



GENOMA FUNCIONAL

Uso de arranjos de DNA em náilon para a análise da expressão gênica em larga escala

Juliana de Maria Felix

Mestranda em Biologia Celular e Estrutural
(UNICAMP)
felix@unicamp.br

Rodrigo Duarte Drummond

Mestrando em Genética e Biologia Molecular
(UNICAMP)
rduarte@unicamp.br

Fábio Tebaldi Silveira Nogueira

Doutorando em Genética e Biologia Molecular
(UNICAMP)
tebaldi@unicamp.br

Vicente Eugenio de Rosa Junior

Doutorando em Genética e Biologia Molecular
(UNICAMP)
vicentej@unicamp.br

Renato Atílio Jorge

Professor Depto. de Físico-Química
Instituto de Química - (UNICAMP)
rjorge@iqm.unicamp.br

Paulo Arruda

Professor Doutor - Depto. de Genética e
Evolução
Pesquisador do Centro de Biologia Molecular e
Engenharia Genética (CBMEG-UNICAMP)
parruda@unicamp.br

Marcelo Menossi

Professor Doutor Depto. Genética e Evolução
Pesquisador do Centro de Biologia Molecular e
Engenharia Genética (CBMEG-UNICAMP)
menossi@unicamp.br

Os organismos respondem a diferentes estímulos ambientais por meio de alterações no proteoma, conjunto de proteínas das células. Esse processo pode permitir a adaptação a condições adversas, como, por exemplo, choque térmico, exposição a drogas, ataque de patógenos, etc. O fluxo da informação gênica do DNA nos cromossomos (genoma) até o proteoma, é intermediado pelo conjunto das moléculas de RNA (transcritoma). Assim, a concentração relativa dos transcritos de um determinado gene em uma célula é um indicativo do quanto esse gene está sendo expresso, isto é, do quanto a célula está investindo do seu maquinário bioquímico para produzir a proteína codificada pelo gene.

Em vista disso, foram desenvolvidas metodologias visando a medir a concentração relativa dos transcritos dos genes em células e tecidos. Até pouco tempo atrás, a análise da expressão gênica era feita com metodologias que avaliavam poucos genes de cada vez: *Northern blot*, *dot blots*, *RT-PCR*, entre outros. O crescimento exponencial do número de novos genes descobertos nos projetos genoma e o desenvolvimento dos arranjos de DNA propiciaram uma nova abordagem nos estudos da regulação gênica, tornando possível o monitoramento dos níveis de transcritos de um grande número de genes simultaneamente. O presente artigo enfoca a tecnologia dos arranjos de DNA em náilon, que empregam metodologias rotineiras, robustas e de fácil implementação na maioria dos laboratórios de biologia molecular, sem a necessidade de investimentos vultoso

se e longos períodos de padronização.

Os arranjos de DNA são suportes sólidos, comumente vidro ou náilon, aos quais estão fixadas, de forma ordenada, seqüências completas ou parciais de genes. A Figura 1 mostra as etapas de uma análise de expressão gênica em larga escala empregando arranjos em náilon. O DNA dos genes é depositado em membranas réplicas, contendo os mesmos genes nas mesmas posições, em densidades que variam de 10 a 1000 genes/cm². A partir do RNA das células em estudo são produzidas sondas de cDNA via transcrição reversa na presença de um nucleotídeo radioativo, que permite sua detecção posterior. As sondas são hibridadas contra os arranjos e quanto maior a expressão de um gene em uma determinada condição, maior será o número de moléculas de RNAm, sendo maior o número de cDNAs desse gene na sonda sintetizada. Conseqüentemente, maior será o valor de intensidade do sinal derivado da sonda hibridada na região do arranjo que contém a seqüência desse gene.

Caso o exemplo esquematizado na Figura 1 fosse feito com arranjos em vidro, o procedimento seria muito similar. No entanto, uma das duas sondas seria sintetizada na presença de outro nucleotídeo modificado, não radioativo, mas fluorescente. A outra sonda seria sintetizada com um segundo nucleotídeo que fluoresce num comprimento de onda distinto do utilizado na primeira sonda. Dessa forma, ambas as sondas poderiam ser hibridadas num mesmo arranjo, o que é uma clara vantagem. A detecção da intensidade das duas sondas hibridadas no arranjo é

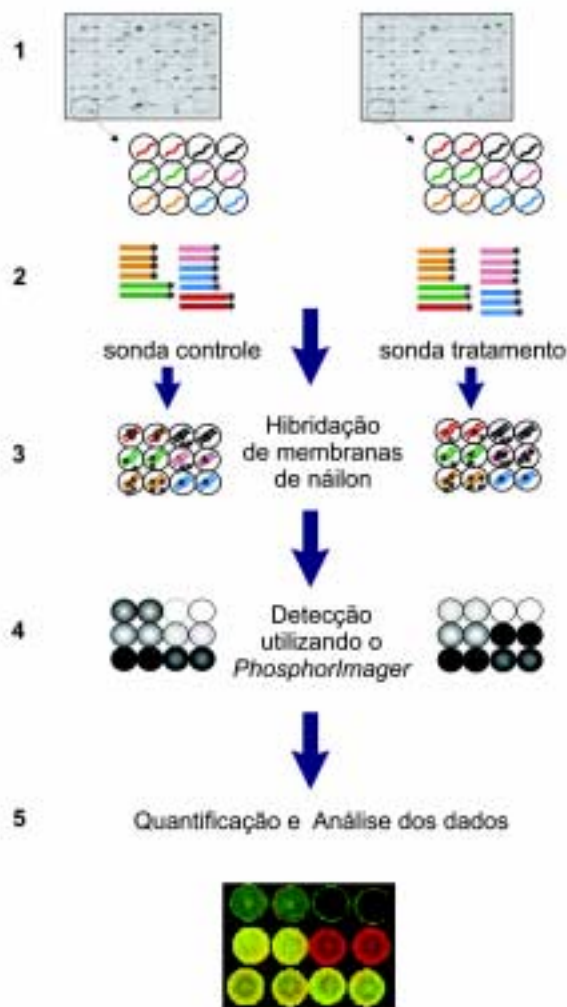


Figura 1:
Representação esquemática de um experimento utilizando macroarranjos
(1). O DNA plasmidial contendo os ESTs do projeto SUCEST é fixado nas membranas de náilon, em duplicata (cada cor representa um gene distinto); **(2).** A partir das amostras de RNA extraídas de diferentes tecidos ou tratamentos, é feita uma transcrição reversa em presença de α^{33} -dCTP, produzindo as sondas de cDNA; **(3).** As membranas são hibridadas com as sondas de cDNA radioativas; **(4).** A radioatividade emitida de cada *spot* é detectada utilizando-se um *phosphorimager*; **(5).** Os dados são comparados utilizando-se programas específicos para identificação do perfil de expressão de cada gene

realizada com um equipamento especial. A densidade de genes em arranjos de vidro pode chegar a ser uma ordem de magnitude maior. Para uma revisão sobre os arranjos em vidro, veja Freeman *et al.*, 2000. Apesar dessas vantagens, a hibridação de arranjos em vidro requer investimentos muito maiores que os necessários para arranjos em

al., 1995), a identificação de genes expressos diferencialmente em função da presença ou ausência de luz em *Arabidopsis* (Desprez *et al.*, 1998), a análise de 21.500 genes de ovário para identificação de genes superexpressos em carcinomas (Schummer *et al.*, 1999), o monitoramento do perfil de expressão gênica em raízes de toma-

náilon. Além disso, é um consenso que a padronização dessa metodologia é muito mais complexa.

Convém ressaltar que existem diversos termos empregados para descrever os arranjos de DNA: *glass arrays*, *DNA chips*, *biochips* e *microarray* por um lado, os quais, geralmente, refletem arranjos em vidro, e *nylon array*, *filter arrays*, *high density membranes* e *macroarray*, que se referem a arranjos em náilon. Neste artigo utilizaremos os termos arranjos em náilon e arranjos em vidro, uma vez que o suporte empregado é determinante para a seleção das metodologias de deposição do DNA no arranjo, síntese de sonda, hibridação e detecção dos níveis de expressão gênica.

Os arranjos em náilon têm sido empregados com sucesso desde o trabalho de Lennon & Lehrach (Lennon & Lehrach, 1991), os quais identificaram genes diferencialmente expressos em diferentes tratamentos. A partir daí, vários estudos vêm sendo realizados empregando-se essa metodologia, como, por exemplo, a análise da expressão gênica em 2.505 genes no cérebro humano (Zhao *et*

teiro durante suplementação com nitrogênio (Wang *et al.*, 2001), entre outros. O primeiro relato na literatura sobre o uso de arranjos em vidro foi publicado pelo grupo de Patrick Brown, da universidade de Stanford (Wenzl *et al.*, 2001). A partir daí, vários estudos foram feitos utilizando-se essa metodologia (para mais detalhes veja revisão de Passos *et al.*, 1999 e de Freeman *et al.*, 2000).

O nosso grupo tem estudado o perfil de expressão gênica da cana-de-açúcar em resposta a estresses abióticos empregando macroarranjos (*macroarrays*), que são arranjos em náilon com média densidade de genes (ao redor de 20 genes/cm²). Os genes utilizados são provenientes de bibliotecas de cDNA do projeto genoma da cana-de-açúcar, SUCEST (<http://sucest.lad.ic.unicamp.br>).

Análises genômicas empregando macroarranjos

Os arranjos em náilon são uma opção interessante aos arranjos em vidro, tanto pelo custo como pela facilidade de implementação. A seguir estão descritas as diversas etapas que realizamos em nosso laboratório para implementação dessa metodologia. Para maiores detalhes, veja *link* na Tabela 1.

1. Confeção dos macroarranjos

As amostras de DNA são preparadas a partir de bactérias contendo os plasmídeos de bibliotecas de EST, empregando lise alcalina em placas de 96 poços, prática comum nos laboratórios de seqüenciamento em larga escala. O DNA é fixado às membranas manualmente, com o auxílio de um replicador composto por 96 pinos (V&P Scientific, EUA, <http://www.vp-scientific.com>), que possibilita a transferência simultânea de ~0,1µL de DNA de cada posição de uma placa de 96 poços, equivalente a 10 ng de DNA em cada spot.

Com esse procedimento, o DNA de dezesseis placas (1.536 amostras de DNA) pode ser transferido para uma membrana de 12 x 8 cm (Fig. 1.1). Normalmente cada EST é depositado em duplicata, para maior confiabilidade da análise, com o qual os macroar-

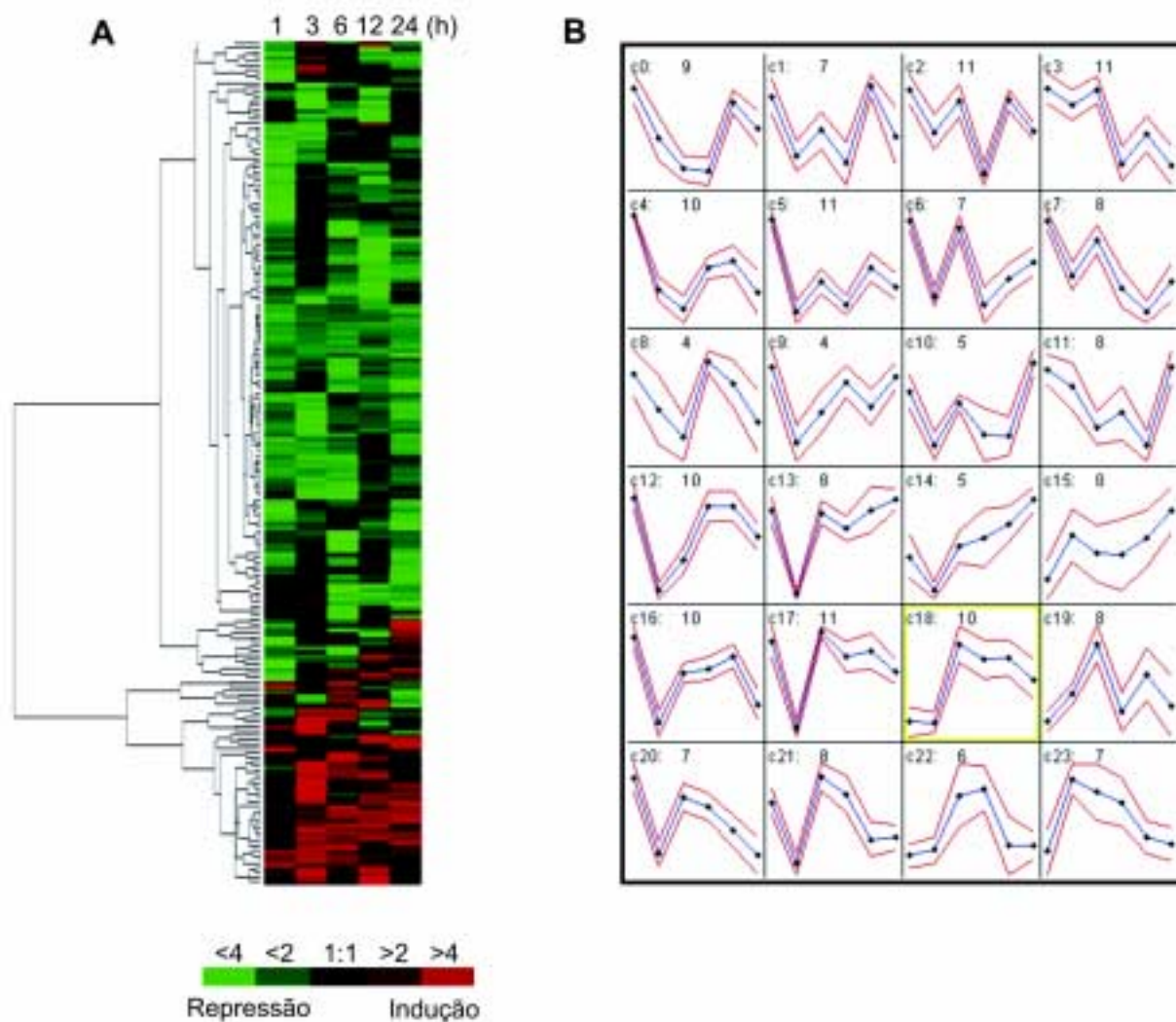


Figura 2:
Formas de visualização dos dados de macroarranjos

(A). Dendrograma mostrando os padrões de expressão ao longo do tempo em resposta a metil jasmonato. Cada linha representa um gene e as colunas representam os tratamentos. A expressão de cada gene é calculada a partir da razão entre o sinal no tratamento e no controle, sendo representada com diferentes cores, de acordo com a referência mostrada na parte inferior da figura. (B). Agrupamentos obtidos com o algoritmo de *Self Organizing Maps* (SOM). Os genes foram agrupados em 24 conjuntos (c0 a c23) e o número de genes em cada grupo está indicado à direita do nome do cluster. A linha central de cada gráfico representa a média do sinal de expressão dos genes do grupo nos seis pontos experimentais (controle, 1, 3, 6, 12 e 24h). As linhas vermelhas representam o desvio padrão no grupo para cada valor de expressão

ranjos produzidos com o replicador manual contém 768 ESTs (aproximadamente 8 ESTs/cm²). A confecção de membranas de maior densidade pode ser conseguida com o auxílio de robôs, podendo-se obter uma densidade de genes dez vezes maior. O BCCCenter (<http://www.bcccenter.fcav.unesp.br>), que gerencia os clones do projeto SUCEST, tem previsão de disponibilizar para a comunidade científica membranas de 22,2 x 22,2 cm contendo 27.648 ESTs em duplicata.

2. Síntese de sonda e hibridação das membranas de náilon

As sondas são produzidas a partir de RNA total através de reação de transcrição reversa, na presença α 33-dCTP (Fig. 1.2). Os cDNAs marcados radioativamente são hibridados contra as membranas de alta densidade (Fig. 1.3) em fornos de hibridação comuns, empregando um protocolo muito similar ao de *Southern blot*, técnica robusta e de uso corrente na maioria

dos laboratórios de biologia molecular.

3. Obtenção das imagens

Após as lavagens para eliminar a hibridação inespecífica, a membrana de náilon é posta em contato com uma placa composta por material sensível à radioatividade (*imaging plate*). Após 96 h, a placa é “lida” em um aparelho do tipo *Phosphorimager* que a armazena na forma de uma imagem digitalizada (Fig. 1.4). Esse aparelho é o

Tabela1. Relação de programas utilizados para análise de perfil de expressão gênica

Softwares para Banco de Dados	
ARGUS	Brigham and Women's Hospital and Harvard Medical School (http://vessels.bwh.harvard.edu/software/argus/default.htm)
GeneX	Banco de dados aberto para expressão gênica (http://genex.ncgr.org/)
MAExplorer	Padrões de DataMining e Expressão Gênica (http://www.lecb.ncicrf.gov/MAExplorer/)
MAXDSQL	Manchester University, baseado em requerimentos do MIAME-EBI (http://www.bioinf.man.ac.uk/microarray/maxd/maxdSQL/)
MGED	Grupo de Banco de dados de expressão gênica em Microarrays (http://www.mged.org/)
Exemplos de Banco de Dados para Expressão Gênica	
Microarray Centre	Centro de Microarray, The Ontario Cancer Institute (http://www.oci.utoronto.ca/services/microarray/index.html)
NIH	Banco de dados de expressão gênica, Molecular Pharmacology of Cancer (http://discover.nci.nih.gov/nature2000/naturemain.html)
SMD	Banco de Dados de Microarrays, Stanford University (http://genome-www4.stanford.edu/MicroArray/SMD/index.html)
yMGV	Visão global sobre Microarray de levedura (http://transcriptome.ens.fr/ymgv/)
Softwares para Análises	
ArrayViewer	Visualização e análise de dados de Microarrays – gratuito (http://www.tigr.org/softlab/)
BRB ArrayTools	Pacote integrado para visualização e análise estatística de dados de expressão gênica em Microarrays; Richard Simon & Amy Peng, National Cancer Institute, USA – gratuito (http://linus.nci.nih.gov/BRB-ArrayTools.html)
Cluster	Software de clusterização; Michael Eisen, Eisen Lab, University of California, Berkeley - gratuito (http://rana.lbl.gov/EisenSoftware.htm)
Expression Profiler	Ferramenta para clusterização, análise e visualização de expressão gênica. Análises on line – gratuito (http://ep.ebi.ac.uk/)
Multi Experiment Viewer	Aplicação Java que permite análise de dados de Microarrays para identificar padrões de expressão e genes expressos diferencialmente – gratuito. (http://www.tigr.org/softlab/)
SAM	Análise de Significância de Microarrays – gratuito (http://www-stat.stanford.edu/~tibs/SAM/)
SOM	Aplicação de Self Organizing Maps; Gavin Sherlock, Stanford University – gratuito (http://genome-www.stanford.edu/~sherlock/SOMviewer.html)
Spotfinder	Detecta e quantifica spots – gratuito (http://www.tigr.org/softlab/)
Treeview	Visualização e clusterização de dados providos do software Cluster; Michael Eisen, Eisen Lab, University of California, Berkeley - gratuito (http://rana.lbl.gov/EisenSoftware.htm)
X-Cluster	Software de clusterização; Gavin Sherlock, Stanford University – gratuito (http://genome-www.stanford.edu/~sherlock/SOMviewer.html)
Protocolos para arranjos de DNA	
LGF	Home page do laboratório de Genoma Funcional, CBMEG, UNICAMP, na qual estão disponibilizados protocolos completos para arranjos de DNA (http://cafe.cbmeg.unicamp.br)

equipamento mais custoso da metodologia (preço ao redor de US\$ 40.000). A imagem é então quantificada utilizando-se programas específicos (veja Tabela 1) que geram tabelas de dados nas quais cada *spot* (região da membrana contendo um gene) recebe um valor numérico de acordo com a sua intensidade na imagem. A inspeção visual em filmes de raio-X também é possível, porém indicada apenas em ensaios qualitativos, uma vez que a quantificação das intensidades dos *spots*

fica muito prejudicada.

4. Análise de dados

Cada gene está representado por dois *spots* no macroarranjo e a média desses dois dados é considerada a expressão do gene no macroarranjo, na condição experimental testada. Com a tecnologia disponível hoje para os arranjos em náilon e vidro, em boas condições experimentais, geralmente toma-se como genes induzidos aque-

les que apresentaram um sinal pelo menos duas vezes maior no tratamento comparado ao controle, enquanto os reprimidos possuem uma intensidade de sinal no tratamento menor que a metade do controle. No entanto, essa premissa pode não ser verdadeira, sendo necessário avaliar a variabilidade dos dados obtidos nos diversos tratamentos.

As metodologias para análise estatística dos dados de arranjos de DNA estão em franco desenvolvimento,

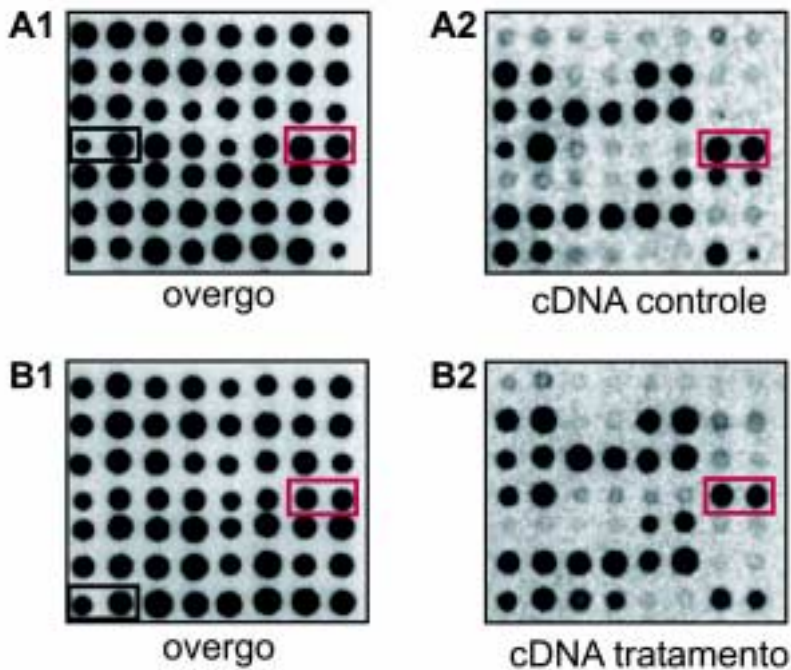


Figura 3:
Variações no sinal detectado devido a diferentes quantidades de DNA fixadas no spot

Macroarranjos réplicas contendo genes em duplicata foram hibridados com a sonda overgo (**A1 e B1**) e, em seguida, com sondas de cDNA de células do controle (**A2**) e de um tratamento (**B2**). Em A1 e B1, os retângulos em negro mostram spots réplicas com diferentes massas de DNA depositadas na mesma membrana, enquanto que os retângulos em vermelho mostram spots com variações entre membranas. Em A2 e B2, os retângulos em vermelho mostram os sinais das duas sondas de cDNA. Observa-se uma menor intensidade na sonda do tratamento. Essa aparente repressão do gene está relacionada com a diferença de massa nos spots

existindo diversas estratégias que atribuem níveis de confiança às diferenças observadas entre tratamentos. Em tais análises, é necessário obter o logaritmo dos valores de expressão, já que esses dados não apresentam distribuição normal.

O volume de dados gerados em nossos experimentos, com aproximadamente, 3.000 genes e 5 a 6 tratamentos, ainda é passível de ser analisado com planilhas de cálculo, como o MS Excel. Nessas análises, calculamos os logaritmos das razões entre a expressão de cada gene em um tratamento e a expressão observada no controle. Como é esperado que a maioria dos genes não altere sua expressão, a média dessas razões deve ser próxima a 1 (com o que a média dos logaritmos das razões deve ser próxima a zero). O indicador da variabilidade

de experimental é o desvio padrão dos logaritmos das razões de todos os genes de um macroarranjo. Em cada experimento, consideramos induzidos os genes cujo logaritmo da razão seja superior à média de todos os genes somada a 1,65 vezes o desvio-padrão (o inverso vale para os genes reprimidos). Esse fator multiplicador do desvio é escolhido com base na suposição de que os valores de logaritmos apresentam distribuição normal, de modo que a chance de um gene aparecer como falso positivo é menor que 5%. Por fim, selecionamos aqueles genes que foram induzidos (ou reprimidos) em duas repetições do experimento. Com esse critério, a chance de um gene ser erroneamente identificado como sendo expresso diferencialmente em função do tratamento é de $(5\%)^2$, o que garante um nível de

significância de mais de 99% nos resultados.

Uma vez selecionados os genes que variam a expressão, são feitas outras análises que envolvem programas baseados em algoritmos de agrupamento (tabela 1). Esses programas identificam grupos de genes com um padrão de expressão semelhante nos tratamentos avaliados, permitindo a visualização dos dados através de dendrogramas e gráficos. A Figura 2 mostra duas formas de agrupamento e visualização de dados de expressão gênica em plântulas de *Arabidopsis thaliana* expostas ao hormônio metil jasmonato ao longo de um período de 24 h (dados calculados a partir dos resultados de Sasaki *et al.*, 2001). Através dessa análise, podemos identificar genes ligados a um mesmo processo fisiológico ou por via metabólica. Assim, é possível inferir o papel bioquímico de genes desconhecidos, caso existam no seu agrupamento genes cuja função já tenha sido descrita.

Uma outra forma de representar os genes induzidos e reprimidos pode ser obtida sobrepondo-se duas imagens digitalizadas (tratamento e controle), e gerar, assim, uma única imagem falsamente colorida (Fig. 1.5). Nessas imagens, similares às imagens oriundas dos experimentos de arranjos em vidro, que utilizam dois fluoróforos, os spots em vermelho representam genes induzidos, os spots em verde representam genes reprimidos, os spots em amarelo representam genes que não alteraram sua expressão em resposta ao tratamento, enquanto que os spots em negro representam genes com níveis de expressão próximos a zero.

Banco de dados: O destino dos dados de expressão gênica

O grande volume de informação gerado pelos projetos de análise de transcriptomas tem tornado cada vez mais complexo o armazenamento e a análise dos dados. Para contornar tal dificuldade, devem ser implementados, bancos de dados, que disponibilizem, de modo confiável, os dados e ferramentas de análise. Em muitos casos, esses bancos são abertos, o que aumenta ainda mais a aplicabilidade da pesquisa.

Alguns programas de bancos de dados estão disponíveis gratuitamente

te (Tabela 1). No momento, nossa equipe está-se empenhando em implementar o banco GeneX (Mangalam *et al.*, 2001). O GeneX é um banco de dados com código fonte aberto, que possibilita sua adaptação às necessidades específicas de cada grupo de pesquisa, além de possuir uma interface gráfica tipo web que pode ser acessada remotamente, facilitando a utilização de ferramentas de análises e a interação entre diversos grupos.

Padronização da metodologia

Diversos experimentos de padronização foram realizados visando à implementação da técnica de macroarranjos de DNA em nosso laboratório. Vários fatores podem afetar a sensibilidade, a reprodutibilidade e a confiabilidade da técnica, destacando-se: as variações na massa de DNA presente nos *spots* fixados na membrana; as flutuações que ocorrem durante as etapas de transcrição da sonda, hibridação e lavagem das membranas; além das variações no *background* (Schuchhardt *et al.*, 2000; Schummer *et al.*, 1999). Uma vez que esses fatores dificultam a implementação da técnica, nós detalharemos a seguir alguns dos procedimentos adotados em nosso grupo:

1. Variação na quantidade de DNA depositada nas membranas

O sinal obtido nas hibridações de ácidos nucleicos é proporcional não só à massa de transcrito presente na sonda, mas também à massa de DNA depositada na membrana. Tal como foi mostrado na Figura 1, para comparar diferentes tratamentos, é necessário empregar macroarranjos réplicas. Assim, existe o risco de considerar variações de sinal, que refletem apenas as diferentes massas de DNA nos *spots* nas membranas réplicas, não relacionadas com diferenças de expressão entre os tratamentos avaliados. Uma forma de se estimar flutuações na massa de DNA em cada *spot* é a hibridação das membranas de náilon com uma sonda que reconhece uma região específica do plasmídeo (chamada *overgo*), comum a todos os ESTs analisados (Perret, 1998). Uma vez que todos

os *spots* competem igualmente pela sonda *overgo*, a hibridação entre a sonda e o *spot* é proporcional à quantidade de DNA nele depositada. Ao hibridar os macroarranjos com uma sonda que reconhece o gene da ampicilina do plasmídeo, nós observamos variações nos sinais entre *spots* de uma mesma membrana e entre membranas réplicas, que, teoricamente, deveriam ter a mesma quantidade de DNA (Fig. 3A e 3C). Usualmente, o dados de *spots* réplicas (com o mesmo EST) que apresentaram uma variação de mais de duas vezes entre seus respectivos sinais ou entre membranas não são considerados nas análises (J. Amselem, INRA – Versailles, com. pessoal).

Uma estratégia para diminuir a variação no sinal entre *spots* réplicas é a transferência do DNA em etapas, fixando-o mais de uma vez em cada *spot* (L. Reis, Instituto Ludwig, com. pessoal). Ao fixar uma só vez, caso ocorra alguma imperfeição na coleta do DNA ou na sua deposição na membrana, não há como corrigir. Ao fixar mais de uma vez, a probabilidade de que o erro aconteça duas vezes no mesmo *spot* é menor. Esse procedimento foi avaliado em um experimento no qual um macroarranjo contendo 768 ESTs em duplicata, depositados com o replicador manual, foi hibridado com a sonda *overgo*. Nós observamos que a variabilidade na quantidade de DNA depositada foi reduzida em 50% quando o DNA é transferido em duas etapas, sendo que a transferência em três etapas não causou maior redução (dados não mostrados). Aplicando duas transferências, cerca de 98% do total de *spots* réplicas tiveram suas razões de intensidade de sinal entre 0,5 - 2,0, resultado muito similar ao obtido com arranjos comerciais.

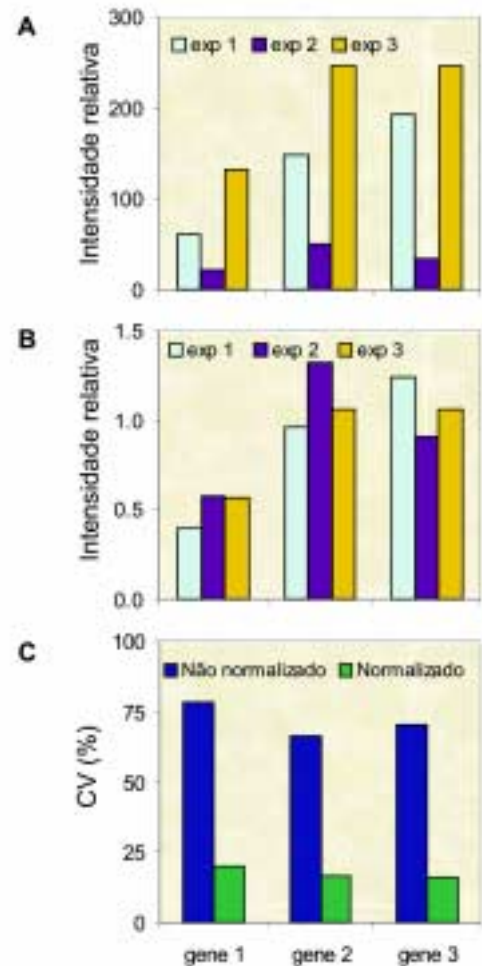


Figura 4: Correção de variações do sinal entre experimentos empregando a normalização com a mediana de todos os genes

Três membranas réplicas foram hibridadas com sondas de cDNA obtidas a partir da mesma amostra de RNA de folha de cana-de-açúcar. Os sinais de 3 genes são mostrados. **(A)**. Sinal dos três genes observados em três membranas réplicas hibridadas com sondas de cDNA, sem normalização. **(B)**. Sinal observado em (A), após normalização utilizando as medianas de todos os genes contidos em cada membrana. **(C)**. Coeficiente de variação (CV) dos genes antes e após a normalização

2. Subtração do background

Após a captação da imagem e a quantificação dos sinais obtidos para cada *spot*, é feita a subtração do valor referente à emissão de fundo (*ba-*

ckground), não específica, presente em toda membrana. Essa subtração procura remover o nível de “ruído”, permitindo fazer a comparação entre os sinais específicos de cada *spot*. Existem várias maneiras de determinar o *background*, como a inserção de spots contendo plasmídeos sem inserto de cDNA. Porém, existem variações do nível de *background* ao longo da membrana, podendo afetar mais alguns *spots* que outros. Dessa forma, a subtração do *background* local, ao redor de cada *spot* é a maneira mais confiável para a obtenção dos dados. A maioria dos programas de quantificação dos sinais de arranjos de DNA já subtrai esse valor, gerando tabelas que contêm os valores locais de *background*, os valores brutos para cada *spot* e esses valores subtraídos do *background*, que serão os utilizados nas próximas etapas de normalização.

3. Normalização do sinal causado por flutuações experimentais

A eficiência do processo de hibridação é influenciada por diversos parâmetros experimentais, principalmente no que se diz respeito a qualidade e quantidade de RNA utilizado na síntese de cDNA (Schuchhardt *et al.*, 2000). Assim, os sinais de todos os genes obtidos da hibridação com uma determinada sonda podem ser maiores que aqueles obtidos com uma segunda sonda. Um parâmetro empregado para reduzir essa variação é a mediana do sinal de todos os genes da membrana, pois ela reflete as flutuações de hibridação, a síntese de sonda, etc. Nessa estratégia, divide-se a intensidade de sinal de cada *spot* pela mediana do sinal de todos os *spots* daquela membrana (Schummer *et al.*, 1999). Uma alternativa à mediana é a média de todos os valores; no entanto, como a mediana é o valor

central do sinal de todos os *spots*, ela é menos sensível à presença de valores extremos de intensidade.

Para avaliar essa estratégia de normalização, três macroarranjos idênticos foram hibridados individualmente com três sondas de cDNAs obtidas de reações de transcrição reversa independentes, porém da mesma amostra de RNA total de cana-de-açúcar. O sinal para cada gene deveria ser similar entre os três macroarranjos, porém observamos uma flutuação no sinal detectado (Fig. 4A). Essa flutuação é um provável reflexo das diferenças no processo de hibridação ou, alternativamente, na síntese de cDNA. Houve uma menor flutuação no sinal normalizado com suas respectivas medianas (Fig. 4B), com conseqüente redução do coeficiente de variação (Fig. 4C). Esses resultados comprovam que, em uma determinada membrana, a divisão do sinal de cada gene pela mediana de todos os genes reduz a flutuação nos dados observados, aumentando a confiabilidade da técnica. Porém, essa estratégia não se aplica a tratamentos nos quais a maioria dos genes tem diferentes níveis de expressão. Nesse caso, uma alternativa é identificar genes

que não variam entre os tratamentos e usá-los como controle.

4. Sensibilidade para detecção de transcritos pouco abundantes

A maioria dos genes apresenta baixos níveis de expressão, com cerca de 10 moléculas de mRNA por célula, enquanto que os genes intermediários e abundantes (aprox. 300 e 12.000 moléculas/célula, respectivamente) são uma pequena fração do transcrito (Huang *et al.*, 1999). Assim, é fundamental que o método de análise de expressão tenha sensibilidade suficiente para detectar genes pouco expressos. Os macroarranjos construídos em nosso laboratório apresentaram um bom nível de sensibilidade, suficiente para detectar genes com expressão muito

reduzida, que correspondem a 0.004% dos transcritos na célula (Felix *et al.*, manuscrito em preparação).

Identificação de genes envolvidos na resposta a estresses abióticos, utilizando a técnica de macroarranjos

Após a padronização e normalização dessa metodologia, nosso grupo vem-se empenhando na identificação de padrões de expressão gênica em cana-de-açúcar em resposta a estresses abióticos, como baixas temperaturas e a hormônios como o elicitador metiljasmonato e o ácido abscísico.

Os estresses abióticos representam um dos principais fatores limitantes para a produtividade agrícola no mundo inteiro, além de representarem uma barreira para a introdução de espécies cultiváveis em áreas que ainda não são utilizadas para agricultura (Cherry, 1994). Como exemplo desses estresses, podemos destacar o decorrente de baixas temperaturas. Diversas plantas desenvolvem tolerância ao frio quando expostas a baixas temperaturas, processo denominado aclimação ao frio. Embora já se conheçam alguns genes cuja expressão é alterada nesse processo, existem ainda muitas dúvi-

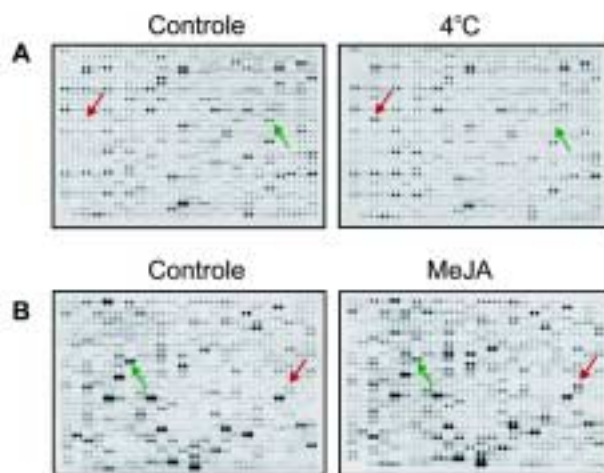


Figura 5: Identificação de genes envolvidos na resposta a estresses abióticos utilizando macroarranjos. Macroarranjos contendo 768 ESTs em duplicata foram hibridados com sondas de folhas jovens de cana-de-açúcar em condições de controle e tratados com MeJA (A) e controle, e expostas a 4°C (B). As setas indicam os genes induzidos (seta vermelha) e reprimidos (seta verde)

das quanto às rotas de sinalização envolvidas na resposta ao frio.

Sendo assim, com o intuito de descobrir novos genes induzidos por baixas temperaturas, macroarranjos de DNA contendo 1.536 ESTs de cana-de-açúcar foram hibridados com sondas de cDNA sintetizadas a partir de 30 µg de RNA total de plantas expostas ou não a 4 °C, durante vários períodos (Fig. 5A). Um total de 60 genes com expressão diferencial foram identificados. Dentre eles, destacam-se genes que participam da fixação de CO₂, dobramento correto de proteínas, metabolismo antioxidante, entre outros (Nogueira *et al.*, manuscrito em preparação).

Por outro lado, as plantas estão sujeitas a outros estresses, tais como a radiação ultravioleta, vento e granizo e ao ataque de patógenos e pragas que acarretam em severos danos celulares. Sob tais condições, as plantas promovem a biossíntese de ácido jasmônico (JA) e seu metil éster volátil, metil-jasmonato (MeJA). Esses reguladores de crescimento participam de diversos processos fisiológicos como, por exemplo, senescência, mecanopercepção, morfogênese, defesa a patógenos e pragas, entre outras (Wasternack & Parthier 1997; Glazebrook, 1999).

Para avaliar os efeitos do MeJA na expressão gênica em plântulas de cana-de-açúcar, macroarranjos contendo 1.536 ESTs oriundos do projeto SUCEST foram hibridados com sondas de cDNA sintetizadas a partir de RNA total extraído de tecidos foliares de cana-de-açúcar após diversos períodos de tratamento com MeJA (Fig. 5B). Os resultados permitiram a identificação de 37 genes induzidos e 44 reprimidos, entre os quais se encontram genes que participam da biossíntese do ácido jasmônico como a desaturase, três fatores de transcrição que participam da transdução de sinais, e as enzimas catalase e superóxido dismutase, que participam de estresse oxidativo, além de genes sem similaridade no *GenBank*. Esses resultados podem ajudar na compreensão dos efeitos produzidos por esse potente octadecanóide, além de revelar novos genes que podem ser importantes para a fisiologia e/ou defesa da planta a estresses.

Perspectivas e Conclusões

A tecnologia dos arranjos de DNA

têm permitido um enfoque genômico para elucidar a regulação gênica dos mais diversos processos fisiológicos, colocando-nos diante de uma nova era na ciência. Em particular, os arranjos de DNA em náilon tal como descritos neste trabalho são uma alternativa robusta e pouco dispendiosa para análises de transcritomas.

Agradecimentos:

Gostaríamos de agradecer aos técnicos do projeto SUCEST e especialmente a Renato V. dos Santos pela ajuda na bioinformática; e à Cooperativa de Produtores de Açúcar e Alcool do Estado de São Paulo (Copersucar) pelo cultivo e manutenção das culturas de cana-de-açúcar. O trabalho em nosso grupo é financiado pela Fapesp, Comunidade Econômica Européia, CNPq/PADCT e Capes.

Referências Bibliográficas

- Cherry JH (1994). **Biochemical and cellular mechanisms of stress tolerance in plants**. 604 pp Springer-Verlag
- Desprez T, Anselem J, Caboche M, Hofte H (1998) Differential gene expression in Arabidopsis monitored using cDNA arrays. *Plant Journal* 14: 643-652
- Freeman WM, Robertson DJ, Vrana KE (2000) Fundamentals of DNA hybridization arrays for gene expression analysis. *Biotechniques* 29: 1042-1055
- Glazebrook J (1999) Genes controlling expression of defense responses in Arabidopsis. *Current Opinion in Plant Biology* 2: 280-286
- Huang GM, Ng WL, Farkas J, He L, Liang HA, Gordon D, Yu J, Hood L (1999) Prostate cancer expression profiling by cDNA sequencing analysis. *Genomics* 59: 178-186
- Lennon GG e Lehrach H (1991) Hybridization analyses of arrayed cDNA Libraries. *Trends in Genetics* 7: 314-317
- Mangalam H, Stewart J, Zhou J, Schlauch K, Waugh M, Chen G, Farmer AD, Colello G, Weller JW (2001) GeneX: An Open Source gene expression database and integrated tool set. *Ibm Systems Journal* 40: 552-569
- Passos GAS, Nguyen C, Jordan B (1999)

Projeto Transcriptoma, Análise da expressão gênica em larga escala usando DNA-arrays. *Biociência & Desenvolvimento* 12: 34-37

Perret E, Ferrán EA, Marinx O, Liauzun P, Dumont X, Fournier J, Kaghad M, Ferrara P, and Caput D. (1998) Improved differential screening approach to analyze transcription variations in organized cDNA libraries. *Gene* 208, 103-115.

Sasaki Y, Asamizu E, Shibata D, Nakamura Y, Kaneko T, Awai K, Amagai M, Kuwata C, Tsugane T, Masuda T, Shimada H, Takamiya X, Ohta H, Tabata S (2001) Monitoring of methyl jasmonate-responsive genes in Arabidopsis by cDNA macroarray: Self-activation of jasmonic acid biosynthesis and crosstalk with other phytohormone signaling pathways. *Dna Research* 8: 153-161.

Schuchhardt J, Beule D, Malik A, Wolfski E, Eickhoff H, Lehrach H, and Herzog H. (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Research* 28, e47.

Schummer M, Ng VLV, Baumgarner RE, Nelson PS, Schummer B, Bednarski DW, Hassell L, Baldwin RL, Karlan BY, Hood L (1999) Comparative hybridization of an array of 21 500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. *Gene* 238: 375-385

Wang YH, Garvin DF, Kochian LV (2001) Nitrate-induced genes in tomato roots. Array analysis reveals novel genes that may play a role in nitrogen nutrition. *Plant Physiology* 127: 345-359

Wasternack C, Parthier B (1997) Jasmonate signalled plant gene expression. *Trends in Plant Science* 2: 302-307

Wenzl P, Patino GM, Chaves AL, Mayer JE, RAO IM (2001) The high level of aluminum resistance in signalgrass is not associated with known mechanisms of external aluminum detoxification in root apices. *Plant Physiology* 125: 1473-1484

Zhao Nd, Hashida H, Takahashi N, Misumi Y, Sakaki Y (1995) High-density cDNA filter analysis - a novel approach for large-scale, quantitative-analysis of gene-expression. *Gene* 156: 207-213. 🌱