

# Expressão de Genes Humanos

# 11

## CONCEITOS PRINCIPAIS

- As formas e os comportamentos característicos de diferentes tipos celulares resultam de diferentes padrões de expressão do mesmo conjunto de genes. A expressão é regulada tanto em nível transcricional como em nível traducional.
- A RNA-polimerase II transcreve todos os genes que codificam proteínas e muitos que codificam RNAs não codificantes. Um grande complexo multiproteico precisa ser montado no promotor gênico antes que a transcrição possa ser iniciada.
- Um conjunto fixo de fatores gerais de transcrição forma a base do aparato de transcrição basal, com um grande elenco de fatores de transcrição gene-específicos ou tecido-específicos, coativadores e correpressores que modulam a atividade do promotor.
- A atividade do promotor é afetada por proteínas de ligação ao DNA sequência-específicas que se ligam diretamente ao promotor ou a posições mais distantes (reforçadores, repressores e regiões controladoras de *loci*). Alças de DNA aproximam sítios distantes do promotor. Coativadores e correpressores não se ligam ao DNA, mas são recrutados aos promotores para interações proteína-proteína.
- Estudos em larga escala de transcritos humanos têm demonstrado que a maioria do DNA genômico humano é transcrita. A função (se houver) da maioria dos transcritos é desconhecida. Essas implicações não são atualmente claras, mas uma revisão substancial da visão tradicional de genes discretos esparsamente distribuídos ao longo do DNA genômico pode ser necessária.
- A conformação da cromatina local pode ser mais importante do que a sequência de DNA local na definição de promotores. Esta é governada pela ação combinada entre enzimas modificadoras de histonas, complexos de remodelamento da cromatina dependentes de ATP e metiltransferases que metilam citosinas em dinucleotídeos CpG. As moléculas de RNA podem também possuir um papel neste processo.
- Histonas em nucleossomos podem sofrer muitas modificações covalentes diferentes. O padrão destas modificações constitui uma camada adicional de informação herdável (um código de histonas) que tem um papel principal, mas não exclusivo, na determinação da conformação da cromatina.
- Efeitos herdáveis na expressão gênica que não são causados por uma mudança na sequência de DNA são denominados mudanças epigenéticas. A herdabilidade se deve a padrões herdáveis de metilação das sequências CpG.
- Para poucas dúzias de genes humanos, a expressão depende da origem parental devido ao *imprinting* genômico. Este é um processo epigenético por meio do qual alelos em um *locus* carregam uma impressão da sua origem parental que governa o processo de expressão.
- A maioria dos genes codificadores de proteínas codifica múltiplos polipeptídeos em função do uso alternativo de promotores e/ou de *splicing* alternativo. Em muitos casos, as isoformas das proteínas resultantes têm especificidade tecidual, propriedades biológicas e funções diferentes.
- A expressão gênica é também regulada por meio do controle, independentemente se um mRNA será traduzido. Isso muitas vezes depende das proteínas e microRNAs (miRNAs) que se ligam às sequências da região 3' não traduzida do mRNA.
- O papel da miríade de diferentes miRNAs é ainda incerta, mas acredita-se que eles realizem a regulagem fina da expressão de muitos genes. Mudanças principais na expressão gênica devem-se frequentemente a controles que afetam a transcrição.

Todas as células do corpo humano, afora os gametas, são derivadas de mitoses repetidas do ovo fertilizado original. Todas aquelas células, com poucas exceções, portanto, contêm exatamente o mesmo conjunto de genes. As diferenças em anatomia, fisiologia e comportamento entre células, descritas nos Capítulos 4 e 5, são o resultado de diferentes padrões de expressão gênica. Durante o desenvolvimento normal, as células seguem trajetórias ramificadas nas quais padrões de expressão gênica sucessivamente mais especializados definem as transições de totipotência para pluripotência em direção à diferenciação terminal. Essas decisões de desenvolvimento, em sua maioria, são irreversíveis no contexto da fisiologia humana normal, embora haja grande interesse em revertê-las por meio da manipulação artificial, para indução de células pluripotentes (ver Capítulo 21) ou para promover a regeneração de tecidos ou órgãos danificados. Além desses padrões fixos, as células têm um repertório de mudanças flexíveis de curto prazo na expressão gênica que as capacita a responder a seus ambientes flutuantes.

Controles na expressão gênica operam em vários níveis. Na mais alta escala, regiões de cromatina com milhões de bases em torno dos centrômeros e em outros lugares formam heterocromatina condensada e geneticamente inerte que pode ser observada nos cromossomos metafásicos, utilizando coloração Giemsa (banda G). As bandas G escuras contêm algumas sequências expressas, mas a maioria dos genes expressos é localizada nas regiões que coram claramente em uma preparação de bandeamento G padrão (ver Figura 2.14). A localização de um cromossomo dentro do núcleo celular interfásico pode também ter efeitos profundos nos seus níveis de expressão.

Este capítulo explora como a expressão gênica é controlada desde o simples nível gênico, desde a seleção sobre quais genes transcrever, passando pelo processamento do transcrito, até a tradução para produzir o produto proteico inicial.

Como descrito no Capítulo 8, técnicas para o estudo da expressão gênica têm evoluído rapidamente. Estudos recentes usam RT-PCR para documentar a expressão ou não de genes únicos de interesse em tecidos diferentes ou em diferentes estágios de desenvolvimento. A expressão gênica foi quantificada por RT-PCR em tempo real. Assim como outros aspectos da genética, a ênfase mudou para estudos do genoma, usando microarranjos para perfis de expressão. Geralmente os experimentos comparam perfis de expressão em dois ou mais tecidos ou estados celulares. Agora o foco está avançando para o sequenciamento direto de cDNA em uma escala massiva. Em alguns aspectos, isto é apenas um refinamento da geração em larga escala de identificação de sequências expressas (ESTs, do inglês *expressed sequence tags*) por sequenciamento parcial de bibliotecas de cDNA. Entretanto, as novas máquinas de sequenciamento paralelo massivo permitem uma profundidade de sequenciamento sem precedentes e, conseqüentemente, uma quantificação sem viés, até mesmo de transcritos muito raros, e a detecção de isoformas raras de *splicing*.

Esforços para compreender os fatores que controlam a expressão gênica humana têm sido alvo do projeto ENCODE (Enciclopédia de Elementos de DNA, do inglês *Encyclopedia of DNA Elements*). Esta é uma tentativa sistemática para identificar todos os elementos de sequência funcionais no DNA humano (ver <http://genome.ucsc.edu/ENCODE>). Uma fase piloto, lançada em 2003, concentrada em 44 regiões genômicas, totalizou 30 Mb ou 1% do genoma humano. Em 2007, o projeto passou para uma fase de produção, estudando o genoma total. Ao mesmo tempo, esforços paralelos (ModENCODE; <http://www.modencode.org>) tiveram como alvos dois organismos-modelo importantes, a mosca-da-fruta *Drosophila melanogaster* e o verme nematoide *Caenorhabditis elegans*. Dados do projeto ENCODE compõem muito da discussão deste capítulo.

Há uma crescente conscientização de que a maneira pela qual o genoma funciona é muito mais complexa do que se aparentava anteriormente. O conhecimento inicial certamente se focou nas sequências codificadoras de proteínas espaçadamente distribuídas e foi guiado pela ideia de que cada gene codificava um único polipeptídeo. Sabe-se agora que a maioria dos genes que codifica proteínas codifica múltiplos polipeptídeos. Além disso, embora menos de 1,5% do DNA humano codifique proteínas, muito do restante é transcrito, e células são inundadas com moléculas de RNA recentemente identificadas, das quais possuímos apenas um entendimento muito limitado sobre sua função (se houver). O projeto Genoma demonstrou que os humanos têm mais genes codificadores de proteínas do que o verme *Caenorhabditis elegans*, de 1 mm de comprimento. Parece inevitável que nossa maior complexidade deve ser o resultado de uma organização e regulação mais complexas de um conjunto amplamente similar de genes. Um pouco da nova visão foi apresentada no Quadro 9.5, o qual revisa a discussão do conceito de gene. Ainda se está no estágio inicial da revelação dos mecanismos que controlam a expressão gênica humana, mas este capítulo esboça os conceitos correntes e o cenário para novos dados emergentes.

## 11.1 PROMOTORES E OS TRANSCRITOS PRIMÁRIOS

A expressão gênica depende da transcrição. Como apontado no Capítulo 1, humanos possuem três RNA-polimerases nucleares (e na verdade uma quarta, usada apenas nas mitocôndrias). As três RNA-polimerases nucleares têm funções distintas e usam diferentes tipos de promotores:

- Polimerase I (pol I) transcreve especificamente os genes de RNA ribossomal que existem como arranjos em *tandem* nos braços curtos dos cromossomos acrocêntricos 10 (ver Figura 1.22).
- Polimerase II (pol II) transcreve todos os genes codificadores de proteínas e muito dos genes que codificam RNAs funcionais, incluindo o snoRNAs (pequeno RNA nucleolar, do inglês *small nucleolar RNAs*), que modificam o RNA ribossômico e os microRNAs que regulam a tradução. Este capítulo trata da expressão de genes que são transcritos pela pol II.
- Polimerase III (pol III) é utilizada para transcrever os genes do RNA transportador e o RNA ribossomal 5S, junto com alguns outros pequenos RNAs (ver Figura 1.15).
- As mitocôndrias têm uma RNA-polimerase totalmente diferente, codificada pelo gene nuclear *POLRMT*. Surpreendentemente, considerando a suposta origem bacteriana da mitocôndria, esta não é uma polimerase do tipo bacteriana, mas uma polimerase de polipeptídeo único relacionada às RNAs-polimerase de bacteriófagos.

Os eventos moleculares que ocorrem nos promotores ativos são complexos em seus detalhes, mas o princípio é bastante simples. Estes eventos serão considerados, primeiramente, e, após, a difícil questão sobre o motivo de uma sequência de DNA ser mais ativa como um promotor do que outra.

### Transcrição pela RNA-polimerase II é um processo de múltiplos passos

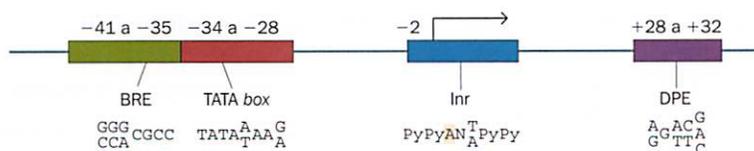
Embora a RNA-polimerase seja uma enzima que sintetiza o transcrito, ela não age sozinha. A transcrição é um processo altamente regulado. Decisões nas quais partes do genoma devem ser transcritas e como fazê-lo ativamente são cruciais para permitir a uma célula responder apropriadamente ao seu ambiente e às demandas do organismo como um todo. Estas decisões são governadas por um grande elenco de fatores regulatórios agindo em um processo complexo e de múltiplos estágios.

### Definindo o promotor-núcleo e o sítio de início da transcrição

Em termos de sequência de DNA, os promotores-núcleo da pol II são altamente heterogêneos. Vários elementos-consenso foram identificados, incluindo os TATA, GC e CCAAT *boxes* descritos no Capítulo 1 (ver Figura 1.14), mas nenhum deles é essencial, e muitos promotores não possuem todos eles. O espaçamento de vários elementos é importante (Figura 11.1), mas muito depende da ligação de ativadores *locus*-específicos, tanto no promotor-núcleo como em sítios mais distantes. Modificações de histonas características (ver a seguir) são melhores do que as sequências de DNA na predição da atividade de promotores, mas o modo como tais modificações são direcionadas para sequências promotoras não está bem compreendido. Uma inspeção de 1.031 promotores-núcleo de pol II descobriu que apenas 32% dos genes portavam um TATA *box* e que tais promotores são normalmente utilizados por genes com expressão altamente tecido-específica. Em contrapartida, promotores de genes *housekeeping*\*, ou aqueles com padrões de expressão complexos, tendem a ter ilhas CpG (ver a seguir), mas não TATA *box*. Os sítios de início da transcrição são fortemente restringidos aos promotores TATA *box*, mas são amplamente distribuídos sobre uma região de 50 a 100 pb nos promotores associados a ilhas CpG. O real sítio de início é quase sempre a purina de um dinucleotídeo pirimidina-purina. O

**Figura 11.1 Elementos de sequência encontrados nos promotores-núcleo das RNA-polimerases II.** As caixas mostram elementos de sequência de DNA que são frequentemente encontrados nos promotores das RNA-polimerases II, junto com suas sequências-consenso e posições relativas ao sítio de início de transcrição. O TATA *box* liga a subunidade da proteína de ligação à TATA do fator de transcrição geral TFIID. O elemento iniciador (Inr) define o sítio de início da transcrição (o A em destaque) quando localizado a 25 a 30 pb de um TATA *box*. O elemento do promotor-núcleo a jusante (DPE, do inglês *downstream core promoter element*) é funcional apenas quando posicionado precisamente em +28 a +32 pb relativo ao A em destaque do elemento Inr. O fator de transcrição geral TFIIB liga-se ao BRE (elemento de reconhecimento do TFIIB). Entretanto, nenhum destes elementos é necessário ou suficiente para promover atividade, e muitos promotores de polimerase II ativos não possuem todos eles. Py, nucleotídeo de pirimidina (C ou T); N, qualquer nucleotídeo. [Adaptada de Smale ST & Kadonaga JT (2003) *Annu. Rev. Biochem.* 72, 449-479. Com permissão da Annual Reviews.]

\* N. de T.: Referente a "mantenedores da casa", ou seja, sempre ativos.



### QUADRO 11.1 Técnicas para identificação dos sítios do início de transcrição

**5' RACE** (Amplificação rápida das extremidades do cDNA, do inglês *Rapid amplification of cDNA ends*) é uma forma de RT-PCT que permite a amplificação da sequência de mRNA inteira entre um *primer* interno e a extremidade 5' de um RNA (Figura 1). Este é um método de baixo rendimento, mas é particularmente útil para a identificação dos éxons não suspeitos a montante de um gene, resultantes da iniciação em um promotor novo. No projeto ENCODE, 5' RACE foi sistematicamente utilizado para examinar transcritos de 399 *loci* em 12 tecidos diferentes. Produtos de RACE foram identificados com o uso de microarranjos de oligonucleotídeos. Fragmentos novos foram encontrados em 90% de todos os *loci* testados.

**CAGE** (Análise de *cap* da expressão gênica, do inglês *Cap analysis of gene expression*) é um método de alto rendimento que sequencia apenas os primeiros 18 nucleotídeos adjacentes à extremidade 5' *cap* de todos os mRNAs em uma amostra. A sequência permite que o sítio de início da transcrição seja identificado, e o número de ocorrências dessa sequência é uma contagem direta do número de moléculas daquele mRNA na amostra original. O método depende de três passos básicos:

- cDNAs que correspondem à extremidade 5' *cap* dos mRNAs são selecionados. Isto é conseguido por uma reação química que biotinila o *cap* do mRNA, seguida pela seleção com contas magnéticas "magnetic beads" recobertas com estreptavidina.
- Um adaptador contendo o sítio de reconhecimento para as enzimas *MmeI* do tipo II é ligado ao cDNA. Enzimas de restrição do tipo II reconhecem uma sequência de DNA específica, mas cortam o DNA em outro lugar, em um número definido de pares de bases distantes. A digestão com *MmeI* libera um fragmento de cDNA de 18 nucleotídeos com duas bases sobressalentes (uma extremidade aderente).
- Os fragmentos de 18 bases com suas extremidades aderentes são ligados um ao outro para formar longos concatêmeros, os quais são então sequenciados.

O método é uma variante da técnica de **SAGE** (Análise em série da expressão gênica, do inglês *Serial analysis of gene expression*), que primeiramente introduziu o uso de enzimas de restrição do tipo II e de sequenciamento de etiquetas (*tags*) concatenadas. As pequenas etiquetas são suficientes para identificar a molécula, e a concatenação possibilita o sequenciamento de um grande número de etiquetas, permitindo uma quantificação precisa sem precedentes de abundâncias relativas de diferentes mRNAs. A proposta do uso de enzimas do tipo II para geração de etiquetas curtas reduz o montante de sequenciamento necessário. Com a introdução de tecnologias de sequenciamento paralelo massivo, isso pode não ser mais uma consideração relevante, e pode ser mais simples apenas sequenciar cDNAs selecionados pela presença de *cap* em uma escala massiva.

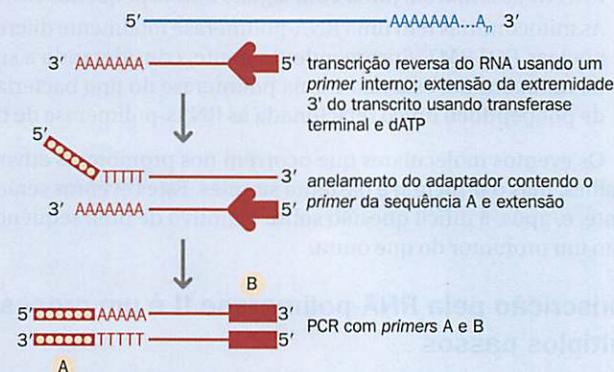


Figura 1 Princípio do 5' RACE

Quadro 11.1 descreve dois dos principais métodos utilizados para definir os sítios de início dos genes.

### Montando o aparato de transcrição basal

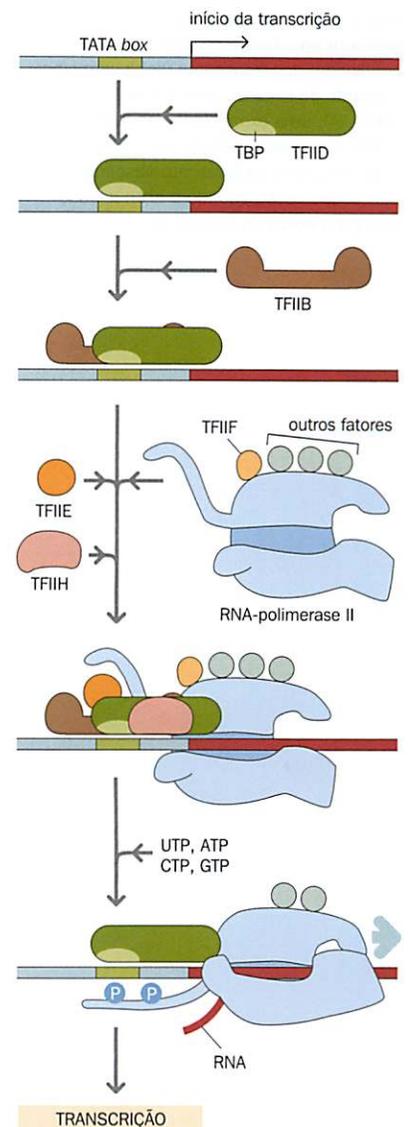
Devido a um ambiente local favorável de cromatina, a maquinaria de transcrição basal será montada sobre o DNA em um promotor. Isto compreende 27 subunidades polipeptídicas: RNA-polimerase II e 5 fatores gerais de transcrição multissubunidades denominados TFIIB, D, E, F e H (Tabela 11.1). A maquinaria monta sequencialmente (Figura 11.2). O primeiro componente a ligar é o TFIID, um complexo multiproteico compreendendo a proteína de ligação à TATA (*TBP*, do inglês *TATA-binding protein*), e em torno de 11 TAFs (fatores associados à TBP, do inglês *TBP-associated factors*) promotores que incluem um *TATA box* são ligados pelo componente TBP do TFIID; algum dos outros componentes

TABELA 11.1 Os fatores gerais de transcrição necessários para a iniciação da transcrição pela RNA-polimerase II

Nome	Nº. de subunidades	Funções na iniciação da transcrição
TFIID		
subunidade TBP	1	reconhece o <i>TATA box</i>
subunidades TAF	~11	reconhece outras sequências de DNA próximas ao ponto de início; regula a ligação de TBP ao DNA
TFIIB	1	reconhece o elemento BRE nos promotores; posiciona precisamente a RNA-polimerase no sítio de início da transcrição
TFIIF	3	estabiliza a interação da RNA-polimerase com TBP e TFIIB; auxilia a atrair o TFIIE e TFIIH
TFIIE	2	atrai e regula TFIIH
TFIIH	9	despiraliza o DNA no ponto de início da transcrição; fosforila a Ser 5 do domínio C-terminal da RNA-polimerase; libera a RNA-polimerase do promotor

TBP, proteína de ligação à TATA; TAF, fatores associados a TBP; BRE, elemento de reconhecimento de TFIIB.

**Figura 11.2 Iniciação da transcrição em um promotor de pol II.** Antes que a RNA-polimerase II possa iniciar a transcrição, vários fatores gerais de transcrição – TFIID, TFIIB, TFIIE, TFIIIF e TFIIH – são requisitados para montar seqüências no promotor. O promotor aqui ilustrado inclui um TATA *box*, ao qual o TFIID se liga. Muitos promotores não possuem um TATA *box* e utilizam outros elementos de seqüência para montagem dos fatores de transcrição e RNA-polimerase II. Quando este aparato de transcrição basal estiver totalmente montado, TFIIH abre a dupla-hélice de DNA no sítio de início da transcrição e inicia uma mudança conformacional na polimerase por meio da fosforilação de seu domínio C-terminal saliente. Isto possibilita à polimerase iniciar a transcrição. Adicionalmente, fatores de transcrição que se ligam ao DNA e coativadores que se ligam a proteínas (não mostrados) controlam o nível de atividade de um promotor. TBP, proteína de ligação a TATA; CTD, domínio C-terminal da RNA-pol II. [De Alberts B, Johnson A, Lewis J et al. (2007) *Molecular Biology of the Cell*, 5th ed. Garland Science/Taylor & Francis LLC.]



presumivelmente direciona a ligação aos promotores que carecem de um TATA *box*. Uma vez que o TFIID tenha sido posicionado no promotor, os fatores TFII remanescentes e a RNA-polimerase II juntam-se para formar o complexo de iniciação.

### Alongamento do transcrito

Ao contrário das DNA-polimerases, RNA-polimerases não necessitam de um *primer* para começar, mas ações adicionais dos fatores TFII são necessárias antes que a transcrição possa iniciar. Uma subunidade da TFIIH é uma DNA-helicase. Esta utiliza energia da hidrólise de ATP para abrir as duplas-hélices de DNA, fornecendo acesso da pol II à fita-molde. A polimerase então começa a síntese de RNA, mas inicialmente falha em progredir para longe do promotor. Pequenos oligorribonucleotídeos são produzidos em um processo de iniciação abortiva até que ocorra um conjunto de mudanças conformacionais que permitam à pol II se mover no modo de alongamento. Tais mudanças são desencadeadas pela fosforilação dependente de TFIIH de resíduos de serina do domínio C-terminal de pol II. Alguns dados sugerem que esta transição é a parte mais crucial do controle da transcrição. Acredita-se que a maioria dos promotores, ativos ou inativos, seja ocupada por um complexo de pré-iniciação, mas apenas aqueles sobre genes ativos são capazes de se mover para um modo de alongamento.

A polimerase move-se então ao longo da fita-molde, deixando para trás a maioria dos fatores gerais de transcrição, mas recrutando novas proteínas, os **fatores de alongamento**, tais como o complexo Elongina. A velocidade média é em torno de 20 nucleotídeos por segundo, embora ocorram pausas e arrancadas. Nesta velocidade ela poderia levar mais do que 24 horas para transcrever o gene da distrofina de 2,4 Mb!

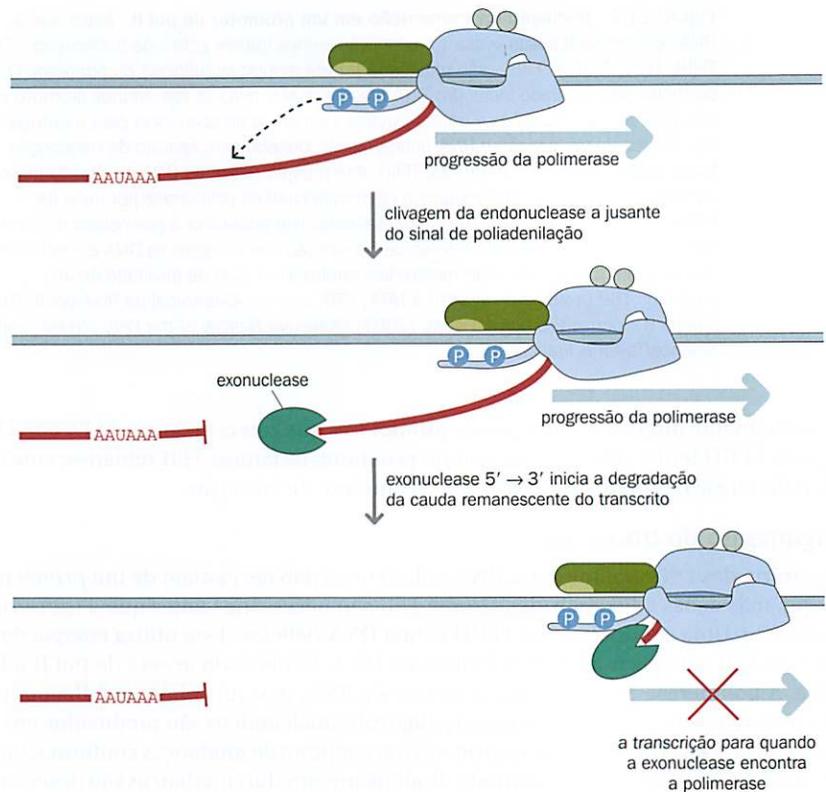
### Término da transcrição

Parecem não existir seqüências específicas que ajam como sinais de parada para RNA-polimerase; não há análogos transcricionais dos códons de parada traducionais UGA, UAG e UAA. Para muitos genes, a terminação ocorre em várias posições, em diferentes transcritos individuais. Verifica-se que o término da transcrição é ligado à clivagem do transcrito. Uma característica bem conhecida dos mRNAs é a do sítio de adição da cauda de poli(A)-AAUAAA ou uma seqüência muito similar a esta. Como descrito no Capítulo 1 (ver Figura 1.21), o mRNA é clivado a poucos nucleotídeos a jusante deste ponto. Esta clivagem também inicia os eventos que levam ao término da transcrição. O corte é feito por um complexo proteico que é fisicamente encadeado à RNA-polimerase e que inclui uma exonuclease. Uma vez que a clivagem tenha produzido uma extremidade livre de RNA, a exonuclease move-se ao longo do RNA que ainda está emergindo da polimerase, degradando-o na direção 5' → 3'. Uma espécie de corrida ocorre entre a polimerase, prosseguindo ao longo do DNA e alongando a cauda do transcrito, e a exonuclease que vem logo atrás da polimerase consome o transcrito (Figura 11.3). A exonuclease é a mais rápida das duas, e quando ela alcança a polimerase, a transcrição termina.

### Muitas outras proteínas modulam a atividade do aparato de transcrição basal

A eficiência do início da transcrição é altamente influenciada por proteínas adicionais que se ligam ao DNA ou ao complexo de início da transcrição. Elas podem ser divididas em duas classes de acordo com sua capacidade de ligação ou não ao DNA. **Fatores de transcrição** são proteínas de ligação ao DNA seqüência-específica que se ligam no pro-

**Figura 11.3 Término da transcrição.** Após a clivagem (flecha pontilhada) dos transcritos (linha vermelha) a jusante do sinal de poliadenilação (AAUAAA), uma exonuclease encadeada à RNA-polimerase cumpre seu curso ao longo da cauda do transcrito nascente. A transcrição cessa quando a exonuclease alcança o corpo da polimerase.



motor núcleo ou próximo a ele. Muitos são tecido-específicos ou regulados pelo desenvolvimento. Promotores são frequentemente inativos na ausência dos fatores de transcrição apropriados. Outras proteínas de ligação ao DNA se ligam a sequências específicas (reforçadores - *enhancers* - e silenciadores) localizadas um pouco distantes do promotor. As alças do DNA unem-as nas proximidades do promotor. **Coativadores** e **correpresores**, em contraste, não se ligam ao DNA: eles dependem de interações proteína-proteína para se associarem com o complexo de início da transcrição e modular sua atividade.

### Proteínas de ligação ao DNA sequência-específicas podem ligar próximo ao promotor ou em locais mais remotos

Fatores de transcrição contêm um domínio de ligação ao DNA que confere especificidade e um domínio de ativação que medeia o efeito funcional. As principais classes de motivos de proteína de ligação ao DNA foram descritas no Capítulo 4: hélice-alça-hélice, hélice-volta-hélice, zíper de leucina e motivos de dedo de zinco (ver Quadro 4.2). Estes se ligam ao DNA de uma maneira sequência-específica. Sequências consenso de ligação podem ser identificadas por experimentos que modificam sistematicamente oligonucleotídeos em busca de um que é mais fortemente ligado. Entretanto, sequências reais no genoma que são ligadas por um dado fator de transcrição não são facilmente predizíveis. Imuno-precipitação de cromatina (ver a seguir) pode ser utilizada para identificar as sequências realmente ligadas *in vivo*, e estas nem sempre correspondem à sequência ideal de ligação identificada pelo primeiro método.

A ligação de um fator de transcrição ao DNA não é determinada simplesmente pela presença de uma sequência de ligação adequada em qualquer parte do genoma. Sítios de ligação para a maioria das proteínas de ligação ao DNA são bastante curtas, geralmente entre quatro e oito nucleotídeos, e alguma variação na sequência é tolerada. Possíveis sítios de ligação para qualquer fator estão presentes em milhares de cópias por meio do genoma, mas apenas uma pequena minoria deles é utilizada. Dois fatores governam a seleção de sítios reais dentre um grande número de sítios potenciais. Primeiro, a ligação é combinatória. A montagem do complexo de iniciação da transcrição depende de interações cooperativas entre a RNA-polimerase e um conjunto completo de fatores de transcrição e coativadores. Um promotor típico tem múltiplos sítios de ligação potenciais, em geral para várias proteínas diferentes. Segundo, proteínas podem apenas ligar-se ao DNA

se a sequência relevante está acessível. O acesso ao DNA é fortemente restrito quando este está enrolado em torno dos nucleossomos. Promotores ativos precisam ter uma estrutura da cromatina relaxada e aberta na qual o DNA esteja relativamente livre de nucleossomos. Considera-se a estrutura da cromatina a seguir.

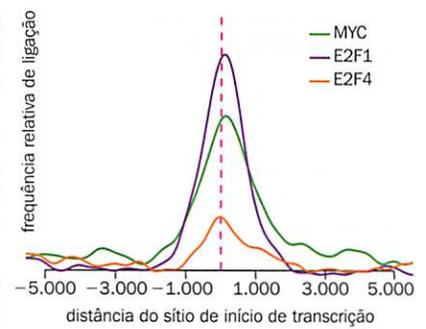
Acreditava-se que sítios de ligação para fatores de transcrição encontravam-se a montante do sítio de início da transcrição. Entretanto, dados do projeto ENCODE mostraram que entre todos os genes estudados, em média, a ligação é simetricamente distribuída em torno do sítio de início da transcrição. Fatores de transcrição podem ligar-se tanto a montante como a jusante do sítio de início de transcrição (Figura 11.4).

Sequências de DNA distais onde a ligação de proteínas sequência-específicas afetam a transcrição são denominadas **reforçadores** ou **silenciadores**, dependendo se ativam ou reprimem a transcrição. Eles podem estar em ambos os lados do promotor. As alças da cromatina curvam-se, possibilitando o contato físico direto com proteínas ligadas ao promotor (Figura 11.5). Genes do desenvolvimento, em particular, frequentemente possuem um conjunto complexo de reforçadores muito distantes. Normalmente estes têm sido descobertos por acaso, quando uma translocação cromossômica separa um gene de um reforçador. Este produz um fenótipo associado com mutações de perda de função do gene, embora a sequência codificante esteja intacta. Exemplos típicos incluem os seguintes:

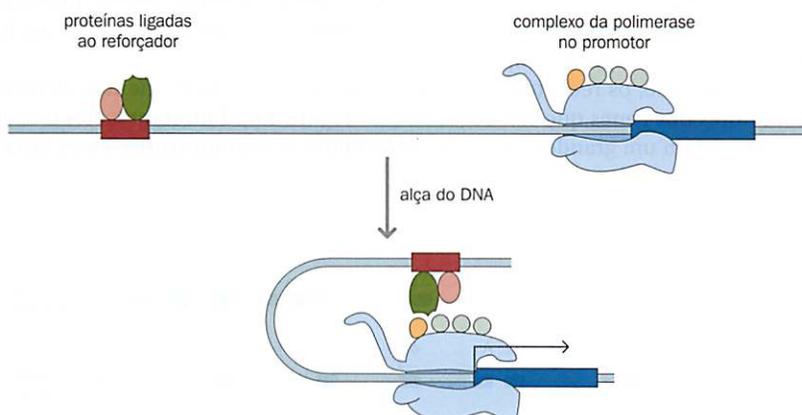
- Mutações com perda de função heterozigotas no gene *PAX6*, no cromossomo 11p15, causa aniridia (ausência da íris do olho; OMIM 106210). Alguns pacientes com aniridia não possuem mutações no *PAX6*, mas translocações com pontos de quebra de até 125 kb a montante do último éxon do *PAX6*. Experimentos com camundongos transgênicos localizaram um grupo de elementos regulatórios essenciais nesta região (Figura 11.6A). Existem vários reforçadores tecido-específicos diferentes neste grupo.
- O papel do gene *SHH* (*Sonic hedgehog*) no desenvolvimento dos membros foi descrito no Capítulo 5. Outro gene, *LMBR1*, é localizado a 1 Mb de distância do gene *SHH* no cromossomo 7q36. Várias deleções, inserções ou translocações no gene *LMBR1* em humanos ou camundongos causam várias anormalidades nos membros (Figura 11.6B). Naturalmente assumiu-se que estas eram o resultado de uma perda de função de *LMBR1* – mas mutações de ponto em *LMBR1* não têm efeito no desenvolvimento dos membros. Em vez disso, tais mudanças rompem uma série de reforçadores de *SHH* que ocorrem em íntrons de *LMBR1*.

Quando um rearranjo cromossômico suprime a expressão de um gene aparentemente intacto pela separação deste de um promotor distante, como no caso do *PAX6* descrito acima, este é frequente mas incorretamente descrito como um *efeito de posição*. Tal termo foi cunhado para descrever o silenciamento gênico observado em *Drosophila* quando um rearranjo cromossômico reposicionou um gene em uma posição próxima a uma região heterocromática (ver Capítulo 16). Efeitos de posição refletem a tendência da dispersão da heterocromatização ao longo de um cromossomo. Sem dúvida, o único efeito de posição conhecido em humanos é o silenciamento ocasional de *loci* autossômicos quando uma translocação de autossomo-X o reposiciona para um cromossomo X inativado (ver a seguir).

Quando se descobriu que grupos de elementos regulatórios distais eram tanto necessários como suficientes para garantir a expressão correta tecido-específica de um gene, eles foram denominados **regiões controladoras de locus** (LCRs, do inglês *locus control*

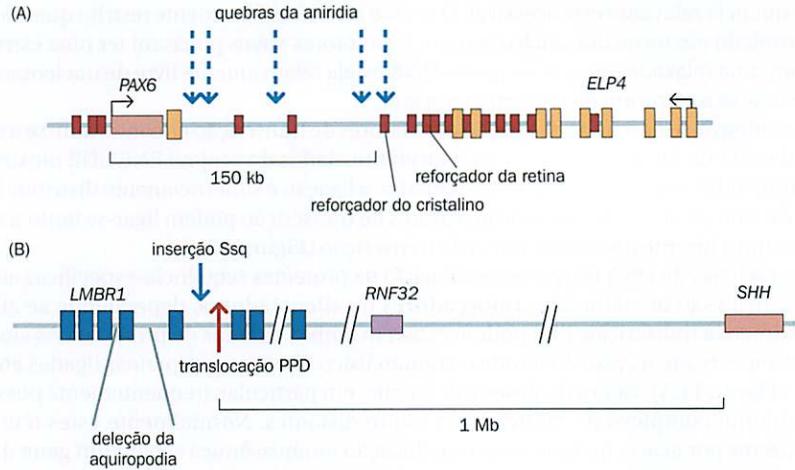


**Figura 11.4** Os fatores de transcrição podem ligar-se tanto a jusante como a montante do sítio de início da transcrição. As curvas mostram a distribuição dos locais nos quais os fatores de transcrição MYC, E2F1 e E2F4 se ligam ao promotor de um conjunto de genes diferentes expressos ativamente relativo ao sítio de início da transcrição (linha pontilhada vertical). As curvas são aproximadamente simétricas em torno do sítio de início da transcrição, mostrando que o sítio de ligação para um destes fatores de transcrição em um gene em particular tem a mesma probabilidade de ocorrer a jusante do sítio de início como a montante na localização convencional do promotor. [Adaptada de ENCODE Project Consortium (2007) *Nature* 447, 799-816. Com a permissão de Macmillan Publishers Ltd.]



**Figura 11.5** Alças do DNA colocam em contato direto proteínas ligadas a reforçadores ou silenciadores com proteínas ligadas ao promotor. Elementos regulatórios como os reforçadores (caixa vermelha) podem estar localizados a centenas de quilobases a montante ou a jusante do gene que eles controlam (caixa azul). A alça do DNA permite interações físicas diretas entre proteínas ligadas a estes elementos distais e algumas das muitas proteínas ligadas aos promotores. Para simplificar, apenas a RNA-polimerase é mostrada no promotor.

**Figura 11.6 Controles de amplo espectro na expressão gênica.** (A) Pontos de quebra de translocação (flechas pontilhadas verticais) de até 125 kb a jusante da extremidade 3' de um gene *PAX6* intacto causa a perda da função de *PAX6* em pacientes com aniridia. Elementos regulatórios essenciais para expressão de *PAX6* devem localizar-se distalmente a todos estes pontos de quebra. Sítios hipersensíveis à DNase (caixas vermelhas) marcam trechos de DNA onde os nucleossomos são ausentes ou instáveis, os quais são, portanto, disponíveis para interação com proteínas de ligação ao DNA. Dois destes sítios foram identificados como reforçadores específicos da retina e específicos do cristalino. As sequências regulatórias de *PAX6* localizam-se dentro de introns do *ELP4* (caixas amarelas marcam os éxons deste gene), um gene não relacionado que é transcrito na orientação oposta. (B) A função do gene *SHH* (*Sonic hedgehog*) no desenvolvimento dos membros depende de reforçadores localizados a 1 Mb de distância, dentro de introns do gene *LMBR1* (caixas azuis mostram os éxons deste gene). Um gene adicional, *RNF32*, fica entre o *SHH* e esses elementos regulatórios. Posições de uma deleção, uma inserção e um ponto de quebra de translocação são apresentados, e todos causam fenótipos devido ao controle anormal da expressão de *SHH*. Ssq, camundongo mutante Sasquatch, do inglês *Sasquatch mouse mutant*; PPD, polidactilia pré-axial humana, do inglês *human preaxial polydactyly* (OMIM 174500); aquiropodia, síndrome humana da ausência bilateral de mãos e pés (OMIM 200500). [De Kleinjan DA & van Heyningen V (2005) *Am. J. Hum. Genet.* 76, 8-32. Com permissão de Elsevier.]



regions). Em sistemas experimentais, regiões controladoras de *loci* podem conferir uma forte expressão independente de posição em transgenes, sendo requeridas se o transgene funcionar como uma unidade transcricional totalmente independente. O exemplo humano de controle de expressão mais conhecido é o dos dois grupos de genes de globina. O grupo da  $\alpha$ -globina no cromossomo 16p e o grupo da  $\beta$ -globina no 11p, LCRs distantes permitem a expressão tecido-específica e a comutação, ao longo do desenvolvimento, de genes nos grupos (Figura 11.7). As LCRs incluem reforçadores e elementos insuladores – o último prevenindo sequências mais distais de afetarem a expressão de genes controlados pela LCR. A deleção da LCR previne a expressão de genes que esta controla, mesmo que as sequências dos genes, incluindo os promotores, estejam intactas. Muitos exemplos de genes são agora conhecidos, nos quais as sequências regulatórias estão mais dispersas, e então a ideia de regiões controladoras de *locus* distintas tem se tornado menos útil.

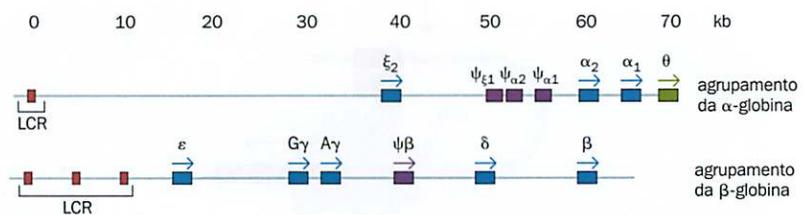
### Coativadores e correpressores influenciam promotores sem ligação ao DNA

Os componentes dos complexos de início da transcrição são reunidos por interações proteína-proteína. Muitas proteínas auxiliam na junção do complexo sem a ligação delas mesmas ao DNA. Portanto, alguns coativadores e correpressores funcionam simplesmente como plataformas para a montagem de um complexo; outros, entretanto, têm atividades enzimáticas. O complexo correpressor CoREST ilustra este caso. CoREST é recrutado aos promotores pela proteína de ligação ao DNA REST, onde este forma uma plataforma de montagem para os outros componentes do complexo correpressor CoREST. Este inclui múltiplas enzimas modificadoras de histonas que coletivamente reprimem a transcrição.

Os receptores nucleares exemplificam a forma na qual coativadores e correpressores geram redes regulatórias elaboradas. Os 48 receptores nucleares humanos são fatores de transcrição cujas atividades são induzíveis. No Capítulo 4, será descrito o seu papel permitindo às células responderem a sinais externos. Em muitos casos, a atividade é induzida pela ligação de um esteroide, tireoide, retinoide ou outro hormônio lipídico solúvel que possa atravessar a membrana celular externa. Alternativamente, um receptor nuclear pode ser ativado indiretamente como uma consequência a jusante da ligação de um ligante extracelular a um receptor da superfície celular.

Uma vez ativado, os receptores nucleares associam-se com elementos de resposta nos promotores dos genes que eles regulam (ver Figura 4.8). Entretanto, eles não agem sozinhos: recrutam um grande elenco de coativadores e correpressores. Estes são com-

**Figura 11.7 Regiões controladoras de loci para os grupos da  $\alpha$ - e  $\beta$ -globina.** As regiões controladoras de *loci* (LCRs) são marcadas por um ou mais sítios hipersensíveis a DNase I (caixas vermelhas) localizadas a montante de um grupo, onde nucleossomos são ausentes ou instáveis, e o DNA é portanto disponível para interação com proteínas de ligação ao DNA. Esses sítios são hipersensíveis em células eritroides, onde os genes da globina são expressos, mas não em outras linhagens celulares. Caixas azuis mostram genes expressos; caixas roxas, pseudogenes. O status funcional do gene da  $\theta$ -globina (caixa verde) é incerto. Flechas mostram a direção do transcrição de genes expressos.



plexos multiproteicos que incorporam uma gama de atividades enzimáticas relevantes ao controle da transcrição. Por sua vez, tais atividades enzimáticas são reguladas por uma série de modificações covalentes reversíveis (fosforilação, metilação, acetilação, etc.) de proteínas nos complexos coativadores e correpressores – modificações que podem ser realizadas por outros coativadores e correpressores. Estas interações geram redes muito extensivas para controlar e coordenar respostas celulares múltiplas a sinais externos.

Entre eles, fatores de transcrição, coativadores e correpressores controlam a atividade de um promotor. No entanto, eles não explicam por que uma sequência do DNA mais do que outra é escolhida para funcionar como um promotor. Como mencionado acima, existem milhares de sítios de ligação potenciais para fatores de transcrição por meio do genoma, mas apenas um limitado subconjunto é realmente utilizado. A escolha depende principalmente da configuração local da cromatina, a qual será descrita na próxima seção.

## 11.2 CONFORMAÇÃO DA CROMATINA: A METILAÇÃO DO DNA E O CÓDIGO DE HISTONAS

Quando totalmente estendido, o DNA de células diploides humanas poderia alcançar dois metros de comprimento. Como descrito no Capítulo 2, a dupla-hélice exposta está sujeita a diversos níveis de empacotamento. O nível mais básico é o nucleossomo, no qual 147 bp de DNA são enrolados em torno de um complexo de oito núcleos de moléculas de histonas – normalmente duas de cada de H2A, H2B, H3 e H4. Um trecho de comprimento variável de DNA livre separa nucleossomos adjacentes, e este é estabilizado por uma molécula da histona conectora H1. Histonas são proteínas pequenas (com geralmente 130 aminoácidos de comprimento) e altamente básicas, ricas em lisina e arginina. O ponto isoelétrico das histonas é 11 ou maior, tanto que em pH intracelular típico elas possuem uma forte carga positiva. Isto dá a elas uma afinidade pelo DNA negativamente carregado. Uma molécula de histona tem um corpo globular, bem como caudas N- e C-terminal que sobressaem do corpo dos nucleossomos. Modificações covalentes de aminoácidos nas caudas governam o comportamento dos nucleossomos.

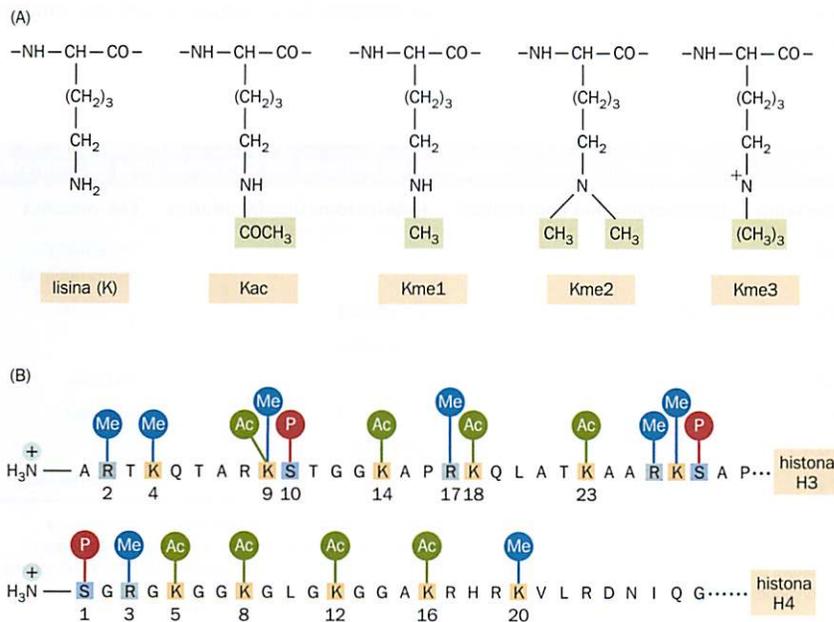
### Modificações de histonas nos nucleossomos podem compor um código de histonas

Histonas em nucleossomos estão sujeitas a muitas modificações diferentes que afetam resíduos de aminoácidos específicos nas caudas. A nomenclatura utilizada para descrevê-las é explicada no **Quadro 11.2**. Modificações comuns incluem a acetilação e monometilação, dimetilação ou trimetilação de lisinas (**Figura 11.8**) e fosforilação de serinas. Estas são efetuadas por grandes famílias de enzimas: histona-acetiltransferases (HATs),

#### QUADRO 11.2 Nomenclatura para modificação de histonas

Um código abreviado é utilizado para modificações de histonas. Resíduos de aminoácidos específicos são identificados pelo tipo de histona, o código de aminoácido de uma letra e a posição do resíduo, contando da extremidade N-terminal. Assim, H3K9 é lisina 9 na histona H3.

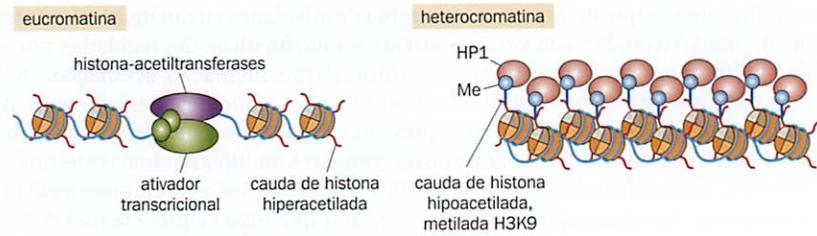
Modificações são então descritas usando ac para acetilação, me para metilação, ph para fosforilação, ub para ubiquitinação e su para sumoilação. Por exemplo, H4K12ac é lisina 12 acetilada da histona H4, e H3K4me3 é a lisina 4 trimetilada da histona H3.



**Figura 11.8** Modificações de histonas. (A) O grupo  $\epsilon$ -amino dos resíduos de lisina pode ser modificado pela acetilação ou adição de 1 a 3 grupos metil. A nomenclatura-padrão das lisinas modificadas é apresentada. (B) As caudas N-terminal das histonas H3 e H4 são os sítios de muitas modificações que controlam a estrutura da cromatina. A sequência de aminoácido é apresentada no código de aminoácido de uma letra, e modificações potenciais são indicadas. Ac, acetilação; Me, metilação; P, fosforilação.

**Figura 11.9 Eucromatina e heterocromatina.**

Na eucromatina, os nucleossomos são frouxamente empacotados, com regiões livres de nucleossomos que podem ligar proteínas regulatórias. Na heterocromatina, os nucleossomos são densamente empacotados e associados com a proteína da heterocromatina 1 (HP1, do inglês *heterochromatin protein 1*). Diferentes modificações de histonas marcam os dois estados. Em comparação com a eucromatina, a heterocromatina é pobre em genes, contém muitas repetições de seqüências de DNA e replica mais tarde, durante a fase S do ciclo celular. [De Grewal SI & Elgin SC (2007) *Nature* 447, 399-406. Com a permissão de Macmillan Publishers Ltd.]



histona-metiltransferases (HMTs) e histona-cinases. Histona-desacetilases (HDACs), histona-demetilases e histona-fosfatases revertem estes efeitos. Modificações menos frequentes incluem monometilação, dimetilação ou trimetilação de argininas, fosforilação de treoninas H3T3 e H3T11 e ubiquitinação de lisinas H2AK119 e H2BK120.

**Cromatina aberta e fechada**

O empacotamento dos nucleossomos em estruturas cada vez mais complexas é crucial na determinação da atividade dos genes. Existem duas variantes básicas. Heterocromatina é uma conformação fechada e inativa; eucromatina é aberta e potencialmente ativa (Figura 11.9). A heterocromatina pode ser constitutiva – ou seja, mantém tal estrutura por meio de todo o ciclo celular – ou facultativa. As formas facultativas da heterocromatina são reversíveis durante a vida da célula como parte do controle da expressão gênica, como ilustrado pelo *status* do cromossomo X em fêmeas (ver abaixo). Como mencionado no Capítulo 9, o DNA da heterocromatina constitutiva consiste principalmente de repetições e é amplamente desprovido de genes. Modificações de histona específicas diferenciam os principais tipos de cromatina (Tabela 11.2). Várias técnicas permitem que a conformação da cromatina possa ser investigada em uma escala local ou genômica (Quadro 11.3).

O conceito de **código de histonas** implica que combinações particulares de modificações de histonas definem a conformação da cromatina e conseqüentemente a atividade do DNA contido naquele lugar. Embora haja indubitavelmente uma correlação geral forte, nenhuma modificação de histona simples é completamente preditiva do estado da cromatina ou da atividade do DNA. As várias modificações de histonas influenciam a estrutura e função da cromatina agindo como sítios de ligação para uma grande gama de proteínas efetoras não histônicas (normalmente a cromatina contém, aproximadamente, quantidades iguais de proteínas histônicas e não histônicas). Estas proteínas contêm domínios que reconhecem modificações de histonas específicas: **bromodomínios** reconhecem lisinas acetiladas, e **chromodomínios** reconhecem lisinas que são metiladas. Proteínas particulares em cada classe reconhecem resíduos específicos de lisina particulares. Outros domínios frequentes, tais como o PHD (homeodomínio em planta, do inglês *plant homeodomain*), interagem de uma forma generalizada com a cromatina. Muitas proteínas de ligação à cromatina contêm vários destes domínios, capacitando-as a lerem modificações de histonas de uma forma combinatória.

**TABELA 11.2** Características das modificações de histonas em diferentes tipos de cromatina

Aminoácido	Heterocromatina constitutiva	Heterocromatina facultativa	Eucromatina
H3K4			trimetilada <sup>a</sup> , monometilada <sup>b</sup>
H3K9	trimetilada	dimetilada	acetilada
H3K27		trimetilada	
H4R3			metilada
H4K5			acetilada
H4K12	acetilada	acetilada	
H4K20	trimetilada		monometilada

Dentro de cada categoria de cromatina existem subvariedades e padrões variantes de modificação, mas aqueles apresentados são as modificações mais frequentes em cada categoria ampla de cromatina. Ver Quadro 11.2 para nomenclatura dos aminoácidos nas histonas. Caixas brancas significam que não há um padrão claro. <sup>a</sup>Nos promotores.

<sup>b</sup>Nos reforçadores.

Um bom exemplo da importância das modificações de histona para a regulação gênica é fornecido pela metilação de H3K4. Formas dimetilada e trimetilada de H3K4 aparecem em picos discretos no genoma que se sobrepõem precisamente com regiões promotoras. Por sua vez, essa modificação é reconhecida por um domínio PHD na subunidade TAF3 da maquinaria de transcrição basal TFIID. Assim, a metilação H3K4 constitui um *panorama específico da cromatina* que coincide com promotores, e esta modificação contribui diretamente no recrutamento da maquinaria de transcrição da RNA-pol II. As metiltransferases de DNA, as quais metilam citosinas nos sítios CpG e operam como complexos de

**QUADRO 11.3 Técnicas para estudo da conformação da cromatina**

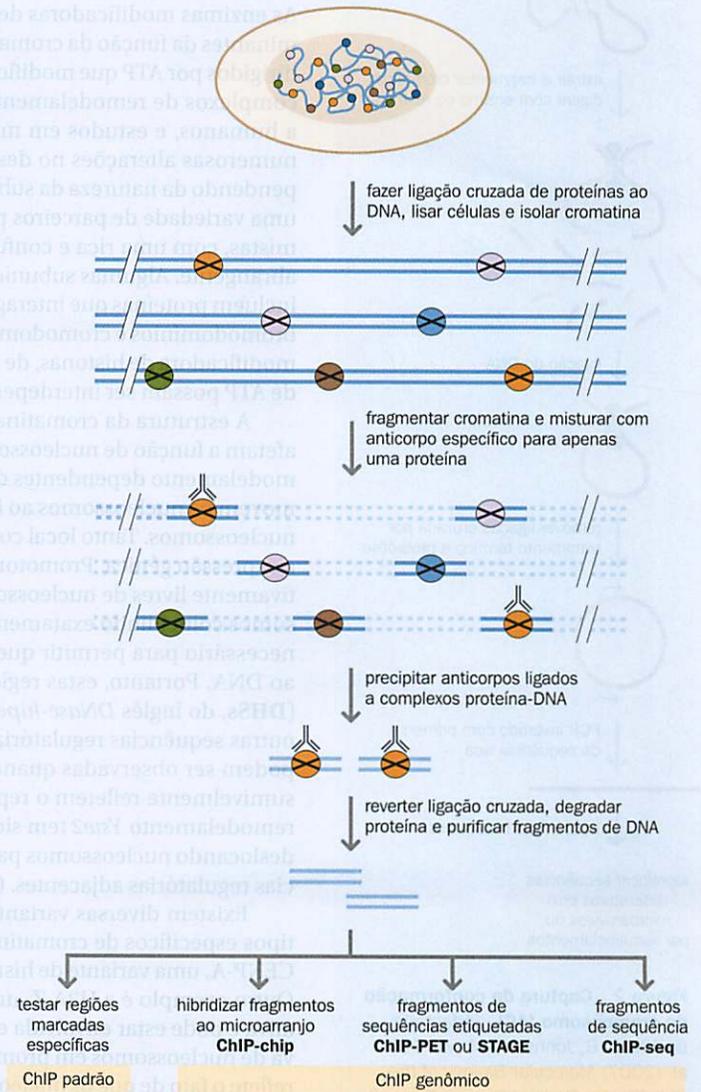
Na **imunoprecipitação da cromatina (ChIP)**, células são tratadas com formaldeído para formar ligações covalentes entre o DNA e suas proteínas associadas. As células são lisadas, a cromatina é fragmentada, e um anticorpo contra alguma proteína de interesse é utilizado para precipitar os fragmentos de cromatina que contêm aquela proteína. Exemplos típicos incluem os fatores de transcrição ou histonas que possuem alguma modificação específica (anticorpos estão disponíveis contra muitas histonas modificadas). A incubação a 65°C reverte a ligação cruzada entre o DNA e a proteína, e a proteína é removida por digestão com proteinase. O DNA que estava associado com a proteína é então identificado. PCR com *primers* gene-específicos podem ser utilizados para verificar quais proteínas estavam ligadas próximo a um gene de interesse. Normalmente o DNA recuperado é hibridizado a um microarranjo para identificar associações genômicas – a chamada técnica ChIP-chip.

A ChIP-chip pode dar uma visão genômica global e tem sido amplamente utilizada, mas possui várias limitações. Ela requer vários microgramas de DNA para obter um bom sinal, então os resultados são uma média do estado de milhões de células. Sequências repetidas e variações alélicas são difíceis de estudar devido à hibridização cruzada, e os resultados são apenas semi-quantitativos em consequência do possível viés na extensiva amplificação de PCR necessária para obter DNA suficiente. Uma abordagem alternativa é o sequenciamento em larga escala do DNA recuperado sem amplificação. Ocorrências de cada gene individual no imunoprecipitado são simplesmente contadas. Algumas vezes, apenas etiquetas curtas de cada extremidade são sequenciadas [ChIP-PET (*paired end tags*) ou STAGE (*sequence tag analysis of genomic enrichment*)]. Progressivamente, o fragmento inteiro é sequenciado (ChIP-seq). A nova geração de sequenciadores paralelos massivos está tornando este um método atraente para obtenção de dados não enviesados para qualquer nível de resolução desejado.

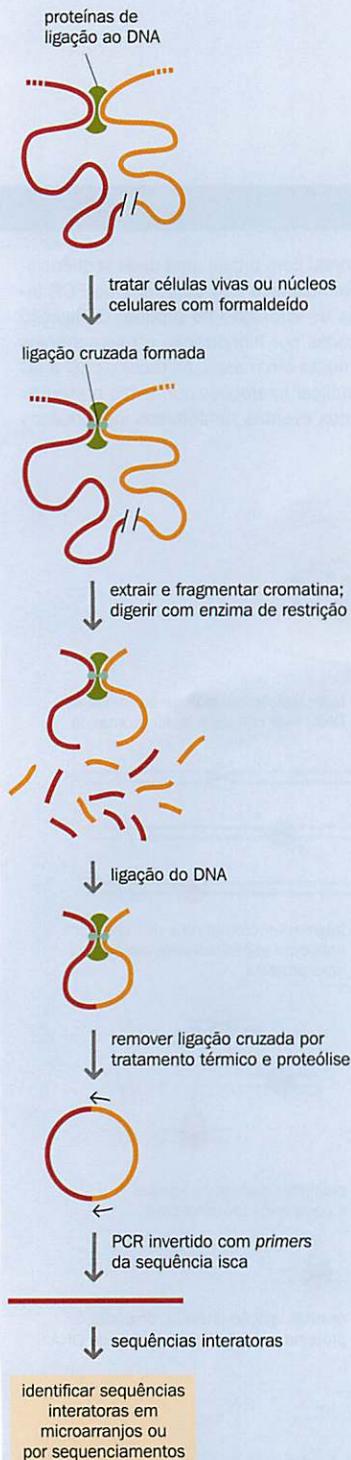
A **captura da conformação do cromossomo (3C, 4C)** (Figura 2, no verso) é utilizada para identificar sequências de DNA que podem ser amplamente separadas na sequência genômica, mas que estão fisicamente próximas dentro do núcleo celular. Células vivas ou núcleos isolados são tratados com formaldeído para realizar uma ligação cruzada entre regiões da cromatina que estão fisicamente em contato uma com a outra. A cromatina ligada é solubilizada e digerida com uma enzima de restrição. Sequências de interação distante serão representadas por uma cromatina com ligação cruzada contendo um fragmento de DNA de cada uma das sequências de interação. Como resultado da digestão com enzimas de restrição, os dois fragmentos possuirão extremidades ligantes compatíveis. A DNA-ligase é usada para tentar ligar as extremidades de duas diferentes regiões de DNA. Os fragmentos de DNA ligado são liberados da cromatina como antes. Eles então podem ser identificados por quaisquer dos vários métodos.

O método 3C original utilizou amplificação por PCR quantitativo do produto de ligação para testar a frequência com a qual duas sequências de DNA especificadas se associam. Um *primer* de cada par era específico para cada uma das duas sequências escolhidas. O 4C estendeu esta ideia para produzir uma lista de

sequências não enviesada (interatores) com a qual uma dada sequência-isca se associa. *Primers* da sequência-isca são utilizados no PCR invertido para amplificar sequências de interação no produto de ligação circular. Estas são então identificadas por hibridização e microarranjos ou, cada vez mais, por sequenciamento em massa. As técnicas 3C e 4C são quantitativas: o objetivo é identificar interações que estão presentes mais frequentemente do que muitos eventos randômicos identificados por estas técnicas.



**Figura 1** Imunoprecipitação da cromatina.

QUADRO 11.3 *continuação*

**Figura 2** Captura da conformação do cromossomo (4C). [Adaptada de Alberts B, Johnson A, Lewis J et al. (2007) *Molecular Biology of the Cell*, 5th ed. Garland Science/Taylor & Francis LLC.]

silenciamento potente (ver a seguir), também creditam o *status* de H3K4 na cromatina por meio de um módulo dedicado de ligação de histonas. Em contraste com TAF3, as DNA-metiltransferases podem ligar-se apenas a nucleossomos que não são metilados em H3K4. Assim, a metilação de H3K4 contribui não apenas para o recrutamento da maquinaria de transcrição, mas também para a proteção das regiões promotoras contra o silenciamento não desejado. Isto é particularmente importante para genes *housekeeping* transcritos de promotores associados a ilhas CpG, os alvos naturais para DNA-metiltransferases.

Uma rede de interações afeta o resultado final, e nenhum fator único é totalmente determinante. As proteínas que reconhecem especificamente histonas modificadas podem elas mesmas ter atividade de modificação de histonas, produzindo ciclos de retroalimentação positiva – por exemplo, algumas proteínas que contêm cromodomínios têm atividade de desacetilação de histonas. Assim, modificações de histonas podem ser interdependentes. Certas modificações dão seguimento a outras. A fosforilação da serina na H3S10 promove a acetilação da H3K9 adjacente e inibe sua metilação. A ubiquitinação da H2AK119 é um pré-requisito para a dimetilação e trimetilação de H3K4. Os complexos de proteínas muitas vezes também incluem componentes de complexos de remodelamento da cromatina dependentes de ATP e DNA-metiltransferases que metilam o DNA, ambos descritos a seguir. A expressão gênica depende de um balanço de efeitos estimulatórios e inibitórios mais do que um simples código individual de histonas.

### Complexos de remodelamento da cromatina dependente de ATP

As enzimas modificadoras de histona descritas acima constituem um conjunto de determinantes da função da cromatina; outro conjunto compreende complexos multiproteicos dirigidos por ATP que modificam a associação de DNA e histonas. Os componentes desses complexos de remodelamento de cromatina são fortemente conservados, de leveduras a humanos, e estudos em muitos organismos têm demonstrado seu envolvimento em numerosas alterações no desenvolvimento. Eles podem ser agrupados em famílias, dependendo da natureza da subunidade ATPase (Tabela 11.3). Cada ATPase associa-se com uma variedade de parceiros para formar grandes complexos apresentando combinações mistas, com uma rica e confusa nomenclatura. A Tabela 11.3 está longe de ser uma lista abrangente. Algumas subunidades são tecido-específicas. Os complexos frequentemente incluem proteínas que interagem especificamente com histonas modificadas, por meio de bromodomínios e cromodomínios, por exemplo. Outras subunidades podem ter atividade modificadora de histonas, de modo que fatores de modificação de histonas e dependentes de ATP possam ser interdependentes.

A estrutura da cromatina é dinâmica. Enquanto enzimas modificadoras de histonas afetam a função de nucleossomos por meio de mudanças covalentes, os complexos de remodelamento dependentes de ATP mudam a ocupação dos nucleossomos no DNA. Eles movem os nucleossomos ao longo do DNA ou promovem a montagem ou desmonte dos nucleossomos. Tanto local como globalmente, o posicionamento dos nucleossomos afeta a expressão gênica. Promotores de genes ativos e outras sequências regulatórias são relativamente livres de nucleossomos, com declínio característico na ocupação dos nucleossomos coincidindo exatamente com o sítio de início da transcrição. Isto é provavelmente necessário para permitir que a RNA-polimerase e fatores de transcrição tenham acesso ao DNA. Portanto, estas regiões aparecem como **sítios de hipersensibilidade a DNase (DHSs, do inglês DNase-hipersensitive sites)**, os quais geralmente marcam promotores e outras sequências regulatórias. Mudanças na sensibilidade a DNase em *loci* particulares podem ser observadas quando células se diferenciam ou mudam seu estado; estas presumivelmente refletem o reposicionamento de nucleossomos. Em leveduras, o fator de remodelamento *Ysw2* tem sido apresentado agindo como repressor geral de transcrição, deslocando nucleossomos para longe das sequências codificadoras em direção a sequências regulatórias adjacentes. O análogo humano age provavelmente de maneira similar.

Existem diversas variantes de histonas, as quais substituem as histonas padrão em tipos específicos de cromatina ou em resposta a sinais específicos. A mais conhecida é a CENP-A, uma variante de histona H3 que é encontrada em heterocromatina centromérica. Outro exemplo é a H2A.Z, uma variante H2A que é associada com regiões de cromatina ativa e pode estar envolvida em conter a dispersão da heterocromatina. A depleção relativa de nucleossomos em promotores ativos e sequências regulatórias, mencionadas acima, reflete o fato de que os nucleossomos nesses locais tendem a conter H2A.Z e outra variante de histona, H3.3. Isto as torna instáveis, permitindo assim um melhor acesso das proteínas de ligação ao DNA a essas sequências. Outra variante de H2A, macro-H2A, é característica

**TABELA 11.3** Exemplos de complexos de remodelamento da cromatina dependentes de ATP

Família	Complexo	Nº de unidades	Subunidade ATPase	Subunidade de interação com histonas
SWI/SNF	BAF	10	BRG1 ou BRM	BAF155 (cromodomínio)
ISWI	NuRF	4	SNF2L	BPTF (bromodomínio)
	CHRAC	3	SNF2H	–
	ACF	2	SNF2H	ACF1 (bromodomínio)
CHD	NuRD	6	CHD3/4	HDAC1/2 (cromodomínio)
INO80	TIP60	15	DOMINO	TIP60 (cromodomínio)

BAF, fator associado a Brm ou Brg, do inglês *Brm- or Brg-associated factor*; NuRF, fator de remodelamento de nucleossomos, do inglês *nucleosome remodeling factor*; CHRAC, complexo de acessibilidade à cromatina, do inglês *chromatin accessibility complex*; ACF, fator de montagem da cromatina dependente de ATP, do inglês *ATP-dependent chromatin assembly factor*; NuRD, remodelamento de nucleossomos e desacetilação de histonas, do inglês *nucleosome remodeling and histone deacetylase*; TIP, proteína interatora Tat, do inglês *Tat-interactive protein*.

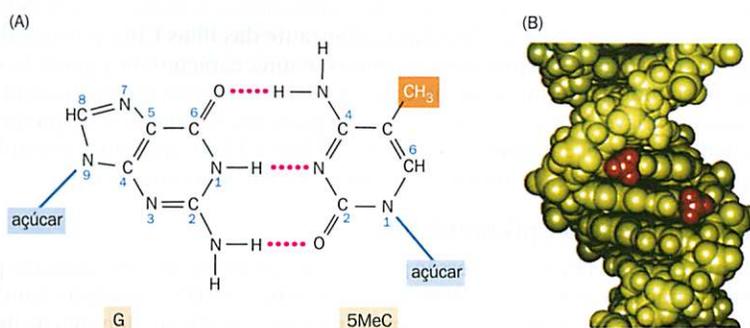
de nucleossomos no cromossomo X inativo. A incorporação destas variantes de histona depende da habilidade de alguns complexos de remodelamento de cromatina afrouxarem a estrutura dos nucleossomos e promoverem a permuta de histonas. Por exemplo, o complexo TIP60 (ver Tabela 11.3) tem uma função na substituição de H2A por H2A.Z.

### A metilação do DNA é um controle importante na expressão gênica

Como mencionado no Quadro 9.1, o DNA é algumas vezes modificado pela metilação da posição 5 das bases de citosinas. Apenas citosinas as quais o vizinho a jusante é uma guanina – denominadas sequências CpG – são sujeitas à metilação. O grupo metil fica na fenda principal sobre o lado externo da dupla-hélice de DNA; isto não interfere no pareamento de base. Assim, a 5-metilcitosina pareia com guanina da mesma forma que citosinas não modificadas (Figura 11.10), mas o grupo metil age como um sinal que é reconhecido por proteínas específicas de ligação à meCpG. Estas têm um papel na regulação da estrutura da cromatina e na expressão gênica. MeCpG também tem um papel importante na memória epigenética, a qual é discutida abaixo. O estado de metilação de citosinas individuais pode ser investigado examinando-se o resultado do tratamento de DNA com bissulfito de sódio (Quadro 11.4).

A metilação de citosinas é realizada pelas DNA-metiltransferases (DNMTs). Humanos possuem três DNMTs funcionais (Tabela 11.4); uma quarta proteína, DNMT3L, auxilia a direcionar dimetilases para as sequências apropriadas, enquanto uma quinta, DNMT2, acaba por ser uma enzima de metilação do RNA, a despeito de sua estrutura similar às DNA-metiltransferases.

Em mamíferos, a grande maioria (em torno de 70%) de todas as sequências CpG é metilada. Apenas em plantas e muitos invertebrados, a metilação CpG é concentrada em sequências repetitivas, incluindo repetições-satélite características de heterocromatina pericêntrica e transposons dispersos. A metilação do DNA em sequências repetidas serve provavelmente para reprimir a transcrição, ilustrando seu papel como mecanismo de defesa genômico. A metilação é também esporadicamente distribuída no corpo dos genes, tanto em íntrons como em éxons, e em sequências intergênicas. Como discutido no Quadro 9.1, o dinucleotídeo CpG é o menos frequente no genoma humano devido à tendência das citosinas metiladas de desaminarem-se em timinas. Ilhas CpG, as quais são encontra-



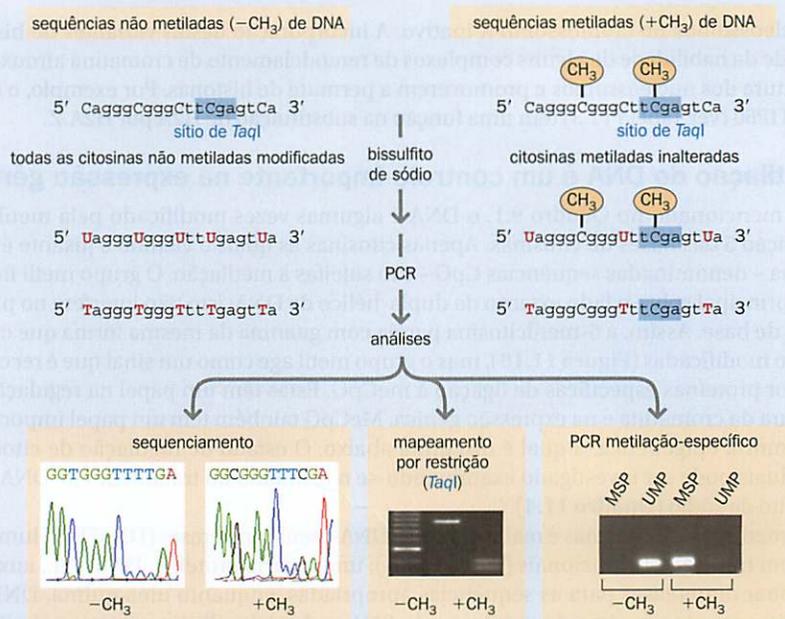
**Figura 11.10** 5-Metilcitosina. (A) A 5-metilcitosina pareia com guanina da mesma forma que uma citosina não modificada. (B) Há uma sequência CpG simetricamente metilada no centro desta molécula. Os grupos metil (vermelho) ficam na cavidade principal da dupla-hélice. [(B) Criada por Mark Sherman. Cortesia de Arthur Riggs e Craig Cooney.]

### QUADRO 11.4 Modificação de bissulfito do DNA

Modificação de bissulfito é um método para a identificação do estado de metilação das citosinas no DNA genômico. Quando a fita simples do DNA é tratada com bissulfito de sódio ( $\text{Na}_2\text{SO}_3$ ) ou metabissulfito ( $\text{Na}_2\text{S}_2\text{O}_5$ ) sob condições cuidadosamente controladas, a citosina é desaminada à uracila, mas a 5-metilcitosina permanece sem modificações. Quando o DNA tratado é sequenciado ou amplificado por PCR, a uracila é lida como timina. Comparando a sequência antes e após o tratamento com bissulfito, é possível identificar quais citosinas foram metiladas.

Vários métodos têm sido utilizados para identificar quaisquer mudanças induzidas pelo tratamento. O estado de específicas sequências

CpG pode ser examinado por sequenciamento, digestão por enzimas de restrição ou PCR. Uma mudança  $\text{C} \rightarrow \text{T}$  pode criar ou destruir um sítio de restrição (na Figura 1, o tratamento destrói um sítio TCGA de *TaqI* se a C não está metilada). PCR metilação-específico utiliza *primers* que são específicos tanto para citosinas modificadas (MSP) como para não modificadas (UMP) nos sítios de ligação aos *primers*. Perfis de metilação genômica podem ser obtidos pela análise do DNA tratado com bissulfito em microarranjos de oligonucleotídeos especialmente desenhados que contêm sondas para encontrar tanto cada sequência normal como seu análogo modificado por bissulfito.



**Figura 1** Modificação do DNA por bissulfito de sódio. MSP, *primer* específico para a sequência metilada; UMP, *primer* específico para a sequência não metilada.

das nos promotores de muitos genes humanos, representam uma exceção à regra, retendo uma proporção relativamente alta de dinucleotídeos CpG.

Estes densos agrupamentos de sequências CpG tendem a ser protegidos da metilação do DNA independentemente se o gene associado está ativo ou inativo, e eles provavelmente estão sob forte pressão seletiva para reter a atividade do promotor. Ocasionalmente, entretanto, ilhas CpG podem sofrer metilação do DNA, a qual inevitavelmente resulta na parada da transcrição, como observado no cromossomo X inativo ou em genes que sofreram *imprinting* (ver a seguir). Metilação aberrante das ilhas CpG, particularmente aquelas associadas a genes supressores tumorais, é uma característica geral de células cancerosas (discutido posteriormente no Capítulo 17). Uma minoria das ilhas CpG apresenta metilação variável que se correlaciona com o silenciamento gênico, frequentemente como parte de alterações no desenvolvimento. A Figura 11.11 apresenta exemplos das distribuições da metilação CpG em diferentes tecidos e em diferentes *loci*.

#### Proteínas de ligação a CpG Metiladas

O efeito da metilação do DNA na expressão gênica pode ser observado diretamente, pois alguns fatores de transcrição tais como YY1 falham em se ligar ao DNA metilado. A metilação do DNA pode também ser interpretada por proteínas que contêm um domínio de ligação a

**TABELA 11.4** Metiltransferases de DNA humanas

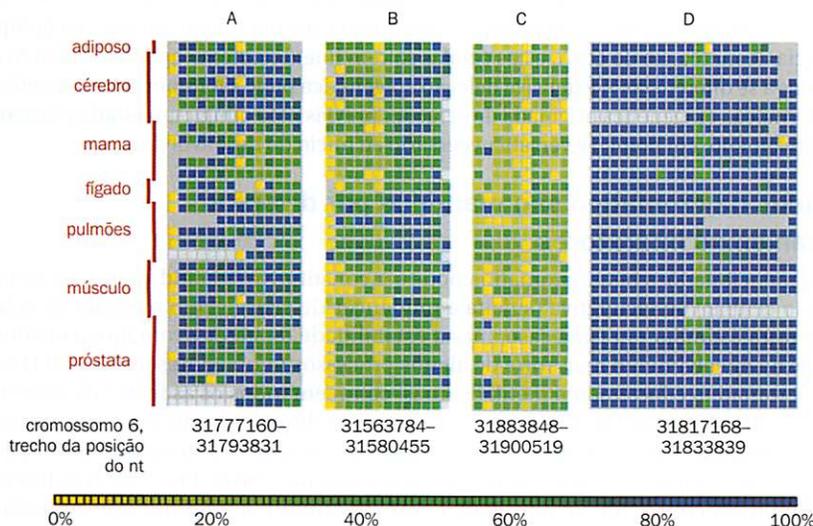
Enzima	Nº no OMIM	Funções principais <sup>a</sup>	Proteínas associadas
DNMT1	126375	Manutenção da metilase	PCNA (forquilha de replicação); histona-metiltransferases; histona-desacetilases; HP1 (heterocromatina); proteínas de ligação a DNA metilado
DNMT2 <sup>b</sup>	602478	Metilação da citosina 38 no RNAt <sup>Asp</sup> ; sem atividade de metilação do DNA	
DNMT3A	602769	Metilase <i>de novo</i>	histona-metiltransferases; histona-desacetilases
DNMT3B	602900	Metilase <i>de novo</i>	histona-metiltransferases; histona-desacetilases
DNMT3L	606588	Liga-se à cromatina com H3K4 não metilada e estimula a atividade de DNMT3A/3B	DNMT3A, DNMT3B; histona-desacetilases

PCNA, antígeno nuclear de proliferação celular, do inglês *proliferating cell nuclear antigen*; HP1, proteína de heterocromatina 1. <sup>a</sup>Para a distinção entre *de novo* e manutenção de metilases, ver Seção 11.3 na p. 365. <sup>b</sup>DNMT2 acabou por ter RNA mais do que DNA como seu substrato, mas sua estrutura é a de uma DNA-metiltransferase, e é incluída aqui porque é listada assim em muitas publicações.

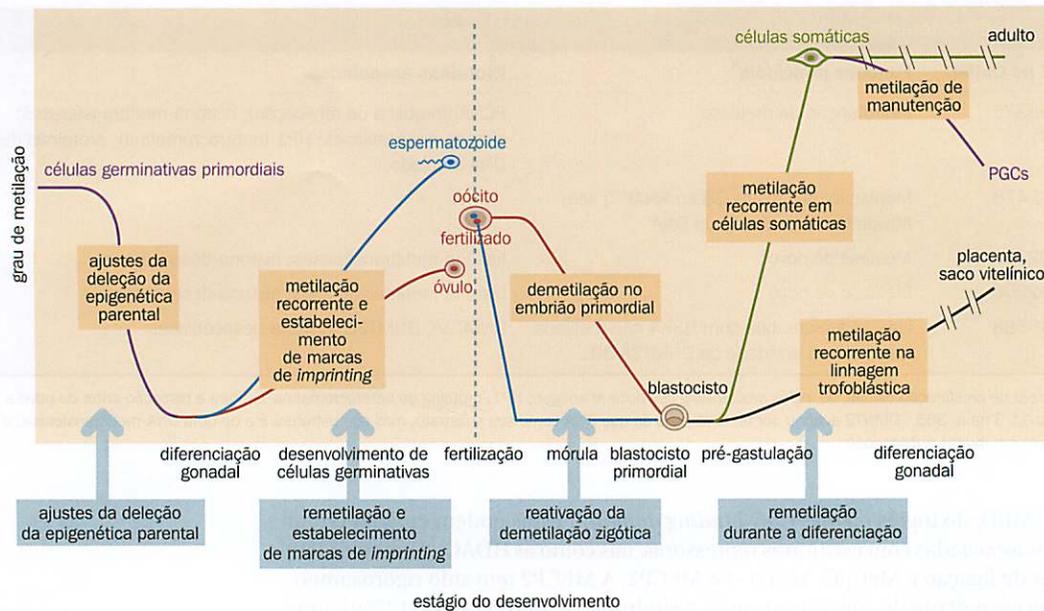
CpG metilada (MBD, do inglês *methyl-CpG-binding domain*). Estas podem então recrutar outras proteínas associadas com estruturas repressoras, tais como as HDACs. Humanos têm cinco proteínas de ligação à MeCpG, MBD1-4 e MECP2. A MECP2 tem sido rigorosamente estudada porque a perda de sua função causa a síndrome de Rett (OMIM 312750), uma condição singular em que meninas heterozigotas se desenvolvem normalmente nos seus primeiros anos, mas então regridem de uma maneira bastante característica. O gene que codifica MECP2 é ligado ao X, e mutações em *MECP2* são normalmente letais em machos. A metilação do DNA ocorre normalmente em pacientes com síndrome de Rett, mas como resultado da ausência da proteína MECP2 em células que têm o cromossomo X normal inativo, alguns sinais não são interpretados corretamente. A função de MECP2 é particularmente necessária em neurônios maduros. O perfil de expressão gênica no hipotálamo de camundongos que tanto carecem ou superexpressam MECP2 demonstraram que a expressão de muitos genes foi afetada, mas, inesperadamente, em 85% dos genes afetados, MECP2 aparentemente agiu mais para promover do que para reprimir a expressão.

### Metilação do DNA no desenvolvimento

A metilação do DNA apresenta mudanças impressionantes durante o desenvolvimento embrionário (Figura 11.12). Óvulos e, especialmente, espermatozoides têm o DNA fortemente metilado; os perfis de metilação de células germinativas são muito diferentes daqueles de qualquer outra linhagem somática. O genoma do oócito fertilizado é um agregado dos genomas do espermatozoide e do óvulo, com diferenças na metilação dos alelos paternos e maternos de muitos genes. Mesmo antes da fusão de dois genomas parentais, o pronúcleo paterno é sujeito a um processo de demetilação ativa do DNA por meio de uma atividade enzimática ainda não caracterizada residente no citoplasma do oócito. Uma vez



**Figura 11.11** Padrões de metilação CpG podem ser específicos para cada indivíduo, tecido ou locus. Cada quadrado representa uma sequência CpG dentro do complexo principal de histocompatibilidade no cromossomo 6, com os nucleotídeos intervenientes omitidos. As sequências são de quatro sub-regiões, como indicado. Os quadrados apresentam código de cores para indicar a extensão da metilação (ver escala na parte inferior). As linhas apresentam o padrão de metilação visto em diferentes amostras de tecidos. A região A apresenta diversos padrões de metilação tanto entre diferentes tecidos quanto entre diferentes amostras do mesmo tecido, enquanto a região B apresenta uma metilação mais tecido-específica. A região C é amplamente não metilada, e a região D é quase que completamente metilada. [Dados de [www.sanger.ac.uk/PostGenomics/epigenome](http://www.sanger.ac.uk/PostGenomics/epigenome), como retratado por Hermann A, Gowher H & Jeltsch A (2004) *Cell. Mol. Life Sci.* 61, 2571-2587. Com permissão da Springer.]



**Figura 11.12 Mudanças na metilação do DNA durante o desenvolvimento em mamíferos.** Mudanças drásticas e muitas vezes tecido-específicas na metilação geral acompanham a gametogênese e o desenvolvimento embrionário inicial. Seu papel causal permanece incerto, embora camundongos que são especificamente incapazes de metilar o DNA do espermatozoide sejam inférteis. Note as quebras (barras) nas linhas do desenvolvimento. PGCs, células germinativas primordiais, do inglês *primordial germ cells*.

que os dois pronúcleos tenham se fundindo, o genoma inicial do zigoto sofre uma demetilação genômica adicional até os estágios de mórula e início da blástula no embrião pré-implantado. Iniciando o estágio de blastocisto, e simultaneamente com eventos de diferenciação inicial, o genoma sofre novamente uma onda de metilação do DNA. A extensão desta metilação varia em diferentes linhagens celulares: linhagens de células somáticas são fortemente metiladas, enquanto linhagens derivadas do trofoblasto (dando origem à placenta, saco vitelínico, e assim por diante) são relativamente submetiladas. As células germinativas primordiais do embrião (das quais os gametas, em última análise, serão derivados - ver Capítulo 4) sofrem uma segunda onda de reprogramação genômica, com uma demetilação progressiva ocorrendo primeiro na entrada das células germinativas migratórias primordiais na crista genital, seguida por uma remetilação genômica assim que as células germinativas começam a se desenvolver.

Essas mudanças são muito drásticas, mas é discutido o quanto são causadoras dos eventos que elas acompanham, além de serem resultados a jusante dos mecanismos que causam diferenciação. A metilação do DNA é uma característica dos vertebrados, plantas com flores e alguns fungos, mas é praticamente ausente em muitos dos organismos-modelo mais bem estudados, incluindo leveduras, *C. elegans* e *Drosophila* - embora os mecanismos de diferenciação sejam fundamentalmente similares em todos estes organismos. A metilação do DNA é certamente requerida para o desenvolvimento em vertebrados, porque animais deficientes na atividade da metiltransferase de DNA morrem em vários estágios do desenvolvimento. Similarmente, embora células tronco embrionárias pluripotentes derivadas do estágio de blastocisto do embrião possam crescer normalmente, mesmo na completa ausência de metilação do DNA, estas células sofrem imediatamente apoptose quando estimuladas a se diferenciar. A extensiva deleção epigenética pode ser importante à redefinição da informação epigenética originada de gerações prévias para atingir um estado pluripotente relativamente em branco e versátil necessário ao início do desenvolvimento.

### Estados da cromatina são mantidos por diversos mecanismos iteradores

Como mencionado, a heterocromatina pode ser constitutiva (mantida como tal ao longo do ciclo celular) ou facultativa (sujeita a mudanças de acordo com o estado da célula). A estrutura é similar em cada caso, mas a estabilidade da heterocromatina constitutiva se deve ao envolvimento mais forte de alguns processos reforçadores mútuos. O DNA da heterocromatina é fortemente metilado, mas, como em todos os aspectos da conformação da cromatina, a metilação do DNA não é o único desencadeador para a formação da heterocromatina. Camundongos mutantes que são incapazes de metilar seu DNA ainda possuem uma heterocromatina centromérica claramente visível. Em contraste, humanos com mutações com perda de função de *DNMT3B* apresentam a síndrome autossômica

recessiva ICF (imunodeficiência, instabilidade centromérica, anomalias faciais, do inglês *Immunodeficiency, Centromeric instability, Facial anomalies*; OMIM 242860), na qual os satélites heterocromáticos pericentroméricos 2 e 3 dos cromossomos 1, 9 e 16 são demetilados e instáveis.

A metilação do DNA e a modificação de histonas interagem (Figura 11.13). DNA-metiltransferases e algumas proteínas que ligam o DNA metilado (proteínas de ligação à MeCpG) recrutam complexos repressores que contêm HDACs e/ou histona-metiltransferases. A H3K9 metilada pode ligar-se a proteínas com cromodomínios tais como a HP1 (ver a seguir), as quais podem, em contrapartida, recrutar as DNA-metiltransferases. Estes efeitos de reforço mútuo podem constituir a base de estados da cromatina repressivos de longo prazo. Além disso, a heterocromatina constitutiva, mas não a facultativa, é muito rica em DNA repetitivo. Na levedura *S. pombe*, isso invoca os mecanismos dependentes de RNA descritos a seguir, sendo que o mesmo pode ser verdadeiro em mamíferos.

### O papel da proteína HP1

A HP1 é uma pequena proteína não histônica, altamente conservada ao longo da evolução, que é geralmente encontrada na heterocromatina. Uma característica da heterocromatina é sua tendência a dispersar, e isto pode ser mediado por HP1. HP1 contém um cromodomínio que capacita sua ligação a histonas modificadas pela metilação de H3K9. Nucleossomos contendo H3K9me3 recrutam HP1, a qual então liga H3K9 metiltransferase. Esta pode metilar H3K9 no nucleossomo vizinho, levando à ligação de HP1 e então mediando a dispersão da metilação e HP1 ao longo de uma série de nucleossomos. O espalhamento é limitado por sítios, conhecidos como elementos limitadores ou insuladores, nos quais a cromatina é especialmente rica em histonas acetiladas e H3K4 metilada, as quais são marcadoras de transcrição ativa.

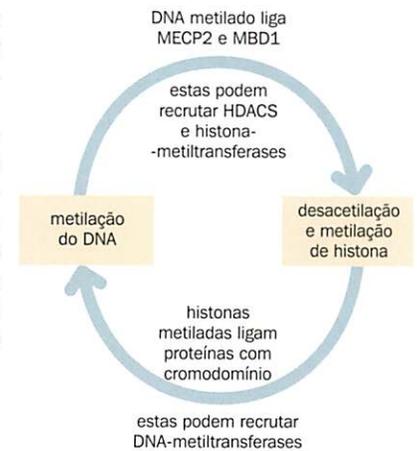
Mais comumente, HP1 age como uma plataforma na qual uma variedade de proteínas modificadoras da cromatina podem ser montadas. Essas incluem tanto HDACs como histona-metiltransferases. A multiplicidade dos possíveis parceiros provavelmente explica por que HP1 é também encontrada em regiões eucromáticas, onde ela normalmente reprime mas algumas vezes ativa a expressão gênica. Em *Neurospora* e talvez em outros exemplos, HP1 pode recrutar uma DNA-metiltransferase, fornecendo outro vínculo entre metilação de histonas e metilação de DNA.

### Um papel para pequenas moléculas de RNA

Um surpreendente achado na levedura *Schizosaccharomyces pombe* foi o de que quando genes envolvidos no mecanismo de interferência de RNA (descrito no Quadro 9.7) eram inativados, afetavam a montagem da heterocromatina. Regiões heterocromáticas são cobertas com componentes do complexo de silenciamento transcricional induzido por RNAi (RITS, do inglês *RNAi-induced transcriptional silencing complex*). Um componente deste complexo é uma proteína contendo um cromodomínio, o que pode explicar como os RITS poderiam ser direcionados para cromatina contendo H3K9 metilada. Paradoxalmente, silenciamento da heterocromatina requer a transcrição de elementos repetidos dentro da região, e estes transcritos são processados para pequenos RNAs de interferência (siRNAs, do inglês *small interfering RNAs*). Em alguns exemplos, os siRNAs podem desencadear a metilação de histonas, fornecendo um mecanismo de reforço mútuo, como a metilação de reforço mútuo das histonas e DNA descrito acima. Em mamíferos, pequenos RNAs gerados de sequências repetidas são similarmente envolvidos na nucleação da heterocromatina. Mecanismos com base em RNA introduzem a possibilidade de que a hibridização sequência-específica tenha um papel nas decisões sobre a conformação da cromatina – uma interessante área para especulação, dado que os promotores são pobremente definidos pela sequência local de DNA.

### Um papel para a localização nuclear

O núcleo interfásico celular é altamente estruturado. A síntese do RNA ribossômico é confinada ao nucléolo; a maior parte da transcrição ocorre em um número limitado de fábricas de transcrição, e o *splicing* do RNA ocorre principalmente em estruturas chamadas manchas nucleares. Cromossomos individuais ocupam territórios distintos que são característicos de um tipo celular particular (ver Figura 2.9). Genes ativos tendem a estar localizados no lado de fora do território do cromossomo e a situar-se em direção ao centro do núcleo. Genes inativos são mais prováveis de estar na periferia do núcleo. Anexar artificialmente uma região de cromatina à lâmina nuclear afeta a expressão dos genes que ela contém.



**Figura 11.13** Reforço mútuo das metilações de histona e DNA em cromatina inativa. A metilação do DNA atrai proteínas que modificam proteínas histônicas associadas. A desacetilação e a metilação de histonas atraem proteínas que metilam o DNA associado. Além dos efeitos apresentados aqui, algumas histonas-metiltransferases podem interagir diretamente com DNA-metiltransferases. Os pequenos RNAs de interferência endógenos podem ser importantes na nucleação da heterocromatina. HDACs, histona-desacetilases; MECP2, MBD1, proteínas de ligação à citosina metilada.

Muitos relatos documentam como sequências que são localizadas distante uma da outra em um cromossomo ou até mesmo em diferentes cromossomos, mas que estão envolvidas em interações regulatórias, podem se colocalizar dentro do núcleo – um fenômeno descrito como *gene kissing*, ou beijo gênico. A técnica de captura de conformação cromossômica (3C ou 4C) é utilizada para estudar tais efeitos (ver Quadro 11.3). Por exemplo, em células-tronco embrionárias em diferenciação de camundongos fêmeas, as duas cópias da sequência *Xic* (**centro de inativação do X**, do inglês *X-inactivation center*) localizam-se próximas uma da outra apenas antes de o processo de inativação do X iniciar, quando os cromossomos X estão sendo considerados, mas não em outros momentos.

### Sem causa única principal?

A metilação de DNA, modificação de histona e posicionamento de nucleossomo apresentam correlações claras com diferentes estados da cromatina, mas ainda não foi possível identificar um controle principal. A heterocromatina contém DNA metilado, nucleossomos fortemente empacotados e H3K9 dimetilada ou trimetilada. Sítios de início de transcrição contêm DNA não metilado e são marcadas por H3K4 monometilada, dimetilada ou trimetilada e acetilação de H3K9 e H4K5, 8, 12 e/ou 16. Adicionalmente, nucleossomos nesses sítios frequentemente contêm as histonas variantes H2A.Z e H3.3, as quais os tornam instáveis. Genes ativos são descritos por terem um gradiente de modificações 5'→3'. Próximo à extremidade 5', lisinas 4, 9 e 27 da H3 são todas monometiladas. Em trechos mais distantes dos genes ativos, H3K36 é trimetilada, e o DNA é altamente metilado com frequência. Em contrapartida, genes inativos têm trimetilação da H3K4 e H3K27. Conhecimentos atuais sugerem que os diferentes estados da cromatina são o resultado de vários processos diferentes, cada um contribuindo individualmente para apenas uma pequena parte do efeito. Ciclos de retroalimentação provocam combinações específicas de pequeno efeito que determinam o estado geral da cromatina.

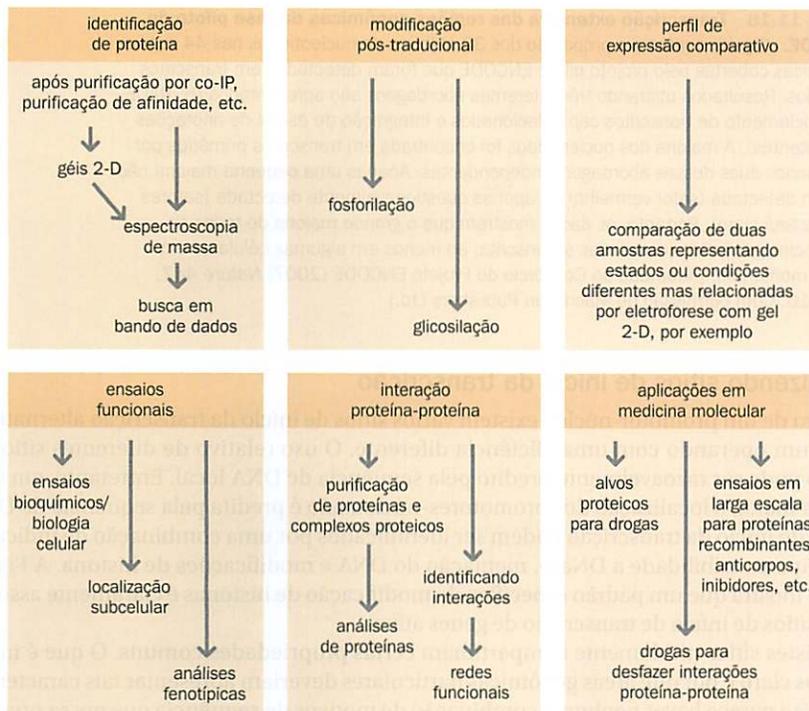
### O projeto ENCODE busca dar uma visão geral compreensiva da transcrição e de seu controle

Como discutido no Capítulo 10, em torno de 5% do DNA humano são conservados entre muitas espécies de mamíferos, mas sequências que codificam proteínas contabilizam apenas 20-25% das sequências conservadas. Boa parte do restante provavelmente tem função regulatória. O projeto ENCODE, mencionado no início deste capítulo, tem como último objetivo a definição de todos os elementos de sequência de DNA funcional no genoma humano. Na fase piloto (<http://genome.ucsc.edu/encode/pilot.html>), 44 regiões genômicas, totalizando 30 Mb ou 1% do genoma humano, foram escolhidas para o estudo. Metade das 44 regiões foi escolhida porque elas já estavam relativamente bem caracterizadas; isso poderia permitir que os resultados do projeto pudessem ser checados contra conhecimento preexistente. A outra metade foi randomicamente selecionada. Após o sucesso desta fase, o trabalho foi estendido para cobrir o genoma completo.

Assim como no Projeto Genoma Humano, este foi um grande projeto de colaboração envolvendo grupos de muitos países. Entretanto, a divisão do trabalho foi diferente. No Projeto Genoma, a cada grupo foi atribuída sua parte do genoma, e todos usaram a mesma tecnologia de sequenciamento capilar. No projeto ENCODE piloto, cada um dos 35 grupos colaboradores estudou todas as 44 regiões, mas os diversos grupos utilizaram uma ampla gama de técnicas diferentes. O objetivo global foi lançar cada técnica possível nas regiões-alvos e combinar os resultados. A **Figura 11.14** resume as principais abordagens e as questões que eles procuraram abordar, e a **Figura 11.15** apresenta um pequeno subconjunto do montante massivo de dados gerados.

### A transcrição é muito mais extensiva do que previamente imaginada

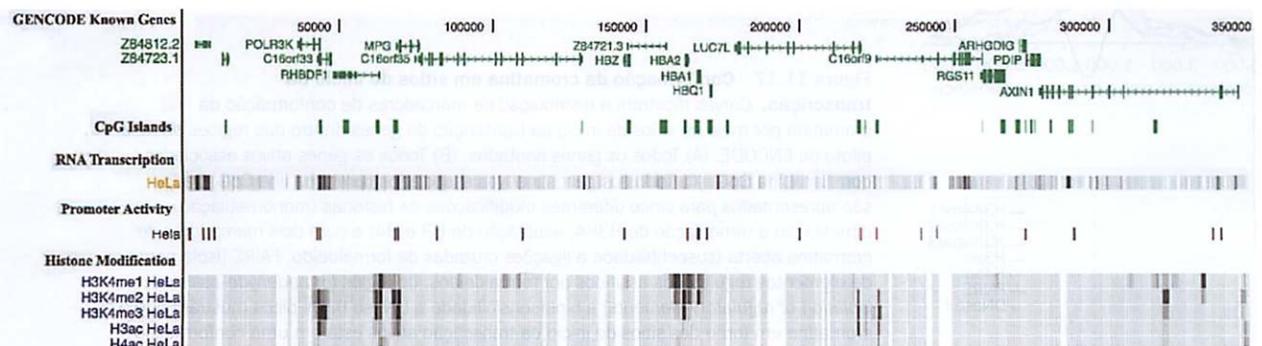
Uma grande descoberta do projeto ENCODE piloto foi que a maioria do DNA genômico nas 44 regiões era transcrita, ao menos em algumas células e em certos momentos. Frequentemente ambas as fitas eram transcritas. De todos os nucleotídeos das 44 regiões do ENCODE, 93% foram observados ao menos duas vezes em um transcrito primário, 74% deles por duas ou mais técnicas independentes (**Figura 11.16**). Isto está em forte contraste com a visão tradicional de unidades de transcrição isoladas relativamente pequenas distribuídas muito esparsamente pelo genoma e separadas por grandes regiões intergênicas não transcritas. Entretanto, os dados não foram completamente inesperados. Vários estudos recentes têm sugerido que muito mais do genoma é transcrito do que a visão tradicional poderia sugerir. Muito dos novos transcritos foram não codificantes, mas outros



**Figura 11.14 Principais facetas do projeto ENCODE.** Uma ampla variedade de técnicas é utilizada para tentar usar todos os elementos de seqüências funcionais do DNA no genoma humano. Isto envolve análise de transcriptoma intensivo, esforços combinados para identificar tanto elementos *cis*-regulatórios do DNA como suas respectivas proteínas *trans*-regulatórias e esforços para definir a estrutura da cromatina e algumas de suas regulações epigenéticas. co-IP, coimunoprecipitação; 2-D, bidimensional.

estenderam genes conhecidos pela inclusão de novos éxons a montante. Os novos éxons eram, em geral, 50 a 100 kb a montante do sítio de início da transcrição previamente documentado, sendo 20% localizados mais do que 200 kb a montante. Houve também evidência de transcritos contendo éxons de mais do que um gene. O que isso significa em termos de função celular ainda precisa ser determinado.

Na luz desses achados, é difícil manter a ideia tradicional de genes como trechos definidos de DNA que constituem unidades de transcrição claramente determinadas. Os resultados levantam questões importantes para pesquisas futuras. Os novos transcritos representam novos RNAs funcionais? E, se sim, quais são suas funções? Ou eles representam um requerimento para a transcrição de grande parte de DNA intergênico, mas não para o próprio transcrito? – como sugerido por estudos de alguns *loci* com *imprinting* (ver a próxima seção). Alternativamente, os dados significam que as células, assim como as pessoas que as estudaram, acham difícil identificar quais partes do genoma são funcionais e quais são lixo? O exame dos novos sítios de início da transcrição demonstrou que eles têm propriedades indistinguíveis de sítios de iniciação conhecidos, os quais sugerem que os novos transcritos não são apenas algum tipo de ruído no sistema. Parece provável que a visão tradicional do genoma esteja carecendo de uma revisão maior, mas as implicações disto estão longe de ser claras no presente (ver também Quadro 9.4).



**Figura 11.15 Um exemplo dos dados do ENCODE.** Este pequeno subconjunto dos dados globais apresenta genes, ilhas CpG, transcritos, promotores candidatos e modificação de histona em 350 kb de DNA genômico adjacente ao telômero do cromossomo 16p de células HeLa. O objetivo é identificar e mapear todos os fatores que, agindo em combinação, compõem o panorama da cromatina e governam a expressão gênica. [Adaptada do Consórcio do Projeto ENCODE (2007) *Nature* 447, 799-816. Com permissão de Macmillan Publishers Ltd.]

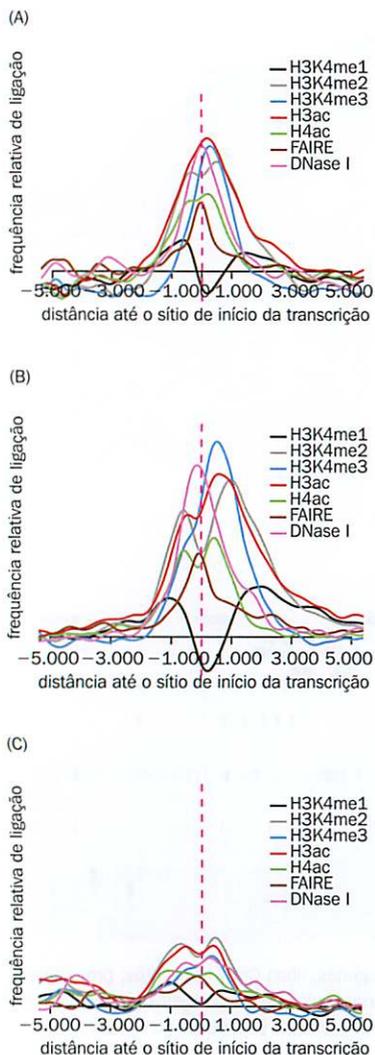


**Figura 11.16 Transcrição extensiva das regiões genômicas da fase piloto do ENCODE.** O gráfico mostra a proporção dos 30 milhões de nucleotídeos nas 44 regiões genômicas cobertas pelo projeto piloto ENCODE que foram detectadas em transcritos primários. Resultados utilizando três diferentes abordagens são apresentados (5' RACE, sequenciamento de transcritos *cap*-selecionados e integração de dados de anotações preexistentes). A maioria dos nucleotídeos foi encontrada em transcritos primários por pelo menos duas dessas abordagens independentes. Apenas uma pequena maioria não foi nem detectada (setor vermelho) ou apenas questionavelmente detectada (setores verde e azul claro). Portanto, os dados mostram que a grande maioria de todas as sequências genômicas estudadas é transcrita, ao menos em algumas células e em certos momentos. [Adaptada do Consórcio do Projeto ENCODE (2007) *Nature* 447, 799-816. Com permissão de Macmillan Publishers Ltd.]

**Predizendo sítios de início da transcrição**

Dentro de um promotor-núcleo existem vários sítios de início da transcrição alternativos, cada um operando com uma eficiência diferente. O uso relativo de diferentes sítios de início pode ser razoavelmente predito pela sequência de DNA local. Entretanto, em uma escala maior, a localização dos promotores-núcleo não é predita pela sequência de DNA. Sítios de início da transcrição podem ser identificados por uma combinação de indicadores: hipersensibilidade a DNase, metilação do DNA e modificações de histona. A **Figura 11.17** mostra que um padrão específico de modificação de histonas é claramente associado a sítios de início de transcrição de genes ativos.

Estes sítios nitidamente compartilham certas propriedades comuns. O que é muito menos claro é por que áreas genômicas particulares deveriam apresentar tais características. Não parece haver nenhuma combinação de motivos de sequência que possa prever se um trecho de DNA particular irá adotar tais características. A estrutura da cromatina é seguramente crucial, mas o que a determina? Experimentos que interferem com metiltransferases histônicas ou DNA-metiltransferases não causam a ruptura global catastrófica da transcrição que poderia ter sido predita se a metilação do DNA ou histonas fossem os determinantes primários da transcrição. Talvez os determinantes estejam com estes fatores, mas sua interação é muito complexa, e nenhum fator é importante. Ou talvez muito da cromatina não repetitiva possa potencialmente adotar esta configuração, e um conjunto muito heterogêneo de fatores desencadeie regiões diferentes a fazê-la. Alternativamente, talvez a resposta resida em ordens superiores da estrutura da cromatina, possivelmente no nível das alças do DNA e da matriz ou arcabouço nucleares. Certamente, o projeto ENCODE confirmou trabalhos prévios que demonstraram a existência de grandes (centenas de quilobases) domínios cromossômicos funcionais. Embora agora se pode descrever e manipular muitos elementos de controle gênico, com frequência permanece-se incapazes de explicar por que aquelas sequências têm adquirido tal função.



**11.3 MEMÓRIA EPIGENÉTICA E IMPRINTING**

Mudanças epigenéticas são mudanças duradouras na expressão gênica que não envolvem quaisquer mudanças na sequência de DNA. Elas podem permanecer por períodos variáveis - dentro de uma célula de vida longa, de uma célula para suas células-filhas durante

**Figura 11.17 Conformação da cromatina em sítios de início da transcrição.** Curvas mostram a distribuição de marcadores de conformação da cromatina por meio de sítios de início da transcrição de genes dentro das regiões da fase piloto do ENCODE. (A) Todos os genes. (B) Todos os genes ativos associados com uma ilha CpG. (C) Todos os genes inativos associados com uma ilha CpG. Dados são apresentados para cinco diferentes modificações de histonas (monometilação, dimetilação e trimetilação de H3K4, acetilação de H3 e H4) e para dois marcadores de cromatina aberta (susceptibilidade a ligações cruzadas de formaldeído, FAIRE [isolamento de elementos regulatórios assistidos por formaldeídos, do inglês *formaldehyde-assisted isolation of regulatory elements*] e hipersensibilidade a DNase I). Os picos mostram que a cromatina em torno dos sítios de início da transcrição ativos está em uma conformação aberta e possui uma combinação particular de modificação de histonas. As assinaturas compreendem H3K4me3 e H3K9/14ac localizadas logo a jusante do sítio de início da transcrição, com H3K4me2 um pouco mais a jusante. Há um pico mais fraco e difuso de H4 acetilada. [Adaptada do Consórcio do Projeto ENCODE (2007) *Nature* 447, 799-816. Com permissão de Macmillan Publishers Ltd.]

o desenvolvimento ou algumas vezes do parental para a descendência. Essas mudanças dependem de modificações na conformação da cromatina que, dependendo da escala de tempo, pode ser capaz de persistir por meio da replicação do DNA e mitose ou mesmo por meio da meiose. As mudanças epigenéticas herdáveis mais conhecidas são a inativação do X e o *imprinting*, os quais resultam em expressão gênica monoalélica – ou seja, apenas um dos dois alelos do *locus* é expresso. Alguns outros casos de expressão monoalélica são descritos no final desta seção.

As modificações epigenéticas responsáveis pela inativação do X e o *imprinting* são herdáveis da célula para suas células-filhas, mas não de pai para filho. Alguns efeitos epigenéticos mal compreendidos parecem também ser herdados de pai para filho. Estes são algumas vezes denominados **paramutações** e são discutidos a seguir. Algumas pessoas acreditam que exista um mundo inteiro de efeitos epigenéticos transgeracionais esperando para serem descobertos, embora muitos possam contestar esta visão.

### A memória epigenética depende da metilação do DNA e possivelmente dos grupos de proteínas policomb e tritorax

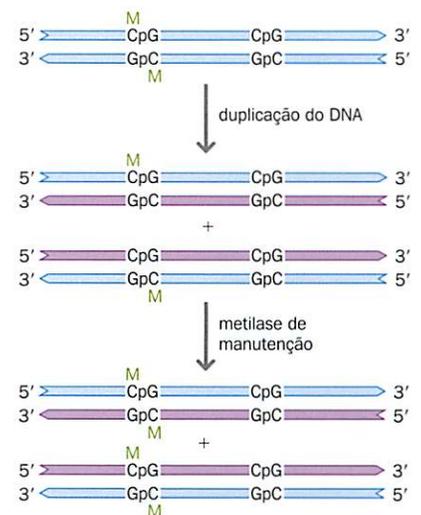
As DNA-metiltransferases DNMT3A e DNMT3B (ver Tabela 11.4) são metiltransferases recorrentes – isto é, elas podem metilar qualquer sequência CpG apropriada. DNMT1 é diferente. Ela reconhece especificamente sequências CpG hemimetiladas. Estas surgem principalmente quando o DNA metilado é replicado. A sequência CpG é simétrica – ou seja, 5'-CpG-3' de uma fita é pareada a 5'-CpG-3' de outra. A metilação do DNA é um fenômeno pós-replicativo, de modo que, quando o DNA é replicado, se uma fita inclui uma CpG metilada, sua nova fita complementar sintetizada terá também CpG, mas não metilada. A manutenção da metiltransferase DNMT1 converte tais sequências hemimetiladas em seus equivalentes completamente metilados (**Figura 11.18**). Esta manutenção da metilação é o principal (ou, ao menos, o mais bem compreendido) mecanismo de memória epigenética. Ela capacita os padrões de metilação CpG a se manterem intactos ao longo da replicação do DNA e serem transmitidos por meio da mitose para as células-filhas.

As células realizam decisões de desenvolvimento em resposta a sinais transientes, mas elas e suas células-filhas precisam recordar tais decisões muito depois dos sinais terem desaparecido. Os genes dos grupos policomb (PcG, do inglês *Polycomb Group*) e tritorax (TrxG, do inglês *Trithorax Group*) foram identificados por meio de mutantes de *Drosophila* que tomaram decisões de desenvolvimento corretas mas que parecem ter esquecido a decisão posteriormente. Humanos possuem ao menos dois complexos repressores de grupos policomb, PRC1 e PRC2. Estes metilam H3K27 em genes-alvo, os quais são predominantemente genes de desenvolvimento que codificam fatores de transcrição. O grupo de proteínas tritorax funciona em complexos nos quais os componentes coincidem com o conjunto de complexos de remodelamento da cromatina SWI/SNF (ver Tabela 11.3). Um exemplo notável é o MLL (linhagem misturada de leucemia, do inglês *Mixed-Lineage Leukemia*), um gene envolvido em vários rearranjos cromossômicos que criam genes quiméricos, os quais originam leucemias (ver Capítulo 17). Complexos do grupo tritorax auxiliam na ativação da transcrição metilando H3K4.

Estes sistemas agem para manter estados transcricionais que foram estabelecidos por outros mecanismos. Eles fornecem uma memória do estado transcricional, e ao menos alguns dos efeitos podem ser mantidos por meio da mitose. Entretanto, nota-se que efeitos de longa duração destes genes foram descritos principalmente em *Drosophila*. É importante manter em mente que *Drosophila* não usa metilação CpG em regulação gênica; pode ser que em mamíferos haja menos dependência das proteínas PcG e TrxG para regulação epigenética de longo prazo.

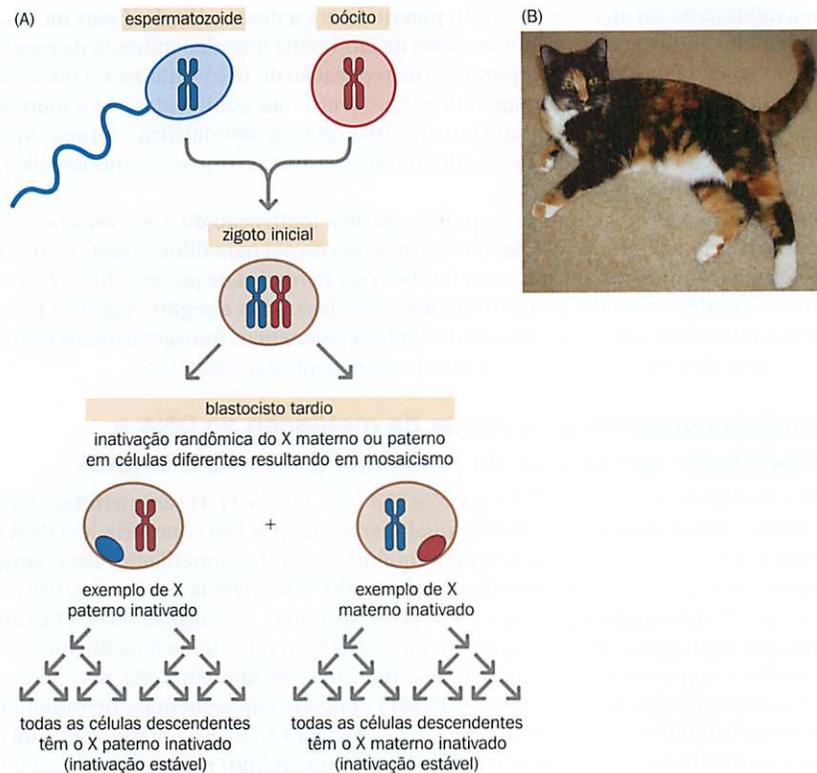
### Inativação do X: uma mudança epigenética herdável das células para suas células-filhas, mas não de pai para filho

No Capítulo 3, será descrito o fenômeno da inativação do X. Para recapitular brevemente: no princípio da embriogênese, células humanas de alguma maneira contam quantos cromossomos X elas contêm e então inativam permanentemente todos exceto um selecionado aleatoriamente. Nos estágios muito primordiais, ambos os cromossomos X são ativos, mas a inativação do X é iniciada logo que as células começam a se diferenciar de linhagens toti ou pluripotentes, o que ocorre no final do estágio de blástula em camundongos e muito provavelmente em humanos também. Cromossomos X inativos permanecem condensados durante todo o ciclo celular e podem ser vistos como o corpúsculo de



**Figura 11.18** Manutenção da metilação do DNA. Uma sequência 5'-CpG-3' de uma fita de DNA pareada com uma sequência 5'-CpG-3' de uma outra fita. Quando o DNA contendo uma sequência CpG metilada é replicado, a CpG na nova fita sintetizada é inicialmente não metilada. A metilase de DNA DNMT1 metila especificamente sequências CpG que estão pareadas com CpG metilada, preservando assim o padrão de metilação que estava presente no DNA antes da replicação.

**Figura 11.19 Inativação do X.** (A) Em embriões humanos, um cromossomo X escolhido é randomicamente inativado em cada célula de um embrião 46,XX, mas uma vez feita, a escolha é transmitida através de todos os ciclos de mitose. Na oogônia de uma fêmea, ambos os cromossomos X são ativos; cada um tem uma chance igual de ser transmitido pelo ovo. (B) O gato cálico (tartaruga e branco) é heterozigoto para cor de pelagem em um locus ligado ao X. Um alelo específica pelagem de cor preta, o outro, laranja. As diferentes manchas de cores refletem clones nos quais diferentes cromossomos X são inativados. As manchas brancas são resultado de um gene não relacionado para cor de pelagem. Gatos cálicos são sempre fêmeas, exceto para machos ocasionais XXY. [Adaptada de Migeon BR (1994) *Trends Genet.* 10, 230-235. Com permissão de Elsevier.]



Barr ou cromatina sexual na periferia do núcleo celular (ver Figura 3.8). Uma célula 46,XX pode inativar o cromossomo X materno ou paterno, mas seja qual for o escolhido para inativação, o mesmo é inativado em todas as células-filhas (Figura 11.19A). O corpo de uma fêmea adulta é, portanto, um mosaico de clones celulares, cada clone retendo o padrão de inativação do X que foi estabelecido em suas células progenitoras no princípio da vida embrionária. A Figura 11.19B mostra um exemplo impressionante do gato *tortoiseshell*, ou “gato tartaruga”. A inativação é estável através da mitose, mas não através de gerações. Um cromossomo X materno de uma mulher pode ter sido igualmente o ativo ou o inativo na sua mãe, e este tem a mesma chance de ser inativado quanto o cromossomo X paterno em suas próprias células.

### Iniciando a inativação do X: o papel do XIST

A inativação é iniciada no centro de inativação do X (*XIC*, do inglês *X-inactivation center*) no Xq13 e então se prolonga por todo o comprimento do cromossomo no que pode ser um exemplo extremo da tendência de a heterocromatina se espalhar. A dispersão depende da continuidade física do cromossomo e também de alguma característica especial desconhecida do DNA do cromossomo X. Como descrito no Capítulo 2, se o cromossomo X inativo é dividido em dois por uma translocação do X para um autossomo, a inativação é limitada ao segmento que inclui o *XIC*. Esta não pode pular para a porção destacada do cromossomo X. Ela pode se espalhar a partir de *XIC* de algum modo para a sequência autossômica vizinha, mas apenas abrangendo uma curta distância; ela não pode propagar-se ao longo de todo o segmento autossômico.

Como antes mencionado, o evento inicialmente observado é o de um pareamento transiente de duas sequências *XIC*. Este é provavelmente o mecanismo pelo qual os cromossomos X são contados, porque a contagem é interrompida se as sequências *XIC* são deletadas, replicadas ou translocadas. O *XIC* codifica um grande RNA não codificante, *XIST* (transcrito específico de inativação do X, do inglês *X-inactivation-specific transcript*), o qual é expresso apenas de um cromossomo X inativo. Seu transcrito primário sofre *splicing* e poliadenilação para gerar um RNA não codificante maduro de 17 kb. Acredita-se então que o RNA recruta alguns fatores de proteínas que organizam a cromatina em uma conformação fechada, transcricionalmente inativa. Em células diferenciadas que já tenham sofrido inativação do X, a perda do *XIST* não causa reativação. Assim, *XIST* é requerido para o estabelecimento da inativação do X, mas não a mantém.

Em camundongos, a expressão de *Xist* no cromossomo X ativo é silenciada por um transcrito antissenso (*Tsix*) que altera a conformação da cromatina no *locus Xist*, possivelmente por meio de um mecanismo de RNAi induzido por RNA dupla-fita *Xist-Tsix*. Humanos possuem um gene *TSIX*, mas ele tem pouca homologia estrutural ou de sequência ao *Tsix* murino e é muito menos claro se um mecanismo opera em humanos. Entretanto, em ambas as espécies o RNA *XIST/Xist* de certa maneira está envolvido na dispersão da heterocromatinização ao longo do cromossomo X que está para ser inativado. O RNA *XIST* vem para revestir todo o X inativo. A cromatina do X inativo carrega modificações típicas de heterocromatina. A H3K9 é dimetilada ou trimetilada, H3K4 não é metilada e H4 é desacetilada. H3K27 é também trimetilada, refletindo o envolvimento de genes do grupo policomb. As ilhas CpG nos promotores de genes inativados são metiladas. Além disso, muitos nucleossomos possuem a variante de histona H2A macro-H2A.

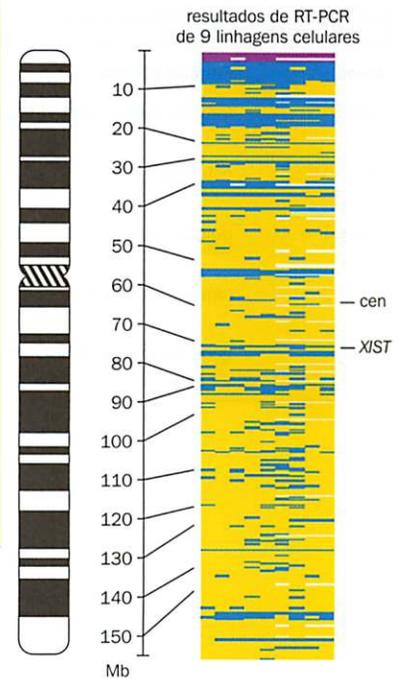
### Escapando da inativação do X

A inativação do X não é uma inativação trivial do cromossomo inteiro. As duas regiões pseudoautosômicas (ver Capítulo 2) escapam da inativação, mas mesmo fora destas regiões a inativação do X é desigual. Alguns exemplos específicos de genes que fogem da inativação foram mencionados no Capítulo 10 (ver p. 326). Alguns destes são genes que possuem um análogo funcional no cromossomo Y, para os quais não haveria necessidade de compensação de dosagem. Muitos, entretanto, não possuem tais análogos. Em um estudo, diferentes células híbridas contendo cópias independentes de um cromossomo X inativo humano foram utilizadas para investigar a transcrição de 612 genes ligados ao X: 458 dos genes estavam inativados em todas ou na maioria das linhagens celulares, mas 94 estavam expressos. Os 60 genes remanescentes apresentaram um padrão variável de expressão em diferentes linhagens celulares (Figura 11.20). Afora os requerimentos gerais de continuidade física, a dispersão da inativação do X é evidentemente capaz de saltar sobre estes genes excluídos e continuar seu progresso.

### Em loci que sofreram imprinting, a expressão depende da origem parental

No Capítulo 2, foram descritas evidências de que, para alguns genes, a origem parental importa. Zigotos com dois genomas maternos ou paternos desenvolvem-se anormalmente, como teratomas ovarianos ou verrugas hidatiformes, respectivamente. Para alguns segmentos cromossômicos, embora não para outros, a *dissomia uniparental* (UPD, do inglês *uniparental disomy*; ver Seção 2.5, p. 50) é patogênica. Sondagens sistemáticas foram feitas para identificar regiões cromossômicas que sofreram *imprinting* em camundongos. Ao contrário de humanos, todos os cromossomos de camundongos são acrocêntricos, então translocações robertsonianas (ver Figura 2.23) podem permitir cruzamentos que produzem proles com ambas as cópias de qualquer cromossomo completo escolhido derivado de um único progenitor especificado. Estas revelam que a UPD para alguns cromossomos não tem efeito fenotípico, enquanto para outros produz fenótipos anormais. Os fenótipos anormais são algumas vezes complementares para diferentes origens paternas; por exemplo, supercrescimento é geralmente visto em UPD paterna, e o retardo no crescimento, em UPD materna. Para alguns cromossomos, a UPD é letal. Estes efeitos são causados por UPD para pequenos números de genes específicos no cromossomo relevante, em vez de o cromossomo completo. Um trabalho adicional identificou em torno de 100 *loci* que sofreram *imprinting* em 11 dos cromossomos de camundongos. Os resultados estão resumidos em [www.har.mrc.ac.uk/research/genomic\\_imprinting.maps.html](http://www.har.mrc.ac.uk/research/genomic_imprinting.maps.html). Em humanos, uma investigação sistemática similar de *imprinting* não é possível, mas sabe-se agora que muitos dos genes de camundongos que sofreram *imprinting* também sofreram *imprinting* em humanos.

Para confirmar o *imprinting* de um gene, é necessário identificar um indivíduo o qual é heterozigoto para uma variante de sequência presente no mRNA maduro; mRNA de diferentes tecidos pode então ser conferido para expressão mono ou bialélica, e a origem de cada alelo pode ser determinada pela tipagem dos pais. Para alguns genes, este tipo de análise tem demonstrado que o *imprinting* é confinado a certos tecidos ou a certos estágios do desenvolvimento (Tabela 11.5). Em humanos, genes que sofreram *imprinting* foram identificados nos cromossomos 6q24-q26, 7q21-q22, 7q32 (e talvez 7p12), 8q24, 11p15, 14q32, 15q11-q13, 19q13 e 20q13. A Figura 3.18 apresentou exemplos de genealogias envolvendo mutações patogênicas em genes que sofreram *imprinting*.



**Figura 11.20** Genes que escapam da inativação do X. Colunas no retângulo apresentam resultados de testes de RT-PCR sistemáticos para expressão de genes ligados ao X e nove células somáticas híbridas independentes. Cada híbrido contém um único cromossomo X humano inativo. Barras azuis identificam genes que estavam expressos no cromossomo X inativo em um híbrido particular, e barras amarelas marcam genes que não estavam expressos. Muitos genes, dispersos ao longo do cromossomo X, escapam da inativação em um ou mais dos híbridos. Alguns genes, como o *XIST*, escapam da inativação em todos os nove híbridos, enquanto outros apresentam um padrão mais desigual. Cen, centrômero. [De Carrel L & Willard HF (2005) *Nature* 434, 400-404. Com permissão de Macmillan Publishers Ltd.]

**TABELA 11.5** Genes humanos apresentando *imprinting* tecido-específico ou estágio do desenvolvimento-específico

Gene	Função gênica	Alelo reprimido	Diferença nos padrões de expressão
<i>IGF2</i>	fator de crescimento do tipo insulina	materno	sofreu <i>imprinting</i> em muitos tecidos, mas foi bialelicamente expresso no cérebro, fígado adulto, condrócitos, etc.
<i>PEG1/MEST</i>	esterase	materno	sofreu <i>imprinting</i> em tecidos fetais, mas foi bialelicamente expresso no sangue adulto
<i>UBE3A</i>	ubiquitina ligase	paterno	sofreu <i>imprinting</i> exclusivamente no cérebro; foi expresso bialelicamente em outros tecidos
<i>KvLQT1</i>	canal de potássio	paterno	sofreu <i>imprinting</i> em vários tecidos, mas foi expresso bialelicamente no coração
<i>WT1</i>	gene tumoral de Wilms	paterno	sofreu <i>imprinting</i> frequentemente em células da placenta e adulto, mas foi expresso bialelicamente nos rins

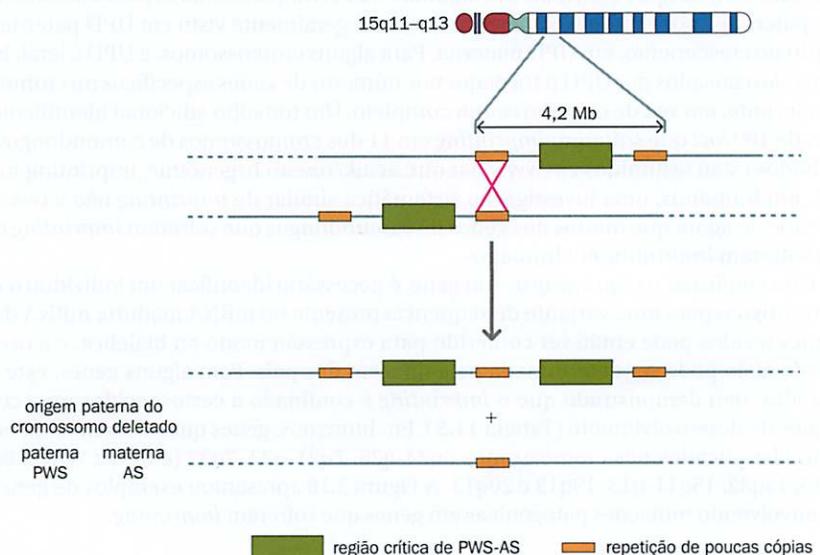
### Síndromes de Prader-Willi e Angelman são exemplos clássicos de *imprinting* em humanos

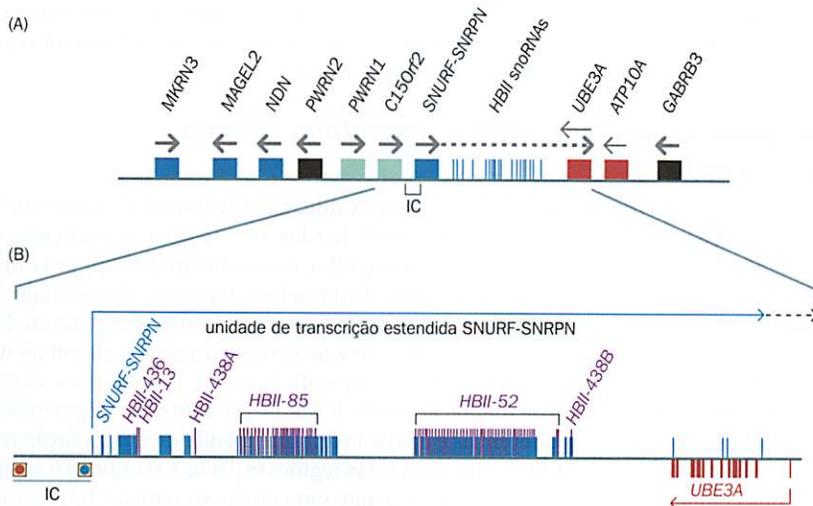
Um exemplo impressionante de efeitos do *imprinting* em humanos vem de duas raras condições congênicas, síndrome de Prader-Willi (PWS; OMIM 176270) e síndrome de Angelman (AS; OMIM 105830). Fenotipicamente, as duas síndromes são muito diferentes. Bebês com PWS são hipotônicos (flácidos), meninos apresentam hipogonadismo e crianças são mentalmente retardadas e possuem apetites incontroláveis, levando à obesidade grave, a menos que os pais sejam muito rigorosos. Crianças com AS parecem completamente diferentes: são pequenas, mais retardadas, carecem da fala e têm um padrão de comportamento característico. Fora essas diferenças, 70 a 80% das crianças com uma das duas síndromes têm uma microdeleção recorrente do 15q11-q13. É natural assumir que, em nível molecular, as duas deleções poderiam revelar o envolvimento de diferentes conjuntos de genes – mas, na verdade, elas são idênticas, mediadas por *recombinação homóloga não alélica* (ver Capítulo 13) entre sequências repetidas que flanqueiam a região deletada (Figura 11.21).

A chave para a diferença entre PWS e AS demonstra estar na origem parental do cromossomo que carrega a deleção. Se a deleção foi paterna na origem, a criança tem PWS, se materna, AS. Em alguns casos, de uma das duas condições que carecem da deleção característica houve uma dissomia uniparental – para o cromossomo 15 materno e para o homólogo em AS. Portanto, em todos os casos, PWS foi o resultado da falta de uma cópia paterna do 15q11-q13, enquanto AS foi o resultado da falta de uma cópia materna.

Essas síndromes direcionam o caminho da pesquisa de *imprinting* em humanos. A investigação intensiva do conteúdo genético e epigenético da região 15q11-q13 revelou uma história complexa notável. Um grupo de genes que sofreu *imprinting* oposto foi en-

**Figura 11.21** Uma microdeleção recorrente causa a síndrome de Prader-Willi ou a síndrome de Angelman. A recombinação homóloga não alélica entre repetições de poucas cópias geram deleções ou duplicações recorrentes de uma região de 4 Mb no cromossomo 15q11-q13. Pessoas heterozigotas para deleção apresentam duas condições muito diferentes, dependendo da origem parental do cromossomo deletado. Se o cromossomo paterno tem a deleção, o resultado é a síndrome de Prader-Willi (OMIM 176270; retardo mental, hipotonia e apetite incontrolável levando à obesidade). Se o cromossomo é o materno, o resultado é a síndrome de Angelman (OMIM 105830; microcefalia, retardo mental severo sem fala e movimentos atáxicos bruscos). Portadores da duplicação recíproca apresentam uma combinação de retardo mental e problemas de comportamento menos severos e menos bem definidos.





**Figura 11.22** O grupo de genes com *imprinting* no 15q11-q13 deletado na síndrome de Prader-Willi ou na síndrome de Angelman. (A) Genes no agrupamento com *imprinting*. Setas mostram a direção da transcrição. As azuis representam genes expressos paternalmente, azuis claras genes expressos predominantemente paternalmente, vermelhas genes expressos matematicamente e pretas genes expressos bialelicamente. IC marca um centro de controle de *imprinting*, uma sequência que é diferencialmente metilada em machos e fêmeas e que parece ter o controle geral da região de *imprinting*. (B) Parece haver sub-regiões separadas do IC que controlam o *imprinting* em machos e fêmeas. O notável transcrito originado no locus *SNURF-SNRPN* compreende ao menos 148 éxons e se estende por pelo menos 460 kb, se sobrepondo à unidade de transcrição *UBE3A* na fita de DNA oposta. Os éxons 1-10 codificam as proteínas *SNURF* e *SNRPN*; o restante é não codificante, mas introns contêm agrupamentos de genes (agrupamentos *HBII*, roxo) *snoRNA* (pequeno RNA nucleolar). A falta de cópias paternas destes é a causa da síndrome de Prader-Willi, enquanto a falta da cópia materna de *UBE3A* causa a síndrome de Angelman.

contrado (Figura 11.22). A síndrome de Angelman é causada pela falta de uma cópia materna funcional do gene *UBE3A*, o qual codifica uma ubiquitina ligase. Alguns pacientes com AS apresentam apenas mutações de ponto na cópia materna desse gene, sem anormalidades cromossômicas. Na maioria dos tecidos, *UBE3A* é expresso bialelicamente, sem *imprinting*, mas no cérebro apenas o alelo materno é expresso. Portanto, o *imprinting* pode ser tecido-específico.

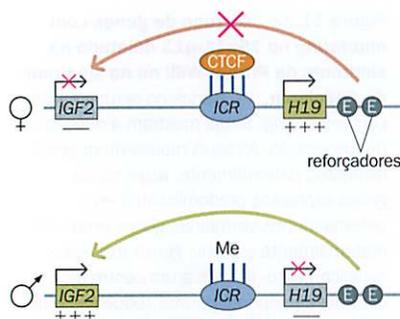
A síndrome de Prader-Willi foi por um longo período mais do que um mistério. Não foi encontrado paciente algum no qual a única anormalidade seja uma mutação de ponto em qualquer dos genes que codificam proteínas na região. A pesquisa com pacientes raros com pequenas deleções muito excepcionais tem mostrado que a causa de PWS é a falta do transcrito com *imprinting* *SNURF-SNRPN*, o qual é produzido apenas a partir do cromossomo paterno. Embora esse transcrito codifique as proteínas *SNURF* e *SNRPN*, o evento patogênico real é a deficiência de alguns RNAs nucleolares pequenos que estão codificados em íntrons do grande transcrito (ver Figura 11.22). Tal deficiência poderia afetar o processamento do RNA em múltiplos *loci* a jusante.

Dentro da região crítica de PWS-AS estão pequenas sequências diferencialmente metiladas nos cromossomos paterno e materno. Em pacientes ocasionais, o padrão de metilação não corresponde com a origem paterna do cromossomo. Estudos marcantes demonstraram que tais pacientes herdaram um cromossomo 15 de cada progenitor normalmente, mas ambos os cromossomos possuem o mesmo padrão de metilação, sendo materno em um paciente com PWS e paterno em um paciente com AS. Estudos de pacientes com esses defeitos de *imprinting* têm fornecido uma maior compreensão sobre o mecanismo de *imprinting*. Os defeitos de *imprinting* são causados por deleções de quaisquer duas pequenas sequências intergênicas localizadas próximas ao final da região 5' do gene *SNRPN*, a qual tem sido identificada como o **centro de controle de *imprinting*** (IC, do inglês *imprinting control center*).

A região PWS-AS apresenta características comuns a várias regiões que sofreram *imprinting*:

- Regiões que sofreram *imprinting* geralmente contêm agrupamentos gênicos, alguns que sofreram *imprinting* paternalmente e outros matematicamente. Pode também haver genes sem *imprinting* (bialelicamente expressos) no agrupamento. O *imprinting* de alguns genes pode ser tecido-específico.
- Há, muitas vezes, sobreposição e transcritos-senso e antissenso que sofreram *imprinting* opostamente, dos quais apenas um pode ser expresso a qualquer momento. Isto aponta para algum tipo de mecanismo de *flip-flop*, no qual a transcrição de uma fita de DNA previne a transcrição da fita oposta. O transcrito antissenso real frequentemente parece não ter função própria; o que importa é o evento da transcrição.
- Em geral, o *imprinting* é controlado por uma ou mais regiões pequenas que são diferencialmente metiladas nos cromossomos maternos e paternos.

Em apenas alguns casos, as regiões diferencialmente metiladas controlam a expressão gênica pela permissão ou pelo bloqueio do acesso de um promotor a um reforçador.



**Figura 11.23** Os genes *H19* e *IGF2* no cromossomo 11p competem por um reforçador. A ligação de uma proteína insuladora CTCF a regiões diferencialmente metiladas determina o resultado da competição. No cromossomo materno, a região controladora de *imprinting* (ICR, do inglês *Imprinting Control Region*) não é metilada, o que permite que ela se ligue a CTCF (elipse laranja). Isto previne o gene *IGF2* de acessar o reforçador, então o reforçador direciona a expressão de *H19*. No cromossomo paterno, a metilação da região controladora de *imprinting* previne a ligação de CTCF, permitindo assim que *IGF2* possa competir com *H19* pelo acesso ao reforçador. [Adaptada de Wallace JA & Felsenfeld G (2007) *Curr. Opin. Genet. Dev.* 17, 400-407. Com permissão de Elsevier.]

Essas regiões ligam-se à proteína insuladora CTCF quando não metiladas, mas não quando metiladas. A CTCF pode prevenir um promotor de ter acesso a um reforçador se, mas apenas se, ela estiver situada entre os dois (Figura 11.23).

### Duas questões surgem sobre o *imprinting*: como e por que é feito?

O fenômeno do *imprinting* levanta duas questões difíceis. A primeira é: como isto funciona? Ele precisa ser reversível: um homem pode herdar uma sequência particular com um *imprinting* materno, mas, se ele o passar a seu filho, este então irá possuir um *imprinting* paterno. O *imprint* consiste provavelmente de dinucleotídeos CpG diferencialmente metilados em centros de controle de *imprinting* nos genomas materno e paterno. Estes são impostos por metilação *de novo* sexo-específica recorrente durante a gametogênese. É desconhecido o motivo por que certas regiões específicas são escolhidas para sofrerem *imprinting*. Embora os padrões de metilação sejam bastante diferentes em espermatozoides e óvulos, a maioria dessas diferenças é apagada durante a onda de demetilação no zigoto inicial (ver Figura 11.12). Entretanto, para estas regiões especiais, o *imprint* é apagado apenas nas células germinativas primordiais, e não em células somáticas. Interessantemente, ao menos em camundongos, a DNA-metiltransferase *Dnmt1* usa promotores diferentes nos espermatozoides, oócitos e nas células somáticas, sendo que cada um contém uma isoforma diferente da enzima. Talvez isoformas gameta-específicas permitam este controle diferencial da metilação.

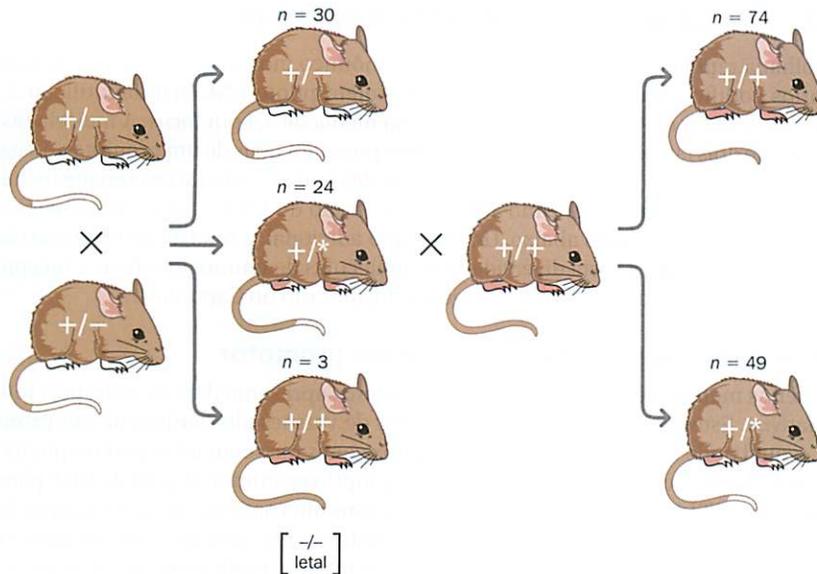
A segunda questão se refere à função do *imprinting*. Uma teoria conhecida é baseada no conflito de interesse evolucionário entre pais e mães. A teoria do gene egoísta sugere que genes paternos podem ser mais bem propagados pela garantia de que a prole nasça tão robusta quanto possível, mesmo às custas da mãe – um homem pode ser pai por meio de muitas mães diferentes. Em contraste, genes maternos são mais bem propagados se a mãe permanece capaz de engravidar posteriormente. Então, genes paternos programam o feto a extrair nutrientes à maior taxa possível da mãe por meio da placenta, enquanto genes maternos agem para limitar as depredações do feto parasita. Isto coincide com muitos *loci* que sofrem *imprinting*, mas é improvável que seja a explicação completa. Por que deveriam alguns *imprintings* ser tecido-específicos? E nem todos os genes com *imprinting* têm funções relacionadas a crescimento intrauterino ou sofrem *imprinting* na direção predita pela hipótese do conflito parental.

### Paramutações são mudanças epigenéticas transgeracionais

Em milho e outras plantas, há vários exemplos bem estudados de efeitos epigenéticos transgeracionais. Estes começam com uma planta heterozigota. Um alelo (o alelo paramutagênico) de alguma forma modifica o outro (o alelo paramutável), de modo que a prole que herda o alelo paramutável, todavia, expressa o fenótipo do alelo paramutagênico. O alelo paramutável torna-se paramutagênico na prole, e a modificação pode persistir através de muitas gerações, embora, dependendo do *locus*, ela possa no fim desaparecer. Onde os mecanismos foram identificados, envolvem modificação da cromatina e RNAs não codificantes.

Poucos casos similares foram relatados em camundongos. Talvez o mais claro envolva o *locus Kit* (Figura 11.24). Heterozigotos para o alelo com perda de função têm pintas brancas. Quando estes heterozigotos eram inter cruzados ou retrocruzados com os homozigotos selvagens, a proporção da prole homozigota selvagem apresentava as pintas brancas características dos heterozigotos, a despeito de seu genótipo selvagem. O efeito persistia por algumas gerações, mas com diminuição da intensidade. Apenas alguns alelos com perda de função apresentaram efeito. Uma mutação de ponto não tem efeito, mas dois mutantes nos quais o gene foi inativado pela inserção de diferentes cassetes em diferentes posições têm, sugerindo que algum grau de mau pareamento meiótico foi parte do mecanismo. Há evidências de que o RNA contido nos gametas estava envolvido – mesmo espermatozoides contêm quantidades significativas de RNA.

Quão ampla é a ação de tais mecanismos em mamíferos (incluindo humanos) é completamente desconhecida – embora o sucesso geral da genética mendeliana sugira que esses não podem ser os principais efeitos. Em geral, os efeitos das paramutações relatadas são mais quantitativos do que qualitativos (como um reostato mais do que um interruptor), o que os torna mais difíceis de ser identificados em humanos.



**Figura 11.24 Paramutações no locus *Kit* em camundongos.** Camundongos heterozigotos para uma mutação *Kit* (+/-) apresentam pintas brancas. Quando camundongos heterozigotos foram cruzados, a maioria (24 de 27) da prole genotipicamente selvagem (+/+) apresentou as pintas brancas características dos animais heterozigotos. O alelo selvagem herdado do progenitor heterozigoto foi modificado de alguma forma (paramutado, indicado por asterisco). A mudança é instável: quando camundongos +/\* foram cruzados com animais selvagens, menos de 50% da prole predita apresentou o fenótipo +/\* esperado. [Dados de Rassoulzadegan M, Grandjean V, Gounon P et al. (2006) *Nature* 441, 469-474. Com permissão de Macmillan Publishers Ltd.]

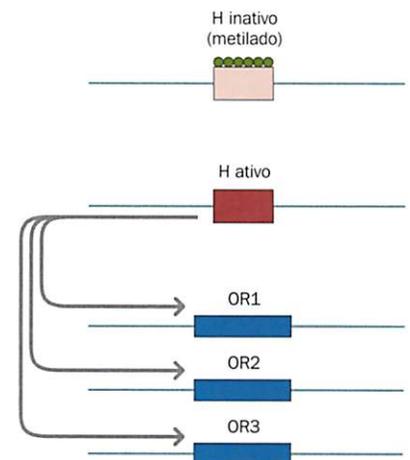
### Alguns genes são expressos em apenas um alelo, mas independentemente da origem parental

A inativação do X garante a expressão monoalélica da maioria dos genes ligados ao X. Entretanto, a expressão monoalélica, independentemente da origem parental, é também característica de uma surpreendente alta proporção de genes autossômicos (talvez chegando até mesmo a 5-10%). Três processos podem ser distinguíveis.

Primeiro, os genes da imunoglobulina e do receptor de célula T apresentam expressão monoalélica após rearranjos de DNA programados. Cada linfócito B e T expressa apenas um único alelo do gene relevante. Como descrito no Capítulo 4, os genes da imunoglobulina funcional e do receptor de célula T são montados a partir de grandes baterias de segmentos gênicos potenciais por meio de uma complicada série de rearranjos de DNA. Alguns elementos randômicos dentro do processo significam que a possibilidade de um rearranjo produtivo é bastante baixa; entretanto, apenas uma proteína funcional foi produzida, um mecanismo de retroalimentação inibe rearranjos de DNA posteriores.

A competição por um reforçador de cópia única fornece um segundo mecanismo. Os genes de receptores olfativos são um exemplo bem estudado (Figura 11.25). Como mencionado no Capítulo 9, genes de receptores olfativos são a maior família gênica em humanos e camundongos. Camundongos possuem 1.300 deles, e humanos, pelo menos 900. No entanto, cada neurônio olfativo expressa apenas um alelo único de um gene de receptor único; então, este dispara apenas em resposta a um odor específico. Em camundongos, e presumivelmente também em humanos, a expressão de qualquer gene de receptor olfativo depende absolutamente de uma sequência reforçadora (H) presente como uma cópia única em cada genoma. Experimentos de captura da conformação cromossômica demonstram que uma cópia única de H pode se associar com qualquer um dos 1.300 genes de receptores, independentemente da sua localização cromossômica. Embora um neurônio olfativo diploide contenha duas cópias de H, apenas um é ativo; o outro é inativado pela metilação (aparentemente mais em sequências CpA do que CpG).

Finalmente, a grande maioria dos casos é randômica. Em clones celulares independentes, um gene é monoalelicamente expresso em alguns clones, mas ambos os alelos são expressos em outros. Em culturas celulares, os efeitos são estáveis dentro de uma linhagem clonal. Para um dado gene, alguns clones podem apresentar expressão paternal e outros, expressão maternal. Em um indivíduo, não há consistência de origem parental entre diferentes genes expressos monoalelicamente, mesmo quando eles estão no mesmo cromossomo. A variação polimórfica no nível de expressão dos alelos é comum e herdável (expressão de *loci* de traço quantitativo ou e-QTLs, do inglês *expression quantitative trait loci*), então a expressão monoalélica pode ser o final da cauda de distribuição. A causa e o significado de tudo isto são desconhecidos, mas provavelmente e-QTLs e a expressão monoalélica randômica explicam muito da diversidade fenotípica de nossas espécies.



**Figura 11.25 Competição por um único reforçador leva à expressão monoalélica de receptores olfativos.** A expressão de qualquer um das centenas de genes de receptores olfativos (OR1, OR2, OR3, ...) distribuídos ao redor do genoma depende da associação com um reforçador de cópia única, H. Um dos dois alelos de H em uma célula diploide é inativado pela metilação (círculos verdes), então há uma única cópia ativa de H em cada célula.

## 11.4 UM GENE – MAIS DE UMA PROTEÍNA

Uma olhada rápida com qualquer navegador genômico mostra que a maioria dos genes humanos codifica mais de um transcrito. O projeto piloto ENCODE identificou 2.608 transcritos a partir de 487 *loci* examinados, uma média de 5,4 por *locus*. A maioria disto é resultado de dois fatores. Primeiro, muitos genes possuem mais de um promotor, e, segundo, *splicing* alternativo produz mRNAs nos quais alguns éxons são variavelmente incluídos ou negligenciados. Um mecanismo adicional, a edição do RNA, o qual opera apenas em certos genes, é mencionado abaixo. Um exemplo adicional, a notável série de rearranjos complexos que geram a vasta diversidade de moléculas de imunoglobulinas e receptores de células T de um pequeno número de genes, foi descrito no Capítulo 4.

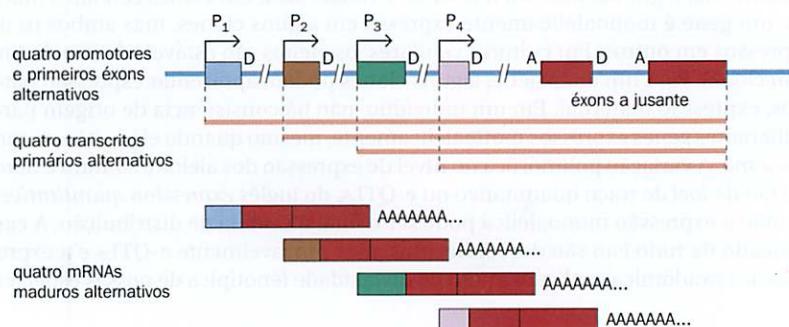
### Muitos genes apresentam mais de um promotor

Pelo menos metade de todos os genes de mamíferos apresenta dois ou mais promotores alternativos. Estes guiarão a transcrição a partir de versões alternativas de um primeiro éxon, o qual pode ou não ser codificante. Alguns genes possuem arranjos completos dos primeiros éxons alternativos, cada um com seu próprio promotor. O gene da UDP glicosiltransferase *UGT1A1* no 2q37 tem 13 primeiros éxons alternativos. No cromossomo 5q31, cada um dos genes da protocaderina  $\alpha$  e  $\gamma$  consiste em grandes arranjos em *tandem* de 2.400 bp de éxons alternativos iniciais que codificam uma parte substancial da proteína. Estes são unidos a três pequenos éxons invariantes que codificam a parte C-terminal da proteína. Podem ser feitas algumas analogias às imunoglobulinas, pois em ambos os casos ocorrem regiões N-terminal variáveis, mas C-terminal constantes, embora estas tenham sido geradas por meios completamente diferentes.

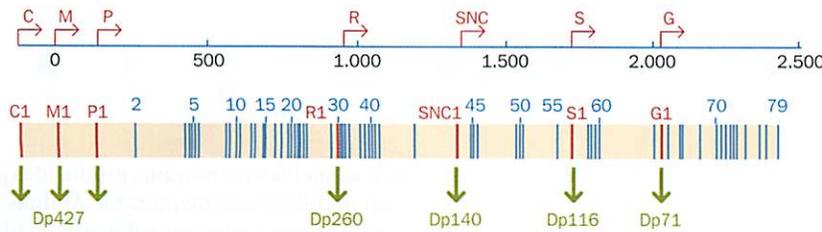
A extremidade 3' de cada primeiro éxon de um gene tem uma sequência doadora para *splicing*, mas a extremidade 5' não apresenta qualquer sinal aceptor para *splicing*, de modo que a maquinaria de *splicing* pode saltar, se necessário, sobre qualquer éxon alternativo inicial no transcrito primário. Assim, éxons alternativos iniciais podem ser unidos em cada caso a um conjunto comum de éxons a jusante (**Figura 11.26**).

O projeto piloto do ENCODE identificou éxons adicionais a montante em 273 dos 399 genes investigados, a maioria dos quais poderia ser alternativa aos primeiros éxons previamente anotados. Observa-se, entretanto, que em torno de metade destes já havia sido anotada como éxons de um diferente gene a montante, de modo que podem não ter sido produzidos a partir de promotores alternativos verdadeiros. Alguns promotores alternativos são internos a um gene, mais do que a montante. Em alguns casos, éxons a montante do promotor interno não são incluídos naquele transcrito. O gene da distrofina, descrito abaixo, possui diversos exemplos.

Os promotores alternativos podem servir a duas propostas. Primeiro, podem conter diferentes elementos regulatórios. Promotores alternativos tecido-específicos ou promotores do estágio do desenvolvimento-específicos podem permitir regulação diferencial da expressão gênica em diferentes circunstâncias. É, portanto, significativo que o primeiro íntron de um gene seja, geralmente, de longe o mais extenso (tamanho médio genômico de 14.186 pb, em contraste com 4.847 pb para íntrons internos). Portanto, promotores estão, com frequência, bem a montante dos éxons internos de um gene, e promotores alternativos são, em geral, bem separados uns dos outros. Os novos éxons identificados a montante pelo projeto ENCODE eram localizados em média a 186 kb a montante do éxon 5' previamente anotado. Isso permitiria aos promotores ficarem em um ambiente diferente de cromatina, talvez em uma alça cromossômica diferente uma da outra, e do centro do gene.



**Figura 11.26 Um gene com vários promotores alternativos.** O gene tem quatro promotores alternativos (P<sub>1</sub>-P<sub>4</sub>) e primeiros éxons (retângulos coloridos), mais dois éxons a jusante (retângulos vermelhos). As posições dos sítios do doador (D) e aceptor (A) de *splicing* são apresentadas. Cada junção de *splicing* deve ser feita unindo um sítio doador e um aceptor. O mRNA maduro possui uma cauda de poli-A.



**Figura 11.27 Promotores alternativos para o gene da distrofina.** As posições dos sete promotores alternativos são apresentadas no topo da figura. C, cortical; M, músculo; P, células de Purkinje; R, retina, cérebro e células musculares cardíacas; SNC, sistema nervoso central; S, células de Schwann; G, geral (quase ubiqüamente expresso, mas não detectável no músculo esquelético completamente diferenciado). Cada promotor utiliza seu próprio primeiro éxon, mostrado em vermelho no diagrama da estrutura gênica, junto com os éxons compartilhados a jusante, apresentados em azul. Cada produto proteico é nomeado pelo seu peso molecular; Dp427 é a proteína da distrofina de 427 kD, por exemplo.

Segundo, o uso de primeiros éxons alternativos pode resultar em isoformas de proteínas com diferentes propriedades. Elas podem ter diferentes localizações subcelulares (isoformas solúveis e ligadas à membrana, por exemplo) ou diferentes funções. Por exemplo, o gene do receptor de progesterona *PGR* no cromossomo 11q22 utiliza promotores alternativos para produzir duas isoformas, PRA e PRB, que diferem em 165 aminoácidos N-terminais. Ambas isoformas são fatores de transcrição, membros da família de receptores nucleares descritos no Capítulo 4, mas têm como alvo diferentes genes de resposta e apresentam diferentes efeitos fisiológicos. Um exemplo extremo é o gene *CDKN2A*, mostrado na **Figura 17.16**. Neste caso, dois primeiros éxons alternativos contêm um sítio de início de tradução cada um. Dependendo de qual seja utilizado, a sequência do éxon a jusante compartilhada é traduzida em diferentes fases de leitura. Assim, os mesmos éxons a jusante codificam proteínas completamente diferentes, dependendo de qual promotor seja utilizado.

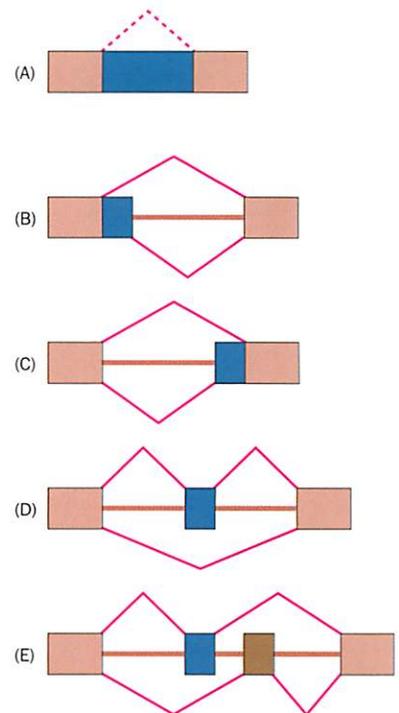
Um exemplo impressionante de uso diferencial de promotor em humanos refere-se ao gene gigante da distrofina. Este contém 79 éxons distribuídos sobre 2,4 Mb de DNA em Xp21. Pelo menos sete promotores alternativos podem ser utilizados. Três destes são localizados próximos ao sítio de início convencional. Em adição ao promotor músculo-específico, há um promotor cerebral córtex-específico localizado 100 kb a montante e um promotor que é utilizado nas células de Purkinje do cerebelo 100 kb a jusante (**Figura 11.27**). Transcritos utilizando estes promotores codificam grandes isoformas da proteína da distrofina (Dp, do inglês *distrophyn protein*) com um peso molecular de 427 kD. As três isoformas Dp427 diferem em suas sequências de aminoácidos N-terminal porque os três primeiros éxons alternativos incluem sequências codificantes. Entretanto, além destes três promotores, pelo menos quatro outros promotores internos alternativos podem ser utilizados, gerando transcritos muito menores que não possuem muitos dos éxons das isoformas Dp427.

**O splicing alternativo permite que um transcrito primário codifique múltiplas isoformas de proteínas**

Transcritos que sofreram *splicing* alternativo (isoformas de *splicing*) podem ser identificados por quase todos os genes humanos. Estes podem saltar um ou mais éxons, incluir éxons internos adicionais ou variar o comprimento de um éxon pelo reposicionamento da junção éxon-íntron (**Figura 11.28**). Pode ser que muitas dessas isoformas apenas reflitam a imprecisão de uma complexa maquinaria de *splicing* e não sejam funcionalmente significantes - mas numerosos exemplos são conhecidos nos quais o *splicing* alternativo é claramente funcional.

A básica maquinaria de *splicing* foi descrita no Capítulo 1. Entretanto, nem todos os sítios de *splicing* são iguais. A escolha de onde o *spliceossomo* é montado no transcrito primário depende de um balanço de fatores positivos e negativos. A sequência que circunda as sequências invariantes GT...AG pode ser um melhor ou pior ajuste à sequência. Próximo a isso, reforçadores de *splicing* ligam proteínas SR (serina-arginina) que ajudam a ancorar o *spliceossomo* no local, enquanto supressores de *splicing* ligam proteínas hnRNP (ribonucleoproteínas heterogêneas), que apresentam um efeito oposto. Portanto, sítios de *splicing* podem ser fortes ou fracos. Sítios fracos podem ser saltados algumas vezes em favor de um alternativo. Tecidos podem diferir no seu repertório de proteínas SR ou hnRNP, ou eles podem conter proteínas que ligam reforçadores ou supressores de *splicing* tecido-específicos, de modo que padrões de *splicing* são frequentemente tecido-específicos.

O *splicing* alternativo funcional pode ter uma variedade de resultados. Éxons que sofreram *splicing* alternativo podem introduzir códons de parada, mudanças no quadro de leitura ou sítios para modificações pós-traducionais importantes, tais como a fosforilação de serinas. Eles podem codificar sinais governando diferentes localizações intracelulares.



**Figura 11.28 Tipos de eventos de splicing alternativo.** (A) Um índon (azul) pode ser tanto retido como excluído. (B) O uso de sítios doadores de *splicing* alternativo resulta na inclusão ou exclusão da sequência em azul. (C) O uso de sítios aceptores de *splicing* alternativo resulta na inclusão ou exclusão da sequência em azul. (D) O éxon em azul pode ser tanto excluído como saltado (um éxon cassete). (E) Éxons alternativos: o mRNA maduro inclui o éxon em marrom ou o éxon em azul, mas nunca ambos ou nenhum. Retângulos em rosa representam éxons que são sempre incluídos no mRNA maduro.

**Figura 11.29 Splicing alternativo do transcrito da neurexina-3 no sistema nervoso.** Existem dois promotores alternativos,  $\alpha$  e  $\beta$  (barras vermelhas). Os éxons 3, 4, 5, 12 e 20 (azuis) podem tanto ser incluídos como saltados. O éxon 7 (verde claro) pode ser incluído, usando qualquer um dos dois sítios aceptores de *splicing* 5' alternativos, ou ele pode ser completamente saltado. O éxon 22 (roxo) tem dois sítios doadores de *splicing* 3' alternativos. O éxon 23 (rosa) tem dois sítios aceptores de *splicing* 5' alternativos que utilizam diferentes fases de leitura, uma das quais leva a um códon de parada em fase dentro deste éxon. A proteína produzida desta variante carece dos domínios transmembrana e citoplasmático codificados pelo éxon 24. O éxon 24 (verde escuro) apresenta três sítios aceptores de *splicing* 5' alternativos. Utilizando diferentes combinações de variantes, esse simples gene poderia potencialmente codificar em torno de mil diferentes proteínas.



Algumas proteínas apresentam isoformas solúveis e ligadas à membrana, produzidas pela inclusão ou omissão de um éxon que codifica um domínio transmembrana. As duas formas podem competir. A competição por um ligante entre as isoformas solúveis e ligadas à membrana pode regular receptores da superfície celular, por exemplo. O *splicing* pode ser tecido-específico, de modo que diferentes tecidos contêm variantes diferentes.

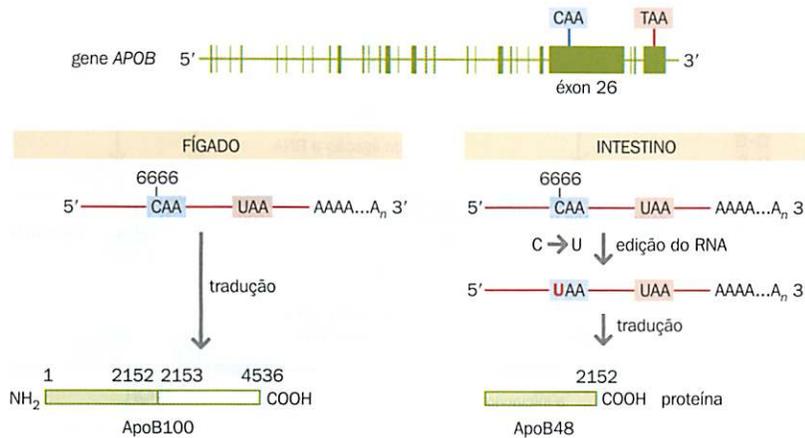
Como parte mais complexa do corpo humano, o sistema nervoso central pode necessitar um proteoma mais complexo, sendo que o *splicing* alternativo é particularmente evidenciado em neurônios. Muitos genes globalmente expressos têm isoformas de *splicing* neurônio-específicas. Já em 1994 uma compilação feita por Stamm e colaboradores listou quase 100 exemplos de *splicing* neurônio-específicos, incluindo todos os tipos mostrados na **Figura 11.28**. Muitas das isoformas variantes de canais iônicos e receptores são conhecidas por serem funcionalmente importantes. Alguns genes codificam um número extraordinário de diferentes transcritos. O exemplo das protocaderinas  $\alpha$  e  $\gamma$ , com suas baterias de promotores alternativos, foi mencionado anteriormente. A **Figura 11.29** apresenta o gene da neurexina-3 (*NRXN3*). Este grande gene no cromossomo 14q codifica uma molécula receptora de adesão celular que está presente nas sinapses do sistema nervoso. Ela possui dois promotores e 24 éxons a jusante. Sete dos éxons podem ser incluídos ou excluídos individualmente nos transcritos; um éxon tem sítios doadores de *splicing* alternativo, e dois outros têm aceptores de *splicing* alternativo. Um dos éxons com *splicing* alternativo inclui um códon de parada, sendo que o uso desse poderia produzir uma proteína sem domínios transmembrana e citoplasmático. Potencialmente este único gene poderia codificar mil proteínas diferentes.

Além da geração de diferentes isoformas de proteínas, o *splicing* alternativo pode também resultar em diferentes sequências 5' ou 3' não traduzidas. Estas também podem ser funcionalmente significativas, devido especialmente à região 3' não traduzida (UTR), que é o principal sítio por meio do qual microRNAs afetam a expressão gênica, como descrito posteriormente neste capítulo. O uso de sinais de poliadenilação alternativos é também bastante comum em RNA humano, e diferentes tipos de poliadenilação alternativos têm sido identificados. Em vários genes, dois ou mais sinais de poliadenilação são encontrados na 3' UTR, e os transcritos poliadenilados alternativamente podem apresentar especificidade tecidual. Em outros casos, sinais de poliadenilação alternativa podem entrar em cena como uma consequência do *splicing* alternativo.

### A edição do RNA pode modificar a sequência do mRNA após a transcrição

Na contravenção do Dogma Central, existem exemplos nos quais a sequência de DNA de um gene não determina completamente a sequência de seu transcrito. A edição do RNA envolve a inserção, a deleção ou a modificação de nucleotídeos específicos no transcrito primário. Os nucleotídeos podem ser modificados por deaminação ou transaminação enzimática para efetuar conversões C→U, A→I ou U→C (I é inosina; ver a seguir). Isto ocorre em grande escala na mitocôndria e cloroplastos de plantas vasculares. Em mamíferos, não há evidência para inserção ou deleção na edição do RNA, mas modificações dos nucleotídeos foram observadas em um limitado número de genes.

- **Edição C→U** ocorre no gene da apolipoproteína humana *APOB*. No fígado, este gene codifica a grande proteína ApoB100. No intestino, entretanto, a edição do C→U na posição nucleotídica 6666 do mRNA provoca a troca do códon da glutamina CAA por um códon de parada UAA. O mRNA agora codifica um polipeptídeo mais curto, ApoB48 (**Figura 11.30**). A edição do C→U produz similarmente um códon de parada no gene *NFI* (neurofibromina).
- **Edição A→I** é realizada por membros da família de deaminases ADAR (adenosina deaminase agindo no RNA) (**Figura 11.31**). Pares de base de inosina com citosina mais do que timina. A edição de ADAR frequentemente converte códons CAG, codificando glutamina em códons CIG, os quais, como CGG, codificam arginina. (Edição Q/R.) No sistema nervoso, a edição Q/R ocorre em diversos genes que codificam receptores de neurotransmissores ou canais iônicos (*GABRA3*, *GRIA2* e *GRIK2*). No



**Figura 11.30 Edição do RNA: os dois produtos do gene *APOB*.** No fígado, o mRNA de *APOB* codifica um produto proteico de 14,1 kb e 4.536 resíduos, a ApoB100. Entretanto, no intestino, uma citosina de aminase, APOBEC1, converte especificamente a citosina 6666 à uridina, mudando o códon de glutamina CAA da posição 2153 em um códon de parada UAA. O mRNA codifica agora um produto, ApoB48, consistindo apenas dos primeiros 2.152 aminoácidos de ApoB100.

gene do receptor de serotonina *HTR2C*, a edição A→I em sítios de *splicing* modula o *splicing* alternativo.

- **Edição U→C** foi observado no transcrito do éxon 6 do gene do tumor de Wilms.

A edição do RNA possui um modelo pouco inteligente - mas evidentemente satisfaz uma necessidade importante, porque embriões de camundongo, mesmo sendo heterozigotos para o nocaute de *Adar1*, morrem durante a gestação. Curiosamente, humanos heterozigotos para mutações de perda de função no gene ortólogo, *DSRAD*, apresentam apenas uma condição na pele relativamente trivial, discromatose simétrica hereditária (OMIM 127400).

## 11.5 CONTROLE DA EXPRESSÃO GÊNICA NO NÍVEL DE TRADUÇÃO

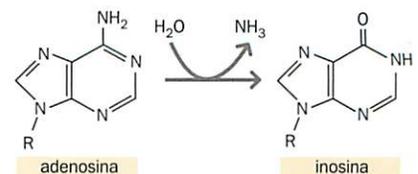
Tradicionalmente, controles no nível transcricional têm sido o principal foco de trabalho para compreender a regulação gênica - mas efeitos no nível de tradução há muito são conhecidos e têm recebido recentemente crescente atenção com a descoberta de pequenos RNAs de interferência (siRNAs) e microRNAs (miRNAs) (ver Capítulo 9, p. 283). O controle traducional da expressão gênica pode permitir uma resposta mais rápida a alterações de estímulos ambientais do que a alterações no nível da transcrição.

### Controles adicionais governam quando e onde um mRNA é traduzido

Os mRNAs recém-sintetizados não emergem simplesmente de um poro nuclear e se conectam ao ribossomo mais próximo. Eles podem ser transportados para locais específicos dentro da célula, traduzidos imediatamente ou mantidos em reserva para uso futuro.

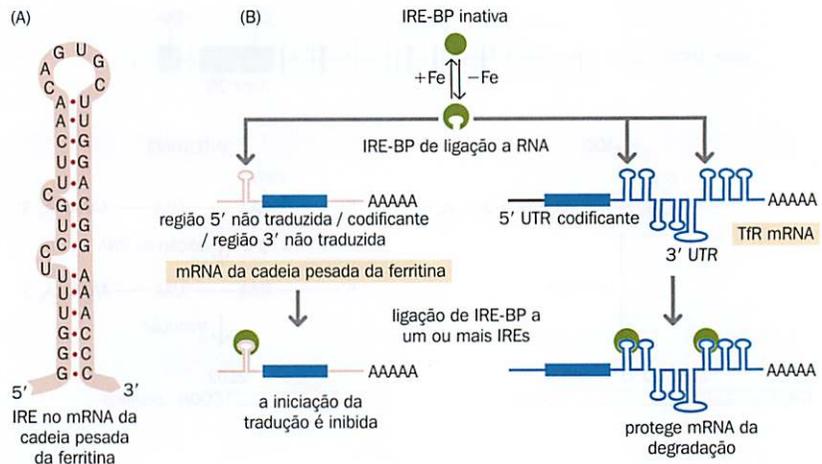
Os mRNAs pode ser transportados como partículas de ribonucleoproteínas a locais particulares dentro de alguns tipos celulares. Em neurônios, por exemplo, o mRNA de tau é localizado na região proximal dos axônios mais do que nos dendritos, onde muitas moléculas de mRNA estão localizadas. O mRNA da proteína básica da mielina é transportado com a ajuda de cinesinas aos processos dos oligodendrócitos. O transporte de RNAs, em vez de produtos proteicos, pode fornecer uma forma mais efetiva de localização de proteínas, porque uma única molécula de mRNA pode dar origem a muitas moléculas proteicas. Proteínas de ligação ao RNA específicas, tais como o grupo ELAV/Hu, reconhecem e ligam sequências ricas em AU na região 3' não traduzida de alguns mRNAs. Os mRNAs relacionados funcionalmente podem estar ligados pelo mesmo fator, o qual pode coordenar sua localização e/ou expressão.

Nem todas as moléculas de mRNA são utilizadas imediatamente para síntese de proteínas. Extrapolações de estudos de organismos-modelo sugerem que uma variedade de mRNAs é estocada em oócitos em uma forma inativa, caracterizada por apresentar uma cauda curta de oligo(a). A cauda curta significa que eles não podem ser traduzidos. Na fertilização ou, depois, no desenvolvimento, os mRNAs inativos estocados podem ser ativados pela poliadenilação citoplasmática, a qual restabelece o tamanho normal da cauda de poli(a).



**Figura 11.31 Deaminação da adenosina.** Enzimas da família ADAR deaminam o grupo amino do carbono 6 da adenosina para produzir inosina.

**Figura 11.32 Elementos de resposta ao ferro nos mRNAs da ferritina e no receptor de transferrina.** (A) Estrutura de alça em pedúnculo (*stem-loop*) de um elemento de resposta ao ferro (IRE) na região 5' não traduzida do mRNA da cadeia pesada da ferritina. (B) Na presença de ferro, uma proteína específica de ligação a IRE (IRE-BP) é ativada, capacitando-a a ligar-se no IRE no gene da cadeia pesada da ferritina e também a IREs na região 3' não traduzida do mRNA do receptor de transferrina. A ligação inibe a tradução de ferritina, mas protege o mRNA do receptor de transferrina da degradação.



Fatores externos podem também regular a tradução de mRNAs, como visto em dois exemplos do metabolismo de ferro. O aumento dos níveis de ferro estimulam a síntese da proteína de ligação ao ferro, a ferritina, sem qualquer aumento correspondente na quantidade de mRNA de ferritina. De forma recíproca, a diminuição dos níveis de ferro estimula a produção do receptor de transferrina (TfR, do inglês *transferrin receptor*), parte da maquinaria de absorção da dieta de ferro, novamente sem algum efeito na produção de mRNA. A extremidade 5' não traduzida de ambas as cadeias leve e pesada do mRNA da ferritina contém um **elemento de resposta ao ferro (IRE, do inglês *iron-response element*)**. Esta é uma sequência regulatória agindo em *cis* que forma uma estrutura em grampo. Várias sequências do tipo IRE são também encontradas na região 3' não traduzida do mRNA do receptor de transferrina. A regulação é efetuada por uma proteína de ligação a IRE específica que é ativada em baixos níveis férricos para ligar IREs (Figura 11.32). Como muitos controles tradicionais, os efeitos dependem da estrutura secundária dos mRNAs.

Até poucos anos atrás, esses exemplos esparsos de controles baseados em RNA na expressão gênica eram considerados interessantes, mas raros. Aquela visão foi modificada com a descoberta do RNAi e dos miRNAs.

### A descoberta de muitos pequenos RNAs que regulam a expressão gênica causou uma mudança de paradigma na biologia celular

Pequenos mRNAs não são raros. Na verdade, alguns o são, mas, por exemplo, uma célula de *C. elegans* adulto contém mais de 50 mil moléculas de miR-2, miR-52 e miR-58 cada. Por que elas não foram descobertas anteriormente? Quando pesquisadores corriam géis de RNA, eles assumiam – geralmente com boas razões – que o borrão de RNA de muito baixo peso molecular na base do gel consistia de produtos de degradação e não era de interesse. Além disso, tais pequenas moléculas são difíceis de estudar: abordagens padrão de bioinformática as omitiam, e tais minúsculos alvos eram raramente encontrados em experimentos de mutagênese. Dessa forma, embora eles possam ter sido observados em muitos experimentos de laboratório, não eram reconhecidos.

Duas linhas de pesquisa aparentemente diferentes nos anos 90, ambas envolvendo o verme *C. elegans*, alertaram biólogos para a insuspeita importância dos RNAs muito pequenos. Andrew Fire e Craig Mello receberam o Prêmio Nobel em Fisiologia ou Medicina em 2006 por seus papéis no entendimento do RNAi – a inibição específica da expressão gênica por pequenas moléculas de RNA de dupla-fita. Suas palestras do Nobel, descrevendo os processos de descoberta, podem ser lidas em [http://nobelprize.org/nobel\\_prizes/medicine/laureates/2006/](http://nobelprize.org/nobel_prizes/medicine/laureates/2006/). Ainda, Victor Ambros e Gary Ruvkun, entre outros, abriram o mundo dos miRNAs em desenvolvimento. Narrativas de como eles fizeram essas descobertas podem ser lidas em dois comentários da *Cell* (ver Leituras adicionais).

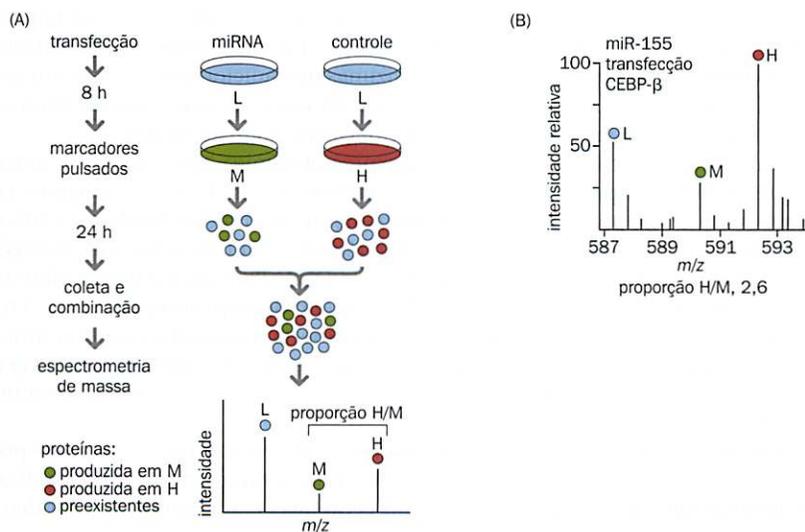
### MicroRNAs como reguladores da tradução

Os mecanismos básicos por meio dos quais pequenos RNAs exercem seus efeitos foram descritos no Capítulo 9 (ver Quadro 9.7 e Figuras 9.16 – 9.18). Conforme anteriormente descrito, vários tipos diferentes de pequenos RNAs são importantes na regulação gênica. Os siRNAs sintéticos são ferramentas extremamente importantes para nocautear seleti-

vamente a expressão de genes no laboratório (ver Capítulos 12 e 20), enquanto endógenos siRNAs são provavelmente envolvidos no estabelecimento da heterocromatina, como descrito acima. Os RNAs que interagem com a proteína Piwi (piRNAs) foram inicialmente descritos em células germinativas, nas quais sua principal função parece ser a de prevenir que os transposons se tornem ativos. MicroRNAs possuem papéis muito mais amplos na regulação gênica em todos os tipos celulares e são considerados em mais detalhes aqui. A maioria, embora nem todos, de seus efeitos relatados estão em nível traducional.

A maioria das sequências alvo reconhecidas por miRNAs e siRNAs está na região 3'UTR de um mRNA. O mRNA-alvo pode ser degradado ou pode permanecer intacto mas não ser traduzido. Quais destes destinos são escolhidos depende do grau de concordância entre as duas espécies de RNA. Pequenos RNAs de interferência possuem pareamento perfeito e degradam seus alvos. MicroRNAs têm pareamentos imperfeitos e normalmente reprimem a tradução, embora eles possam também desencadear a degradação de seu alvo. Em ambos os casos, os nucleotídeos de 2-7 funcionam como uma semente para reconhecimento inicial do alvo. Se o restante do pequeno RNA pode então realizar um pareamento perfeito, o alvo é clivado entre os nucleotídeos pareados com os resíduos 10-11 do pequeno RNA. Se não, o mRNA alvo pode permanecer, intacto, mas a tradução é reprimida. O mecanismo preciso da repressão não é certo. Ele pode funcionar por meio do reconhecimento do *cap* do mRNA ou por um efeito na montagem do ribossomo. Uma proteína, EIF6, que se complexa com a subunidade ribossomal 60S e inibe a montagem do ribossomo completo, foi descrita como um componente de uma RISC modificada. De uma forma ou de outra, miRNAs normalmente operam prevenindo a iniciação da tradução. Quando células são transfectadas com um miRNA, ocorre também uma redução na quantidade do mRNA alvo na célula. Ainda não é certo se esta perda se dá porque o mRNA degrada seu alvo bem como inibe a tradução ou se isto é secundário à repressão da tradução. Vários estudos apoiam a visão de que a perda é secundária, e não um efeito direto do miRNA.

O pareamento imperfeito do miRNA com seu alvo torna mais difícil identificar os alvos por análises de bioinformática. Os resultados dos experimentos em laboratório para identificação de alvos são surpreendentes. Quando a espectrometria de massa foi utilizada para examinar os efeitos no proteoma, a transfecção ou o nocaute de um único miRNA puderam ser vistos afetando a síntese de centenas, ou mesmo milhares, de proteínas (Figura 11.33). Alguma proporção destes efeitos será secundária: o miRNA reprime a produção e reduz a disponibilidade de proteínas, porém afeta a produção de outras proteínas. Entretanto, muitas mudanças foram correlacionadas com a presença da sequência semente para o miRNA relevante na região 3'UTR do mRNA, sendo diretas. Outra observação notável destes experimentos é de quão modestos os efeitos eram. Poucas proteínas apresentaram um aumento maior do que duas vezes quando um miRNA foi nocauteado ou uma repressão maior do que 50% quando a célula foi transfectada com um miRNA. O cenário geral que emerge é de efeitos muito amplos, porém pequenos.



**Figura 11.33 Análise proteômica do efeito da mudança de nível de um miRNA.** (A) Um esquema do procedimento experimental. Células HeLa foram transfectadas com um miRNA em um meio contendo um marcador isotópico leve (L); após 8 horas, elas foram transferidas para um meio contendo um marcador isotópico de peso médio (M). Células HeLa-controle foram transferidas para um meio contendo um marcador isotópico pesado (P). Novas proteínas sintetizadas continham um marcador M ou P. Extratos são combinados e analisados por espectrometria de massa. (B) Um resultado exemplificado: a transfecção com miRNA-155 resultou em uma diminuição de 2,6 vezes na síntese da proteína CEBP- $\beta$  [compare as alturas dos picos de P (controle) e M (experimental)]. [De Selbach M, Schwanhäusser B, Thierfelder N et al. (2008) *Nature* 455, 58-63. Com permissão de Macmillan Publishers Ltd.]

## MicroRNAs e câncer

Como será discutido no Capítulo 17, células cancerígenas apresentam uma expressão modificada de muitos miRNAs. Existe uma depressão geral dos níveis de miRNA, consistente com o estado geralmente desdiferenciado de células cancerígenas (a expressão de miRNA é maior em células diferenciadas do que em vários tipos de células-tronco). Alguns destes efeitos podem ser reproduzidos pelo nocauteamento direcionado das enzimas Dicer ou Drosha, essenciais para o processamento de todos os miRNA a partir de seus transcritos primários (ver Figura 9.16). Quando isto é feito, as células apresentam comportamento tumorigênico pronunciado. Entretanto, certos miRNAs apresentam padrão oposto, sendo super-regulados em muitos tumores. A evidência descrita acima sugere que estas mudanças poderiam afetar a expressão de milhares de genes, assim como podem ser significativas no balanço de um comportamento celular em direção a tumorigênese.

## Algumas questões malresolvidas

MicroRNAs exercem seus efeitos primariamente por meio de interações com moléculas de mRNA dentro do citoplasma. Um número de relatos sugeriu que miRNAs podem também ter funções no núcleo, talvez afetando a transcrição diretamente. Isto permanece altamente especulativo atualmente. Existem mais do que sólidas evidências de que miRNAs podem algumas vezes estimular, mais do que inibir, a tradução. Alguns dos mesmos miRNAs que inibem a tradução durante o crescimento celular normal podem tornar-se reguladores positivos em condições de estresse celular.

## CONCLUSÃO

A expressão gênica é controlada por decisões em todos os níveis: na transcrição ou não de uma sequência, na escolha do sítio de transcrição, no processamento de um transcrito primário que sofreu *splicing* alternativo, em como, quando e onde traduzir a mensagem no mRNA maduro, na modificação e no transporte – dentro ou fora da célula – do polipeptídeo inicial.

Genes codificadores de proteínas e muitos genes que codificam RNAs funcionais são transcritos pela RNA-polimerase II. Antes que a transcrição possa começar, um grande complexo de proteínas precisa ser montado em torno do sítio de início da transcrição. Além da RNA-polimerase, estas incluem os fatores de transcrição TFII (um complexo de pelo menos 27 polipeptídeos individuais), os quais compreendem o aparato basal de transcrição, mais muitas proteínas tecido-específicas. Algumas dessas proteínas reconhecem elementos de sequência de DNA específicos próximos ao sítio de início da transcrição, tais como o TATA *box*. Outras unem o complexo por meio de interações proteína-proteína. Estas incluem proteínas ligadas a elementos regulatórios distantes que são postos em íntima proximidade ao promotor pela formação de alças de DNA.

Os sítios de início da transcrição são definidos mais pela conformação local da cromatina do que pela sequência de DNA. A conformação depende de um conjunto de processos reforçados mutuamente: histonas no centro dos nucleossomos sofrem numerosas modificações covalentes ou são algumas vezes substituídas por moléculas variantes; citosinas em dinucleotídeos CpG podem ser metiladas, e complexos remodeladores da cromatina dependentes de ATP alteram a distribuição dos nucleossomos. As diversas proteínas não histônicas na cromatina incluem enzimas que adicionam ou removem grupos modificados para ou a partir de histonas de DNA. Eles são, portanto, ligados diferencialmente a cromatina, dependendo de sua conformação e estado de modificação.

A conformação da cromatina controla a expressão gênica, tanto em curto prazo, de modo que as células respondam a condições metabólicas e ambientais flutuantes, como em longo prazo, como programas de desenvolvimento estendidos. Mudanças duradouras na conformação são a base dos efeitos epigenéticos – mudanças na expressão gênica que não dependem de mudanças na sequência de DNA. Para alguns genes, efeitos epigenéticos dependem da origem parental – o fenômeno do *imprinting* genético. A transcrição necessita uma conformação da cromatina aberta, enquanto a heterocromatina possui uma conformação fechada e altamente empacotada e reprime a expressão gênica. Pequenas moléculas não codificantes de RNA estão envolvidas no estabelecimento da heterocromatina estável.

Transcritos primários são muitas vezes sujeitos a *splicing* alternativo. Os éxons podem ser saltados ou incluídos, ou sítios doadores e receptores variáveis podem ser utilizados. Alguns genes podem potencialmente codificar centenas de diferentes polipeptídeos em

razão do extenso *splicing* alternativo. O *splicing* é controlado por reforçadores e supressores que ligam proteínas como as SR e hnRNP. O *splicing* alternativo é geralmente tecido-específico. Em muitos casos, as isoformas decorrentes de *splicing* diferentes apresentam claramente funções distintas, embora alguns *splicings* alternativos possam refletir ineficiências randômicas na seleção de sítios de *splicing*.

Os controles no nível da tradução são a regulação refinada da expressão gênica. Estes são menos compreendidos que os controles transcricionais; para a maior parte, eles parecem estar envolvidos no controle da estabilidade de mRNAs. As regiões 3' não traduzidas de muitos mRNAs contêm importantes determinantes de estabilidade, incluindo a maioria dos sítios de ligação a miRNAs. MicroRNAs apresentam efeito amplo mas não muito profundo no metabolismo celular – eles funcionam como reostatos mais do que como interruptores. Entretanto, a importância dos reostatos não pode ser subestimada. Uma das lições das pesquisas de doenças complexas é que os efeitos cumulativos de tais mudanças menores podem ser muito importantes. Isto pode ser especialmente verdadeiro para os efeitos baseados em miRNA, porque eles afetam muitos aspectos amplos do metabolismo celular. No entanto, isto ainda pode subestimar a importância dos controles baseados em RNA. Pequenos RNAs podem apresentar papéis substanciais em decisões majoritárias através de sua função na modificação da estrutura da cromatina.

Uma visão de redes altamente acopladas está emergindo do metabolismo celular mais do que um conjunto de vias lineares. Em sua revisão, Maniatis e Reed (ver Leituras adicionais) citam uma série de evidências que a transcrição, o processamento do RNA e a exportação do mRNA maduro são altamente acoplados. Mais do que uma linha de produção linear, a máquina de expressão gênica é uma rede em que os componentes cujas principais funções estão nos estágios finais do processo também influenciam a eficiência dos estágios anteriores. No Capítulo 12, transitará da expressão para a função gênica, e ficará claro que a maioria dos produtos gênicos funciona como parte de redes muito extensas de proteínas interativas. A emergente ciência da biologia de sistemas estuda tais redes e está ajudando a preencher a lacuna entre a expressão gênica, descrita neste capítulo, e a biologia celular.

## LEITURAS ADICIONAIS

### Promotores e os transcritos primários

- Fuda NJ, Ardehali B & Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461, 186–192. [Uma excelente revisão do controle de início da transcrição, com bons exemplos (mas não em humanos).]
- Kleinjan DA & van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* 76, 8–32. [Fornece detalhes a respeito de *PAX6*, *SHH* e outros exemplos clinicamente importantes.]
- Lonard DM & O'Malley BW (2007) Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. *Mol. Cell* 27, 691–700. [Exemplos de regulação por redes de coativadores e correpresores.]
- Maniatis T & Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* 416, 499–506. [Uma ótima abordagem sobre o controle da expressão gênica.]
- Sandelin A, Carnici P, Lenhard B et al. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* 8, 424–436. [Estrutura do promotor e sítios de início da transcrição; quadros com informações úteis sobre as técnicas.]
- Shiraki T, Kondo S, Katayama S et al. (1999) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Methods Enzymol.* 303, 19–44. [Fornece detalhes a respeito da técnica CAGE descrita no Quadro 11.1.]

- Smale ST & Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449–479. [Resume a maior parte dos trabalhos a respeito dos sítios de início de genes.]
- Visel A, Rubin EM & Pennachio LA (2009) Genomic views of distant-acting enhancers. *Nature* 461, 199–205. [Revisão sobre a natureza e função dos acentuadores, com ênfase em exemplos humanos.]

### Conformação da cromatina: metilação do DNA e o código de histonas

- Barski A, Cuddapah S, Cui K et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. [Mapeamento do genoma completo de metilações de histona em promotores, acentuadores, insuladores e regiões transcritas.]
- ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. [Descobertas do projeto piloto; os detalhes encontram-se em uma série de artigos publicados na revista *Genome Research*.]
- Koch CM, Andrews RM, Flicek P et al. (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* 17, 691–707. [Detalhes do projeto piloto do ENCODE.]
- Lall S (2007) Primers on chromatin. *Nat. Struct. Mol. Biol.* 14, 1110–1115. [Uma boa introdução geral.]

- Ooi L & Wood IC (2007) Chromatin cross-talk in development and disease: lessons from REST. *Nat. Rev. Genet.* 8, 544–554. [Função do complexo CoREST.]
- Wang Z, Zang C, Rosenfeld JA et al. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903. [Um estudo de genoma completo da distribuição de mais de 30 modificações de histona que identificam combinações específicas características de promotores, acentuadores, etc.]

## Heterocromatina

- Bühler M & Moazed D (2007) Transcription and RNAi in heterochromatic gene silencing. *Nat. Struct. Mol. Biol.* 14, 1041–1048. [Envolvimento de pequenos RNAs na formação da heterocromatina.]
- Gaszner M & Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* 7, 703–713. [Uma revisão geral sobre insuladores.]
- Hediger F & Gasser SM (2006) Heterochromatin protein 1: don't judge the book by its cover! *Curr. Opin. Genet. Dev.* 16, 143–150. [Funções da HP1.]
- Trojer P & Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13. [Características de vários tipos de heterocromatina facultativa.]

## Metilação do DNA

- de la Serna IL, Ohkawa Y & Imbalzano AN (2006) Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat. Rev. Genet.* 7, 461–473.
- Eckhardt F, Lewin J, Cortese R et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38, 1378–1385. [Um exemplo de mapeamento em larga escala de assinaturas epigenéticas.]
- Goll ... [Revisão de metiltransferases de DNA.]
- Goll MG & Bestor TH (2005) Eukaryotic DNA cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. [Review of the DNA methyltransferases.]

## Efeitos da localização nuclear sobre a expressão gênica

- Akhtar A & Gasser SM (2007) The nuclear envelope and transcriptional control. *Nat. Rev. Genet.* 8, 507–517.
- Dekker J, Rippe K, Dekker M & Kleckner N (2002) Capturing chromatin conformation. *Science* 295, 1306–1311. [O trabalho 3C original.]
- Fraser P & Bickmore W (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* 447, 413–417.
- Lanctôt C, Cheutin T, Cremer M et al. (2008) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115.
- Linnemann AK, Platts AE & Krawetz SA (2009) Differential nuclear scaffold/matrix attachment marks expressed genes. *Hum. Mol. Genet.* 18, 645–654. [Salienta a dificuldade em definir as regiões de ligação entre matriz/arcaouço; uma boa fonte de referências.]

## Memória epigenética e imprinting

- Cuzin F, Grandjean V & Rassoulzadegan M (2008) Inherited variation at the epigenetic level: paramutation from the plant to the mouse. *Curr. Opin. Genet. Dev.* 18, 193–196. [Discute seus trabalhos a respeito da paramutação *Kit*.]
- Geneimprint. Jirtle Laboratory at Duke University. [www.geneimprint.com](http://www.geneimprint.com) [Um lista de genes com *imprinting* em humanos e outros organismos.]

- Gimelbrant A, Hutchinson JN, Thompson BR & Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318, 1136–1140. [Dados surpreendentes que mostram expressão monoalélica disseminada.]
- Heard E & Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X-chromosome. *Genes Dev.* 20, 1848–1867.
- Horsthemke B & Wagstaff J (2008) Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am. J. Med. Genet.* 146A, 2041–2052.
- Lomvardas S, Barnea G, Pisapia DJ et al. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell* 126, 403–413. [Como o acentuador H garante que cada neurônio olfatório expresse apenas um único receptor de olfato.]
- Ringrose L & Paro R (2007) Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annu. Rev. Genet.* 38, 413–443.
- Wutz A & Gribnau J (2007) X-inactivation Xplained. *Curr. Opin. Genet. Dev.* 17, 387–393.

## Um gene – mais do que uma proteína

- Maniatis T & Tasic B (2002) Alternative splicing and proteome expansion in metazoans. *Nature* 418, 236–243. [Uma boa fonte de exemplos.]
- Stamm S, Zhang MQ, Marr TG & Helfman DM (1994) A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res.* 22, 1515–1526. [Lista de diversos exemplos.]
- Südhof TC (2008) Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature* 455, 903–911. [Discute o *splicing* alternativo do gene da neurexina-3 (*NRXN3*) ilustrado na Figura 11.29.]

## Controle da expressão gênica no nível da tradução

- Baek D, Villén J, Shin C et al. (2008) The impact of microRNAs on protein output. *Nature* 455, 64–71.
- Buchan JR & Parker R (2007) The two faces of microRNA. *Science* 318, 1877–1878. [Revisa as evidências que indicam que miRNAs podem, algumas vezes, em vez de inibir, estimular a tradução.]
- Chendrimada TP, Finn KJ & Ji X (2007) MicroRNA silencing through recruitment of EIF6. *Nature* 447, 823–828.
- Klausner RD, Rouault TA & Harford JB (1993) Regulating the fate of messenger RNA: the control of cellular iron metabolism. *Cell* 72, 19–28. [Fonte dos dados sobre elementos de resposta ao ferro da Figura 11.32.]
- Lee R, Feinbaum R & Ambros V (2004) A short history of a short RNA. *Cell* 116 (2 Suppl.), S89–S92. [Revisão pessoal de seus trabalhos inovadores.]
- Mathonnet G, Fabian MR, Svitkin YV et al. (2007) MicroRNA inhibition of translation initiation in vitro by targeting of the cap-binding complex eIF4E. *Science* 317, 1764–1767. [Uma forma pela qual miRNAs inibem a tradução; mas esta pode não ser toda a resposta.]
- Ruvkun G, Wightman B & Ha Z (2004) The 20 years it took to recognize the importance of tiny RNAs. *Cell* 116 (2 Suppl.), S93–S96. [Uma descrição pessoal deste importante trabalho.]
- Selbach M, Schwanhäusser B, Thierfelder N et al. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63. [Fonte dos dados da Figura 11.33; ilustra uma importante abordagem experimental.]