

O PROJETO GENOMA DO CÂNCER HUMANO

LUDWIG/FAPEP

O Projeto Genoma e a Estratégia Orestes



O Brasil na era da *Genomics*

A área de “genomics” vive hoje uma fase histórica sem precedentes. Vinte e dois genomas bacterianos já se encontram totalmente sequenciados, os genomas de levedura e *C. elegans* foram completados e os genomas de *Drosophila* e de *Arabidopsis thaliana* devem estar finalizados em poucos meses. Recentemente foi publicada a seqüência do primeiro cromossomo 22 do homem e tudo indica que, até o final do ano 2000, estaremos muito próximos de atingir a seqüência completa do genoma humano, fornecendo a base de toda a futura pesquisa na área de genética humana, a “pedra fundamental” da medicina dos próximos anos.

Graças ao esforço feito pela diretoria científica da FAPESP, que pela rede ONSA (Organisation for Nucleotide Sequence and Analysis) tem financiado projetos científicos de larga escala na área de *genomics*, o Brasil ainda pode surgir como um centro de excelência para a execução de trabalhos arrojados nessa área. O programa brasileiro de genomas nasceu com a iniciativa extremamente bem sucedida de sequenciar totalmente o genoma da bactéria *Xylella fastidiosa*. Apesar de os resultados de sequenciamento terem demonstrado que o genoma dessa bactéria é cerca de 30% maior que o esperado, o projeto chegou ao seu final cerca de seis meses antes do prazo determinado. Atualmente este genoma (o pri-

meiro fitopatógeno a ser completamente sequenciado no mundo) encontra-se em fase de publicação e, sem a menor dúvida, constitui um dos mais importantes marcos da ciência nacional.

O impacto do programa ONSA vai muito além do fechamento do genoma de *Xylella*. Um dos principais frutos da iniciativa na área de genomas foi a criação de uma massa crítica respeitável com a capacitação de dezenas de pesquisadores e laboratórios para executar técnicas avançadas de biologia molecular. A difusão da ciência genômica foi fundamental para o surgimento de novos e ousados projetos dentro da rede ONSA. A rede já conta hoje com outros três importantes projetos de pesquisa em larga escala: o SUCEST (Sugar Cane EST project), o projeto genoma de um segundo fitopatógeno, a *Xanthomonas campestris* e o Projeto Genoma do Câncer Humano Ludwig/FAPEP (HCGP). Este último é um dos mais ambiciosos projetos da rede ONSA, pois pretende contribuir para a descoberta gênica e anotação do genoma humano, sem dúvida a área de pesquisa mais competitiva a nível mundial. A entrada do Brasil em um projeto tão competitivo, apenas foi possível com o disponibilidade de aplicação de uma técnica de descoberta gênica desenvolvida no Brasil. Um outro diferencial fundamental foi a parceria entre uma instituição privada de pesquisas (o Instituto Ludwig de Pesquisas sobre o Câncer –

Emmanuel Dias Neto, PhD
Laboratório de Genética do Câncer,
Ludwig Institute for Cancer Research
emmanuel@compbio.ludwig.org.br

Fotos cedidas pelo autor

ILPC) e um órgão estadual de financiamento à pesquisa (a FAPESP), que financiaram em proporções iguais um projeto de cerca de 12 milhões de dólares.

Grupos e estratégias no Genoma Humano

Estimativas atuais indicam que o genoma humano deva ter algo entre 3 a 3,5 bilhões de nucleotídeos com algo entre 80 e 150 mil genes expressos. Com o objetivo comum de desvendar a sequência de todos os genes humanos, diversas estratégias vêm sendo usadas hoje no Projeto Genoma Humano, as três mais importantes são:

1) O sequenciamento ordenado de grandes porções genômicas definidas – estratégia adotada por grupos como Sanger Centre e Washington University. Esta abordagem permitiu o término do sequenciamento do cromossomo

22 humano publicado recentemente.

2) O sequenciamento completo do genoma por “shot-gun”. Estratégia ainda polêmica, aplicada com sucesso em genomas bacterianos que consiste na quebra do genoma humano em fragmentos pequenos que são sequenciados ao acaso até fecharem o genoma com o auxílio de um pesado sistema computacional. Vem sendo adotada pela Celera, uma empresa norte-americana comandada pelos maiores especialistas em genomas bacterianos.

3) Uma abordagem paralela tenta uma identificação prioritária dos genes expressos, pela estratégia de geração de “etiquetas de seqüências expressas” ou ESTs. Esta estratégia vem sendo usada em larga escala nos laboratórios ligados ao CGAP (Cancer Genome Anatomy Project), financiado pelo Instituto Nacional do Câncer, nos USA e se concentra apenas

nos 3% do genoma que representam os genes expressos, permitindo uma grande redução dos custos e um aumento na velocidade de obtenção dos resultados finais.

A estratégia de sequenciamento prioritário dos genes humanos foi escolhida pelo HCGP, adotando a metodologia denominada ORESTES (Open Reading frame ESTs). Nossos resultados indicam que ORESTES permite a obtenção preferencial de regiões codificadoras do genoma humano, favorecendo uma normalização dos genes obtidos. Estes dois fatores são cruciais e vêm determinando o grande sucesso do projeto. A obtenção de ESTs centradas na porção interna dos genes facilita a descoberta gênica, pois as seqüências de DNA podem ser traduzidas e similaridades a nível de proteína são possíveis de serem observadas. Esta característica torna as ORESTES ainda muito mais úteis pois estas se

Tabela I – Centros e laboratórios de sequenciamento do Projeto Genoma do Câncer Ludwig/FAPESP

Centros e laboratórios de sequenciamento	Laboratórios de sequenciamento	Localização
Maria Inês Moura Campos Pardini	IL2	Fac. de Medicina de Botucatu/UNESP - Hemocentro
Marina Pasetto Nóbrega	IL3	Inst. de Pesquisa e Desenvolvimento - UNIVAP
Silvia Regina Rogatto	IL5	Depto de Genética, Instituto de Biociências/UNESP - Botucatu
Fernando Ferreira Costa	CM0	Hemocentro, Fac. de Ciências Médicas, UNICAMP
Christine Hackel	CM1	Depto de Genética Médica, UNICAMP
Helaine Carrer e Dirce Maria Carraro	CM2	CEBTEQ - ESALQ Depto de Química
Maria de Fátima Sonati	CM3	Fac. de Ciências Médicas Depto de Morfologia, UNICAMP
Gonçalo Guimaraes Pereira	CM4	Depto de Genética e Evolução, UNICAMP
Marcelo Ribeiro da Silva Briones	PM0	Depto de microbiologia, imunologia e parasitologia, UNIFESP, SP, SP
Rui Monteiro de Barros Maciel	PM1	Depto de Endocrinologia, UNIFESP
Luis Eduardo Coelho Andrade	PM2	Depto de Reumatologia
Ismael Dale Cotrim Guerreiro da Silva	PM3	Depto de Biofísica
João Bosco Pesquero	PM4	Depto de Ginecologia e Obstetrícia
Sérgio Verjovski de Almeida	QV0	Inst. de Química – USP, SP, SP
Arthur Gruber	QV1	Fac. de Medicina Veterinária e Zootecnia – USP, SP, SP
Mari Cleide Sogayar	QV2	Instituto de Química– USP, SP, SP
Edna Teruko Kimura	QV3	Instituto de Ciências Biomédicas, USP, SP, SP
Hamza Fahmi Ali El-Dorry	QV4	Instituto de Química– USP, SP, SP
Marco Antônio Zago	RC0	Fac. de Medicina, Depto de Clínica Médica, Ribeirão Preto, SP
Enilza Maria Espreafico	RC1	Depto de Morfologia, FMRP
Gustavo Henrique Goldman	RC2	Fac. de Ciências Farmacêuticas de Ribeirão Preto
Maria Luisa Paçó-Larson	RC3	Depto de Morfologia, FMRP
Vanderlei Rodrigues	RC4	Depto de Parasitologia, Microbiologia e Imunologia, FMRP
Eloiza Helena Tajara da Silva	RC5	Instit. Biociências, Letras e Ciências Exatas São José do Rio Preto/UNESP
Sandro Roberto Valentini	RC6	Faculdade de Ciências Farmacêuticas de Araraquara/UNESP
Maria Aparecida Nagai	MR0	Fac. de Medicina, Depto de Radiologia, USP, SP, SP
Angelita Habr-Gama	MR1	Depto Gastroenterologia, FMUSP
Daniel Giannella Neto	MR2	Hosp. das Clínicas – SSSP Depto de Clínica Médica
Suely Kazue Nagahashi Marie	MR3	Depto de Neurologia, FMUSP
Elizabeth Martins e Paulo Lee Ho	MR4	Centro de Biotecnologia, Instituto Butantan

mostram complementares às centenas de milhares de ESTs humanas, geradas por grupos como CGAP, Universidade de Washington e outros. Esta complementaridade se dá pois as ESTs destes grupos são derivadas das extremidades 3' e 5' dos genes, permitindo, com o uso de ORESTES, a formação de "contigs" e a eventual obtenção da seqüência completa dos transcritos humanos. Por outro lado, a capacidade de normalização faz com que genes raros ou pouco expressos, normalmente um ponto fraco das bibliotecas de cDNA, estejam presentes nas bibliotecas e possam ser sequenciados.

Base de ORESTES, tumores estudados e resultados parciais

Para a obtenção das ORESTES nos baseamos na complementaridade parcial entre oligonucleotídeos e seqüências de mRNAs/cDNAs. Sob condições de baixo rigor de hibridização, uma complementaridade parcial entre um oligonucleotídeo e algumas moléculas de diferentes mRNAs já é suficiente para gerar uma interação produtiva, capaz de permitir a síntese de moléculas de cDNA. Estas moléculas de cDNA são então amplificadas, tendo como base o mesmo princípio de complementaridade parcial. Conforme esperado, o uso de iniciadores nestas condições, favorece a prevalência de porções centrais (codificadoras) dos genes expressos, gerando dados de seqüência únicos, não disponíveis com outras metodologias usuais. Se avaliarmos a distribuição posicional de ESTs derivadas de genes conhecidos, podemos avaliar se estas realmente são derivadas da porção central. Esta avaliação é feita tomando-se individualmente cada ORESTES derivada de um gene humano conhecido, e avaliando a região da molécula de mRNA que se encontra representada na EST. Se temos uma molécula de mRNA com 3.000 nucleotídeos e esta apresenta um alinhamento correspondente à base 1.500,

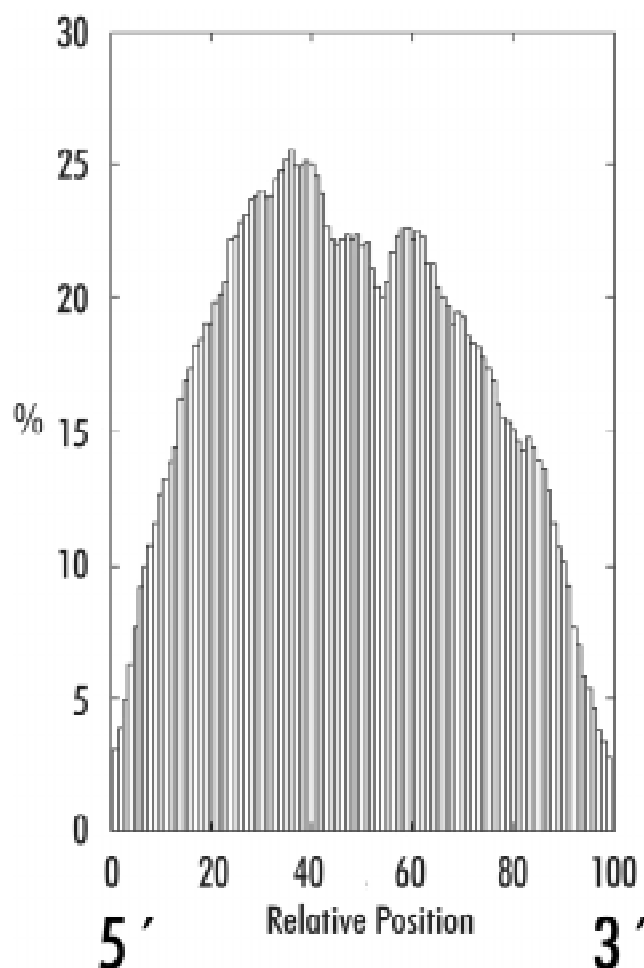


Figura 1 – Distribuição posicional média de ORESTES similares a genes humanos. O ponto zero corresponde à extremidade 5' e o ponto 100 à extremidade 3'

sabemos que esta EST representa o ponto de 50% do gene, a sua porção média. Quando esta distribuição posicional foi avaliada em 7 mil ESTs correspondentes a genes humanos, temos o gráfico mostrado na figura 1. O mesmo padrão foi observado quando avaliamos esta mesma distribuição usando ESTs correspondentes a 20 mil genes humanos.

No HCGP, optamos por utilizar, como a base biológica de descoberta gênica, alguns tumores epidemiologicamente importantes no Brasil. Sendo assim, temos hoje um enfoque principal em tumores de cabeça e pescoço, cólon, mama e estômago. Ao mesmo tempo, temos um enfoque secundário em tumores mais raros, onde o potencial de descoberta gênica deve ser maior.

Entre estes estudamos tumores renais do tipo Wilm's, leiomioma, testículo, PNET (tumor neuroectodérmico primitivo), tumores de sistema nervoso central (gliomas e meningiomas), tumores de cólon de útero e leucemias. Já contamos hoje com mais de 170 mil ESTs disponíveis no banco de dados do projeto e mais de 25.000 diferentes genes já identificados. Quando comparado ao CGAP (Cancer Genome Anatomy Project) americano, lançado em meados de 1997, vemos que o HCGP possui um número mais de 10 vezes maior de ESTs derivadas de tumores de cabeça e pescoço. Também já superamos o CGAP americano em termos de número de ESTs derivadas de tumores de mama, estômago, Wilms, leiomioma, testículo e PNET.

Estrutura do projeto

O projeto tem toda a sua coordenação no Instituto Ludwig de Pesquisas sobre o Câncer em São Paulo. Sob a coordenação geral de Andrew Simpson, que também coordenou o sequenciamento de *Xylella fastidiosa*, trabalham as coordenações de RNA, de bibliotecas, cinco coordenações de sequenciamento e a coordenação de bioinformática. A coordenação de RNA é feita pela equipe de Luis Fernando Reis, responsável pela obtenção de RNA mensageiro com as qualidades requeridas pela técnica (ausência de contaminação com DNA genômico ou degradação de mRNA). Após a assinatura de termo de consentimento informado pelos pacientes, o material tumoral obtido na cirurgia é imediatamente acondicionado em N₂ de modo a evitar a degradação do RNA a ser extraído. O tecido é dissecado pela equipe do Departamento de Anatomia Patológica do Hospital do Câncer (coordenado por Fernando Soares) e tratado com DNase, a fim de remover contaminações com DNA genômico. Após passar por rigorosos controles, que devem demonstrar a ausência de contamina-

ção com DNA genômico e a ausência de degradação, o mRNA é repassado à coordenação de bibliotecas.

A coordenação de bibliotecas, sob responsabilidade de Emmanuel Dias Neto, co-inventor da técnica ORESTES juntamente com Andrew Simpson, é responsável pela execução da parte técnica que assegura as duas maiores vantagens da metodologia: a normalização da população de genes expressos (permite a obtenção de sequências derivadas de genes raros) e a obtenção de fragmentos de cDNA derivados da porção central, codifica-



Detalhes dos dados processados do sequenciamento de uma amostra

dora, dos genes expressos. Esta coordenação é responsável pela síntese e posterior amplificação por PCR das moléculas de cDNA, gerando perfis complexos de bandejamento de cDNAs. Após a avaliação destes perfis e a análise de controles das reações, os perfis de bandejamento são repassados às coordenações de sequenciamento, para a distribuição dos perfis, obtenção das bibliotecas de ORESTES e sequenciamento. A coordenação de bibliotecas também é responsável pelo monitoramento diário dos resultados e pelo constante desenvolvimento de melhorias no processo.

Os cinco centros de sequenciamento de DNA se encontram em algumas cidades do estado de São Paulo. Em São Paulo, capital, temos os coordenadores Maria Aparecida Nagai,



Sequenciador capilar de DNA - Megabace 1000 - Seis destes aparelhos são usados na geração de ESTs no HCGP

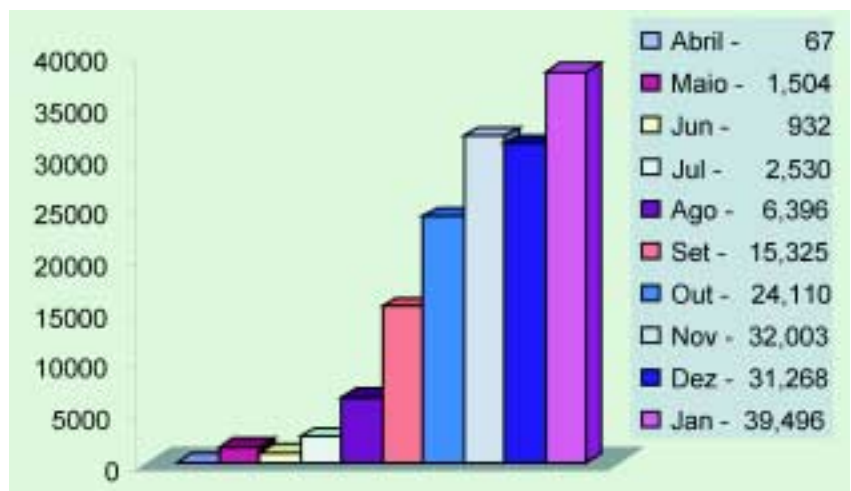
Marcelo Ribeiro da Silva Briones e Sérgio Verjovski de Almeida . Em Ribeirão Preto, temos o grupo de Marco Antônio Zago e em Campinas, o grupo de Fernando Ferreira Costa. Ao redor dos centros, temos os laboratórios de sequenciamento, conforme a tabela 1.

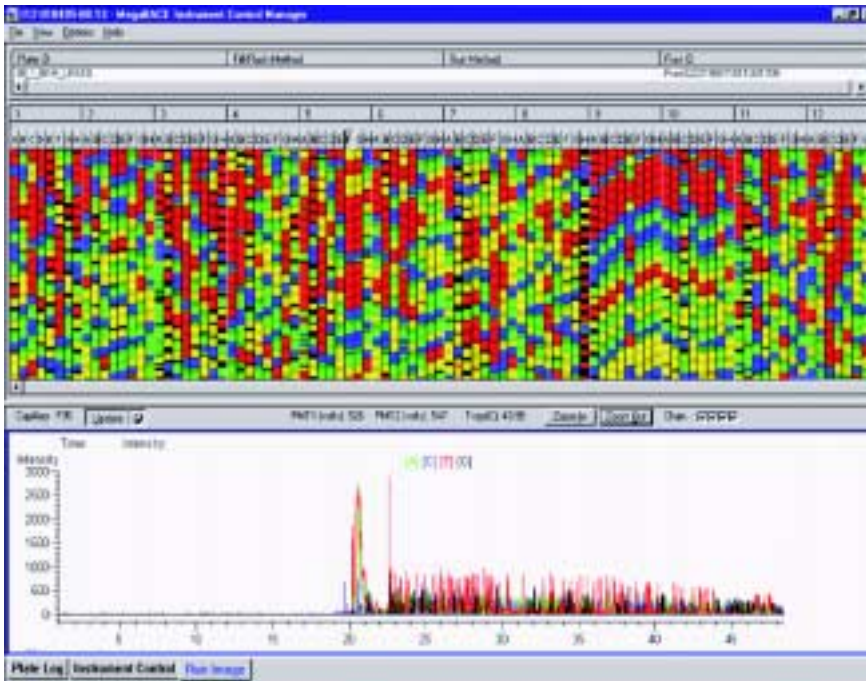
Tecnologia de sequenciamento

Uma outra inovação do HCGP foi a utilização de sequenciadores automáticos de DNA de última geração, que empregam a tecnologia capilar. Tais aparelhos apresentam uma série de vantagens sobre os sequenciado-

res de DNA da geração anterior. As principais vantagens são uma automação quase total no processo de sequenciamento, não havendo a necessidade de preparo de géis de sequenciamento ou do "tracking" das amostras sequenciadas. Estes sequenciadores são também extremamente rápidos, atingindo o sequenciamento de 96 amostras simultaneamente, em apenas 2 horas de operação, com cerca de 700 bases de leitura por amostra. Com o bom desempenho de seis aparelhos MegaBace 1000 adquiridos junto à empresa Molecular Dynamics/Pharmacia, o HCGP está atu-

Geração de mensal de ESTs dentro do HCGP





A parte superior da figura mostra uma visão geral do sequenciamento simultâneo de 96 amostras. A porção inferior mostra em detalhe o cromatograma de uma amostra individual

almente com um ritmo diário de cerca de 1400 ESTs validadas dentro dos critérios de qualidade adotados pelo projeto. Em breve, este valor deve subir para pouco mais de 2.000 novas ESTs/dia. A evolução mensal do projeto pode ser acompanhada no gráfico 1.

Deposição e análise de dados

Após a procedimento das reações de sequenciamento, os dados são enviados na forma de cromatogramas para o Centro de Bioinformática do projeto, também localizado no Instituto Ludwig, em São Paulo, sob a coordenação de Sandro José de Souza. Esta coordenação desenvolveu programas que permitem a avaliação da qualidade das ESTs recebidas e a retirada de regiões de baixa qualidade, vetores e iniciadores usados nas reações. As regiões de alta qualidade são então submetidas a análises de similaridade a nível de nucleotídeos e aminoácidos, permitindo a avaliação dos resultados em termos de expressão de genes conhecidos nos tecidos estudados, observação de formas alternativas de *splicing*, polimorfismos (SNPs), mutações, descoberta gênica e etc...

Os SNPs são de grande valor e

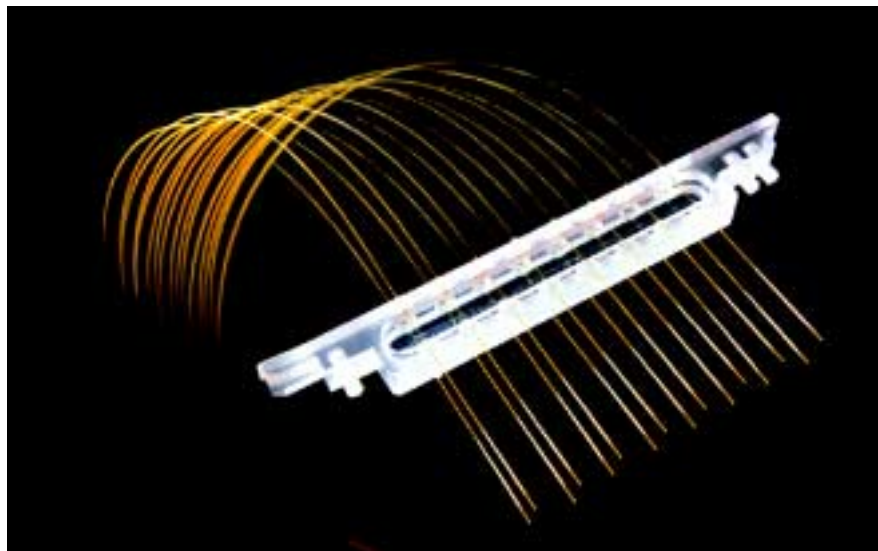
importância, e vêm sendo estudados sistematicamente pelo grupo de Ribeirão Preto. Já os dados relativos à descoberta gênica, compõe um percentual significativo dos achados. Mais de 45 mil de nossas ESTs não apresentam similaridades nos bancos de dados, enquanto que outras cerca de 2.500 ESTs apresentam similaridades com genes descritos em outros organismos

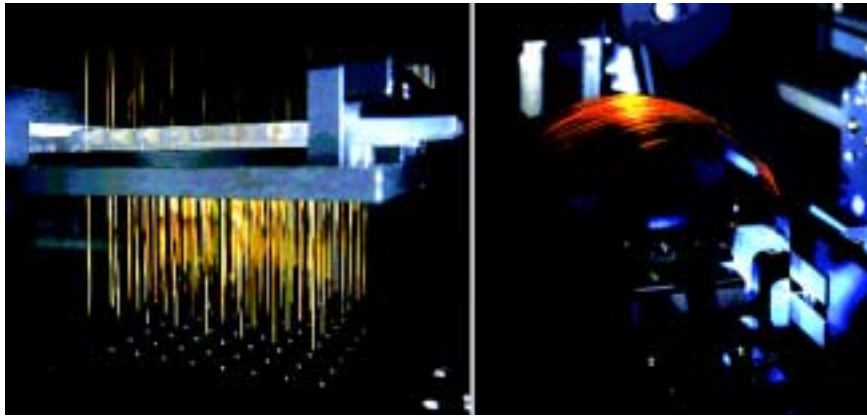
humanas. Sendo assim, este projeto vem contribuir de maneira significativa para a descoberta de novos genes humanos. Algumas destas ESTs já estão sendo estudadas mais a fundo e deverão representar, em breve, alguns dos primeiros genes humanos identificados no Brasil.

A importância da geração de ESTs reveste-se de uma nova importância à medida que o sequenciamento do genoma se completa. Os dados do cromossomo 22 tornaram claro que as informações derivadas de ESTs são cruciais para a identificação de exons e predição de formas alternativas de *splicing*, informação nem sempre óbvia com os programas disponíveis na atualidade. Diante da pequena disponibilidade de ESTs derivadas da porção central dos genes, os dados de ORESTES têm seu valor ainda mais aumentado. Como ficou claro com a avaliação feita na seqüência do cromossomo 22, em diversas ocasiões foi possível fazermos a chamada “anotação” de um gene, graças à natureza única dos dados gerados pelo programa brasileiro.

A importância dos dados foi reconhecida pelo comitê externo de avaliação do projeto. Pessoas de grande importância na pesquisa mundial vêm ao Brasil 2 ou 3 vezes por ano e avaliam a qualidade dos dados gerados e a condução geral do projeto. Alguns dos pesquisadores que avaliaram o projeto são – Phillip Sharp (Prê-

Fotografia de um dos seis conjuntos (cada um com 16 capilares) de capilares de vidro onde é feita a eletroforese dos fragmentos de DNA durante o sequenciamento





Nestes capilares de vidro é feita a eletroforese dos fragmentos de DNA durante o sequenciamento

mio Nobel de Medicina, pela descoberta dos introns); John Sgouros (coordenador de bioinformática do projeto genoma de levedura); Marcelo Bento Soares (autoridade mundial na produção de bibliotecas de cDNA) e Victor Jongeneel (especialista em bioinformática do Ludwig Institute, na Suíça).

Perspectivas finais

Uma das maneiras mais poderosas de estudar esta grande quantidade de

dados é por meio dos chamados biochips de DNA. A obtenção da verba necessária para a montagem do primeiro centro de biochips da América Latina pode ser considerado um dos mais importantes frutos da implantação do HCGP. O centro de produção e análise de micro-chips de DNA, financiado pelo Instituto Ludwig de Pesquisas sobre o Câncer em cerca de 1,1 milhão de dólares, já está funcionando e deverá trabalhar paralelamente com a geração de ESTs no projeto. O centro de Bio-

Visão interna de um dos sequenciadores automáticos utilizados no HCGP. As amostras a serem sequenciadas são injetadas nos capilares por diferença de potencial elétrico



chips do Instituto Ludwig é coordenado por Luis Fernando Reis, com a importante colaboração de Alex Fiorini. Os clones gerados no projeto são armazenados de maneira ordenada em freezers a -70°C . Posteriormente, os clones de interesse são ordenados em membranas de nylon ou sílica, nas quais é feita a hibridização de mRNAs derivados de tumores de interesse. Desta maneira podemos avaliar simultaneamente o perfil de expressão de milhares de moléculas de mRNA. Tal capacidade é de importância crucial na compreensão do câncer humano. Os dados podem ser fundamentais na tentativa de diferenciar tumores agressivos de tumores mais brandos ou ainda tumores com diferentes tipos de resposta à terapêutica.

Se um país pretende ao menos usufruir das pesquisas genéticas derivadas do conhecimento do genoma humano, a disponibilidade de equipamentos e a formação de massa crítica na área de genoma são pré-requisitos fundamentais. Nesse sentido, a existência e a continuidade da ONSA no Brasil são uma garantia de participação ativa de nosso país na geração e adequada exploração das valiosas informações contidas no genoma humano. Essa participação apenas será possível com a disponibilidade de uma tecnologia genômica no país. Diante da imensa biodiversidade brasileira, a nossa entrada na era da pesquisa genômica tem uma importância estratégica. As possibilidades são tão vastas que, com certeza, ainda estão além da nossa imaginação.



Noventa e seis amostras são sequenciadas simultaneamente