

A promessa de seqüenciar o genoma até o ano 2000

Entre os cientistas do mundo da genômica, Craig Venter é dos mais famosos – pela importância de suas contribuições técnicas, e pelo fato de ter anunciado, em maio de 98, a criação de uma nova empresa que, por US\$ 200 milhões, e em três anos, realizaria o seqüenciamento completo do genoma humano. A Celera, situada em Rockville, Maryland, resultou de sua associação com a Perkin Elmer, uma das duas principais fabricantes de seqüenciadores automáticos; e o anúncio de seu objetivo quase paralisou o projeto Genoma Humano, que havia custado, até ali, quase US\$ 1 bilhão, e prometia a seqüência completa apenas para 2005. Mas os participantes do projeto financiado com verbas públicas reorganizaram-se; e o resultado da formação da nova companhia foi um encurtamento geral de prazos.

Agora, ambos os times – o público e o de Craig – pretendem alcançar suas metas dentro de aproximadamente um ano. Na entrevista, realizada em março último, o doutor Venter compara seu trabalho ao do grupo que seqüenciou o genoma da *C. elegans*. Suas observações, e sua personalidade, poderão ser melhor apreciadas quando cotejadas às palavras de Robert Waterston, um dos coordenadores do grupo, que serão publicadas no encarte de agosto.



Craig Venter

TECIDO PRESSO ■ *Muita gente com quem conversei sobre o senhor, nos Estados Unidos e no Brasil, expressam ceticismo e dúvida sobre o anúncio de seu projeto de seqüenciar o genoma humano dentro de 18 meses. Eles têm razão?*

— Para entender a questão, é necessário que se compreenda a tecnologia, e mesmo cientistas da área não entendem os novos passos da tecnologia. Quando Hamilton Smith e eu decidimos seqüenciar o

primeiro genoma da história, em 94, e pedimos ao NIH para financiar o projeto, houve esse mesmo ceticismo. Os assessores que analisaram a proposta disseram que seria impossível, que nunca poderia ser feito, que não se conseguiria montar, que os dados não seriam acurados – enfim, que ia ser terrível. No entanto, em 1995, nós publicamos o primeiro genoma na história, e foi a seqüência mais acurada e precisa jamais publicada. Todas as proteínas foram re-seqüenciadas pela indústria farmacêutica, e é notável a precisão que obtivemos. O método que desenvolvemos ali é o método que vamos usar para fazer o genoma humano. A TIGR seqüenciou até agora 10 genomas completos, incluindo os mais relevantes para a saúde mundial: tuberculose, cólera, malária, sífilis, doença de Lyme, tudo isso foi realizado com o método que desenvolvemos para o *Haemophilus influenzae*, aquele mesmo que diziam que não ia funcionar, que era impossível. Há dúzias de genomas completamente seqüenciados agora, o que as pessoas imaginavam que não seria feito nem nas próximas décadas; tudo por causa da tecnologia que eles disseram que não ia funcionar.

O genoma da *C. elegans* acaba de ser publicado¹, e não se pode dizer que seja muito completo, ou muito preciso. Custou dez anos, centenas de milhões de dólares, e envolveu centenas e centenas e centenas de cien-

¹ A seqüência completa do genoma da *Caenorhabditis elegans* (com cerca de 100 "gaps"), a primeira de um animal, foi publicada em dezembro de 98 pela revista *Science* (colocar referência?). Os dois principais pesquisadores envolvidos no trabalho são John Sulston, do Sanger Center, Inglaterra; e Robert Waterston, da Washington University, que gentilmente nos recebeu para uma entrevista, cuja íntegra será publicada no próximo Encarte.

e sua formulação. Mas quando você começa a falar de dados reais, de organismos reais na natureza, moléculas reais, há toda uma série de complicações para as quais você realmente não foi treinado como matemático. É muito mais difícil do que eu pensava, para alguém que vem da matemática ou das ciências da computação, desenvolver uma compreensão de como os dados são coletados no laboratório, para aprender como pensar os problemas biológicos que de fato importam, e dar espaço às imperfeições do mundo real. Não é fácil. Como se vê, há pontes culturais que devem ser atravessadas de ambos os lados.

■ *A especificidade e a abstração da linguagem da matemática e das ciências da computação não trazem riscos, quando aplicadas à Biologia?*

— É difícil levar em conta todas as complexidades. Nós, da matemática e das ciências da computação, costumamos simplificar os problemas, tentando extrair aqueles que parecem ser os elementos-chaves, e com eles construir algoritmos e procedimentos para calcular. Quando se trata de organismos biológicos, lidamos com sistemas extremamente complexos. Mesmo os biólogos têm que super-simplificar, têm que escolher um aspecto particular do organismo e pensar sobre ele numa forma simplificada, por causa da extrema complexidade. São bilhões de partículas interagindo entre si – a simplificação é necessária. O risco que o ponto de vista da matemática traz é a tentação de olhar para o organismo, para algum aspecto da seqüência, e tentar pensar sobre isso de uma maneira particular, que esteja relacionada com um problema matemático particular que já se conheça. Isto é muito perigoso porque não estamos estudando objetos; na Biologia, são realmente sistemas complexos, que têm muitos aspectos que nós não suspeitamos; é preciso, portanto, manter-se de mente aberta para eles.

■ *O senhor pensa que, a partir desses esforços, vai ser possível responder à questão: o que é a vida?*

— Sim, acho que, em determinado momento, vamos ser capazes de entender o que são os organis-

“A biologia foi o campo ideal para alguém que não queria trabalhar com números. Agora, tudo mudou”

mos. Penso que, de fato, os organismos obedecem às leis da física e da química. Quando entendermos o que eles são nos termos dessas leis, chegaremos a entender, então, os sistemas complexos de moléculas. Como eles se organizam, quais seus componentes, que interações se dão – sempre haverá algum mistério nisto. Talvez a complexidade seja grande demais para ser entendida. Organismos vivos são sistemas complexos de moléculas, interagindo entre si, e que têm a possibilidade de fazer coisas surpreendentes, coisas maravilhosas. É preciso ser capaz de compreender os organismos nestes termos.

■ *Mas esses não são uma abordagem e um ponto de vista reducionistas?*

— Esta é uma crítica pertinente, e minha resposta é sim. Quando se trata de tentar entender qualquer sistema complexo, não é apenas uma questão de identificar o que são as partes. Há questões de mais alto nível, quando se deseja detalhar a operação do sistema. Certamente, interagimos com o meio ambiente. O ambiente é feito de moléculas, ele próprio um sistema complexo, nós somos um componente do complexo meio ambiente. Mas, ainda assim, continuamos a obedecer às leis da física e da química. Há um reducionismo extremo que pode ser criticado – a visão de que, para entender a natureza, basta saber quais são suas partes. Para entender qualquer aspecto da natureza, é necessário entender quais as interações que ocorrem – interações governadas pelas leis da química e da física. O problema principal é o entendimento dos sistemas complexos. Nós podemos determinar quais são os componentes moleculares de um sistema, mas é muito mais difícil entender como esses componentes interagem com o sistema de moléculas. Apesar das dificuldades, acredito que seremos capazes de fazê-lo no futuro.

■ *Como o senhor se sente como o criador das ferramentas mais largamente usadas nos projetos genoma?*

— Um pouco ambivalente. Quando comecei a me debruçar sobre esses assuntos, cinco ou seis anos atrás, eu realmente pensava

“Sempre haverá algum mistério em como os sistemas biológicos se organizam. A complexidade é grande demais”

que seria simples. Ingenuidade minha. Também não via esse assunto como alguma coisa em que eu quisesse trabalhar a longo prazo, porque os problemas me pareciam técnicos: há pouca ou nenhuma biologia envolvida no desenvolvimento de novas ferramentas, por exemplo. Meu propósito ao escolher esse campo era o de gerar conhecimento biológico novo; estou muito mais interessado no problema da interpretação das seqüências. Quer dizer: uma vez que tenhamos uma seqüência, tentar identificar nela detalhes biológicos importantes. Assim, ficar centrado no que tem relação com a montagem e com a determinação das bases me afasta, na verdade, desse meu interesse. Nosso software tem sido bem sucedido; há muitos grupos que dependem de nós para aperfeiçoá-los. Só vamos considerá-lo inteiramente bom quando estivermos num estágio em que não tenhamos mais que melhorá-lo em nada, porque ele vai lidar com todos os problemas automaticamente. Então, gosto de ver nossos softwares em uso, mas é um pouco frustrante que isso me tire do caminho onde estão os problemas mais interessantes.

■ *De que se trata, interpretar seqüências?*

— Interpretar seqüências envolve problemas de mais longo prazo, e mais interessantes. É importante identificar genes, saber dizer onde começam, onde terminam, e identificar os diferentes sinais que estão associados com eles. Há muita gente trabalhando nisto. Mas, presumivelmente, há outro tipo de informação na seqüência, além dos genes. Esta é uma área interessante – identificar esses outros detalhes biológicos na seqüência – e pouco progresso tem sido feito nela. Para enfrentar a questão, é preciso buscar idéias da teoria das probabilidades e da estatística. É preciso desenvolver modelos probabilísticos para tentar entender as seqüências. Basicamente, a razão pela qual a teoria das probabilidades é necessária é que a seqüência de um genoma é produto de bilhões de anos de evolução, e a evolução tem um grande componente aleatório. Muitas das mudanças nas seqüências que ocorreram ao longo do tempo não têm nenhuma implicação funcional, elas são basicamen-

“Nos projetos genoma tentamos identificar diferentes componentes moleculares presentes nas células – as proteínas em particular”

te aleatórias. Então, ao olhar a seqüência genômica de dois organismos diferentes, e tentar entender em que são similares e porque são diferentes, deve-se levar em conta essas diferenças aleatórias. Daí a necessidade da teoria das probabilidades. Então, o que tem sido feito é desenvolver modelos probabilísticos que prevejam como deve ser cada tipo particular de seqüência. Um exemplo: a parte do gene que codifica a proteína tem certas tendências estatísticas que advêm simplesmente do fato de codificar a proteína. Além disso, há fatos estatísticos bastante interessantes, que são apenas parcialmente entendidos. Não sei se não estou ficando muito técnico... O código genético que converte triplas de DNA em aminoácidos é degenerado – há vários codons diferentes para um dado aminoácido. O caso é que eles não são usados com igual freqüência em um dado organismo. Qualquer que seja o organismo que você estuda, se você toma um aminoácido particular que é codificado por mais de um codon, e conta quão freqüentemente cada um desses codons é usado, você encontrará que eles não são usados com a mesma freqüência. Às vezes, há um grande viés nisso. É útil tentar criar modelos estatísticos para classificar determinada seqüência codificadora com o propósito de identificar genes e seqüências. Quando você conhece esses vieses, "viés de uso de codons sinônimos", é assim, que são chamados, então você pode elaborar um modelo estatístico e levá-lo em conta. É realmente muito interessante descobrir, do ponto de vista biológico, porque esses vieses ocorrem. Até agora, isto é entendido apenas parcialmente. Parece que alguns desses codons, em organismos primitivos, são traduzidos mais eficientemente que outros. Como a seqüência proteica é construída a partir do RNA mensageiro, alguns dos codons se reorganizam mais rapidamente

“É preciso desenvolver modelos probabilísticos para tentar entender as seqüências”

e codificam seu aminoácido com mais rapidez e precisão. Parece, então, ter havido alguma seleção durante a evolução para favorecer certos codons. Mas há outros aspectos que não entendemos. Quando se começa a fazer análise estatística de seqüências genômicas, é excitante notar que você começa a notar padrões, aspectos não-aleatórios da seqüência. Então, a questão é: o que isto significa do ponto de vista biológico? Há

■ *O senhor acredita que a genômica poderá responder à pergunta: por que Brabms compunha tão bem? Ou decifrar o sofrimento humano?*

— Vou fazer uma analogia. Nós estamos tentando caracterizar o genoma mínimo de uma pequena bactéria que a equipe da doutora Fraser sequenciou, o *Mycoplasma genitalium*. Há somente 300 genes que são essenciais para a vida, de acordo com o que pensamos no momento; deles, 100 são completamente novos para a ciência, não sabemos o que eles fazem. O que quero destacar é

que ainda não entendemos como 300 genes trabalham em uma única célula; se não entendemos ainda isto, como vamos entender 80 mil genes trabalhando juntos, em dez trilhões de diferentes células e em diferentes combinações, que é o que nos forma? Isto está muito além da compreensão científica para os próximos séculos. Nós vamos precisar de muita inovação tecnológica para chegar até lá. Obviamente, há uma base genética para a personalidade, e há

bases genéticas para a memória e para o pensar, mas ainda não temos como medi-los. Por que pais famosos nem sempre têm filhos famosos? A realidade da vida nos afeta a todos. Nós nunca poderemos prever se alguém virá a ser um notório cientista, ou um bom compositor; isso está baseado também no ambiente onde vivemos ou crescemos e trabalhamos e todas as oportunidades que temos. Tome-se o caso dos gêmeos idênticos, com o mesmo código genético: algumas vezes, eles crescem com personalidades muito diferentes, vidas muito diferentes. Isso depende de quanto você enfatizar o que têm de diferenças entre si, e o que têm em comum. Se você vê gêmeos vestidos iguais, tende a ver o que têm em comum; mas se eles crescem separadamente, vão construir vidas diversas entre si. O que o código genético nos dá são potenciais para a vida; não há determinismo genético que diga que você vai ser boa jornalista e eu, um bom cientista. No período de evolução, esses trabalhos sequer existiam, de forma que não há maneira de haver um componente genético para sua determinação. Poderá determinar uma certa quantidade de inteligência e uma habilidade para lidar com o ambiente, no máximo.

■ *O entendimento da evolução também está fortemente relacionado à genômica, não?*

— Você tem razão. Veja o exemplo do *M. genitalium*.

Estamos eliminando genes entre os 300 de seu genoma para ver os que não são essenciais à vida, e entender qual é esse conjunto mínimo. Note-se que crescem o organismo no laboratório, o que é muito diferente do que é necessário no ambiente. O próximo passo será tentar sintetizar o cromossomo com apenas esses genes mínimos. Decidimos parar aí, e começamos uma discussão ética com um grupo de líderes religiosos, cientistas e gente comum na Universidade da Pensilvânia, para ver se é apropriado ir em frente e sintetizar a vida.

Por isso a TIGR parou, nós não continuaremos os experimentos enquanto a discussão estiver em curso. Vamos publicar um artigo para descrever o trabalho num futuro próximo; depois da discussão pública, decidiremos o que fazer.

■ *O senhor acha que merece ganhar um Prêmio Nobel?*

Tenho mais sobre o que pensar no momento. Estou saindo da TIGR para formar a Celera e sequenciar o genoma humano. Decidi que prefiro passar à história como o sequenciador do genoma humano do que ganhar o Prêmio Nobel. Validação pública é sempre bom; se acontecer, muito bem, mas não é meu objetivo na vida.

■ *O senhor pensa que poderia ficar rico se a genômica não existisse?*

— Sou freqüentemente retratado como um milionário ocasional. Ficar rico não era meu objetivo. Não criei a TIGR ou a Celera com esse objetivo. Mas fica claro, com o caso da recusa do governo em financiar o *H. influenzae*, que é preciso ter recursos próprios. Esta é a primeira razão para se montar uma companhia de biotecnologia. Por ter parte de uma companhia, ganhei alguns milhões de dólares. Agora, de novo, o governo diz que é impossível sequenciar o genoma humano segundo minha proposta; e, felizmente, a Perkin Elmer está investindo centenas de milhões de dólares nisso, e tanto eu quanto a TIGR somos sócios da Celera. Você não precisa sofrer para ser cientista. O fato de trabalharmos às vezes 18 horas por dia mostra que somos dedicados ao que fazemos. Não há incompatibilidade entre a dedicação e ter dinheiro que me permita sair velejando e descansar minhas células cerebrais. É uma questão de motivação.

“O código genético nos dá potenciais para a vida. Não há determinismo que diga que serei um bom cientista”