

PROJETO TRANSCRIPTOMA

Análise da Expressão Gênica em Larga Escala Usando DNA - Arrays



Os projetos *genoma* estão revolucionando a maneira como a biologia e a medicina serão exploradas no próximo século e além (Collins et al., 1998). Com o seqüenciamento total dos genomas humano e de outras espécies e a disponibilidade das bibliotecas de cDNA, abriu-se a possibilidade de uma nova estratégia para a genética, que está sendo chamada de *genômica funcional*, a interpretação da função das seqüências de DNA numa escala genômica.

No curso do seqüenciamento de diversos genomas, demonstrou-se que muitos genes são descobertos somente quando suas seqüências totais tornam-se conhecidas. Assim, a descoberta de novos genes fica dependente do trabalho de seqüenciamento do DNA.

Entretanto, o conhecimento da estrutura de um gene ou de outro elemento funcional é somente parte da resposta que precisamos.

A necessidade vai mais além; está na elucidação da função dos genes e isso resultará da comparação de dados, como interação de genomas com o ambiente ou níveis de expressão gênica com o desenvolvimento normal e patológico.

Isto será um desafio para toda a biologia do futuro.

Os recursos rotineiros atuais para abordarmos a função dos genes numa escala genômica são por comparação e análise de padrões de seqüência com expressão gênica (RNAs mensageiros).

Mas são necessárias novas estratégias para podermos relacionar a expressão dos genes diretamente com os fenômenos biológicos. Isso nos aproximará mais da função desses genes.

A caracterização em larga escala de transcritos (RNAs) e suas proteínas re-

sultantes está bem dentro da análise funcional do genoma.

É bastante razoável pensarmos que o seqüenciamento e a análise dos níveis de expressão de todos os genes de um organismo devem ser tomados como alta prioridade.

Por analogia com o termo *genoma*, que representa o conjunto completo dos genes (DNA) de um organismo, criou-se o termo *transcriptoma*, que representa o conjunto completo dos transcritos (RNAs). Seguindo a idéia, temos também o *proteoma*, que representa todas as proteínas.

O novo desafio para a biologia está, então, na análise global desses sistemas: genoma, transcriptoma e proteoma.

Recursos metodológicos para enfrentar esses desafios estão sendo padronizados com sucesso.

Dentro da análise dos transcriptomas, o primeiro recurso disponível foram as bibliotecas de DNAs complementares (cDNA), ou seja, um conjunto de clones (de fagos ou plasmídeos) que abrigam cópias, se possível, de todos os RNAs mensageiros de uma linhagem celular ou órgão na forma de cDNA.

Seqüências de cDNA e conjuntos validados de clones com marcadores únicos são muito úteis na análise da expressão gênica.

Atualmente, uma das “últimas palavras” em tecnologia para o estudo dos transcriptomas são os *DNA-arrays*.

Aliando-se a disponibilidade das bibliotecas de cDNA com a robótica de alta precisão na deposição de pequenas amostras em superfícies sólidas, tornou-se possível a preparação dos *arrays* (arranjos) de clones de cDNA ou até de produtos de PCRs (reação de polimerização em cadeia) em membranas de nylon ou em lâminas de vidro (figura 1).

Geraldo A. S. Passos

Professor Associado (Genética),
Faculdade de Odontologia de
Ribeirão Preto e
Grupo de Imunogenética Molecular,
Faculdade de Medicina de Ribeirão Preto, USP
passos@rgm.fmrp.usp.br

Catherine Nguyen

Bertrand Jordan
Pesquisadores do Centre d'Immunologie
Marseille-Luminy, Equipe TAGC,
INSERM-CNRS de Marseille, França
nguyen@ciml.univ-mrs.fr
jordan@ciml.univ-mrs.fr

Fotos cedidas pelos autores

Vejamos o princípio do método dos *DNA-arrays* e das *sondas complexas de cDNA* (Jordan, 1998).

Essencialmente, os métodos das *assinaturas de hibridação* usam sondas complexas preparadas a partir de RNAs mensageiros de uma dada linhagem celular, tecido ou amostra cirúrgica por transcrição reversa e marcação radioativa. A sonda contém muitas espécies de RNAs mensageiros que podem variar muito em quantidade; por exemplo, no cérebro, podemos encontrar espécies variando de 1.000 a 30.000 moléculas de RNA mensageiro por célula.

Sugere-se que uma célula de mamífero típica contenha, aproximadamente, 300.000 moléculas individuais de RNAs mensageiros. Algumas delas estão presentes em abundância que variam de um a vários por cento.

Uma parcela razoável de RNAs mensageiros são de espécies raras, representados a uma molécula por célula ou mesmo menos (uma espécie de RNA mensageiro presente somente uma vez a cada dez células poderá, mesmo assim, ter significado biológico).

O seqüenciamento sistemático de cDNAs poderá proporcionar os dados de caracterização do transcriptoma, especialmente se feito em bibliotecas de cDNA (Okubo et al., 1992).

Entretanto, a quantificação de baixos níveis de expressão envolve uma quantidade não usual de seqüenciamentos: para seqüenciar um dado RNA mensageiro abundante de 1:10.000 faz-se necessário o seqüenciamento de 50 a 100.000 clones. Para cada biblioteca !

Os métodos que envolvem os *tags* (etiquetas moleculares de cDNAs) diminuem o seqüenciamento pela redução de cada cDNA, mas o trabalho para padronizar esses métodos é enorme e as limitações estatísticas ainda persistem.

Por outro lado, os métodos de *hibridação molecular*, usando sondas complexas e um grande arranjo de alvos de DNA (*DNA-arrays*) têm seu poder derivado do fato de que cada experimento individual proporciona uma grande quantidade de informação; ainda não há método rival para as medidas de expressão gênica em grande escala.

É claro que o seqüenciamento do



Figura 1.

Utilização de bibliotecas de cDNA, produtos de PCR ou oligonucleotídeos na confecção dos DNA-arrays e suas alternativas de análise

DNA é e será o “método final” para a caracterização da estrutura gênica. Com ele podemos deduzir a seqüência dos resíduos de aminoácidos da proteína mesmo sem a termos isolado.

Mas fica claro que o caminho mais rápido e lógico para identificarmos os genes implicados nos fenômenos biológicos normais e patológicos está sendo o uso dos *DNA-arrays*.

O poder dos métodos que também estão sendo chamados de *assinaturas de hibridação* se deve ao fato de que milhares de níveis de expressão podem ser mensurados num único experimento.

Brevemente, uma sonda complexa é hibridada com um *array* consistindo de muitos “alvos de DNA”, cada um representando um gene em particular. Seguiremos a definição adotada por Jordan (1998), onde o DNA imobilizado é considerado como alvo.

Os DNA alvos podem ser tanto colônias de bactérias fixadas em membranas, produtos PCR de clones de cDNA ou oligonucleotídeos sintéticos desenhados para ensaiar um gene em particular.

O ponto importante é que a combinação de uma sonda complexa contendo muitas espécies de RNAs mensageiros com um grande arranjo (*array*) de alvos, o que permite a coleção de um

grande volume de informações em paralelo.

Nesse ponto, temos que salientar que as diferenças entre esse tipo de hibridação com as clássicas hibridações Southern- e northern-blot. Nestas últimas, trabalhamos com um excesso de sonda, o DNA alvo é hibridado até a saturação, no final da incubação.

Nos experimentos de medidas de expressão gênica, entretanto, cada seqüência individual na sonda complexa está presente em pequenas quantidades em relação ao(s) seu(s) alvo(s) no

array.

Determinações realizadas pelo grupo de um dos autores (B. Jordan, ver Nguyen et al., 1995) para esses *arrays* indicaram uma proporção de 30 ng de inserto de cDNA de uma colônia de *E. coli* crescida sobre a membrana de nylon para menos de 0,1 ng de RNA mensageiro (cDNA).

Nessas condições, a cinética de hibridação é linear e a quantidade da sonda hibridada a um dado alvo será proporcional à abundância da seqüência correspondente na sonda complexa e, portanto, ao nível de expressão do gene na célula ou tecido do qual a sonda foi preparada.

Embora o princípio do uso de *DNA-arrays* tenha sido discutido tão cedo como, em 1991 (Lennon & Lehrach, 1991), somente quatro anos após é que apareceram os primeiros trabalhos mostrando a possibilidade de quantificação em *DNA-arrays* (Nguyen et al., 1995; Zhao et al., 1995; Pietu et al., 1996).

A aquisição dos dados usando sondas radioativas ou fluorescentes envolve outras necessidades como algoritmos para detecção dos pontos (*spots* de hibridação), subtração do *background*, atribuição dos sinais às entidades corretas (Granjeaud et al., 1996) e, geralmente, procedimentos corretos para o manuseio de milhares de dados gerados em cada experimento individual.

Os requerimentos para um sistema ideal são, principalmente, a habilidade de quantificar níveis de expressão abaixo do nível de uma molécula por célula (1/300.000); dar uma resposta linear dentro da faixa de abundância (1/300.000 a 1/100); capacidade de mensuração de

muitos alvos ao mesmo tempo, idealmente o complemento total dos 100.000 genes humanos e o uso de pequenas quantidades de material de partida para a sonda, já que os interesses envolvem o estudo de populações de células difíceis de coletar e também de biópsias clínicas.

Tudo isso deve ser conseguido com uma reprodutibilidade aceitável, pois uma diferença de 2 entre duas assinaturas de hibridação pode ser altamente significativo. É claro que a aparelhagem deve ser compacta, com alta performance e fácil de usar.

Nessas aplicações, clones de cDNA (ou produtos PCR) são arranjados em filtros de nylon com auxílio de um robô (figura 2); as sondas complexas são marcadas com fósforo radioativo (^{32}P ou ^{33}P) e, então, hibridadas com os filtros.

Os resultados são adquiridos com auxílio de sistemas de placas de imagens (*phosphor imagers*).

O uso de sondas fluorescentes, que permitem a miniaturização devido à sua muito boa resolução, iniciou-se em 1995 (Schena, 1995) e foi posteriormente desenvolvidas a ponto de padronizarem os *micro-arrays* contendo vários milhares de amostras de DNA em superfícies de um centímetro quadrado, os quais têm sido produzidos por laboratórios acadêmicos e firmas (figura 3).

Também desenvolveram os *oligonucleotide-chips*, sendo o setor privado norte americano o principal produtor. Tais *chips* contém 64.000 diferentes oligonucleotídeos sintetizados numa superfície de dois centímetros quadrados (Chee et al., 1996).

O *DNA-chip* resultante é então usado para aplicações tipo “quase-seqüenciamento”, como, por exemplo, a detecção de mutações ao longo do gene p53.

Embora originariamente projetado para esse fim, os *DNA-chips* podem também ser aplicados aos estudos de medidas de expressão gênica (Lockhart et al., 1996). Devido ao enorme potencial de miniaturização, o *DNA-chip* poderá, a longo prazo, representar a principal via para

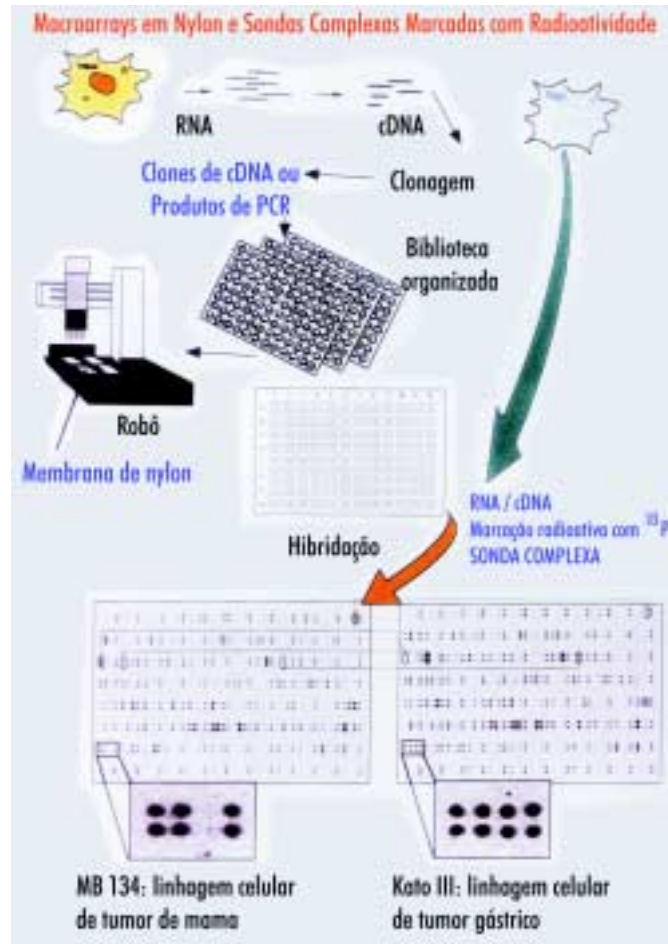


Figura 2.

Utilização dos *macro-arrays* e sondas complexas de cDNA radioativas na identificação de genes exclusivamente expressos numa dada linhagem tumoral

medidas sistemáticas de expressão gênica.

Entretanto, as membranas de nylon de alta densidade (membranas HD) são

Sites na Internet sobre *DNA-arrays* e expressão gênica:

<http://tagc.univ-mrs.fr/>
www.ncbi.nlm.nih.gov/dbEST/index.html
www-bio.llnl.gov/bbrp/image/image.html
www.clontech.com/clontech/Catalog/Hybridization/Atlas.html
www.genome-systems.com/GDA/
www.synteni.com
<http://cmgm.Stanford.EDU/pbrown/>

a opção mais razoável atualmente para os estudos de expressão gênica. As colônias de *E.coli* são crescidas sobre as membranas que são então lisadas e fixadas por métodos usuais.

Alternativamente, DNA obtido por PCR a partir desses clones também pode ser fixado nas membranas; isso pode contornar o problema de crescimento desigual das colônias de bactérias. Os produtos PCR podem ser aplicados às membranas como são, só necessitando de ser dosados e diluídos de maneira uniforme antes de ser aplicados. Mas realizar milhares de PCRs e depois dosar cada uma das reações não é tarefa fácil e muito demorada.

Por isso é que o crescimento de colônias bacterianas sobre as membranas ainda é a melhor opção.

Quanto às sondas complexas de cDNA, estas são preparadas a partir dos RNAs mensageiros (RNA poli A+) ou a partir de RNA total por transcrição reversa e marcação radioativa simultânea dos cDNAs resultantes (Bernard et al., 1996). A marcação radioativa é feita com ^{32}P ou, preferencialmente, com ^{33}P , que permite maior resolução quando se trabalha com mem-

branas HD.

A aquisição dos dados quantitativos das hibridações é obtida com as *imaging plates* de aparelhos leitores de fósforo (*phosphor imagers*). As *imaging plates* associadas com *scanners* a laser proporcionam um formato conveniente, já que muitas membranas podem ser simultaneamente expostas e um *scanning* demora em média cinco minutos.

A imagem da hibridação, uma vez adquirida, é salva como arquivo de computador e pode ser quantificada, isto é, os *spots* de hibridação são definidos assim como o *background* e valores numéricos são atribuídos a cada *spot*. Agora cada valor é associado ao clone correspondente para análises posteriores.

A sensibilidade atual das membranas HD está na faixa de 1/10.000, isto é, os níveis de expressão podem ser quantificados ao que corresponde

a espécies de RNA presentes nessa proporção em relação à população total de RNAs mensageiros.

As sondas são preparadas geralmente a partir de um a alguns microgramas

de RNA poli A+, ou a partir de 25 microgramas de RNA total para uso em *macro-arrays* (membranas HD de 8 x 12 cm) ou 5 microgramas de RNA total para uso em *micro-arrays* (1 cm²).

Esse “screening diferencial quantitativo” tem sido aplicado na procura de “novos” genes diferencialmente expressos em células cancerosas e em tecidos normais (Zhao et al., 1995) para estudar genes especificamente expressos no músculo (Pietu et al., 1996) e para estudos sobre a maturação do timo do camundongo (Rocha et al., 1997).

Nosso *Grupo de Imunogenética Molecular* situado no Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, USP, trabalha em colaboração com a *Equipe TAGC (Technologies Avancées pour le Génome et la Clinique)* no *Centre d'Immunologie de Marseille-Luminy (Marseille, França)*, por intermédio do Acordo INSERM-FAPESP. Esse acordo prevê a ida de pesquisadores brasileiros para a Equipe TAGC e a vinda dos pesquisadores franceses para o nosso laboratório.

Estamos trabalhando com análise da expressão gênica em larga escala durante o desenvolvimento do timo do camundongo, com o intuito de descobriremos “novos” genes implicados na maturação dos linfócitos T e na recombinação V(D)J dos genes de receptores de antígenos (TCR) (Passos et al., 1999).

Essa tecnologia tenderá a tornar o processo muito mais prático quando dispusermos de toda a expressão de genomas (o transcriptoma) numa só membrana, provavelmente isso acontecerá com os *micro-arrays* e *DNA-chips* e detecção por fluorescência.

O que, há dois anos, ainda era um ponto de “estrangulamento” nessa tecnologia, hoje, é realidade, ou seja, já encontramos no mercado todo um conjunto de aparelhos, que vão desde o robô para repicar os clones nas placas de microtitulação até a preparação dos *micro-arrays* acoplados a um sistema para hibridação e a um *scanning* que identifica e quantifica os níveis de hibridação com *softwares* dedicados.

Em resumo, os *micro-arrays* são definitivamente a etapa lógica nos estudos de expressão gênica em larga escala,

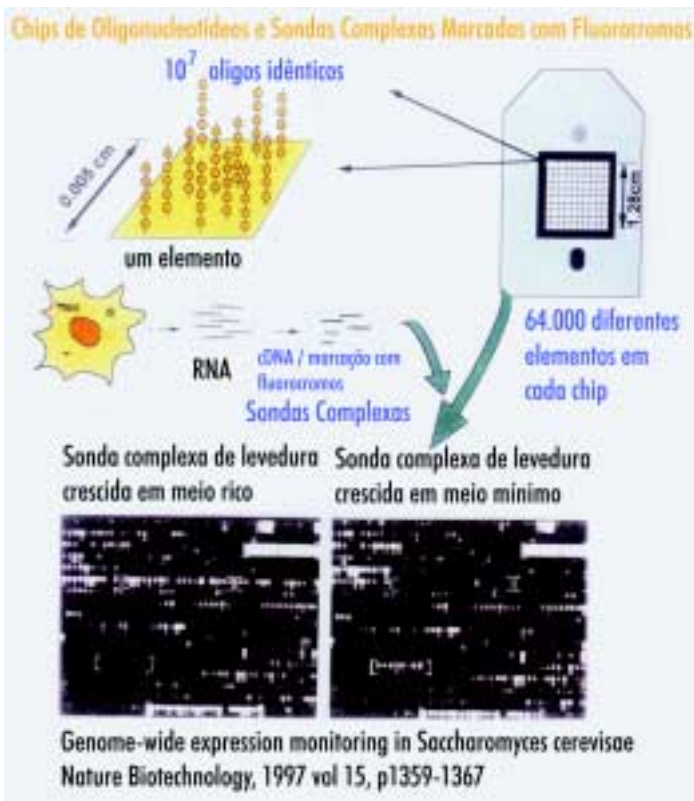


Figura 3.

Utilização dos *DNA-chips* na identificação de genes preferencialmente expressos pela levedura crescida em meios de cultura de diferentes composições

embora estejam rapidamente sendo alcançados pelos *DNA-chips* que foram originariamente desenvolvidos para um propósito diferente.

As realizações que apresentamos revelam o imenso potencial dos *DNA-arrays* e o futuro promete a essa tecnologia uma revolução comparável àquela que ocorreu com o advento da *reação de polimerização em cadeia (PCR)*.

Enfim, a padronização de *DNA-chips* para explorar *n* genes em *n* situações abre um futuro de combinações muito mais sofisticadas, do tipo *lab on a chip*.

No momento em que as necessidades do seqüenciamento massivo de genomas orientou a biologia em direção aos grandes equipamentos, no modelo da física de partículas, vemos o nascimento de uma solução metodológica proporcionada por uma ferramenta que cabe na palma de nossa mão!

Referências Bibliográficas:

Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A.M., Jordan, B.R. e Nguyen, C. (1996) *Nucleic*

Acids Res. 24: 1435-1443.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L. et al. (1998) *Science* 282: 682-689.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. e Fodor, S.P. (1996) *Science* 274: 610-614.

Granjeaud, S., Nguyen C., Rocha, D., Luton, R. e Jordan, B. (1996) *Genetic Anal. Biomol. Eng.* 12: 151-162.

Jordan, B.R. (1998) *J. Biochem. (Tokyo)* 124: 251-258.

Lennon, G.G. e Lehrach, H. (1991) *Trends Genet.* 7: 314-317.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mitmann, M., Wang, C., Kobayashi, M., Horton, H. e Brown, E.L. (1996) *Nature Biotechnol.* 14: 1675-1680.

Nguen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P. e Jordan, B. (1995) *Genomics* 29: 207-215.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. e Matsubara, K. (1992) *Nature Genet.* 2: 173-179.

Passos, G.A.S., Junta, C.M., Espagnol, A.R., Macedo, C., Jordan, B., Nguyen, C., Victorero, G., Liorod, B., Granjeaud, S. (1999) *Genetics Mol. Biol.* 22, *supp.*: 312.

Pietu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariage-Samson, R., Houlgate, R., Soularue, P. e Auffray, C. (1996) *Genome Res.* 6: 492-503.

Rocha, D., Carrier, A., Naspetti, M., Victorero, G., Anderson, E., Botcherby, M., Nguyen, C., Naquet, P. e Jordan, B. (1997) *Immunogenetics* 46: 142-151.

Schena, M., Shalon, D., Davis, R.W. e Brown, P.O. (1995) *Science* 270: 467-470.

Zhao, N., Hashida, H., Takahashi, N., Misumi, Y. e Sakaki, Y. (1995) *Gene* 156: 207-213.

Agradecimentos:

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e ao Institut National de la Santé et de la Recherche Médicale (INSERM, França) por possibilitar o trabalho em colaboração entre nossas equipes (Proc. No. 98/09789-4).