

BIOLOGIA MOLECULAR

# Preenchendo as lacunas

Brasileiros criam estratégia de pesquisa genômica e completam a seqüência de 211 genes humanos

**E**m fevereiro de 2001, ao ser apresentado publicamente, o genoma humano foi comparado a uma paisagem com extensos desertos entremeados por esparsas cidades. Os desertos representavam os longos trechos do DNA, tecnicamente chamados de íntrons, que aparentemente não faziam nada – não levavam à produção de proteínas que formam os seres vivos. As cidades seriam os trechos funcionais do DNA, chamados éxons. Mas ainda havia muita neblina e, num primeiro momento, era impossível saber o que era deserto e o que eram as cidades, nem quantas eram, nem onde estavam. Tãmanha era a incerteza que as estimativas do número de genes variavam de 35 mil a 120 mil.

Em uma corrida internacional em busca de um número exato, da localização precisa, do tamanho e da estrutura dos genes, grupos de pesquisa dos Estados Unidos, do Japão e da Alemanha lotaram salas com dezenas de seqüencia-

dores de DNA, que funcionavam dia e noite. Mesmo sem tantos equipamentos, os pesquisadores de universidades e institutos paulistas não se deixaram abater. Adotaram uma estratégia própria e ambiciosa – com uma análise exaustiva dos dados públicos sobre o genoma aliada a testes de laboratórios – e, quatro anos depois, conseguiram completar a seqüência de 211 genes, dos quais antes só havia fragmentos, além de mostrar onde eles se encontram no genoma – as cidades ganharam uma localização exata no meio da paisagem desértica. Os quase cem pesquisadores de 31 laboratórios de universidades paulistas e do Instituto Ludwig de Pesquisa sobre o Câncer também descobriram cerca de 40 genes novos, que ainda não haviam sido descritos por nenhum outro grupo. Os resultados desse trabalho, coordenado por Anamaria Camargo, do Ludwig, e por Mari Cleide Sogayar, do Instituto de Química da Universidade de São Paulo (USP), foram publicados on-line no final do

mês passado e saíram no dia 1º deste mês na versão impressa da revista *Genome Research*.

“Essa é a ciência que resulta de uma parceria entre a FAPESP e o Instituto Ludwig”, comenta José Fernando Perez, diretor-científico da FAPESP. A Fundação e a filial paulista do Ludwig conduziram durante dois anos, de 1999 a 2001, o Projeto Genoma Humano do Câncer, para o qual cada instituição destinou o equivalente, hoje, a R\$ 30 milhões. O trabalho conjunto terminou com o saldo de aproximadamente 1,2 milhão de seqüências de genes associados a vários tipos de câncer – eram trechos centrais dos genes, caracterizados por meio de uma metodologia criada no país, a Orestes, sigla de Open Reading Expressed Sequence Tags, que em português significa algo como etiquetas da fase aberta de leitura de seqüências expressas. Em uma abordagem complementar, outros grupos de pesquisa haviam seqüenciado os extremos dos trechos de genes, com outra técnica, a EST, de

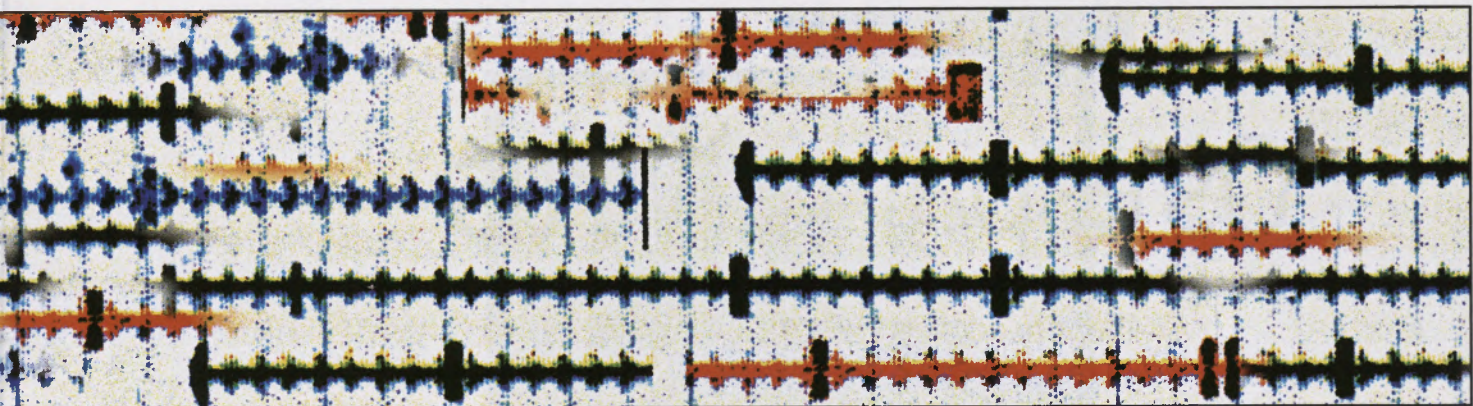


ILUSTRAÇÃO HÉLIO DE ALMEIDA

Expressed Sequence Tags ou etiquetas de seqüências expressas.

Por meio do Projeto Transcriptoma do Câncer, que começou no final de 2000, com investimentos de cerca de R\$ 4 milhões da FAPESP e R\$ 1,5 milhão do Ludwig, os pesquisadores tentaram unir os dois conjuntos de seqüências, as do meio e as dos extremos dos genes. Ambos eram formados apenas por éxons, as partes ativas dos genes, mas nem sempre eram o bastante para completar os genes – restavam muitos espaços vazios. “A princípio qualquer grupo de pesquisa poderia ter feito esse trabalho, já que todos os dados eram públicos”, afirma Sandro José de Souza, coordenador da equipe de bioinformática do Ludwig. “Nossa vantagem foi unir grupos com vocações diferentes.” A empreitada começou oficialmente em 2001 e mobilizou cinco equipes de bioinformática – do Ludwig, da Universidade de Ribeirão Preto (Unaerp), da Universidade Federal de São Paulo (Unifesp) e da Faculdade de Medicina de Ribeirão Preto e do Instituto do Coração, ambos da Universidade de São Paulo (USP).

**Candidatos a genes** - Os bioinformatas, como são chamados, sobrepunham as seqüências do miolo e dos extremos dos genes com as informações que chegavam dos projetos internacionais de seqüenciamento do genoma humano – eram longas listas de nucleotídeos, as unidades do DNA, sem que ninguém tivesse a menor idéia de onde estavam os íntrons e os éxons. “Centramos a atenção nos genes incompletos, reunindo as seqüências de Orestes e de outras ESTs”, conta Souza. Saíam daí listagens com candidatos a genes, selecionados com a ajuda de programas de computador,

que eram testados experimentalmente pelas equipes de 31 laboratórios do Instituto Ludwig, da USP, da Unifesp, da Universidade Estadual Paulista (Unesp), da Universidade de Campinas (Unicamp) e da Universidade do Vale do Paraíba (Univap).

**O**s primeiros resultados demonstrando a viabilidade da técnica saíram em outubro de 2001 no *Proceedings of the National Academy of Sciences (PNAS)* e, em um comentário de duas páginas, ganharam o reconhecimento de duas autoridades mundiais em genoma humano, Robert Strausberg e Gregory Riggins, ambos do Instituto Nacional do Câncer (NCI em inglês), dos Estados Unidos.

Ter encontrado o caminho tornou as coisas apenas um pouco mais fáceis. Luciana Oliveira Cruz, pesquisadora da equipe de Mari Cleide, trabalhou muito, durante meses, cultivando 20 linhagens de tecidos humanos – de útero,

testículo, fígado, entre outras –, preparando amostras de DNA complementar (cDNA), que corresponde aos genes ativos em cada tecido, e as distribuindo aos laboratórios que testavam os 488 candidatos a genes previamente selecionados para ver se eram realmente genes. Cada gene era submetido à reação em cadeia de polimerase, técnica conhecida pelas iniciais PCR, com dois *primers* específicos. *Primers* são seqüências de nucleotídeos sintéticos, feitas, no caso, a partir de dois trechos conhecidos de DNA (Orestes ou ESTs) – são os *primers* que delimitam os extremos de um fragmento de DNA a ser copiado milhares de vezes.

“Sabemos que os dois trechos antes descritos como seqüências individuais pertencem a um só gene quando, a partir dos *primers*, se obtém a cópia do cDNA, numa demonstração de que se tratava de pedaços de uma única molécula”, diz Luciana. Essa estratégia – alinhamento dos trechos de DNA e teste com *primers* – foi chamada de Iniciativa de Finalização de Transcritos (TFI, de Transcript Finishing Initiative) e determinou o que era íntron e o que era éxon nos genes incompletos. Com uma eficiência de 43%, revelou 211 novos genes, muitos deles descritos por outros grupos de pesquisa, com outras técnicas, no transcorrer do projeto. Restaram cerca de 40 inéditos, apresentados no artigo da *Genome Research*.

O Projeto Transcriptoma terminou no final do ano passado, embora restem milhares de brechas a serem preenchidas no genoma humano. Ainda não há um consenso sobre o número total de genes, mas já foram descritos cerca de 25 mil genes completos – ou quase completos.

## O PROJETO

*Caracterização de Genes Humanos Completos – Uma Extensão do Genoma Humano do Câncer*

**MODALIDADE**  
Projeto Especial

**COORDENADORAS**  
ANAMARIA ARANHA CAMARGO – Instituto Ludwig de Pesquisa sobre o Câncer  
e MARI CLEIDE SOGAYAR – Instituto de Química da USP

**INVESTIMENTO**  
R\$ 540.000,00 e R\$ 550.000,00