

Quebra-cabeças da complexidade

Conhecimento da seqüência de um genoma não basta para entender nossa carga genética

EMMANUEL DIAS NETO*

O estudo da biologia talvez seja tão antigo quanto o surgimento das estruturas cerebrais que permitiram o estabelecimento da linguagem e o desenvolvimento do pensamento consciente. Esse processo elaborativo fez com que a humanidade se ocupasse, há milhares de anos, com a compreensão de suas origens e também com os processos relacionados à vida, doença e morte. Nos últimos 50 anos, foi possível acompanhar descobertas fantásticas em diversas áreas do conhecimento. Na biologia, tivemos um salto tremendo, que culminou com o seqüenciamento completo do genoma humano alcançado recentemente. A habilidade de acumular uma vasta quantidade de informação genética nos permitiu conhecer a base estrutural de diversos genomas, desde as bactérias mais primitivas (chamadas de arqueobactérias) até o homem, passando por fungos, parasitas, vermes, plantas e modelos como a mosca-das-frutas e o camundongo.

Esse conhecimento acumulado, permite reconstruir a relação evolutiva de uma boa parcela dos seres vivos, recontando a história da vida em nosso planeta. A capacidade de ler genomas alterou de maneira fundamental o estudo da biologia, da medicina e de diversos campos associados, influenciando um variado grupo de indústrias, que incluem a química fina, a farmacologia, a agroindústria e outras. A grande quantidade de informação produzida representa uma rica fonte de informação que deve ser cuidadosamente estudada, de modo a permitir o avanço mais pro-

veitoso de nosso conhecimento. O maior impacto desses achados ainda está por vir, e certamente virá, quando realmente tivermos conseguido decifrar, compreender e associar as informações contidas em nosso genoma.

Para isso devemos, antes de tudo, ter consciência de que o conhecimento da seqüência completa de um genoma, apesar de ser uma peça importante, está longe de permitir, apenas por si, a montagem do intrincado quebra-cabeça de nossa complexidade. Para entender o que nos permite ter essa maravilhosa complexidade – nosso enorme repertório comportamental, a habilidade de ter ações conscientes, nossa capacidade criativa, musical e científica, a capacidade de aprender, a nossa memória, entre outras – não poderemos contar apenas com nossa carga genética de 3,2 bilhões de nucleotídeos e um número de genes não muito superior ao de uma mosca.

Devemos ter a consciência de que o domínio de um genoma significa a posse de um mapa. No caso do genoma humano, um mapa complexo, ainda não totalmente conhecido, que auxiliará enormemente na busca pelas origens das doenças, com base nas variações genéticas, na diversidade, complexidade e no comportamento das proteínas no interior das células. Estudos de neuropsiquiatria mostram que, em gêmeos monozigóticos separados ao nascimento e criados em ambientes distintos, a concordância para o desenvolvimento de doenças neuropsiquiátricas é de cerca de 50%.

Isso mostra que em certas circunstâncias existe um balanço entre a im-

portância dos genes e do ambiente na determinação de determinadas condições. Se por um lado a genética tem grande peso no desenvolvimento de doenças, os fatores ambientais também têm importância considerável. A genética não é absoluta. Ainda temos muito o que aprender no estudo das interações do genoma com o ambiente, além de conhecer e desvendar as sutilezas de nosso genoma.

Já sabemos muito, mas é pouco

É curioso observar que, mesmo após o entusiasmo gerado pela finalização dos rascunhos do seqüenciamento do genoma humano, um tremendo esforço ainda precisa ser feito para que possamos conhecer o significado da imensa maioria das seqüências obtidas. Uma das primeiras questões que surgem é: como identificar as regiões importantes do genoma? Como determinar sua função no organismo? As regiões do genoma com função mais óbvia são os genes, que se acham comprometidos com a produção de proteínas. No entanto, essas regiões estão restritas a cerca de 3% do nosso genoma. O número de genes preditos nos badalados trabalhos que descreveram o genoma humano ficou, segundo as estimativas mais baixas, na faixa de 30 mil – alguns estudos indicam haver até 120 mil genes. Mesmo com esse pequeno número, apenas para a metade deles encontramos algum tipo de domínio que permita a predição de atividade fisiológica.

Enquanto se acredita que as estimativas do número de genes devam

* EMMANUEL DIAS NETO é pesquisador do Instituto de Psiquiatria da Faculdade de Medicina da Universidade de São Paulo e um dos criadores do método ORESTES.

crescer, com o desenvolvimento de melhores programas computacionais de predição gênica e com o acúmulo de mais dados experimentais, está muito claro que o número de genes é apenas um dos mecanismos que criam a diversidade bioquímica necessária para fazer as proteínas. Em nosso genoma, os genes são formados por blocos de informação intercambiáveis, chamados exons, que são separados por blocos sem informação protéica, conhecidos por introns. Os exons podem ser recombinados como se combinam sílabas para formar uma palavra, formando mensagens distintas. Desse modo, a seqüência de um único gene pode começar e terminar em regiões

Blocos que são como sílabas de palavras

diferentes e sua porção interna pode ser montada alternando diferentes blocos, gerando proteínas com características funcionais distintas. Essas combinações (conhecidas como *splicing* alternativo) representam um eficiente mecanismo de geração de diversidade sem a necessidade de manter um imenso número de diferentes genes funcionais. Além dos eventos de *splicing* alternativo, diversos mecanismos conhecidos como epigenéticos, tais como metilação do DNA ou modificação de histonas, podem modular a expressão de um gene. Esses eventos de regulação epigenética regulam a atividade de genes silenciando sua atividade ou remodelando a estrutura dos cromossomos, expondo ou escondendo determinados genes de acordo com a necessidade de sua expressão. Dessa maneira, um complexo sistema de regulação intracelular é disparado, ligando ou desligando genes em determinados tecidos ou em fases específicas do desenvolvimento.

DNA lixo?

Os genes são distribuídos de modo desigual nos nossos cromossomos. Dados de seqüenciamento mostram que alguns cromossomos, como o 17, o 19 e o 22, são ricos em genes, quando comparados com os cromossomos 4, 8, 13, 18 e o cromossomo Y. O empacotamento do material genético no

núcleo de nossas células é um processo complexo, pois o DNA de uma única célula humana tem cerca de 2 metros de comprimento. Há alguns anos, descobriu-se que a distribuição dos cromossomos dentro das células, no processo de empacotamento, é extremamente organizada. Na periferia do núcleo celular ficam os cromossomos com menor densidade gênica, enquanto os cromossomos mais ricos se situam na porção mais interna do núcleo.

Foi demonstrado que essa distribuição cromossomal é regulada há pelo menos 30 milhões de anos, pois é conservada nos primatas. Essa conservação indica um papel funcional importante. Alguns pesquisadores sugerem que os cromossomos que possuem mais genes se encontrem na porção mais central dos núcleos, e os demais, ao seu redor, protegendo-os de agentes mutagênicos externos. Além disso, diversos trabalhos demonstram que ocorre um freqüente movimento dos cromossomos nos núcleos celulares. Essa dança dos cromossomos mostra que a estrutura do DNA e do seu empacotamento nas células não é nada rígida. Os cromossomos parecem se movimentar, possibilitando a troca de material genético entre si e a exposição de genes que devem ser ativados em determinadas circunstâncias. O estudo da distribuição cromossomal tem inclusive sido sugerido como um possível critério de diagnóstico para o câncer. Se apenas 3% do genoma codifica proteínas, será que o restante de nosso DNA é um resquício evolutivo que serve somente para proteção?

Uma das maneiras de analisar nosso genoma é compará-lo com o de outros organismos. É a chamada genômica comparativa. Esses estudos partem da premissa de que um bloco de DNA conservado por milhões de anos deve ter alguma função importante, que poderia ser comprometida se a seqüência fosse muito alterada. É a chamada conservação fisiológica. Estudos de genômica comparativa demonstraram que aproximadamente 95% de nosso genoma é muito parecido (possuindo, de fato, cerca de 99% de identidade) com o de um chimpanzé. No entanto, nosso tempo de divergência (o período de tempo que nos separamos de um ancestral comum) com os chimpanzés é de apenas 5 milhões de anos. Talvez esse

período não tenha sido longo o suficiente para que regiões não funcionais tivessem se diferenciado, e teríamos uma conservação passiva. Quando aprofundamos as comparações e investigamos as semelhanças que possuímos com o genoma do camundongo, cujo último ancestral comum com o homem existiu há 145 milhões de anos, vemos que uma significativa parcela desse DNA lixo ainda é conservada.

Se os chimpanzés são geneticamente muito próximos e não é possível distinguir conservação passiva de conservação funcional, o camundongo é muito distante, o que impede detectar mudanças no DNA adquiridas mais recentemente. Enquanto a comparação com o genoma de um organismo distante, como o do camundongo, oferece um painel importante de regiões genômicas com potencial funcional, a longa divergência entre as duas espécies não permite identificar algumas sutilezas. As regiões do genoma que se modificaram entre essas espécies e nos permitiram evoluir como primatas – e posteriormente como *Homo sapiens* – não estão no genoma de camundongo e deveriam ser encontradas de outra maneira.

Em um artigo publicado na revista *Science* no final de fevereiro (Boffelli *et al.*, 2003), um grupo de pesquisadores comparou regiões não-codificadoras do genoma humano com áreas semelhantes do genoma de outros primatas não-humanos. Os cientistas encontraram diversas regiões conservadas, mesmo quando foram usadas as espécies de primatas tropicais, muito distantes de nossa espécie. Conseguiram identificar elementos conservados e provar que têm atividade funcional: atuam na regulação da expressão gênica entre as diferentes espécies. Para isso, foram utilizadas diversas espécies de primatas, incluindo-se aí o DNA de vários primatas brasileiros. Fica clara a importância da biodiversidade para decifrar nosso genoma.

No entanto, cada genoma possui características únicas, que por seus reflexos funcionais permitem diferenciar as espécies. Como investigar as regiões funcionais (com atividade fisiológica) únicas do genoma humano? Sabemos que elas não estão restritas aos



genes ou às regiões conservadas em outros primatas. São características únicas da nossa espécie. Um trabalho muito interessante nesse sentido foi feito por pesquisadores de uma empresa norte-americana, em conjunto com um pesquisador do Instituto Nacional do Câncer, nos Estados Unidos (Kapranov *et al.*, 2002).

Usando a sequência dos cromossomos 21 e 22 humanos, os cientistas desenharam pequenos fragmentos de DNA artificial, cobrindo toda a sequência desses cromossomos em curtos intervalos de 35 nucleotídeos, os blocos constituintes do DNA. Os milhões de fragmentos produzidos foram usados para investigar se linhagens celulares humanas estariam produzindo RNA complementar a esses fragmentos. A estratégia comprovou a atividade funcional de novas regiões e permitiu uma análise transcricional em larga escala desses dois cromossomos humanos. Para surpresa de todos, um altíssimo percentual desses fragmentos mostrou estar associado a RNAs maduros das linhagens celulares. Os

autores demonstraram que as regiões ativamente transcritas de nosso genoma são, pelo menos, dez vezes mais extensas do que poderíamos imaginar. Talvez essas regiões contenham genes muito raros, ainda não demonstrados por nenhuma técnica, ou moléculas regulatórias ainda não conhecidas, mas de importância central para o conhecimento da fisiologia de nosso genoma. Dessa maneira, se antes imaginávamos que 3% do genoma continha genes, esse trabalho sugere que talvez esse percentual seja muito maior.

Genes e novas drogas

Dentre a fração de genes conhecidos atualmente, algumas centenas codificam proteínas potencialmente poderosas para o tratamento de doenças. Diversas dessas proteínas, assim como drogas baseadas em anticorpos monoclonais, estão em fase final de experimentação e algumas já são testadas em humanos. Sendo assim, buscam-se hoje mecanismos mais eficientes e de menor custo de produção de medica-

mentos. Uma das promessas é a manipulação genética de alimentos. Produzir um feijão mais nutritivo, milho com hormônio do crescimento humano ou cenouras com vacinas são sonhos que rondam as cabeças dos pesquisadores há anos.

Esses sonhos estão cada dia mais próximos, e um passo importante nesse sentido foi anunciado há algum tempo por uma empresa norte-americana após associação com o renomado Instituto Escocês Roslin (o mesmo que assombrou o mundo com a clonagem da ovelha Dolly), para dessa vez produzir drogas dentro de ovos de galinhas. Enquanto animais, como cabras, vacas, ovelhas e coelhos vêm sendo usados para produzir medicamentos em seu leite, a tecnologia de trabalho com aves surge com a promessa de ser mais rápida, barata e praticamente ilimitada graças à capacidade de produção de ovos. O primeiro produto deverá ser um anticorpo monoclonal dirigido ao combate do melanoma, um dos mais agressivos e comuns tumores que ocorrem no Brasil. O domínio dessa tecnologia, aliado à descoberta da totalidade dos genes humanos e à determinação de sua função biológica, permite imaginar um futuro promissor para essa nova forma de produção de medicamentos.

**Sonhos
estão cada
vez mais
próximos
do real**

Polimorfismos de DNA

Cada um dos bilhões de seres humanos de nosso planeta – com exceção dos gêmeos monozigóticos – possui seu próprio e único genoma. Apesar de serem únicos, os genomas de dois seres humanos não aparentados têm uma identidade média de 99,9%. Diante de um genoma de cerca de 3,2 bilhões de bases, a sutil diferença de 0,1% representa uma coleção de alguns milhões de nucleotídeos, responsáveis pela nossa fabulosa diversidade. A maioria dessas diferenças tem a forma de substituições ou polimorfismos de nucleotídeos únicos (conhecidos em inglês como SNPs, ou *Single Nucleotide Polymorphisms*). Os SNPs constituem um elemento-chave para compreen-

dermos a variabilidade genética humana e sua associação com diversas doenças. Recentemente houve um aumento explosivo no número de SNPs depositados nos bancos de dados públicos. Há cerca de um ano, apenas o banco dbSNP, ligado ao Instituto Nacional de Saúde dos Estados Unidos, tinha cerca de 4 milhões de SNPs depositados. Hoje esse número cresceu em cerca de 50%, ultrapassando os 6 milhões de SNPs. No entanto, apenas 0,3% dos polimorfismos já foi estudado de maneira mais aprofundada e numerosos polimorfismos ainda restam por serem descobertos.

No futuro, se poderá prevenir em vez de só curar

A literatura científica demonstra preocupação de que haveria um número imenso de polimorfismos ainda não revelados, mas de grande relevância, que não seriam encontrados apenas pelo uso de estratégias computacionais. Essa preocupação se deve ao fato de que a porção central dos genes estaria pouco representada (nos dados das chamadas ESTs ou etiquetas de seqüências expressas), enquanto os dados do Projeto Genoma Humano são baseados em um número muito limitado de indivíduos, levando a uma redução da população de polimorfismos. Nesse sentido, a iniciativa brasileira de geração de ESTs do tipo ORESTES foi extremamente positiva. No contexto do Human Cancer Genome Project (HCGP, financiado pelo Instituto Ludwig de Pesquisa do Câncer e pela FAPESP), o grupo brasileiro produziu 1,2 milhão de ESTs, derivadas da porção interna dos genes, um dos maiores conjuntos de dados mundiais. Além de serem derivadas de diversos tipos de tumores humanos, essas amostras são de grande valor, por derivarem de uma população com alto índice de mistura étnica, contribuindo com variações dificilmente encontradas em populações mais homogêneas. Os dados nacionais serão essenciais para atingirmos uma abrangente análise de polimorfismos clinicamente relevantes, buscando associações entre polimorfismos de DNA e doenças e avaliando os níveis de polimorfismos de nosso genoma.

Esses polimorfismos devem ser a chave para a predisposição ou proteção ao desenvolvimento de numerosas doenças, além de estar diretamente associados à maneira como diferentes pessoas respondem a drogas. Já sabemos, por exemplo, que a diferença de uma única base na seqüência do gene APOE confere maior suscetibilidade ao desenvolvimento do mal de Alzheimer e a doenças cardiovasculares. O conhecimento do efeito dessas alterações na nossa resposta a drogas abre um tremendo espaço para a farmacogenética. Sabemos que o custo para o desenvolvimento de uma nova droga está na faixa de US\$ 600 milhões. Por vezes, drogas extremamente eficazes para a imensa maioria da população devem ser retiradas do mercado, pois provocam efeitos graves em algumas pessoas. Esses estudos sobre polimorfismos alteram a evolução da medicina.

Nas terapias medicamentosas podemos imaginar o fim da abordagem de tentativa e erro. Na área clínica, podemos prever que a prevenção será privilegiada em relação ao tratamento das doenças. Para isso, a interação entre a pesquisa científica e a iniciativa privada é fundamental, devendo permitir a tradução e incorporação dos achados científicos no dia-a-dia das pessoas comuns. Apesar de o genoma ainda oferecer uma série de descobertas pela frente, o caminho já percorrido pela pesquisa científica permite uma série de possibilidades práticas atuais, prontas para serem implementadas na rotina de nossa sociedade.

Os estudos de polimorfismos também nos permitiram calcular a diversidade de DNA existente entre indivíduos das diferentes etnias humanas. Ao compararmos o DNA de dois indivíduos de uma mesma etnia, vemos que o número de diferenças encontradas é tão freqüente quanto as diferenças observadas entre indivíduos de etnias diferentes. Desse modo a ciência demonstrou que o conceito de raça, visto a partir das diferenças de DNA, não faz o menor sentido. Diversidade e individualidade são características das mais fundamentais de cada ser humano. Os dados do genoma humano permitem

que essas características sejam visualizadas de maneira muito clara e nos mostram que, por mais que sejamos seres únicos no Universo, temos muito em comum com o resto da humanidade. Nosso genoma pode ser visto como um patrimônio da nossa espécie. Deixando de lado uma postura antropocêntrica, o DNA mostra que todas as formas de vida conhecidas são codificadas pela mesma matéria-prima básica, os nucleotídeos que formam os genomas. Isso nos permite uma visão conceitual de que o evento do surgimento da vida no planeta deve ter sido único. A visão filosófica e até mesmo poética nos diz que somos todos membros de uma grande família, composta pelas mais diversas formas de vida do planeta. Todos tivemos um ancestral comum, e isso nos deve levar a refletir sobre a nossa postura diante de aspectos como poluição, degradação do ambiente e preservação da vida no planeta.

O futuro

Apesar de sempre podermos ser surpreendidos em nossas previsões, o estágio atual das pesquisas permite vislumbrar como será o cenário em um intervalo de 15-20 anos. Dentre amplas possibilidades, algumas poucas coisas parecem certas:

- Teremos uma lista abrangente de produtos gênicos humanos provendo um enorme potencial de drogas de reposição (de modo semelhante à insulina ou ao hormônio do crescimento recombinante hoje disponíveis), com um dramático efeito preventivo e curativo em diversas doenças.
- O prontuário médico em breve irá conter uma lista com o status de diversos polimorfismos ligados à farmacogenética, assim como a propensão ao desenvolvimento de uma série de doenças.
- A obtenção da seqüência genômica completa de um indivíduo deverá ser possível em poucos anos e devemos estar prontos para lidar com a manutenção dessa confidencialidade de maneira responsável.
- A terapia gênica deve tornar-se realidade para doenças causadas por alterações em um único gene. Genes de-



CLAUDIUS

feituosos poderão ser repostos por versões funcionais.

- A compreensão da base genética das doenças complexas permitirá o design de drogas racionais, dirigidas a vias metabólicas de funcionamento inadequado, eventualmente possibilitando a modelagem de estratégias preventivas.
- O conhecimento de alterações genéticas específicas de certos tumores permitirá o diagnóstico precoce da maioria dos tumores humanos.
- A indústria da farmacogenômica se estabelecerá de modo crescente, gerando uma medicina personalizada na qual drogas serão elaboradas de acordo com as feições genéticas de diferentes grupos de indivíduos.

Acredita-se que, por volta de 2010, marcadores genéticos efetivos estejam disponíveis para uma grande série de doenças e condições humanas. Estima-se que o custo para um teste diagnóstico, incluindo uma grande lista de marcadores, custe por volta de US\$ 100. À medida que testes que permitam avaliar a predisposição genética a cer-

tas doenças se tornem possíveis, a sociedade enfrentará questões que envolvem a disponibilidade dessas informações a empregadores ou a seguros-saúde. As leis devem proteger os cidadãos do uso inadequado dessas informações, e devemos questionar a validade de uso dessas informações no momento de decisões de contratação de pessoal.

Nos próximos anos, o público terá mais e mais oportunidades de fazer testes genéticos e especular sobre seu destino genético. É urgente que a legislação acompanhe os avanços científicos, incorporando e usufruindo as descobertas e impondo limites nas áreas mais delicadas. Sem debate público e controles apropriados, as pessoas poderiam ser discriminadas por causa de suas características genéticas. É preciso discutir o que a genética pode e não pode realizar, e que tipo de sociedade queremos.

A dupla hélice, com sua beleza e simplicidade, uniu de modo definitivo a bioquímica, a fisiologia e a genética. Sua estrutura oferece uma explicação imediata para os processos de cópia do DNA, mecanismos de herança genéti-

ca, mutação e diversidade. No entanto, a dupla hélice não esclareceu ainda os detalhes da interação entre a genética e o meio ambiente. A individualidade humana revelada pelo DNA faz com que diversos conceitos venham a ser revistos e a medicina passe novamente a focar o indivíduo. Predições em nível populacional não têm o mesmo poder preditivo em nível individual. Depois de termos caminhado tanto no conhecimento do “livro da vida”, talvez este seja o momento de re-avaliarmos nossas expectativas e o próprio conceito de nossa existência. Se por um lado temos nossa individualidade genética, nós também somos todos muito semelhantes e semelhantes às outras formas de vida do planeta.

Termino este texto com um repente com o qual tive contato durante uma viagem à cidade do Recife. A sabedoria popular nos surpreende, e acredito que o número de genes proposto nos versos deva estar muito mais próximo da realidade do que os 30 mil sugeridos nos trabalhos da *Nature* e da *Science*.

*“O mundo se encontra bastante avançado,
A ciência alcança progresso sem soma,
Na grande pesquisa feita no genoma,
Todo o corpo humano já foi mapeado,
E lá neste mapa já foi tudo contado,
Oitenta mil genes se pode contar,*

*A ciência faz chover e molhar,
Faz clone de ovelha,
Faz cópia completa,
Mas duvido a ciência fazer um poeta,
Cantando galope na beira do mar...”*

REPENTISTA GERALDO AMÂNCIO,
PERNAMBUCO, BRASIL.

Referências

- Boffelli, D, McAuliffe, J, Ovcharenko, D, Lewis, KD, Ovcharenko, I, Pachter, L, & Rubin, EM. 2003. *Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. Science* 299: 1391-1394.
- Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR. 2002. *Large-scale Transcriptional Activity in Chromosomes 21 and 22. Science* 296: 916-919.