia pode ser tecido-específica, e o tecido examinado (normalmente sangue ou músculo) pode não ser crítico na patogênese.
- Precocemente no desenvolvimento da linhagem germinativa, as células passam por um estágio no qual elas contêm apenas um número reduzido de mitocôndrias (o gargalo mitocondrial). Por um erro aleatório de amostragem neste estágio, em uma mãe heteroplásmica a proporção de mitocôndrias normais e mutantes pode ser significativamente distinta após esse gargalo. Assim sendo, o grau de heteroplasmia pode variar amplamente entre a mãe e o filho.
- Duplicações e deleções no genoma mitocondrial frequentemente evoluem com o passar dos anos em um único indivíduo afetado; de modo similar, a proporção de mitocôndrias mutadas em pessoas heteroplásmicas pode variar ao longo do tempo.
- Muitas funções mitocondriais são codificadas por genes nucleares, de modo que uma variação nuclear pode ser uma causa importante ou um modificador de um fenótipo mitocondrial.

O *MITOMAP database of mitochondrial mutations* (<http://www.mitomap.org>) possui uma boa discussão geral, além de extensas tabelas de dados mostrando a dimensão do desafio de se prever fenótipos mitocondriais.

A variabilidade dentro das famílias é uma evidência da atuação de genes modificadores e de eventos do acaso

Muitas condições mendelianas são clinicamente variáveis mesmo entre membros afetados de uma mesma família, os quais carregam exatamente a mesma mutação. A variabilidade intrafamiliar deve ser causada por alguma combinação dos efeitos de outros genes não ligados (genes modificadores) e efeitos ambientais (incluindo eventos do acaso). Humanos são uma população natural típica, que apresentam maior diversidade genética e uma variedade de ambientes muito mais ampla que os animais de laboratório, de modo que não é surpreendente que correlações entre genótipo e fenótipo sejam muito mais fracas em humanos do que em camundongos de laboratório. Fenótipos dependentes de haploinsuficiência são especialmente sensíveis aos efeitos dos modificadores. A síndrome de Waardenburg é um exemplo típico: a **Figura 13.23** apresenta a prova de que esta condição dominante é causada por haploinsuficiência, e a **Figura 13.17** apresenta a variação intrafamiliar típica.

Correlações entre genótipo e fenótipo passariam a ser significativamente mais precisas se os genótipos considerassem os modificadores tanto quanto as mutações primárias. No entanto, a identificação e a caracterização de genes modificadores em humanos é extremamente desafiadora. Dados os limites inevitáveis da análise genética em humanos, uma abordagem mais produtiva poderá vir com a identificação de todos os componentes das rotas bioquímicas envolvidas em uma doença.

CONCLUSÃO

Neste capítulo foram descritas as muitas formas pelas quais os genomas de dois humanos podem diferir. Isso pode ocorrer principalmente por meio de polimorfismos de nucleotídeo único, números variáveis de mutações em *tandem* e variações em larga escala do número de cópias. Estudos sobre estas diferenças são diretamente aplicáveis à compreensão de diversas características das populações humanas. A estrutura das populações em termos de endogamia e isolamento reprodutivo, as relações entre populações distintas, assim como as origens históricas e geográficas das populações – todos esses aspectos são esclarecidos por estudos de variabilidade genética.

Quando se volta para estudos de saúde e doença, no entanto, não é tão simples. Conectar variações de sequência com saúde e doença requer abordar a questão por ambos os lados. Foi visto como é possível tentar decidir se uma variante particular pode ou não ser patogênica. Em alguns casos a resposta é clara, mas muitas vezes a simples verificação da sequência de DNA não fornece uma resposta clara. Mesmo quando se pode ter confiança de que uma mudança de sequência será patogênica, raramente se pode prever especificamente os sintomas clínicos que ela irá produzir. Para relacionar DNA com doença também precisa-se iniciar pelo lado “doença” dessa relação. O estudo de doenças genéticas nos permite identificar as variantes patogênicas subjacentes. Esta abordagem tem tido amplo sucesso com doenças mendelianas, conforme apresentado nos próximos capítulos. Havendo famílias adequadas, é relativamente fácil identificar a localização cromossômica de um gene que carrega um caractere mendeliano (Capítulo 14), o que geralmente permite que os pesquisadores continuem até a identificação do gene relevante (Capítulo 16). Frequentemente permanece enigmático porque uma disfunção em um gene particular deveria causar um determinado conjunto de sintomas, mas pelo menos as variantes patogênicas podem normalmente ser identificadas com precisão.

Muitas variantes genéticas não causam condições mendelianas, mas podem, mesmo assim, apresentar influência sobre a saúde. Fatores de suscetibilidade genética são determinantes importantes de muitas doenças comuns, embora não sejam os únicos. Relacionar variantes específicas com determinadas suscetibilidades a doenças tem sido um desafio difícil, ainda muito distante de ser concluído. Os Capítulos 15 e 16 revisarão o progresso nessa área. Espera-se, por fim, alcançar uma posição que nos permita derivar informações preditivas úteis em saúde, a partir da análise da sequência genômica de um indivíduo – o que irá levantar toda uma série de questões sobre testes genéticos e triagem populacional. Essas questões serão abordadas nos Capítulos 18 e 19.

LEITURAS ADICIONAIS

Tipos de variação entre genomas humanos

Bentley DR, Balasubramanian S, Swerdlow HP et al. (2008)

Accurate whole-genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. [Em sua maior parte, uma abordagem técnica sobre o uso da tecnologia Illumina para o sequenciamento de um genoma individual.]

DbSNP. www.ncbi.nlm.nih.gov/projects/SNP [As 12 milhões de entradas no Build 126 incluem algumas variantes patogênicas raras e algumas alterações mais complexas do que polimorfismos de um único nucleotídeo.]

Decipher database of pathogenic large-scale variants. <http://decipher.sanger.ac.uk/>

Den Dunnen JT & Antonarakis SE (2001) Nomenclature for the description of human sequence variations. *Hum. Genet.* 109, 121–124.

Horaitis O, Talbot CC Jr, Phommarinh M et al. (2007) A database of locus-specific databases. *Nat. Genet.* 39, 425. [Um manifesto para bancos de dados de mutações *locus*-específicas.]

Jakobsson M, Scholz SW, Scheet P et al. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451, 998–1003. [Análises de SNPs e CNVs em indivíduos de 29 populações.]

Kidd JM, Cooper GM, Donahue WF et al. (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453, 56–64. [Um estudo sistemático de variantes de 8 kb ou de tamanho superior em indivíduos da Ásia, Europa e África.]

Korbel JO, Urban AE, Affourtit JP et al. (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318, 420–426. [Resultados do uso de um método para a detecção de CNVs de até 3 kb.]

Levy S, Sutton G, Ng PC et al. (2007) The diploid genome sequence of an individual human. *PLoS Biol.* 5, e254. [A sequência de Craig Venter.]

Li JZ, Absher DM, Tang H et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104. [O uso de dados de SNP para inferir relações de populações.]

Lukusa T & Fryns JP (2008) Human chromosome fragility. *Biochim. Biophys. Acta* 1779, 3–16. [Uma revisão de sítios frágeis de cromossomos.]

Bancos de dados de mutações. <http://www.genomic.unimelb.edu.au/mdi/> A *Human Genome Variation Society Mutation Database Initiative* lista detalhes e *links* para diversos bancos

- de dados, inclusive para <http://www.hgvbaseg2p.org> [Banco de Dados da Variação do Genoma Humano], e <http://www.hgvbaseg2p.org> [Human Genome Variation Database], and <http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html> [Banco de Dados de Mutações do Genoma Humano], <http://www.mitomap.org> [banco de dados de mutações mitocondriais.]
- Nomenclatura de mutações. <http://www.hgvs.org/mutnomen/> [Regras para nomear qualquer tipo de alteração de sequência de DNA.]
- Perry GH, Dominy NJ, Claw KG et al. (2007) Diet and the evolution of human amylase gene copy-number variation. *Nat. Genet.* 39, 1256–1260. [Exemplo de variante do número de cópias que afeta o fenótipo e que esteve sujeita à seleção.]
- Redon R, Ishikawa S, Fitch KR et al. (2006) Global copy number variation in the human genome. *Nature* 444, 444–454. [Um importante estudo em variação em grande escala na amostra do HapMap.]
- TCAG Database of Genomic Variants. <http://projects.tcag.ca/variation/> [Variantes de larga escala encontradas em indivíduos saudáveis.]
- Wang J, Wang W, Li R et al. (2008) The diploid genome sequence of an Asian individual. *Nature* 456, 60–65.
- Wheeler DA, Srinivasan M, Ehgol M et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–877. [A sequência de James Watson.]
- Wong KK, deLeeuw RJ, Dosanih NS et al. (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* 80, 91–104. [Variantes de 40 kb ou de tamanho superior em 95 indivíduos de 16 grupos étnicos diferentes.]

Falha ao reparar danos do DNA é a principal causa de grande parte da variabilidade genética

- Barnes DE & Lindahl T (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.* 38, 445–476. [Uma revisão geral.]
- Buchwald M (1995) Complementation: one or more groups per gene? *Nat. Genet.* 11, 228–230. [Interessante discussão a respeito do valor e limitações do teste de complementação, apesar dos questionamentos diretos já terem sido respondidos há muito tempo.]
- Friedberg EC (2003) DNA damage and repair. *Nature* 421, 436–440. [Uma abordagem histórica da evolução do conhecimento a respeito de dano e reparação de DNA.]
- Klein HL (2008) DNA endgames. *Nature* 455, 740–741. [Uma breve revisão do modo como são reparadas quebras duplas de DNA.]
- Kraemer KH, Patronas NJ, Schiffmann R et al. (2007) Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neuroscience* 145, 1388–1396. [A via de reparação por excisão de nucleotídeos e seus defeitos.]
- McKinnon PJ & Caldecott KW (2007) DNA strand break repair and human genetic disease. *Annu. Rev. Genomics Hum. Genet.* 8, 37–55. [Revisão da reparação de quebras de fita simples e quebras de fita dupla e das doenças resultantes desses defeitos.]
- Schumacher B, Garinis GA & Hoeijmakers JHJ (2007) Age to survive: DNA damage and ageing. *Trends Genet.* 24, 77–85.
- Wang W (2007) Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat. Rev. Genet.* 8, 735–748.

Variantes patogênicas de DNA

- Bucciantini M, Giannoni E, Chiti F et al. (2002) Inherent cytotoxicity of polypeptide aggregates suggests a common origin of protein misfolding diseases. *Nature* 416, 507–511.
- Grantham RF (1974) Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864. [A matriz de Grantham quantifica o efeito das substituições de aminoácidos.]
- Groman JD, Hefferon TW, Casals T et al. (2004) Variation in a repeat sequence determines whether a common variant of the Cystic Fibrosis Transmembrane Conductance Regulator gene is pathogenic or benign. *Am. J. Hum. Genet.* 74, 176–179. [Efeito das variantes 5T/7T/9T e (TG)ⁿ no *splicing* do mRNA de *CFTR*.]
- Hardy J & Orr H (2006) The genetics of neurodegenerative disease. *J. Neurochem.* 97, 1690–1699.
- Holbrook JA, Neu-Yilik G, Hentze MW & Kulozik AE (2004) Nonsense-mediated decay approaches the clinic. *Nat. Genet.* 36, 801–808.
- Inoue K, Khajavi M, Ohyama T et al. (2004) Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat. Genet.* 36, 361–369. [SOX10 e MPZ como exemplos de patologias moleculares variáveis devido a NMD.]
- Isken O & Maquat LE (2008) The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.* 9, 699–712. [Uma revisão detalhada das funções do NMD nas células.]
- Komar AA (2007) SNPs—silent but not invisible. *Science* 315, 466–467. [Discussão de dados mostrando que um polimorfismo silencioso no gene *MDR1* altera a especificidade por substrato.]
- Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447, 932–940. [Uma revisão dos possíveis mecanismos da instabilidade de mutações dinâmicas.]
- Orr HT & Zoghbi HY (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30, 575–621. [Uma expressiva revisão sobre doenças e suas patogêneses.]
- Pearson CE, Edamura KN & Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729–742. [Não é de fácil leitura, mas apresenta muitos dados e ideias.]
- Wang GS & Cooper TA (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–761. [Diversos exemplos do efeito patogênico de distúrbios de *splicing*.]

Patologia molecular: entendendo o efeito das variantes

- Karniski LP (2001) Mutations in the diastrophic dysplasia sulfate transporter (DTDST) gene: correlation between sulfate transport activity and chondrodysplasia phenotype. *Hum. Mol. Genet.* 10, 1485–1490.
- Lester HA & Karschin A (2000) Gain of function mutants: ion channels and G protein-coupled receptors. *Annu. Rev. Neurosci.* 23, 89–125. [Diversos exemplos detalhados, nem todos em humanos.]
- Marini JC, Forlino A, Cabral WA et al. (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with binding sites for integrins and proteoglycans. *Hum. Mutat.* 28, 209–221. [Análises da relação genótipo-fenótipo em mais de 800 mutações.]

- Owen MC, Brennan SO, Lewis JH & Carrell RW (1983) Mutation of antitrypsin to antithrombin: α -1-antitrypsin Pittsburgh (358 met-to-arg), a fatal bleeding disorder. *N. Engl. J. Med.* 309, 694–698.
- Rauch F & Glorieux FH (2004) Osteogenesis imperfecta. *Lancet* 363, 1377–1385. [Uma revisão geral clinicamente direcionada.]
- Shenker A, Laue L, Kosugi S et al. (1993) A constitutively activating mutation of the luteinizing hormone receptor in familial male precocious puberty. *Nature* 365, 652–654.
- Wollnik B, Schroeder BC, Kubisch C et al. (1997) Pathophysiological mechanisms of dominant and recessive *KVLQT1* K^+ channel mutations found in inherited cardiac arrhythmias. *Hum. Mol. Genet.* 6, 1943–1949.
- Zlotogora J (2007) Multiple mutations responsible for frequent genetic diseases in isolated populations. *Eur. J. Hum. Genet.* 15, 272–278. [Heterogeneidade alélica limitada devido à combinação dos efeitos do fundador e da seleção.]
- A busca por correlações entre genótipo e fenótipo**
- Carrell RW & Lomas DA (2002) α -1-Antitrypsin deficiency—a model for conformational diseases. *N. Engl. J. Med.* 346, 45–53.
- Khrapko K (2008) Two ways to make an mtDNA bottleneck. *Nat. Genet.* 40, 134–135. [Uma breve revisão sobre o gargalo mitocondrial.]
- Manié S, Santoro M, Fusco A & Billaud M (2001) The RET receptor: function in development and dysfunction in congenital malformation. *Trends Genet.* 17, 580–589.
- Pasini B, Ceccherini I & Romeo G (1996) RET mutations in human disease. *Trends Genet.* 12, 138–144.
- Scriver CR & Waters PJ (1999) Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet.* 15, 267–272.
- Wallace DC, Lott MT, Brown MD & Kerstann K (2001) Mitochondria and neuro-ophthalmologic diseases. In *The Metabolic & Molecular Bases Of Inherited Disease*, 8th ed. (CR Scriver, AL Beaudet, WS Sly, D Valle eds), pp 2425–2509. McGraw Hill.
- Weatherall DJ (2001) Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.* 2, 245–255.
- Weatherall DJ, Clegg JB, Higgs DR & Wood WG (2001) The Hemoglobinopathies. In *The Metabolic & Molecular Bases Of Inherited Disease*, 8th ed. (CR Scriver, AL Beaudet, WS Sly, D Valle eds), pp 4571–4636. McGraw Hill.