



A genética em transformação: crise e revisão do conceito de gene

Leyla MARIANE JOAQUIM & Charbel NIÑO EL-HANI



RESUMO

O conceito de gene tem desempenhado um papel central na biologia desde sua introdução, no início do século xx. Contudo, ao longo do seu desenvolvimento histórico, o conceito tem sido objeto de controvérsia crescente, inicialmente na filosofia da biologia e, depois, na própria biologia. Desafios ao conceito de gene têm levado a uma dificuldade de preservar o chamado conceito molecular clássico, de acordo com o qual um gene é um segmento do DNA que codifica um produto funcional (polipeptídeo ou RNA). As últimas três décadas de estudos experimentais levaram a achados como genes interrompidos, emenda (*splicing*) alternativa, o chamado DNA-lixo, sequências TAR, pseudogenes, regulação pós-transcricional, RNAi e RNAsi, entre outros, os quais colocaram dificuldades inesperadas à compreensão usual do conceito de gene. Neste artigo, discutiremos os principais achados experimentais que desafiaram o conceito molecular clássico de gene. Daremos destaque, em particular, a avanços recentes, que tiveram lugar no Projeto Genoma Humano (PGH) e na Enciclopédia de Elementos de DNA (Encode). Atualmente, é clara a necessidade de uma análise e reformulação cuidadosa desse conceito central para o pensamento biológico. Muitos filósofos da biologia e biólogos, na tentativa de organizar a variedade de definições de gene, apresentaram visões interessantes a respeito desse conceito e de seu papel no conhecimento biológico, assim como propostas de revisão conceitual, que também abordaremos neste artigo. Concluimos que uma definição única de gene não é possível ou necessária. Ao contrário, o pluralismo de modelos e conceitos é provavelmente mais poderoso, desde que os domínios de cada conceito ou modelo sejam claramente definidos.

PALAVRAS-CHAVE • Gene. Conceito molecular clássico. Desafios. Projeto Genoma Humano. Encode.

INTRODUÇÃO

Entre os filósofos da biologia e, mais recentemente, entre os próprios biólogos, há uma visão compartilhada de que o conceito de gene se defronta com grandes dificuldades (cf. Falk, 1986; Fogle, 1990, 2000; Griffiths & Neumann-Held, 1999; Moss, 2001, 2003; Neumann-Held, 2001; Keller, 2002, 2005; Leite, 2006; Neumann-Held & Rhemann-Sutter, 2006; El-Hani, 2007; Waizbort & Solha, 2007; El-Hani; Queiroz & Emmeche, 2009). Este artigo discute os principais achados empíricos que levaram à crise do conceito de gene, por terem tornado evidente a diversidade estrutural do gene molecular, sobretudo em eucariotos, levando à dissolução da ideia de genes como unidades de

estrutura e/ou função. Em particular, abordaremos os avanços recentes que tiveram lugar nos contextos do Projeto Genoma Humano (PGH) e da Enciclopédia de Elementos de DNA (Encode).

É importante ter clareza de que a crise do conceito de gene refere-se, na verdade, a um modelo específico dos genes e de suas funções nos sistemas biológicos, expresso no chamado conceito molecular clássico, de acordo com o qual um gene é um segmento de DNA que codifica um produto funcional (polipeptídeo ou RNA). Abordaremos os estudos experimentais que conduziram às dificuldades enfrentadas pelo conceito molecular clássico de uma perspectiva histórica, com o intuito de entender como eles desafiaram esse modo usual de compreender o gene. É importante lembrar que os campos da genética e biologia molecular desenvolvem-se rapidamente e, portanto, acaba sendo imprescindível o tratamento de achados bastante recentes. Isso nos obriga a fazer uma história do presente, mesmo reconhecendo as dificuldades de tal tarefa, como nos alerta Keller:

(...) caímos em todos os tipos de armadilhas ao tentarmos ser historiadores do presente. Mas, talvez, a mais séria, especialmente em tempos excitantes como o nosso, é que a história pode acontecer muito mais rápido do que um acadêmico (pelo menos, um acadêmico como eu) pode escrever (Keller, 2005, p. 3).

Destacaremos dois projetos recentes que reúnem importantes achados da bioquímica e biologia molecular das últimas décadas: o Projeto Genoma Humano (PGH) e a Enciclopédia de Elementos de DNA (Encode). O PGH teve efeitos surpreendentes sobre o pensamento biológico, trazendo, em particular, suspeitas com relação à visão reducionista predominante na biologia da segunda metade do século xx, algo inesperado em um projeto que era, em boa medida, uma culminação do reducionismo (cf. Keller, 2002). Esse megaprojeto, realizado por meio de uma iniciativa pública e uma privada, constituiu um esforço mundial para a análise do genoma humano. No site oficial do projeto, há um *link* para pesquisas educacionais no qual encontramos um glossário.¹ No glossário, a definição encontrada para gene é a seguinte:

A unidade física e funcional fundamental da hereditariedade. Um gene é uma sequência ordenada de nucleotídeos localizada em uma posição particular em um cromossomo particular que codifica um produto funcional específico (isto é, uma proteína ou molécula de RNA).

¹ Cf. <http://www.ornl.gov/sci/techresources/Human_Genome/glossary/glossary_g.shtml>.

A definição de gene encontrada no glossário associado ao PGH pode ser considerada conservadora, na medida em que constitui um claro exemplo de uso do conceito molecular clássico.² De acordo com a definição, o gene é uma unidade hereditária que possui estrutura, função e localização definidas. Essa concepção sobrepõe o conceito mendeliano de unidade hereditária ao conceito molecular clássico (cf. Fogle, 1990), atualizando em termos moleculares, desse modo, uma visão sobre a existência de uma unidade básica da herança particulada que é anterior ao próprio conceito de gene, sendo encontrada, por exemplo, no uso que Mendel fez de termos como “fator”. Conforme argumentaremos ao longo deste artigo, uma série de achados experimentais tornou esta definição de gene cada vez mais problemática. Contudo, o fato de que, na virada do século XX para o século XXI, encontremos a mesma definição no glossário de um projeto de tal magnitude mostra a tentativa de ainda manter a ideia da unidade estrutural e funcional, mesmo diante dos desafios à noção de tal unidade no genoma. A tendência de identificar o gene com uma unidade no DNA que codifica para alguma função tem lugar no contexto de um hibridismo conceitual que confunde as ideias de codificação, regulação e função. Defenderemos, portanto, que é evidente a necessidade de uma análise cuidadosa e de uma reformulação de nossa compreensão do conceito de gene.

O Encode, por sua vez, é um consórcio público de pesquisa lançado pelo Instituto Nacional de Pesquisa do Genoma Humano (NHGRI, do inglês), em setembro de 2003, com o intuito de identificar todos os elementos funcionais do genoma humano. Por conta do sucesso inicial do projeto Encode, em setembro de 2007, o NHGRI incentivou um maior desenvolvimento do projeto. O Encode está organizado como um consórcio aberto³ e inclui pesquisadores⁴ de diversos campos da investigação e, recentemente, completou a caracterização de 1% do genoma humano, utilizando diversos experimentos e técnicas computacionais para caracterizar os elementos funcionais (The Encode, 2007).

Ao contrário do que podemos concluir no caso do PGH, ao menos à luz da definição de gene encontrada em seu glossário, os cientistas do Encode não ignoram a necessidade de revisão de conceitos centrais da genética, implicados pelos avanços das últimas décadas, em especial, do conceito de gene (cf. Gerstein *et al.*, 2007; Smith & Adkison, 2010). Antes pelo contrário, eles têm feito propostas para essa revisão. Ao discutir, mais abaixo, algumas reações à crise do conceito de gene, consideramos as propostas que surgiram do Encode.

² Como veremos adiante, esta não é a única definição associada ao PGH que encontramos na literatura (cf. Venter *et al.*, 2001).

³ Para a base de dados do ENCODE: <<http://www.genome.gov/10005107#4>>.

⁴ Para os participantes do ENCODE: <<http://www.genome.gov/26525220>>.

I O INÍCIO DA CRISE: A ATUALIZAÇÃO DA VISÃO MENDELIANA

Um dos fatores que levaram à crise do conceito de gene foi a atualização da visão mendeliana dos genes como unidades no conceito molecular clássico. Nas décadas de 1950 e 1960, ao mesmo tempo em que os achados da pesquisa molecular levantavam hipóteses relevantes sobre a estrutura e o funcionamento dos genes, a noção do gene como unidade fundamental de função e herança era mantida. A justaposição dos conceitos mendeliano e molecular foi resultado do esforço para manter a ideia de unidade da estrutura e função diante dos primeiros achados da biologia molecular (cf. Fogle, 1990).

Os problemas enfrentados pela noção de unidade começaram já em 1961, quando Jacob e Monod (1961) revolucionaram a compreensão da função e regulação gênicas com o modelo do operon *lac*. A partir desse modelo, passou-se da ideia, que havia dominado a genética clássica, de que os genes simplesmente agiam para a ideia de que os genes deviam ser ativados, podendo encontrar-se, portanto, em estado inativo na célula (cf. Keller, 2002). Outros avanços-chave do modelo residiam na distinção entre os genes estruturais e os genes regulatórios e na ideia de que o DNA apresenta, além de genes, regiões regulatórias que não são transcritas, mas estão relacionadas à regulação da transcrição (cf. Falk, 1986; Fogle, 1990; Keller, 2002).

O operon *lac* é encontrado em algumas bactérias entéricas – isto é, que vivem no intestino de animais –, como a *Escherichia coli*, e inclui três genes adjacentes e três sequências regulatórias, um promotor, um terminador e um operador. Entre os fatores que regulam o operon *lac*, temos a disponibilidade de lactose. O mecanismo de regulação do operon envolve uma proteína regulatória denominada “repressor *lac*”, que inativa a síntese de RNAs que codificam para enzimas que cumprem diferentes papéis no processo de digestão da lactose. O gene que codifica o repressor encontra-se próximo ao operon *lac*, mas em outra região do DNA, e é sempre expresso. Se não há lactose no meio em que vive a bactéria, o repressor se mantém fortemente ligado ao operador e inibe a síntese das enzimas envolvidas na digestão da lactose. Se a bactéria cresce na presença de lactose, um metabólito da lactose, chamado de alolactose, liga-se ao repressor, causando uma mudança em sua forma, o que impede que ele se ligue ao operador. Assim, as enzimas envolvidas na digestão da lactose são produzidas.

Em princípio, a distinção entre genes estruturais e regulatórios não foi tomada como um desafio de fato ao gene molecular clássico, visto que a regulação também envolve síntese proteica, não parecendo haver problemas em considerar os genes regulatórios como unidades estruturais e funcionais (cf. El-Hani, 2005). Entretanto, o modelo do operon leva a questões importantes sobre a natureza, a estrutura e a função dos genes. Considere-se, primeiro, que nos operons bacterianos é produzido um único RNAm policistrônico, isto é, um único RNAm para todos os genes incluídos em um

operon. A tradução sequencial deste RNAm resulta, então, nos diferentes polipeptídeos (Jacob & Monod, 1961). Isso levanta questões relativas ao que deve contar como “gene” em um operon: no operon *lac*, por exemplo, os três genes estruturais e o gene regulatório podem ser vistos como uma mensagem unitária que resulta em um produto. Desta perspectiva, o operon corresponderia a um único gene. Alternativamente, cada gene pode ser tratado, separadamente, como uma unidade. Este problema, que também é encontrado em eucariotos, começava a mostrar, desde o começo da biologia molecular, uma dificuldade que só cresceria com os avanços dos estudos experimentais: como identificar quais são os genes ao longo dos cromossomos?

Considere-se, agora, um problema ainda mais difícil colocado pelo modelo do operon: qual o *status* do operador e do promotor? Eles têm funções, são herdáveis, sofrem mutações, influenciam o fenótipo e evoluem. Há, assim, muitas propriedades compartilhadas entre as unidades genéticas e as regiões no DNA, as quais, entretanto, não são consideradas genes, simplesmente porque não são transcritas. Por isso, os operons passaram a ser chamados de “região” operadora e “sítio” promotor. Torna-se claro que há certa medida de arbitrariedade envolvida em distinguir unidades genéticas e não genéticas, diante do grande número de propriedades compartilhadas entre os elementos destas duas classes. Portanto, elementos regulatórios, especialmente aqueles mais distantes, tornaram problemático o conceito de gene como um *locus* genético compacto, ou seja, como uma unidade estrutural e funcional bem definida, claramente demarcada (cf. Gerstein *et al.*, 2007).

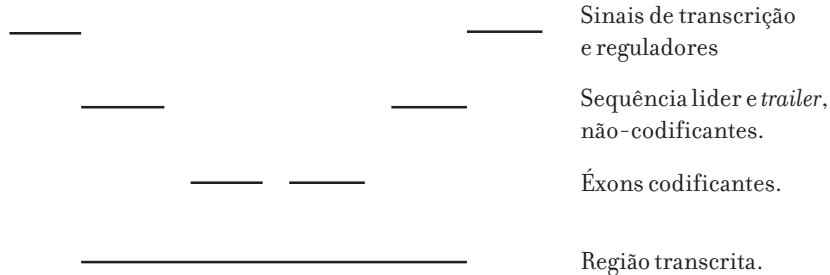
Adicionalmente, hoje em dia sabe-se que muitos elementos regulatórios podem ser encontrados dentro do primeiro éxon, dentro de íntrons ou no corpo inteiro de um gene (cf. The Encode, 2007; cf. Gerstein *et al.*, 2007). Alguns desses elementos podem ser de fato transcritos. Assim, a única diferença que permitia separar unidades genéticas de elementos regulatórios também desaparece em alguns casos. No próprio operon *lac*, por exemplo, elementos regulatórios também residem em regiões transcritas. Estes achados mostram que a situação é ainda mais complexa do que se pensava na década de 1960, uma vez que torna ainda mais arbitrária a distinção entre unidades genéticas e não genéticas, além de tornar problemática a própria distinção entre genes estruturais e regulatórios.

Os desafios mais importantes ao conceito molecular clássico surgiram a partir de pesquisas em organismos eucariotos. Mesmo com todo o sucesso que a pesquisa em procariontes teve em desvendar os mecanismos de controle e regulação da célula, a partir da década de 1970 uma série de estudos mostrou que os genes em organismos eucariotos são distintos em aspectos fundamentais daqueles encontrados nas bactérias. Diferentemente do DNA das bactérias, o DNA dos eucariotos está armazenado no núcleo, onde ocorre a transcrição, enquanto a tradução do RNA originado a partir do

DNA tem lugar no citoplasma, mais precisamente nos ribossomos. Essa diferença entre procariontes e eucariotos pode parecer simples a um olhar menos atento. Entretanto, ela promove um significativo aumento em número e complexidade das etapas dos processos envolvidos na ação gênica. O RNA entra em cena como um elemento atuante em diversas dessas etapas, permitindo a ocorrência de uma imensa gama de processos de controle celular que incidem sobre RNAs. Nas seções seguintes, discutiremos achados da pesquisa em eucariotos que resultaram em anomalias em relação à visão do gene como unidade estrutural e funcional no DNA.

2 OS LIMITES DE UM GENE

Fogle (1990) discriminou quatro modelos estruturais de gene (que chamou de modelos A, B, C e D) e mostrou como todos eles já fracassavam àquela época, diante dos achados sobre os genes e seu funcionamento (Figura 1). Ao longo de nossa argumentação, estes modelos serão usados para auxiliar a discussão sobre os principais achados das últimas décadas e sobre como eles desafiam o conceito molecular clássico de gene (cf. El-Hani, 2007).



Modelos

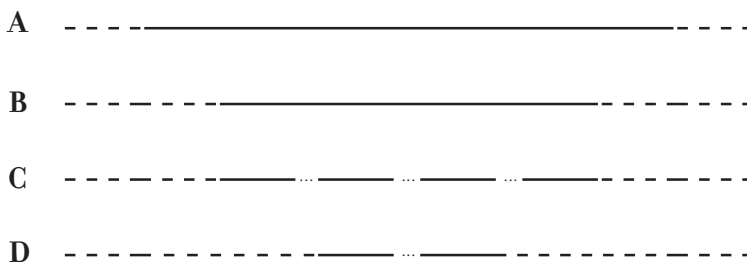


Figura 1. Quatro modelos estruturais para um gene codificador de proteínas (Fogle, 1990).
As linhas sólidas representam as áreas incluídas em cada modelo.

O modelo A incorpora uma tentativa de romper com a distinção entre unidades (genéticas) e não unidades, eliminando o problema da arbitrariedade discutido acima. Este modelo é o mais inclusivo, porque insere nos limites do gene todas as sequências *cis*-regulatórias (isto é, que regulam a expressão genética e encontram-se na mesma fita de DNA dos genes que regulam). Entendido dessa maneira, o gene inclui a região transcrita mais qualquer sequência que influencie a transcrição, como promotores, *enhancers*, silenciadores, terminadores e reguladores, entre outros.

A inclusão de muitos tipos diferentes de elementos regulatórios, que operam em combinações complexas, gera problemas para a delimitação dos genes. O primeiro problema diz respeito à distância entre os elementos regulatórios e os genes regulados. Há sequências *cis*-regulatórias cuja ação é independente da proximidade das sequências codificantes, como no caso de *enhancers* e silenciadores. Isso torna difícil a demarcação das fronteiras de um gene. Outro agravante é que há cooperação entre sequências *cis*-regulatórias, o que dificulta a decisão de quais fatores devemos incluir e quais excluir de um gene. Há ainda vários genes que dependem da mesma sequência *cis*-regulatória, o que faz com que o modelo A aumente substancialmente a superposição de genes no genoma. A conclusão a que podemos chegar é que um modelo estrutural do gene completamente inclusivo fracassa diante das dificuldades e arbitrariedades da decisão quanto à inclusão de sequências *cis*-regulatórias.

Fogle apresenta o modelo B como um meio de contornar os problemas do modelo totalmente inclusivo. Neste caso, considera-se que os limites estruturais do gene são definidos pelo processo de transcrição. O modelo sugere uma relação entre a unidade proposta e a informação necessária para a síntese de um polipeptídeo. Esta noção de gene é amplamente aceita e, por essa razão, merecerá atenção especial. O modelo B fracassa diante do fato de que os transcritos podem resultar em mais de uma unidade de informação pela geração combinatória de sequências, devido à existência de genes interrompidos e ocorrência do *splicing* alternativo, que é uma anomalia que se contrapõe ao gene molecular clássico, porque rompe com a relação 1:1:1 entre gene, produto gênico (transcrito ou polipeptídeo) e função. Como veremos na próxima seção, esses fenômenos resultam em uma relação de uma região no genoma para muitos transcritos e muitas funções, colocando dificuldades para a compreensão dos genes.

2.1 MÚLTIPLOS TRANSCRITOS A PARTIR DE UMA ÚNICA SEQUÊNCIA DE DNA

2.1.1 GENES INTERROMPIDOS, DNA-LIXO E SEQUÊNCIAS TAR

Os bioquímicos norte-americanos David Glover e David Hogness, da *Stanford University School of Medicine*, foram os primeiros a relatar, em fevereiro de 1977, que os genes eucarióticos codificadores de RNAs ribossomais são interrompidos. No mesmo ano, vários pesquisadores, como Richard Roberts do *Cold Spring Harbour Laboratory* (EUA), Philip Sharp, do *Massachusetts Institute of Technology* (MIT, EUA), e Pierre Chambon, do *Centre National de La Recherche Scientifique* (CNRS, França), mostraram independentemente que organismos eucarióticos apresentam genes interrompidos, ou seja, que o DNA nesses organismos não é contínuo, como na maioria das bactérias. Roberts e Sharp mostraram que genes virais codificadores de proteínas estruturais também contêm interrupções. Esses achados foram de suma importância, porque indicaram que os mecanismos de processamento da informação genética poderiam ser muito mais complexos do que se suspeitava anteriormente e, além disso, abriram novos caminhos para a pesquisa, transformando substancialmente nossa compreensão dos genes (cf. Waizbort & Solha, 2007). Foi só um ano depois, em 1978, que Walter Gilbert chamou as regiões não codificadoras que interrompiam os genes eucarióticos de íntrons e as regiões codificadoras, de éxons, e postulou que, diante dos íntrons, o dogma “um gene-uma cadeia polipeptídica” teria desaparecido (cf. Gilbert, 1978).

Na ausência de conhecimento sobre as funções das regiões intrônicas, foi proposto inicialmente que elas seriam um “DNA-lixo” (cf. Ohno *apud* Keller, 2002). Isso foi enfatizado pelo subsequente sequenciamento do genoma humano, no qual foi mostrado que somente cerca de 1,2% das bases de DNA constituem éxons codificantes e, portanto, que a maior parte do genoma seria feita desse suposto “DNA-lixo” (cf. Lander *et al.*, 2001; Venter *et al.*, 2001).

As consequências dos íntrons para o conceito de gene, bem como os novos achados, não pararam por aí. A estrutura descontínua dos genes eucarióticos permite que um gene esteja completamente contido dentro de um íntron de outro gene, ou permite que um gene se sobreponha a outro na mesma fita sem compartilhar qualquer éxon ou elemento regulatório. Isso cria sérias dificuldades para a consideração do gene como uma entidade discreta, na medida em que há genes superpostos e genes aninhados (*nested genes*). Assim, os genes não são nem contínuos nem discretos não só pelo fato de que elementos regulatórios distantes e, muitas vezes, compartilhados são requeridos para a expressão gênica, mas também pela ocorrência de genes superpostos e aninhados.

A existência dos íntrons parece ser uma fonte inesgotável de desafios a concepções tradicionais sobre o que é um gene. Estudos recentes chamaram a atenção para uma descoberta no mínimo intrigante, que coloca em xeque toda a noção do que é gênico e do que é intergênico (The Encode, 2007). Alguns estudos dos cromossomos 21 e 22 indicaram que grandes quantidades de DNA anteriormente tratado como lixo são de fato transcritas, mostrando que muito mais do que os éxons são transcritos no genoma (cf. Kapranov *et al.*, 2002). Podemos dizer que há menos distinções entre as regiões gênicas e as intergênicas do que se pensava anteriormente.

Neste contexto, o chamado “DNA-lixo”, que corresponde à maior parte do genoma, não é mais considerado desprovido de função, como se pensava anteriormente. A função de grande parte dessas regiões é transcrever ncRNA (RNA não codificante), com função regulatória, entre outras (ver a próxima seção). Essas regiões passaram a ser chamadas, então, de regiões transcricionalmente ativas (TARs, *transcriptionally active regions*). Não foram somente genes de ncRNA que foram localizados em íntrons de genes codificantes, mas também pseudogenes transcritos, como veremos mais adiante. Não se pode, assim, perder de vista, em termos funcionais, sequências como os íntrons, porque muitas delas podem influenciar a expressão de seus genes hospedeiros, tanto direta quanto indiretamente.

Um aspecto interessante sobre as chamadas TARs é que o número delas varia bastante entre as espécies, ao contrário do número de genes que, como sabemos, não varia muito. Humanos possuem números de genes semelhantes a outras espécies, mas números de TARs superiores. Com base nisso, tem-se concluído que as sequências podem ter um papel chave na explicação de diferenças entre as espécies e, portanto, o antigo “DNA-lixo” está muito longe de ser descartado de estudos futuros, na medida em que pode ter importante papel evolutivo (cf. Gerstein *et al.*, 2007).

Em suma, apesar de não termos ainda bons modelos para os papéis funcionais dessas regiões comumente chamadas de intergênicas, não podemos ignorar o fato de que elas contêm muitas sequências altamente conservadas ao longo da evolução. Se uma sequência é altamente conservada, isso pode significar que ela cumpre alguma função importante para a sobrevivência e reprodução bem sucedidas dos organismos, mesmo que essas funções ainda não tenham sido identificadas (cf. Gerstein *et al.*, 2007).

Os achados sobre TARs, DNA-lixo, regiões gênicas e intergênicas, de modo geral, sugerem que nosso conhecimento da associação entre sequências de DNA e funções é inadequado, e que muitos estudos sobre essas regiões são atualmente necessários. Tais estudos não podem economizar na cautela com relação à noção cada vez mais problemática de unidades genéticas, uma vez que, afinal, os genes parecem estender-se para o espaço intergênico.

2.1.2 *Splicing* ALTERNATIVO

Como vimos, desde 1978, considera-se que os genes interrompidos contêm regiões codificantes (éxons) e não codificantes (íntrons), sendo estas últimas retiradas durante o processamento de RNA. O DNA é transcrito em uma molécula de RNA precursora longa, a qual precisa ser processada, o que envolve, entre outros processos, a retirada dos íntrons. A partir do RNA precursor, é gerado o RNA mensageiro maduro, o qual é traduzido em proteína. A retirada dos íntrons durante o processamento pode ocorrer de formas alternativas, o que implica que pode ser gerada mais de uma proteína por gene. Esse fenômeno é denominado *splicing* alternativo. A variabilidade em padrões de *splicing* aumenta o número de proteínas expressas por uma região codificante de DNA eucarioto.

Assim, no *splicing* alternativo, várias unidades de mensagem são construídas antes da formação do produto e, portanto, antes de a sequência de DNA exercer sua função. Nesse sentido, a sequência transcrita atua como várias unidades de estrutura e função. Se diferentes proteínas podem ser geradas, é difícil sustentar a ideia de que genes seriam unidades estruturais e/ou funcionais. A relação entre gene, produto gênico e função não é de 1:1:1.

O *splicing* alternativo desafia o conceito de gene de forma muito substancial pelo fato de ser bastante comum. Ele conduz a uma significativa expansão do proteoma de metazoários e é considerado um dos grandes responsáveis pela complexidade funcional do genoma humano, porque permite que grande diversidade proteica coexista com um número relativamente limitado de genes. É difícil estimar a frequência dos processos de *splicing* alternativo nos organismos, porque eles ocorrem, na maior parte das vezes, em momentos precisos do desenvolvimento e em tecidos específicos. Estima-se que a média do número de transcritos diferentes de RNA por *locus* é bem maior do que um, chegando a 5,4 transcritos/*locus* (cf. Gericke & Hagberg, 2007). Modrek e Lee (2002) chegam a relatar que de 35 a 59% dos genes humanos sintetizam RNAs que sofrem *splicing* alternativo. Mesmo que uma porção significativa das variantes preditas de *splicing* não seja funcional (Sorek *et al.*, 2004), ainda assim é o caso de que o *splicing* alternativo é um dos componentes mais significativos da complexidade funcional do genoma dos metazoários.

Para responder ao problema do *splicing* alternativo, Fogle (1990) propôs o modelo C (Figura 1), o qual incorpora a hipótese de que os genes correspondam, de fato, aos éxons ou a conjuntos de éxons que compartilham um transcrito comum.⁵ Portan-

⁵ Cf. Epp (1997). Esta definição de gene foi usada por Venter *et al.* (2001, p. 1317) em um dos artigos que apresentou um esboço da sequência do genoma humano: “Um gene é um *locus* de éxons cotranscritos. Um único gene pode dar

to, as unidades no DNA seriam de menor tamanho, em comparação com os modelos anteriores, A e B. Desse modo, mesmo diante do *splicing* alternativo, o conceito de gene como unidade estrutural poderia ser salvo. O modelo C considera os éxons como unidades estruturais no genoma e, assim, mantém a ideia de que o gene é uma unidade. Porém, o modelo C enfrenta o problema de que há padrões de *splicing* que resultam em transcritos que diferem entre si pela presença ou ausência de éxons correspondendo a sequências *trailers*.

Pode-se propor, então, um novo modelo, D, que modifica o modelo C, enfocando estritamente a região codificante e desconsiderando qualquer diferença no comprimento do transcrito oriunda de regiões não codificantes. Assim, a sequência não codificante (*trailer*) torna-se irrelevante. Porém, o *splicing* alternativo pode afetar também a região codificadora dos éxons, gerando produtos gênicos multifuncionais, como mostraram Schultz e colaboradores (Schulz *et al.*, 1986), em um estudo do gene Eip 28/29 de *Drosophila* (cf. Fogle, 1990). Portanto, o modelo D também fracassa. Fogle conclui que nenhum dos modelos estruturais analisados por ele podia ser sustentado diante dos achados dos estudos moleculares sobre os genes de vinte anos atrás. Se tratarmos esses modelos, como parece plausível fazer, como o conjunto completo dos possíveis modelos estruturais de gene, podemos ver por que a ideia do gene como unidade estrutural está em crise.

Contudo, a existência de padrões de *splicing* que afetam éxons codificantes pode ser vista como um fato que mostra apenas que o modelo D não é inteiramente geral. Como na biologia modelos inteiramente gerais, ou universais, são raros, se é que existem, a ausência de generalidade não parece ser um problema grave para o modelo D, que poderia ser considerado útil, apesar das exceções. Tal conclusão só seria correta, no entanto, se não houvesse muitos desafios adicionais ao conceito molecular clássico, como superposição de genes, *trans-splicing*, edição de mRNA etc. (cf. Falk, 1986; Portin, 1993; Fogle, 1990; Pardini & Guimarães, 1992; Griffiths & Neumann-Held, 1999; Keller, 2002). Com o *splicing* alternativo, ficou claro que os genes não são unidades simples de hereditariedade e função, uma vez que um *locus* gênico pode codificar para múltiplos transcritos de mRNA diferentes. Entretanto, o *splicing* alternativo não é o único meio pelo qual a célula produz proteínas variantes e, portanto, está longe de ser o único problema a resolver na tentativa de salvar a noção de unidade.

origem a transcritos múltiplos e, assim, a proteínas distintas múltiplas com múltiplas funções, por meio de *splicing* alternativo e sítios de iniciação e terminação de transcrição alternativos”. Esta definição de gene é proposta, portanto, por causa dos desafios ao modelo B de Fogle, mas não levam em consideração os problemas enfrentados pelo modelo C (ver abaixo). Vimos acima que, em um glossário associado ao PGH, o conceito molecular clássico de gene continua sendo usado, o que sugere que essa compreensão diferente dos genes pode ser uma característica própria da equipe de Venter.

Sequências de DNA podem ser transcritas em mais do que um transcrito de RNA primário e, conseqüentemente, em diferentes mRNAs chamados de isoformas de transcritos, as quais podem ser produzidas de várias maneiras, por exemplo, mediante o uso de diferentes locais de início de transcrição (transcription start sites, TSS), ou usando regiões promotoras de *loci* gênicos completamente diferentes. Há também o fenômeno de trans-splicing, no qual ocorre a ligação de duas moléculas de RNA separadas em um único transcrito maduro, as quais podem ter sido originadas de fitas de DNA opostas, ou até mesmo de cromossomos diferentes. Adicionalmente, são também conhecidos cerca de 200 tipos diferentes de modificações pós-traducionais. Todos esses achados colocam dificuldades adicionais para a ideia de uma relação 1:1:1 entre gene, produto gênico e função. Nas seções seguintes, examinaremos resultados empíricos mais recentes que trazem ainda mais desafios ao gene molecular clássico.

3 A COMPLEXIDADE DOS MECANISMOS DE EXPRESSÃO GÊNICA

Como vimos, o estudo da natureza da expressão gênica aponta para importantes dificuldades quanto à ideia do gene como unidade de estrutura/função. A ideia de que o gene é uma unidade de função está baseada na noção de que um gene produz um polipeptídeo que, por sua vez, tem uma função singular. Entretanto, a complexidade da expressão gênica, que é altamente dependente do contexto celular, torna bastante difícil manter a ideia de uma relação única entre um gene e sua função.

A complexidade da expressão gênica está especialmente relacionada ao grande número de processos vinculados à regulação. Entre as dificuldades que os processos regulatórios acarretam para as noções simples de função gênica está, como já mencionamos, a própria distinção entre genes estruturais e regulatórios. Já em 1986, Falk relatava que, quanto mais se sabia sobre a atividade de regulação gênica, mais ficava claro que a distinção entre genes estruturais e regulatórios pode ter no máximo um significado instrumental, ou seja, tal distinção é apenas uma ferramenta para auxiliar na compreensão dos fenômenos. De lá para cá, as dificuldades que os processos de regulação gênica acarretam para a concepção de gene como unidade de estrutura/função só aumentaram.

Sabemos hoje que existe uma grande diversidade de RNAs que não havia sido detectada até pouco tempo (cf. Ruvkun, 2001), *dos quais muitos têm papéis regulatórios*. A recente descoberta desses transcritos não codificantes, coletivamente chamados de transcritos de função desconhecida (*transcripts with unknown function* – TUFs), tem confundindo ainda mais os limites físicos das regiões gênicas e a compreensão da organização do genoma. Sequências de transcritos não codificantes frequentemente

sobrepõem-se a genes codificadores de proteínas na mesma fita ou em fitas opostas do DNA. Além disso, essas sequências são frequentemente localizadas em regiões intergênicas (cf. Gingeras, 2007). A seguir, com o intuito de exemplificar a complexidade dos mecanismos regulatórios da expressão gênica, descreveremos duas classes de RNAs não codificantes, os microRNAs e os siRNAs. Escolhemos esses transcritos por eles representarem bem a classe de importantes desafios atuais que a complexidade do reguloma – o conjunto de componentes regulatórios celulares – apresenta para as visões tradicionais sobre os genes, em especial, aquelas que tratam os genes como unidades de estrutura e/ou função. Esses RNAs, devido aos seus papéis relevantes no reguloma, recebem atualmente bastante atenção da comunidade científica.

3.1 EXEMPLOS DE REGULAÇÃO PÓS-TRANSCRICIONAL: MICRORNAs E siRNAs

Os microRNAs e siRNAs não se distinguem pela sua composição química ou pelos seus mecanismos de ação, mas podem se distinguir quanto à sua origem ou quanto aos genes que silenciam, isto é, cuja expressão inibem. Os microRNAs derivam do DNA, enquanto os siRNAs podem derivar do DNA ou de transposons e vírus. Vamos detalhar brevemente cada um deles.

Os microRNAs possuem cerca de 21-23 nucleotídeos e cumprem a função de regular a expressão gênica. Alguns possuem expressão constitutiva, enquanto outros estão sujeitos a controle de expressão temporal-específica e tecido-específica. Ao invés de serem traduzidos em proteínas (como o bem conhecido mRNA), eles são processados a partir de transcritos primários, conhecidos como pré-microRNAs, de modo a apresentarem uma pequena estrutura do tipo “*stem-loop*” e, finalmente, para microRNAs funcionais. Os microRNAs maduros são apenas parcialmente complementares a um ou mais mRNAs. Essa classe de “genes” regulatórios encontra-se em partes do genoma que não codificam proteínas, ou seja, estão escondidos no que era antigamente chamado de “DNA-lixo”.

Isso ilustra como o uso das lentes do conceito molecular clássico, que restringe os genes a regiões que codificam proteínas, pode conduzir a classificar como lixo sequências de imensa importância, tais como as que codificam microRNAs. Afinal, longe de ser descartável, a influência dos microRNAs sobre a síntese proteica tem sido bem documentada nos últimos anos. Trabalhos recentes mostram o grande impacto dos microRNAs no proteoma, como no caso, por exemplo, de um único microRNA que tem, sozinho, capacidade de reprimir a produção de centenas de proteínas (cf. Selbach *et al.*, 2008; Baek *et al.*, 2008). Isso significa que eles podem ter efeito direto ou indireto sobre o funcionamento de centenas a até milhares de genes.

O número de microRNAs identificados continua crescendo. Pesquisadores estão explorando o funcionamento dos microRNAs e caracterizando a especificidade dos tecidos em que são encontrados, assim como as atividades de moléculas individuais de microRNAs. Mudanças nos níveis de microRNAs têm sido correlacionadas com muitas doenças. Muitos trabalhos estão em andamento sobre os efeitos de expressões (baixas ou altas) de microRNAs específicos ao longo do desenvolvimento e na inibição de doenças, principalmente câncer, doenças do coração, desordens neurológicas, diabetes, entre outras (cf. Glaser, 2009).

Os siRNAs (do inglês *small interfering RNAs*) constituem uma classe de RNAs de fita dupla (dsRNAs, do inglês *double strand RNAs*). Os dsRNAs são importantes reguladores da expressão gênica em muitos eucariotos, sendo responsáveis por diferentes tipos de silenciamento gênico (cf. Tuschl & Meister, 2004). Fire *et al.* (1998) mostraram que RNAs de fita dupla iniciam um processo de silenciamento gênico pós-transcricional hoje conhecido como mecanismo de ação do RNA de interferência (RNAi).

A comunidade científica tem reconhecido cada vez mais a importância da regulação gênica por siRNAs. O siRNA desencadeia o fenômeno de RNAi quando proteínas de encaminhamento de RNAi, que têm a sequência complementar a um mRNA, ligam-se ao siRNA. O RNAi que foi direcionado pelo siRNA cliva o mRNA (alvo), silenciando, desse modo, a expressão gênica em um nível pós-transcricional.

MicroRNAs e siRNAs são apenas exemplos da complexidade dos processos de regulação gênica. A cada achado, a complexidade dos processos gênicos e a simplicidade do conceito molecular clássico ficam progressivamente incompatíveis. O problema resultante dessas descobertas sobre os papéis de RNAs regulatórios não é novo: algumas definições de gene referem-se somente a sequências codificantes de proteínas, enquanto outras incluem regiões que não codificam proteínas. Portanto, os RNAs regulatórios são considerados genes à luz de certas definições, e não o são à luz de outras. Eles dramatizam a situação de anarquia em relação ao que devemos chamar de gene e a situação de expansão de entidades genéticas cujo caráter instrumental não pode ser mais disfarçado. Adicionalmente, os RNAs regulatórios enfatizam a dependência que a função gênica tem do contexto celular, o que desafia não só a compreensão usual dos genes, mas também interpretações comuns sobre seu papel nas células, que tendem a hiperbolizar o papel do DNA no controle do funcionamento celular e do desenvolvimento. O tempo e o local nos quais um dado conjunto de genes é ou não ativado dependem crucialmente da regulação e a regulação não é algo que os genes fazem, comandam ou programam, mas algo ao qual eles estão sujeitos (cf. El-Hani *et al.*, 2006). As evidências das pesquisas na genética e na biologia molecular mostram, portanto, a necessidade tanto de uma revisão conceitual do que se entende por gene, quanto de uma reinterpretação do papel do gene no contexto celular.

4 PSEUDOGENES

Pseudogenes são regiões do DNA estruturalmente similares a genes, mas que não são transcritas. Até recentemente, eram entendidos como sequências não funcionais. Foram reconhecidos na década de 1970, quando cientistas começaram a tentar localizar regiões cromossômicas associadas com a produção de moléculas importantes. Ao procurarem pelos genes que codificavam determinadas proteínas, os cientistas acabaram identificando sequências de DNA que se pareciam com tais genes, mas não eram transcritas e, assim, não eram capazes de levar à produção de proteínas. Consequentemente, essas sequências não eram funcionais e, embora possuísem características de genes (como a presença de promotores e sítios de *splicing*), foram chamadas de pseudogenes (cf. Gerstein & Zheng, 2006).

Pseudogenes são formados por processos de duplicação e subsequente mutação de uma sequência gênica, de modo que a cópia que sofreu mutação perde sua função. Devido à ancestralidade compartilhada com um gene funcional, pseudogenes contam, no entanto, histórias evolutivas. A maior parte dos pseudogenes era, de fato, entendida como cópia danificada de genes funcionais, que servia como fósseis genéticos com papéis na evolução do genoma (cf. Gerstein & Zheng, 2006). Essas regiões guardam grande semelhança com o gene original e são tão comuns quanto as sequências codificadoras de proteínas (cf. Gerstein *et al.*, 2007), mas, como não são transcritas, não são consideradas genes de acordo com o conceito molecular clássico, que trata os genes como unidades de estrutura e/ou função.

Assim, como no caso da descoberta de alguns segmentos de DNA que são transcritos, mas não traduzidos (genes de tRNAs, rRNAs etc.), os pseudogenes não pareciam, em princípio, afetar muito o conceito molecular clássico, visto que, como não eram transcritos, não havia problemas em não serem considerados gene, apesar de sua estrutura (cf. El-Hani, 2005). Entretanto, os pseudogenes mostraram-se mais complexos do que era esperado inicialmente.

Evidências recentes sobre a atividade dos pseudogenes sugerem que alguns não estão inteiramente inativos. Como mostram Balakirev e Ayala (2003), há muitas décadas, os papéis funcionais têm sido atribuídos aos pseudogenes em fenômenos como a expressão gênica, a regulação gênica e a geração de diversidade genética. Por exemplo, resultados obtidos em estudos com ovócitos de ratas sugerem que siRNAs são derivados de pseudogenes transcritos, indicando que um dos papéis dos pseudogenes é ajustar os níveis de mRNA, através da interferência de RNA. Ou seja, pseudogenes podem estar relacionados com a regulação gênica (cf. Tam *et al.*, 2008; Watanabe *et al.*, 2008).

O achado de que pseudogenes podem ser funcionais traz novas dificuldades para a conclusão de que eles não são genes. Este não é um problema de pequena monta.

Estima-se que cerca de 19% de todos os pseudogenes são, de fato, transcritos (The Encode, 2007). Assim, se estudos posteriores mostrarem que pseudogenes são também traduzidos, a distinção entre genes e pseudogenes ficará insustentável. Ao que tudo indica, estamos caminhando nessa direção. Por exemplo, o “gene” phosphoglicerato mutase 3 (PGMA3) é considerado um pseudogene em humanos, mas um estudo mostrou que esse pseudogene produz proteínas funcionais em chimpanzés (cf. Betrán *et al.*, 2002). Trata-se, então, de um gene ou de um pseudogene? Uma entidade pode ser considerada um pseudogene em humanos e um gene em chimpanzés? Casos como este mostram claramente como pode ser arbitrário chamar determinada sequência de DNA de gene ou de pseudogene.

5 ALGUMAS REAÇÕES À CRISE DO CONCEITO

Os principais achados moleculares que desafiam o conceito de gene estão sumarizados na tabela 1. Diante da crise do conceito molecular clássico de gene, diversas reações surgiram. Alguns autores, como Keller (2002), Gelbart (1998) e Portin (1993), argumentam até mesmo contra a manutenção do termo ‘gene’ no vocabulário biológico. Keller, por exemplo, escreve que o gene é um conceito problemático e sugere que chegou o tempo de forjar novas palavras e deixar esse conceito de lado. Portin afirma que nosso conhecimento da estrutura e função do material genético ultrapassou a terminologia usada para descrevê-lo e que pode ser o caso de que o termo “gene” não seja mais útil.

Outros autores, menos céticos, propõem que é preciso salvar o conceito de gene através de uma redefinição que não incorpore a ideia de genes como unidades básicas da matéria viva (cf. Fogle, 1990). O que se busca ter em vista, nesse caso, é o fato de que descartar um conceito, que está presente tanto no discurso científico quanto na linguagem cotidiana, é uma tarefa problemática, talvez impossível. Alguns autores apostam no esclarecimento da diversidade de significados do conceito de gene e fazem a partir desse esforço propostas para sua reformulação. A seguir, vamos discutir algumas das principais tentativas de salvar esse conceito central no pensamento biológico. Daremos especial atenção à busca por uma nova definição de gene em trabalhos recentes, produzidos no contexto do projeto Encode (cf. Gerstein *et al.*, 2007) e por Scherrer e Jost (2007a, 2007b). Contudo, discutiremos brevemente uma série de outras abordagens do conceito de gene que emergiram no contexto de sua crise.

Após ter proposto o abandono do conceito de gene, Keller (2005) reformulou sua posição, afirmando que o conceito de gene poderia ser mantido, mas apenas no contexto de uma compreensão das complexas redes informacionais que constituem a

célula e, além disso, de uma maneira mais dinâmica. Para ela, o século XXI será o século dos sistemas genéticos, não do gene. Para que esse conceito não seja abandonado, será preciso, então, enfrentar os desafios colocados pela complexidade biológica e construir novas maneiras de falar. Trata-se de compreender os complexos sistemas de interação entre os processos e as entidades que compõem os sistemas vivos e, para isso, hábitos arraigados de pensamento e linguagem, que dão prioridade às partes do sistema antes que ao sistema vivo como um todo, deverão ser superados. Esses hábitos são muito problemáticos quando genes são tomados como partes, porque genes não têm qualquer significado quando isolados. Keller trata a célula como um sistema de produção de significados que transforma sequências de nucleotídeos em genes. Nestes termos, o conceito de gene pode sobreviver no século XXI, mas apenas se genes passarem a ser entendidos como verbos, e não mais substantivos.

Fenômeno	Descrição	Consequência
LOCALIZAÇÃO E ESTRUTURA GÊNICA		
Genes intrônicos	Um gene dentro de um intron de outro.	Dois genes no mesmo <i>locus</i> .
Genes com quadros de leituras sobrepostos	Uma região de DNA pode codificar para dois produtos proteicos diferentes em diferentes trechos de leituras.	Não há correspondência 1:1 entre DNA e sequência proteica.
<i>Enhancers</i> e Silenciadores	Elementos regulatórios distantes.	Sequências de DNA envolvidas em uma expressão podem estar amplamente separadas uma das outras no genoma. Portanto, genes não possuem fronteiras claras e se superpõem se houver gene entre silenciador/ <i>enhancer</i> e quadro de leitura. Adicionalmente, a relação entre genes e silenciadores/ <i>enhancers</i> é de muitos-para-muitos.
VARIAÇÕES ESTRUTURAIS		
Elementos móveis	Elementos genéticos aparecem em novos locais ao longo das gerações.	Um elemento genético pode não ser constante em sua localização.
Rearranjamentos gênicos/variantes estruturais	O rearranjo do DNA ou <i>splicing</i> em células somáticas resulta em muitos produtos gênicos alternativos.	A estrutura gênica pode ser diferente entre indivíduos e prole, e entre células/tecidos.
ESTRUTURA EPIGENÉTICA E CROMOSSÔMICA		
Modificações epigenéticas, <i>imprinting</i>	A informação herdada pode não ser baseada na sequência de DNA, uma expressão gênica depende da origem (maternal ou paternal) entre outros fatores.	O fenótipo não é determinado estritamente pelo genótipo.

Efeitos da estrutura da cromatina	A estrutura da cromatina, que influencia na expressão gênica, é associada imprecisamente com a sequência particular de DNA.	A sequência de DNA não é suficiente para prever o produto gênico. Eventos pós-transcricionais
EVENTOS PÓS-TRANSCRICIONAIS		
<i>Splicing</i> alternativo de RNA	Um transcrito pode gerar múltiplos RNAs mensageiros, resultando em produtos proteicos diferentes.	Múltiplos produtos a partir de um locus genético; a informação no DNA não é linearmente relacionada com aquela da proteína.
Trans- <i>splicing</i> de RNA	Sequências de DNA distantes podem codificar transcritos ligados por diferentes combinações.	Uma proteína pode resultar de informações combinadas codificadas em múltiplos transcritos.
Edição de mRNA	O RNA é enzimaticamente modificado, isto é, enzimas atuam no processo de edição de partes do RNAm.	A informação no DNA não é codificada diretamente em sequências de RNA.
EVENTOS PÓS-TRADUCIONAIS		
<i>Splicing</i> de proteínas	Produtos proteicos se autoclivam e podem gerar produtos funcionais múltiplos.	Locais de início e final de sequências proteicas não são determinados pelo código genético
Trans- <i>splicing</i> de proteína	Não só os transcritos mas também as proteínas distintas podem sofrer trans- <i>splicing</i>	Locais de início e final de sequências proteicas não são determinados pelo código genético.
Modificação proteica	Proteína é modificada, alterando a estrutura e função do produto final.	A informação no DNA não é codificada diretamente em sequências proteicas.
PSEUDOGENES E RETROGENES		
Retrogenes	Um retrogene é formado por transcrição reversa e pela inserção de um produto de DNA em um genoma.	Fluxo de informação de RNA para DNA.
Pseudogenes transcritos	O pseudogene é transcrito.	Atividade bioquímica de elementos supostamente mortos.

Tabela 1. Fenômenos que apontam anomalias no conceito molecular clássico de gene. Adaptado de Gerstein *et al.*, 2007.

Em perspectiva similar, El-Hani, Queiroz e Emmeche (2009) argumentam que o significado de um gene não está contido na sequência de nucleotídeos do DNA, mas emerge como um processo que envolve o sistema pelo qual os genes são interpretados, incluindo a célula e, em uma série de casos, o ambiente supracelular. Assim, genes não estão dados no DNA, mas são construídos pela célula. Esta visão é, para esses pesquisadores, fundamental para o entendimento de que não é o DNA que controla a célula,

não é o DNA que “faz coisas” com a célula, como se costuma ensinar, mas a célula é que “faz coisas” com o DNA, que é um repositório de informação biológica útil, e não um catalisador de processos, ou um programa de desenvolvimento, ou um controlador da célula. Trata-se de uma molécula relativamente inerte, mas que constitui, sem dúvida, um poderoso meio de transmissão e manutenção de informação nos sistemas vivos. É aí que reside a importância do DNA, sendo importante não exagerar seu papel nos sistemas vivos, na medida em que isso obscurece as complexas redes de controle difuso que caracterizam os organismos, sejam unicelulares ou multicelulares.

Pardini e Guimarães (1992) propuseram um conceito sistêmico de gene, de acordo com o qual “o gene é uma combinação de (uma ou mais) sequências de ácidos nucleicos (DNA ou RNA), definido pelo sistema (a célula inteira, interagindo com o ambiente) que corresponde a um produto (RNA ou polipeptídeo)” (1992, p. 713-7). Essa definição trata o genoma como parte do sistema celular, que “constrói, define e usa o genoma como parte do seu mecanismo de memória, como um banco de dados interativo” (Guimarães & Moreira, 2000, p. 249). Os autores ressaltam a dinâmica da relação entre a informação codificada e o produto da sua codificação, que é muito complexa e varia conforme as condições espaciais e temporais em que ocorre. Eles argumentam que o significado de um segmento de DNA é relativo, dependendo do sistema de expressão gênica no qual ele está inserido. Assim, seu significado pode ser plural; e a natureza plural dos genes, particularmente nos eucariotos, origina-se da dependência da expressão gênica com relação ao contexto celular e supracelular (cf. Pardini & Guimarães, 1992; El-Hani, 2007).

Griffiths e Neumann-Held (1999) levam mais longe a interpretação da dependência da expressão gênica com relação ao contexto bioquímico em que ela ocorre, em seu conceito molecular processual de gene (*process molecular gene concept*) (cf. Neumann-Held, 2001). Esses autores propõem que os genes não sejam tratados como meras sequências no DNA, mas como todo o processo molecular subjacente à expressão de um produto particular (um polipeptídeo ou um RNA). Dessa perspectiva, o “gene” é um processo que ocorre repetidas vezes e conduz à expressão regulada de um produto polipeptídico particular. A proposta de Griffiths e Neumann-Held trata os genes, portanto, como processos, e não entidades físicas no DNA. A natureza processual desse conceito torna possível acomodar anomalias que o modelo molecular clássico tem dificuldade de enfrentar, tal como o *splicing* alternativo e a edição de mRNA. Afinal, o conceito de gene molecular processual simplesmente inclui no gene os processos de *splicing* alternativo e edição de mRNA (cf. El-Hani, 2007; Griffiths & Neumann-Held, 1999; Neumann-Held, 2001). Se, a partir da mesma sequência de DNA, dois produtos proteicos foram sintetizados, em decorrência de diferentes padrões de *splicing*, estaremos frente a frente com dois genes moleculares processuais distintos.

Este conceito tem, contudo, algumas consequências que se afiguram problemáticas (cf. Moss, 2001). Primeiro, torna-se bastante difícil individualizar genes quando eles são tratados como processos, conforme proposto pelo conceito de Griffiths e Neumann-Held, em virtude da extrema dependência da expressão gênica relativamente ao contexto. Segundo, esse conceito aumenta substancialmente o número de genes em eucariotos, por exemplo, como uma decorrência do grande número de isoformas de transcritos e, logo, de polipeptídeos gerados pelo *splicing* alternativo. Terceiro, ele torna necessário incluir nos genes os sistemas multimoleculares associados com a transcrição e o *splicing*, fazendo com que o gene molecular processual salte do nível molecular para um nível superior na hierarquia biológica. Nós retornaremos a esses problemas mais à frente.

Como vimos, para Fogle (1990), o problema com o conceito molecular clássico de gene reside na superposição da ideia mendeliana de unidade. Ele indica, assim, a necessidade de redefinir o gene, de modo a retirar do mesmo a ideia de que ele seja uma unidade de estrutura, função e informação. Ele atribui ao gene o caráter de um “conjunto”, entendendo-o como um produto da reunião de domínios encontrados no DNA. Por “domínios”, ele entende sequências de nucleotídeos que podem ser distinguidas umas das outras com base nas suas propriedades estruturais e/ou atividades funções, como, por exemplo, éxons, íntrons, promotores, intensificadores (*enhancers*), operadores etc. Domínios podem ser combinados de variadas formas para formar genes, ou como escreve Fogle, “conjunto de domínios para a transcrição ativa” (DSAT, do inglês). Essa estrutura de conjuntos elimina a necessidade de encontrar uma unidade genética única que corresponderia a um gene. Um domínio pode fazer parte de mais de um gene. Os genes não se encontram no DNA, mas são construídos pela célula a partir de domínios, estes sim presentes no DNA. Dessa maneira, Fogle considera que se torna mais fácil acomodar fenômenos como genes sobrepostos ou *splicing* alternativo.

Certamente, os domínios são por ele entendidos como entidades reais. Não é tão claro, contudo, o estatuto dos DSATs. Eles são construtos instrumentais, aos quais não se deve exigir uma hipótese de correspondência com alguma entidade real, ou eles são entidades reais? Os DSATs são “objetos epistêmicos”, no sentido explicado por Rheinberger (2000), ou seja, entidades introduzidas como alvos da pesquisa pelos cientistas? Ou eles são encontrados na própria célula, talvez na forma de mRNAs maduros? Ou, quiçá, o estatuto dos DSATs tem caráter intermediário, combinando uma hipótese de realidade com uma ação construtiva dos pesquisadores? A partir do que escreve Fogle, é difícil responder a estas questões.

Em uma tentativa de organizar a variedade de definições de gene encontrada na literatura, Moss (2001, 2003) propôs uma distinção entre dois modos de compreender o gene que são frequentemente confundidos por cientistas, por professores, pela mídia de divulgação e pela opinião pública. Trata-se de uma confusão com importan-

tes consequências sociais, porque dá força ao determinismo genético, à ideia de que uma série de características, mesmo bastante complexas, como vários traços comportamentais, a inteligência, a agressividade etc., são determinadas *apenas* por genes. Moss distingue entre o gene-P (o gene como determinante de fenótipos ou diferenças fenotípicas, sem quaisquer requisitos quanto a sequências moleculares específicas ou à biologia envolvida na produção do fenótipo) e o gene-D (o gene como um recurso desenvolvimental, que é, em si mesmo, indeterminado com relação ao fenótipo).

O gene-P é um conceito instrumental, que foca sobre o efeito distal do gene e a capacidade de previsão de fenótipos a partir de genótipos, sendo empregado em áreas como a biologia evolutiva, o melhoramento genético por seleção de cruzamentos, a genética de populações etc. Trata-se de um instrumento para a realização de algumas tarefas importante na genética, como a análise de genealogias ou heredogramas. Quando se fala de um gene no sentido do gene-P, fala-se como se ele causasse, sozinho, o fenótipo. Por exemplo, quando falamos em genes para olhos azuis, falamos como se houvesse genes que determinassem essa cor de olhos: esse é um gene-P. Contudo, se buscarmos no DNA um gene para olhos azuis, descobriremos que esse gene não existe. Olhos podem ficar menos pigmentados por uma diversidade de problemas na via de síntese de pigmentos na íris, que podem ter origem em mutações em uma diversidade de genes. Não há, pois, um único gene que possa ser caracterizado como o gene para os olhos azuis. Contudo, para entender o resultado de um cruzamento entre um pai de olhos castanhos e uma mãe de olhos azuis, podemos simplificar a situação e falar como se houvesse um gene que determina olhos azuis. Este gene-P é uma ficção útil para realizar essa tarefa da genética, a análise de heredogramas, porque permite a previsão confiável dos resultados de cruzamentos. Para compreender como a utilidade do conceito instrumental não se limita a essa tarefa, basta considerarmos sua importância no melhoramento genético.

Por sua vez, o gene-D é considerado uma entidade real, alguma sequência molecular no DNA. O gene-D é, pois, um conceito realista, que enfoca o efeito proximal do gene e a complexidade dos papéis desempenhados por ele no desenvolvimento e na fisiologia celular, sendo empregado usualmente em áreas como a biologia molecular, genômica, genética do desenvolvimento etc. Ele é um recurso, entre vários recursos igualmente importantes (genéticos, epigenéticos, ambientais), para que ocorra o desenvolvimento de características. Ele não determina, portanto, características fenotípicas. O gene-D cumpre papéis distintos do gene-P, em outras tarefas importantes desempenhadas por geneticistas e biólogos moleculares, como, por exemplo, na anotação e contagem de genes. No caso do gene-D, o conceito mendeliano de unidade não se mostra válido, como se pode depreender das discussões anteriores sobre a dificuldade de identificação das unidades estruturais e funcionais no DNA.

Moss alerta que ambos os modos de compreender os genes são válidos em contextos distintos, mas sua mistura indiscriminada constitui uma das fontes do determinismo genético, na medida em que a ideia de determinação, própria do gene-P, com seu caráter instrumental, é estendida indiscriminadamente ao gene-D. Os dois são conceitos distintos, que apresentam ideias diferentes sobre o que é um gene. Não existe qualquer pedaço de DNA ou qualquer outra coisa que seja simultaneamente gene-P e gene-D. Temos aqui um exemplo de como a hibridização de conceitos e modelos de genes pode ser perigosa: se confundirmos o gene-P, que é um construto instrumental e é pensado como se determinasse características, com o gene-D, acompanhado por uma hipótese de correspondência com alguma entidade real, mas que não determina características, concluiremos que, apesar de toda a complexidade do desenvolvimento, há características determinadas apenas por genes. Nós nos tornaremos convencidos de que o determinismo genético é correto. Uma das razões pelas quais as visões deterministas genéticas têm sido tão divulgadas e aceitas no discurso contemporâneo sobre genes, seja científico ou leigo, reside na hibridização dessas duas ideias.

Nas próximas seções, examinaremos duas tentativas recentes de ressignificação do conceito de gene. A primeira propõe uma definição cujo principal aspecto é satisfazer todos os achados do projeto Encode (cf. Gerstein *et al.*, 2007). A segunda busca preservar a noção dos genes como unidades funcionais (cf. Scherrer & Jost, 2007a, 2007b), no contexto de uma proposta de expansão do vocabulário a seu respeito.

5.1 A DEFINIÇÃO PÓS-ENCODE

O mais recente projeto a ter um impacto importante em nosso entendimento sobre os genes e o genoma é o projeto Encode (Enciclopédia de Elementos do DNA), levado a cabo por um consórcio internacional de cientistas que busca identificar as funções de vários tipos de elementos ou, nos termos de Fogle (1990), domínios conhecidos no DNA, como éxons, íntrons, promotores, terminadores etc.

Em sua discussão sobre as mudanças em nossa compreensão sobre os genes decorrentes do projeto Encode, Gerstein *et al.* (2007) afirmam que qualquer definição de gene deve obedecer aos seguintes critérios:

- (1) ela deve ser compatível com as definições passadas, ou seja, algo que já foi chamado de gene deverá tender a permanecer como gene;⁶

⁶ Os autores do presente artigo discordam entre si quanto a este critério. Leila Joaquim o aceita como um critério necessário para a busca de novas compreensões sobre genes, na medida em que considera que uma mudança conceitual radical pode trazer mais confusão do que benefícios à pesquisa. Charbel El-Hani, por sua vez, considera ser

- (2) deve ser independente do organismo, isto é, deve ser válida para toda a diversidade biológica, das bactérias aos animais;
- (3) deve traduzir uma ideia simples, em vez de listar vários mecanismos e exceções;
- (4) deve ser prática o suficiente para permitir que se responda facilmente quantos genes há em determinado genoma, ou qualquer pergunta oriunda da prática de contagem de genes, a qual é cada vez mais usual na pesquisa genética;
- (5) deve ser compatível com outros vocabulários biológicos, em especial, com o vocabulário associado ao reguloma, que representa o conjunto completo de interações regulatórias em um organismo.

Os cientistas do Encode argumentam que uma sequência, para ser um gene, deve satisfazer às seguintes condições:

- (a) O gene é uma sequência genômica (de DNA ou RNA) que codifica diretamente produtos moleculares funcionais, sejam RNAs ou proteínas;
- (b) Nos casos em que há vários produtos funcionais compartilhando regiões sobrepostas, entende-se como gene a união de todas as sequências genômicas sobrepostas que codificam os produtos funcionais;
- (c) Essa união deve ser coerente, isto é, feita separadamente para os produtos proteicos e de RNA finais, mas não requer que todos os produtos necessariamente compartilhem uma subsequência comum. Considerando essas condições, os cientistas do Encode definiram gene como a união de sequências genômicas que codificam um conjunto coerente de produtos funcionais potencialmente sobrepostos (Gerstein *et al.*, 2007, p. 676-7).

Os critérios que devem ser satisfeitos por uma definição de gene, de acordo com Gerstein *et al.* (2007), mostram-se relevantes, com a possível exceção do primeiro critério, ao menos para um dos autores deste artigo. Porém, não estamos certos de que seja possível a uma única definição obedecer a todos esses critérios. Em particular, questionaremos se a própria definição proposta por Gerstein *et al.* (2007) satisfaz o critério (3), ou seja, argumentaremos que a definição de gene proposta por eles não traduz uma ideia simples e não consegue evitar exceções, porque acaba colocando uma série de aspectos e implicações na definição, o que a torna complexa. Além disso, Scherrer e Jost (2007b) acusam a definição de Gerstein *et al.* de não cumprir o critério (5), por não dar a devida consideração ao reguloma.

esse um critério muito conservador, que limita desnecessariamente a possibilidade de que os novos achados sobre a complexidade e dinâmica dos genomas mudem radicalmente nossa compreensão do que deve contar como genes.

Uma primeira implicação da definição de Gerstein *et al.* que devemos levar em consideração é que diferentes produtos funcionais da mesma classe (proteínas ou RNAs) que se sobrepõem em seus usos de sequências de DNA são combinados no mesmo gene. O foco da definição está nos produtos e, em consequência disso, não existe uma relação de 1:1 entre uma sequência codificadora no nível do DNA e um produto funcional. Por exemplo, os produtos do *splicing* alternativo, como compartilham sequências em comum, são considerados, de acordo com esta definição, produtos de um único gene. Entretanto, diferentes produtos proteicos que se originam de um único e grande transcrito de mRNA policistrônico não são considerados como derivados de um único gene, se os produtos finais não compartilharem qualquer bloco de sequência. Neste caso, temos dois transcritos que se originam a partir do mesmo local de início de transcrição e, portanto, compartilham o mesmo promotor e elementos regulatórios, mas não são considerados produtos do mesmo gene.

Também em decorrência de dar ênfase aos produtos finais de um gene, a definição de Gerstein *et al.* é indiferente aos produtos intermediários originados de uma região genômica que se possam sobrepor. Em tal caso, não importa se um transcrito intrônico, por exemplo, compartilha sequências com um transcrito sobreposto, desde que seus produtos não compartilhem pedaços de sequências. A união dos segmentos define o gene, desde que cada éxon seja compartilhado por no mínimo dois membros desse grupo de produtos.

Outra implicação é que, em eucarioto, um gene pode não estar em um *locus* gênico discreto, ou seja, suas sequências codificantes podem estar espalhadas pelo genoma. Afinal, a definição não restringe os *loci* dos éxons que se combinam para codificar o produto final. Portanto, eles podem estar em diferentes fitas de um cromossomo ou mesmo em cromossomos separados e, ainda assim, pertencer ao mesmo gene. Para essa definição, o gene é um conjunto de sequências compartilhadas pelos produtos, não sendo necessário que essas sequências estejam conectadas, assim como sequências vizinhas podem, por sua vez, não fazer parte do mesmo gene.

Outro aspecto a ser considerado é que as regiões não traduzidas, como as UTRs, não são consideradas partes de um gene. TARs ficam como ‘supostos genes’ e demandam futuras investigações. Pseudogenes, mesmo quando transcritos, são ainda considerados não funcionais e, portanto, não são reconhecidos como verdadeiros genes, de acordo com a definição, a menos que no futuro a pesquisa mostre que eles têm funções. Isso contradiz argumentos discutidos acima, sobre os pseudogenes (cf. Balakirev & Ayala, 2003; Betrán *et al.*, 2002).

Uma última implicação da definição, que se mostra particularmente problemática, está relacionada às sequências regulatórias. Gerstein *et al.* (2007), embora reconheçam e deem importância ao papel crucial das regiões regulatórias na expressão

gênica, sugerem que elas não sejam consideradas na decisão de quais múltiplos produtos pertencem ao mesmo gene. Para eles, “a regulação é simplesmente muito complexa para ser incluída na definição de um gene, e há obviamente uma relação de muitos-para-muitos (ao invés de um-para-um) entre regiões regulatórias e genes” (Gerstein *et al.*, 2007, p. 677). Como regiões regulatórias não são traduzidas, apesar de terem um papel importante na expressão gênica, não são consideradas partes de gene, segundo estes autores. Para fazer referência a essas regiões, eles criaram uma categoria “especial”, referindo-se a elas como “associadas a genes” (Gerstein *et al.*, 2007, p. 678).

Com base nessa série de implicações, consideramos que a proposta de Gerstein *et al.*, por mais interessante e atraente que possa ser, não satisfaz o critério, estipulado por eles próprios, de ser uma definição simples e sem exceções. Ela representa uma tentativa de transferir a ênfase das sequências de DNA para os conjuntos de transcritos, não sendo bem sucedida, em nosso entendimento, quanto ao requisito de simplicidade. Cabe argumentar, contudo, se não seria o caso de que nenhuma definição de gene, para acomodar a complexidade e dinâmica dos genomas, consegue satisfazer esse requisito. Se esse for o caso, o argumento deixa, naturalmente, de ter grande peso na avaliação da proposta de Gerstein e colaboradores.

Outra limitação importante da definição de Gerstein *et al.*, que transparece nos argumentos acima e foi também indicada por Scherrer e Jost (2007a, 2007b), reside no preço a ser pago pelo abandono da noção de gene como unidade codificante e funcional, a saber, pela supressão dos efeitos regulatórios que mediam esses dois aspectos. Mais abaixo, discutiremos a proposta de Scherrer e Jost, na qual as complicações do processo de regulação são colocadas em evidência e retoma-se o esforço de definir o gene como unidade básica de função.

Uma avaliação da definição de Gerstein e colaboradores pode conduzir, contudo, a outra maneira de ver a situação. A complexidade que encontramos no uso da definição, assim como as exceções que podem ser visualizadas mostram a impossibilidade de dar conta da diversidade, complexidade e arquitetura genômicas com base em um único modelo ou conceito de gene. Assim, não nos parece que se possa assumir que qualquer uma das recentes propostas de revisão do conceito de gene possa dominar o cenário da genética, em todos os seus programas de pesquisa e subdisciplinas. Retornaremos a este ponto em nossas considerações finais.

5.2 O GENE VOLTA A SER UNIDADE FUNCIONAL E NÃO ESTÁ NO DNA

Para Scherrer e Jost (2007a, 2007b), a proposta de Gerstein e seus colaboradores tem como objetivo uma descrição sistemática e a classificação de transcritos, o que leva a

um híbrido conceitual, misturando aspectos funcionais e de codificação, sem dar a devida atenção aos aspectos regulatórios. A proposta deles, por sua vez, é centrada na ideia de que, como há dois aspectos distintos envolvidos na produção de polipeptídeos, então dois conceitos são necessários. O primeiro é o conceito de gene, para referir ao aspecto de tradução de trincas de nucleotídeos em aminoácidos, e o segundo é o conceito de “genon”, introduzido por eles para fazer referência à regulação da expressão do conjunto de sequências de trincas, desde a iniciação da transcrição até o mRNA maduro e a tradução. O termo “genon” é uma contração de “gene” e “operon”. Trata-se de uma tentativa de voltar a focar a compreensão dos genes no aspecto funcional, incluindo não somente a distinção entre gene e genon, mas também a utilização de outros conceitos novos, como os de transgenon, protogenon, pré-genon, gene de proteína (P-gene), gene de RNA (R-gene), gene estrutural (s-gene) e gene de regulação (c-gene), como veremos a seguir.

Uma consequência interessante desta abordagem é a de que, como no caso de outras propostas que examinamos acima, não se sustenta que o gene se encontre no DNA. Scherrer e Jost localizam o gene na sequência ininterrupta de ácidos nucleicos que emerge apenas no nível do mRNA, antes da tradução. Eles argumentam que a sequência ininterrupta de mRNA é a unidade de função e de análise genética, uma vez que, ao ser traduzida fielmente, constitui o equivalente da cadeia de polipeptídios produzida. Eles definem gene, então, como “o trecho de ácido nucleico ininterrupto da sequência codificante no mRNA que corresponde a um polipeptídeo ou algum outro produto funcional” (Scherrer & Jost, 2007b, p. 106).

Na sua definição, os autores também adicionam à sequência do mRNA as sequências regulatórias no transcrito e os produtos que atuam sobre a regulação gênica. Para dar conta desses fatores adicionados à sequência codificante, eles cunharam os termos “genon” e “transgenon”. O genon refere-se ao programa associado à sequência codificadora no mesmo cromossomo do qual foi transcrito o mRNA (isto é, em *cis*), que regula a transcrição de um gene. Genons individuais estão contidos no pré-mRNA, formando o pré-genon. Um domínio genômico contém um protogenon, que apresenta os sinais de ativação da transcrição, além do pré-genon encontrado nos transcritos. O conjunto dos fatores regulatórios codificados por outros cromossomos (isto é, em *trans*), que interagem com um dado genon, é chamado por eles de transgenon. Portanto, a expressão gênica é governada por sequências codificantes e pelo genon. O conceito de gene implica, de um lado, o programa em *cis* carregado pelo mRNA durante o processo, o genon, e outro programa em *trans*, constituído pelo transgenon, representando fatores controladores dos processos envolvidos na expressão gênica.

A razão pela qual o gene não pode, em muitos casos encontrados nos eucariotos, mas também em arqueobactérias, ser diretamente identificado no nível do DNA, resi-

de em fenômenos discutidos acima, como os genes interrompidos, o *splicing* alternativo e a edição do mRNA, sendo estes últimos processos regulados pelo transgenon. Após o processamento, ele finalmente emerge como uma sequência ininterrupta de ácidos nucleicos no nível do mRNA, antes da tradução, com uma correspondência fiel à sequência de aminoácidos produzida na síntese de um polipeptídeo. Depois da tradução, o genon termina seu papel e desaparece.

Esta definição separa o gene do conjunto de elementos que é requerido para orquestrar sua expressão, ou seja, para a regulação da expressão gênica. Os elementos são capturados pelos conceitos de genon e transgenon. Scherrer e Jost entendem que, dessa maneira, o gene fica livre dos aspectos regulatórios e, assim, de todas as limitações que o impedem de assumir o papel de unidade funcional. Portanto, nos termos dessa definição, o gene volta a ser uma unidade de função, mas, em contraste com as visões anteriores, ele não possui correspondência com um *locus* no DNA.

Os produtos da expressão gênica podem ser proteínas ou RNAs e podem ter função estrutural ou enzimática ou, ainda, controlar a expressão gênica, desempenhando um papel regulatório. Por essa razão, Scherrer e Jost propõem uma distinção entre genes de proteínas e genes de RNA, P-genes e R-genes, respectivamente, e entre genes estruturais e regulatórios, s-genes (de “*structural*”) e c-genes (de “*control*”), respectivamente.

Consideramos que a proposta de definição de gene de Scherrer e Jost dá um passo à frente, com relação à proposta de Gerstein *et al.* (2007), pelo fato de introduzir outros conceitos para abranger a complexidade do processo de expressão gênica. O fato de o gene não estar localizado ao nível do DNA representa uma ideia inovadora, possivelmente necessária para superar a crise do conceito de gene. Contudo, talvez esta redefinição de gene, ao retirá-lo do nível do DNA, encontre os mesmos obstáculos que a proposta do abandono do conceito de gene encontrou.⁷ Uma das dificuldades de dissociar o gene do DNA é que a associação está muito presente, tanto no discurso científico quanto no leigo, tanto na comunidade científica quanto nas salas de aula e na sociedade, em documentos governamentais e na opinião pública. Em outras palavras, convencer a todos de que o gene não está no DNA não parece ser tarefa fácil. Todavia, não devemos esquecer que o conceito de gene chegou ao discurso público a partir do discurso científico e, portanto, se novos conhecimentos alcançados sobre genes e sua função nos sistemas vivos demandam superar visões que se tornaram socialmente arraigadas, a tarefa da comunidade científica é, precisamente, contribuir para tal superação.

A dissociação dos genes e do DNA pode ser o preço a ser pago para manter a noção do gene como unidade funcional, mas talvez fosse mais fácil simplesmente romper

⁷ A proposta de abandono do conceito de gene não foi aceita, como podemos ver em Coyne, 2000; Magurran, 2000; Maynard Smith, 2000; Hall, 2001; Wilkins, 2002; Moyle, 2002; Judson, 2001.

com a noção de unidade do que efetuaríamos tal dissociação. O caminho da reformulação da definição de gene pode não estar em esforços persistentes para tratar o gene como uma unidade, seja de estrutura ou de função. Esse modo de compreender os genes pode ser, em si mesmo, a principal dificuldade para o devido reconhecimento da complexidade da arquitetura gênica. Por fim, convém considerar se uma proposta como a de Scherrer e Jost, que cria uma série de novos termos para tratar dos genes, de sua expressão e de sua regulação, não enfrenta obstáculos de ordem prática para sua disseminação. Afinal, ela aumenta substancialmente a complexidade do vocabulário da genética.

CONCLUSÃO

O termo “gene” é hoje usado na pesquisa genética não mais para referir-se a uma única entidade, mas como uma palavra de grande plasticidade, definida pelo contexto experimental em que é usada. Os genes tornaram-se objetos epistêmicos, como argumenta Rheinberger (2000). Uma série de descobertas sobre os genes e os processos de expressão gênica dificulta a interpretação do gene como unidade de estrutura e/ou função. Frente à complexidade do genoma e da maquinaria celular, a proposta de uma relação de 1:1:1 entre um gene, um produto proteico e uma função mostra-se insustentável. É evidente, assim, que essas descobertas representam desafios ao conceito molecular clássico.

Contudo, a ideia de que os genes devem ser entendidos como unidades de função ainda permanece no cenário da pesquisa atual, como mostram as definições propostas por Scherrer e Jost (2007a, 2007b). Podemos encontrar nas várias reações à crise do conceito de gene uma convergência para a ideia de que os genes não se encontram no DNA, mas são construídos pela célula a partir de sequências de DNA. Esta é uma ideia que encontramos em Fogle (1990), Pardini e Guimarães (1992), Keller (2005), Scherrer e Jost (2007a, 2007b) e El-Hani; Queiroz e Emmeche (2009). Nestes termos, os genes seriam encontrados em mRNAs maduros, processados, em vez de estarem presentes no DNA.

Para apreciar a proposta, vale a pena considerar um dilema decorrente do fenômeno do *splicing* alternativo. Por um lado, o segmento de DNA, que é transcrito em um único pré-RNA, poderia ser chamado de gene, apesar de gerar muitos produtos polipeptídicos. Essa visão está presente na definição proposta por Gerstein e colaboradores (Gerstein *et al.*, 2007), de acordo com a qual um gene é uma união de sequências genômicas que codificam um conjunto coerente de produtos funcionais potencialmente sobrepostos. Por outro lado, poderíamos chamar de gene cada mRNA processado in-

dividual, assumindo, então, uma relação ‘um mRNA maduro-uma proteína’. É assim como Scherrer e Jost (2007a, 2007b), por exemplo, buscam salvar a ideia do gene como unidade de função.

Uma anomalia enfrentada por essa posição decorre do fato de que o mesmo mRNA maduro pode estar sujeito a modos alternativos de tradução (cf. Pardini & Guimarães, 1992) e, nesse caso, não resultará em uma função única. Mesmo que tratemos, contudo, tais casos como excepcionais, há outras consequências contraintuitivas da ideia de que os genes estariam presentes em mRNAs maduros. Dessa perspectiva, genes existiriam no zigoto apenas como possibilidades e não apresentariam a permanência e a estabilidade tipicamente atribuídas ao material genético. Eles não seriam encontrados nos cromossomos e, por vezes, nem mesmo no núcleo (cf. Keller, 2002). Todos esses aspectos parecem bastante contraintuitivos, diante da ideia de que os genes são unidades de transmissão e herança. Esta é, no entanto, uma ideia anterior ao próprio conceito de gene, que se encontrava, por exemplo, na ideia mendeliana de fator. Suponha-se, contudo, que rejeitemos a ideia de que os genes sejam unidades de transmissão e herança. O que perderíamos com isso? Certamente, precisaremos encontrar outras entidades na célula às quais possamos atribuir o papel de mediar a transmissão e a herança, já que genes serão entidades muito mais transitórias do que costumávamos pensar. Parece-nos que a ideia de genes localizados no mRNA maduro e não no DNA pode tornar-se aceitável, se pensarmos que não são os genes as unidades de transmissão e herança, que são passadas de uma geração a outra de maneira a manter, em grande medida, sua estabilidade, mas as regiões cromossômicas, que raramente estão separadas dos eventos de recombinação que têm lugar na meiose. A convergência que encontramos na literatura, na direção da ideia de que os genes não se encontram no DNA, mas em mRNAs maduros e que são construídos pela célula a partir de sequências de DNA, pode indicar, assim, um caminho promissor e importante para as investigações futuras acerca do significado dos genes.

Outra convergência encontrada nas reações à crise do conceito molecular clássico reside no compromisso com visões processuais, ainda que este seja um ponto de menor acordo. Esta é uma visão sugerida pelo argumento de Keller (2005) de que o conceito de gene pode sobreviver no século XXI caso seja reconceitualizado como verbo. Ela é desenvolvida, contudo, em outros trabalhos, como os de Griffiths e Neumann-Held (1999), Neumann-Held (2001) e El-Hani; Queiroz e Emmeche (2009). Contudo, esta caracterização conceitual enfrenta as dificuldades que foram mencionadas acima.

Para discutirmos essas dificuldades, vamos supor, para fins de argumento, que a comunidade científica tenha chegado à conclusão de que os genes são processos (cf. El-Hani; Queiroz & Emmeche, 2009). Consideremos, então, a dificuldade de individuar genes quando os tratamos como processos, uma das dificuldades apontadas por

Moss (2001). Parece-nos que a visão dos genes como processos não cria de fato um problema para a taxonomia dos genes, mas apenas põe em destaque um problema que já existe na genética atual, diante da dependência da expressão gênica em relação ao contexto. De qualquer modo, caso tivéssemos concluído que genes são processos, este seria simplesmente um problema a ser enfrentado, mediante a construção de um sistema capaz de classificar genes moleculares processuais. Em uma teoria focada em processos, em vez de entidades, é esperado que se enfrentem dificuldades na construção de uma taxonomia de categorias ontológicas altamente contingentes, como os processos. Devemos, contudo, evitar a construção de tal teoria simplesmente porque antecipamos tais dificuldades? Caso uma teoria sobre os genes como processos mostre-se mais poderosa na explicação dos fenômenos estudados pela genética, será preciso construir uma taxonomia operacionalizável de genes como processos. Indicar que a dificuldade aparecerá não é razão para não construir tal teoria, caso se conclua que ela mostra maior poder explicativo do que a atual compreensão dos genes como entidades (ou, mais precisamente, padrões de entidades, como as sequências de DNA).

Quanto ao problema do aumento do número de genes em eucariotos, quando os genes são tratados como processos, podemos construir um argumento semelhante. Se o resultado de uma teoria mais poderosa sobre os genes for o aumento do número de genes eucarióticos, isso realmente importa? Não faria sentido abandonar o projeto de construir tal teoria, caso tivesse ficado claro que ela resulta em significativo avanço na compreensão dos genes, por conta de o número de genes investigados tornar-se maior.⁸

Finalmente, chegamos ao último problema apontado por Moss: a inclusão no gene de processos realizados por sistemas multimoleculares complexos, tais como a transcrição e o *splicing*, faz com que os genes passem do nível molecular para um nível superior da hierarquia biológica. Embora esse problema possa ser considerado importante, mostrando-se mais difícil de enfrentar do que os anteriores, é fundamental ter em conta que hierarquias não se encontram no mundo, pura e simplesmente, mas constituem modelos que são construídos de uma certa perspectiva teórica, para enfrentar problemas de pesquisa específicos. Um sistema pode ser decomposto em diferentes hierarquias ou conjuntos observacionais (cf. O'Neill *et al.*, 1986). Um conjunto observacional é um modo particular de ver os fenômenos de interesse, as medidas específicas e as técnicas de análise de dados. Portanto, não se trata de discutir em qual nível hierárquico genes se encontram no mundo, mas, antes, qual o poder explicativo, preditivo e heurístico, mostrado pelas diferentes propostas de decomposição dos

⁸ Cf. El-Hani; Queizoz & Emmeche, 2009. Os autores argumentam, inclusive, que uma visão dos genes como processos pode ajudar a enfrentar problemas no domínio da contagem de genes, uma tarefa epistêmica cada vez mais importante na pesquisa contemporânea sobre os sistemas genéticos.

sistemas celulares face às questões de pesquisa investigadas pela comunidade de geneticistas. Se chegarmos à conclusão, eventualmente, de que, ao menos em certas tarefas epistêmicas, é mais pertinente e poderoso situar genes acima do nível molecular, devemos, pura e simplesmente, construir modelos nos quais os genes sejam vistos dessa maneira.

Vale a pena considerar, ademais, que se a interpretação de genes como processos enfrenta dificuldades, a situação não é diferente no caso de propostas que situam os genes no DNA ou no RNA, como discutimos no presente artigo. Pareceria uma decisão apressada, portanto, deixar de lado as visões dos genes como processos, diante dos problemas discutidos acima. A resistência a essas visões está relacionada à ideia de que os genes, não importa como os entendamos, são segmentos de DNA. Ela encontra-se profundamente arraigada nas compreensões do conceito construídas ao longo do século xx. Ela pode ser constitutiva, no entanto, do quadro conceitual das dificuldades que enfrentamos para dar conta da complexidade e da dinâmica do genoma. Como Keller escreve:

Virtualmente toda propriedade biologicamente significativa convencionalmente atribuída ao DNA – incluindo sua estabilidade – é, de fato, uma propriedade relacional, uma consequência das interações dinâmicas entre o DNA e muitos processadores proteicos que convergem sobre ele. O próprio significado de qualquer sequência de DNA é relacional – para o propósito de compreender o desenvolvimento ou a doença, são os padrões de expressão gênica que realmente importam, e esses padrões estão sob o controle de um aparato regulatório muito complexo, e não podem ser preditos apenas a partir do conhecimento sobre a sequência (Keller, 2005, p. 4).

A compreensão dos genes como processos é favorecida pelo reconhecimento da dependência do DNA em relação a processos de expressão gênica e complexas redes regulatórias encontradas no ambiente celular e supracelular, caso ele venha a fazer qualquer diferença para os sistemas vivos. Tal interpretação nos distancia de visões que buscam genes como unidades estruturais e/ou funcionais no DNA, rumo a novos modos de pensar sobre a função biológica, nos quais a função não é encontrada em genes particulares, seja no DNA ou no RNA, mas em redes comunicativas, informacionais, encontradas nos sistemas vivos, dos quais DNA, RNA e proteínas são parte (cf. Keller, 2005).

Mesmo que, eventualmente, cheguemos à conclusão de que devemos manter uma visão dos genes como unidades funcionais, quem sabe situando-os no nível do mRNA maduro, parece-nos que ainda precisaremos de algum conceito que capture a visão processual mencionada acima. Neste ponto, podemos retomar a sugestão de Keller de

que, para enfatizar os processos celulares, em vez de entidades, podemos transformar a biologia que construímos a partir de substantivos, ao redor de entidades, em uma nova biologia, edificada a partir de verbos, construída ao redor de processos. Nesses termos, ao tratar do modo como a célula constrói genes a partir de sequências de DNA, podemos, além de situar os genes no mRNA maduro, também construir uma linguagem adequada para explicar como a célula “geneia” (cf. Santos, 2008).

Por fim, com relação à tensão entre as visões instrumentalistas e realistas na história do conceito de gene, concordamos com Falk (1986) em que as últimas décadas assistiram a um aumento da importância da visão instrumentalista. Genes têm sido cada vez mais tratados, explícita ou implicitamente, como construtos instrumentais, elaborados por comunidades de pesquisadores com a finalidade de obtenção dos dados empíricos em suas áreas. O que Falk escreveu há mais de duas décadas parece inteiramente atual:

Hoje o gene não é a unidade material ou a unidade instrumental da herança, mas é, antes, *uma* unidade, *um* segmento que corresponde à função de *uma* unidade, conforme definida pelas necessidades do experimentador (Falk, 1986, p. 169; grifos do autor).

Diante do contexto complexo em que vivemos, na atual genética em franca transformação, marcada por achados que dificultam a manutenção de visões anteriores sobre os genes e sua função, por uma diversidade de reações à crise do conceito molecular clássico, e por uma renovada tensão entre visões instrumentalistas e realistas sobre os genes, parece-nos interessante retomar um argumento formulado por El-Hani (2007). No caso de muitas definições de conceitos biológicos, como os de espécie, organismo, vida, tornou-se claro que a própria demanda por uma generalidade das definições não se mostrava adequada. No caso do conceito de gene, parece-nos apresentar-se a mesma situação. Uma definição de gene não precisa ser inteiramente geral para ser útil. Em ciências tão diversificadas, tais como a genética e a biologia molecular, que abrangem uma diversidade de campos de investigação, parece razoável pensar que uma variedade de modelos e definições de genes, com domínios bem delimitados de aplicação, pode dar conta de maneira mais apropriada das tarefas epistêmicas levadas a cabo pelos pesquisadores, do que uma definição ou modelo universal. Desta perspectiva, trata-se não tanto de escolher uma entre as várias definições/modelos de gene presentes na literatura, e discutidas no presente artigo, mas de organizar a diversidade de definições, de modo a compreender de maneira mais clara e precisa como elas permitem enfrentar as tarefas colocadas para os pesquisadores nos diversos campos da genética e da biologia molecular.☞

Leyla MARIANE JOAQUIM

Doutora pelo Programa de Pós-Graduação em

Ensino, Filosofia e História das Ciências,

Universidade Federal da Bahia /

Universidade Estadual de Feira de Santana, Brasil.

leylamariane@gmail.com

Charbel NIÑO EL-HANI

Professor Doutor do Departamento de Biologia Geral,

Instituto de Biologia, Universidade Federal da Bahia.

Pesquisador do Conselho Nacional de Pesquisa Científica,

CNPq, Brasil.

charbel.elhani@pesquisador.cnpq.br

ABSTRACT

The gene concept has played a central role in Biology since its introduction, in the beginnings of the twentieth century. However, throughout its historical development, this concept has been a matter of increasing controversy, initially in the philosophy of biology and, later, in biology itself. Challenges to the gene concept have resulted in the difficulty of preserving the so called classical molecular concept, according to which a gene is a stretch of DNA encoding a functional product (polypeptide or RNA). The last three decades of experimental studies led to findings such as interrupted genes, alternative splicing, so called junk DNA, TAR sequences, pseudogenes, postranscriptional regulation, RNAi and RNAsi, among others, that posed unexpected difficulties to the usual understanding of the gene concept. In this paper, we address the main experimental findings that challenge the classical molecular gene concept. We focus, in particular, on recent advances that took place in the Human Genome Project (HGP) and the Encyclopedia of DNA elements (Encode). It is now clear for that a careful analysis and reformulation of this central concept for biological thought is necessary. In an attempt to organize the variety of definitions given to genes, many philosophers and scientists presented interesting views about this concept and its role in biological thought, as well as proposals of conceptual revision, which we will also discuss in this paper. We conclude that a single, all-encompassing definition of gene is neither possible nor necessary. Rather, a pluralism of models and concepts is likely to be more powerful, provided that the domains of each concept or model be clearly defined.

KEYWORDS • Classical molecular concept. Challenges. Human Genome Project. Encode.

REFERÊNCIAS BIBLIOGRÁFICAS

- BAEK, D. et al. The impact of microRNAs on protein output. *Nature*, 455, p. 64-71, 2008.
- BALAKIREV, E. S. & AYALA, F. J. Pseudogenes: are they “junk” or functional DNA? *Annual Review of Genetics*, 37, p. 123-51, 2003.
- BETRÁN, E. et al. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Molecular Biology and Evolution*, 19, p. 654-63, 2002.
- BEURTON, P. J.; FALK, R. & RHEINBERGER, H. J. (Ed.). *The concept of the gene in development and evolution*. Cambridge: Cambridge University Press, 2000.
- COYNE, J. A. The gene is dead: long live the gene. *Nature*, 408, p. 26-7, 2000.
- D’OTTAVIANO, I. M. L. & GONZÁLEZ, M. E. Q. (Ed.). *Auto-organização: estudos interdisciplinares*. Campinas: CLE-Unicamp, 2000. v. 2.
- EL-HANI, C. N. Controvérsias sobre o conceito de gene e suas implicações para o ensino de genética. In: NARDI, R. (Ed.). *Atas do V Encontro Nacional de Pesquisa em Educação em Ciências (Enpec)*. Bauru: Abrapec, 2005. p. 178-90.
- _____. Between the cross and the sword: the crisis of the gene concept. *Genetics and Molecular Biology*, 30, 2, p. 297-307, 2007.
- EL-HANI, C. N.; QUEIROZ, J. & EMMECHE, C. A semiotic analysis of the genetic information system. *Semiotica*, 160, p. 1-68, 2006.
- _____; ____ & _____. *Genes, information, and semiosis*. Tartu: Tartu University Press, 2009.
- EPP, C. D. Definition of a gene. *Nature*, 389, p. 537, 1997.
- FALK, R. What is a gene? *Studies in the History and Philosophy of Science*, 17, p. 133-73, 1986.
- FIRE, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391, p. 806-11, 1998.
- FOGLE, T. Are genes units of inheritance? *Biology and Philosophy*, 5, p. 349-71, 1990.
- _____. The dissolution of protein coding genes. In: BEURTON, P. J.; FALK, R. & RHEINBERGER, H. J. (Ed.). *The concept of the gene in development and evolution*. Cambridge: Cambridge University Press, 2000. p. 3-25.
- GERICKE, N. & HAGBERG, M. Definition of historical models of gene function and their relation to students understandings of genetics. *Science and Education*, 16, p. 849-81, 2007.
- GERSTEIN, M. & ZHENG, D. The real life of pseudogenes. *Scientific American*, 295, p. 48-55, 2006.
- GERSTEIN, M. B. et al. What is a gene, post-Encode? History and updated definition. *Genome Research*, 17, p. 669-81, 2007.
- GELBART, W. Databases in genomic research. *Science*, 282, p. 659-61, 1998.
- GILBERT, W. Why genes in pieces? *Nature*, 271, p. 501, 1978.
- GINGERAS, T. R. Origin of phenotypes: genes and transcripts. *Genome Research*, 17, p. 682-90, 2007.
- GLASER, V. Tapping miRNA-regulated pathways. *Genetic Engineering & Biotechnology News*, 28, 5, 2008.
- Disponível em: <<http://www.genengnews.com/articles/chitem.aspx?aid=2382>>. Acesso em: 12 dez. 2009.
- GRIFFITHS, P. E. & NEUMANN-HELD, E. The many faces of the gene. *BioScience*, 49, 8, p. 656-62, 1999.
- GUIMARÃES, R. C. & MOREIRA, C. H. C. O conceito sistêmico de gene – uma década depois. In: D’OTTAVIANO, I. M. L. & GONZÁLEZ, M. E. Q. (Ed.). *Auto-organização: estudos interdisciplinares*. Campinas: CLE-Unicamp, 2000. v. 2, p. 248-80.
- HALL, B. K. The gene is not dead, merely orphaned and seeking a home. *Evolution and Development*, 3, p. 225-8, 2001.
- JACOB, F. & MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. *Molecular Biology*, 3, p. 318-56, 1961.

- JUDSON, H. F. Talking about the genome. *Nature*, 409, p. 769, 2001.
- KAPRANOV, P. et al. Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, 296, p. 916–9, 2002.
- KELLER, E. F. *O século do gene*. Belo Horizonte: Crisálida, 2002.
- _____. The century beyond the gene. *Journal of Biosciences*, 30, 1, p. 101–8, 2005.
- LANDER, E. S. et al. Initial sequencing and analysis of the human genome. *Nature*, 409, p. 860–921, 2001.
- LEITE, M. Retórica determinista no genoma humano. *Scientiae Studia*, 4, 3, p. 421–52, 2006.
- MAGURRAN, A. Backseat drivers, review of the century of the gene by E. F. Keller. *New York Times Book Reviews*, 10 de dezembro, p. 26, 2000.
- MAYNARD SMITH, J. The cheshire cat's DNA. *The New York Review of Books*, 47, p. 43–6, 2000.
- MODREK, B. & LEE, C. A genomic view of alternative *splicing*. *Nature Genetics*, 30, p. 13–9, 2002.
- MOSS, L. Deconstructing the gene and reconstructing molecular developmental systems. In: OYAMA, S.; GRIFFITHS, P. E. & GRAY, R. D. (Ed.). *Cycles of contingency: developmental systems and evolution*. Cambridge: The MIT Press, 2001. p. 85–97.
- _____. *What genes can't do*. Cambridge: The MIT Press, 2003.
- MOYLE, L. Most ingenious: troubles and triumphs of a century of genes. *Biology and Philosophy*, 17, p. 715–27, 2002.
- NEUMANN-HELD, E. Let's talk about genes: the process molecular gene concept and its context. In: OYAMA, S.; GRIFFITHS, P. E. & GRAY, R. D. (Ed.). *Cycles of contingency: developmental systems and evolution*. Cambridge: The MIT Press, 2001. p. 69–84.
- NEUMANN-HELD, E. & REHMANN-SUTTER, C. *Genes in development*. Durham: Duke University Press, 2006.
- O'NEILL, R. V. et al. *A hierarchical concept of ecosystems*. Princeton: Princeton University Press, 1986.
- OYAMA, S.; GRIFFITHS, P. E. & GRAY, R. D. (Ed.). *Cycles of contingency: developmental systems and evolution*. Cambridge: The MIT Press, 2001.
- PARDINI, M. C. & GUIMARÃES, R. C. A systemic concept of the gene. *Genetics and Molecular Biology*, 15, p. 713–21, 1992.
- PORTIN, P. The concept of the gene: short history and present status. *Quarterly Review of Biology*, 56, p. 173–223, 1993.
- RHEINBERGER, H. J. Gene concepts: Fragments from the perspective of molecular biology. In: BEURTON, R.; FALK, R. & RHEINBERGER, H.-J. (Ed.). *The concept of the gene in development and evolution*. Cambridge: Cambridge University Press, 2000. p. 219–39.
- RUVKUN, G. Glimpses of the tiny RNA world. *Science*, 294, p. 797–9, 2001.
- SANTOS, V. C. *Genes, informação e semiose: Do conhecimento de referência ao ensino de biologia*. Salvador, 2008. Dissertação (Mestrado em Ensino). Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências, Universidade Federal da Bahia/Universidade Estadual Feira de Santana.
- SELBACH, M. et al. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455, p. 58–63, 2008.
- SCHERRER, K. & JOST, J. The gene and the genon concept: A functional and information-theoretic analysis. *Molecular System Biology*, 3, p. 1–11, 2007a.
- _____. The gene and the genon concept: Coding versus regulation. A conceptual and information-theoretic analysis storage and expression in the light of modern molecular biology. *Theory in Biosciences*, 126, p. 65–113, 2007b.
- SCHULZ, R. A. et al. Alternative *splicing* generates two distinct Eip28/29 gene transcripts in *Drosophila* Kc cells. *Proceedings of the National Academy of Sciences USA*, 83, p. 9428–32, 1986.
- SMITH, M. & ADKISON, L. Updating the model definition of the gene in the modern genomic era with implications for instruction. *Science Education*, 19, p. 1–20, 2010.

- SOREK, R. et al. How prevalent is functional alternative splicing in the human genome? *Trends in Genetics*, 20, 2, p. 68-71, 2004.
- TAM, O. H. et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, 453, p. 534-8, 2008.
- THE ENCODE PROJECT CONSORTIUM. Identification and analysis of functional elements in 1% of the human genome by the Encode pilot project. *Nature*, 447, p. 799-816, 2007.
- TUSCHL, T. & MEISTER, G. Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431, p. 343-9, 2004.
- VENTER, C. et al. The sequence of the human genome. *Science*, 291, p. 1305-51, 2001.
- WAIZBORT, R. & SOLHA, G. C. F. Os genes interrompidos: o impacto da descoberta dos íntrons sobre a definição de gene molecular clássico. *Revista da Sociedade Brasileira de História da Ciência*, 5, p. 63-82, 2007.
- WATANABE, T. et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453, p. 539-43, 2008.
- WILKINS, A. S. Grappling with developmental complexity. *BioEssays*, 24, p. 1193-5, 2002.

