



---

## **Estudos de expressão gênica utilizando-se *microarrays*: delineamento, análise, e aplicações na pesquisa zootécnica**

**Guilherme Jordão de Magalhães Rosa<sup>1</sup>, Leonardo Bernardes da Rocha<sup>1,2</sup>,  
Luiz Roberto Furlan<sup>3</sup>**

1 - Department of Dairy Science, University of Wisconsin, Madison – WI, USA 53705 email: grosa@wisc.edu

2 - Departamento de Tecnologia, Universidade Estadual Paulista – UNESP, Jaboticabal-SP. Bolsista do CNPq

3 - Departamento de Melhoramento e Nutrição Animal, Universidade Estadual Paulista – UNESP, Botucatu-SP

**RESUMO** – A tecnologia de *microarrays*, ou microarranjos de DNA, possibilita a avaliação simultânea da expressão de milhares de genes em diferentes tecidos em determinado organismo, em diferentes estágios de desenvolvimento ou condições ambientais. *Microarrays* são bastante utilizados em experimentos de genômica funcional com diversas espécies animais e vegetais, e têm sido gradativamente incorporados em diferentes áreas da pesquisa zootécnica, como crescimento e metabolismo, resposta imune a doenças, reprodução e resposta a fatores de estresse não-infecciosos (restrição alimentar, exposição a elementos tóxicos e outras condições ambientais desfavoráveis), bem como melhoramento genético animal. Tais experimentos, entretanto, são ainda consideravelmente caros, como consequência, geralmente são conduzidos com tamanhos amostrais relativamente pequenos. Por outro lado, a realização dos experimentos com *microarrays*, desde a coleta das amostras, até a obtenção das imagens para análise, envolve uma série de procedimentos laboratoriais de alta complexidade, que frequentemente introduzem variações adicionais aos resultados obtidos. Desta maneira, a condução de ensaios com *microarrays* requer cuidadoso delineamento experimental e análise estatística dos dados. Nesta apresentação são discutidos princípios básicos do planejamento de ensaios com *microarrays*, bem como as ferramentas estatísticas e computacionais mais comuns para a análise dos mesmos. São também discutidos alguns exemplos de aplicação de experimentos com *microarrays* em zootecnia e, numa última seção, são traçadas algumas considerações finais envolvendo os tópicos gerais abordados.

**Palavras-chave:** análise estatística, *microarrays*, pesquisa zootécnica, planejamento de experimentos

## **Microarray gene expression studies: experimental design, statistical data analysis, and applications in livestock research**

**ABSTRACT** – Microarray technology allows monitoring thousands of genes simultaneously in a specific tissue of an organism, in different developmental stages or environmental conditions. Microarrays are very common in functional genomics experiments with both animals and plant species, and they have been increasingly used also in different areas of livestock research, such as growth and metabolism, reproduction, immune response to diseases and parasites, response to non-infectious stress factors (such as dietary restriction, exposure to toxic elements and other unfavorable environmental conditions) as well as animal breeding. Such experiments, however, are still considerably expensive and time consuming and, consequently, they are performed with relatively small sample sizes. Nonetheless, microarray experiments are extremely complex, as they involve a number of laboratorial procedures such as sample collection, RNA extraction, reverse transcription and labeling, and the final hybridization. Hence, microarray assays require careful experimental planning and statistical data analysis. In this manuscript, basic principles of experimental design for microarray studies are reviewed, as well as the most common statistical and computational tools used for their analysis. In addition, some examples of application of microarray technology in animal science are discussed, and some concluding remarks are presented afterward.

**Key Words:** animal science, experimental design, microarrays, statistical analysis

## Introdução

A tecnologia de *microarrays*, ou microarranjos de DNA, possibilita a avaliação simultânea da expressão de milhares de genes em diferentes tecidos de um determinado organismo, e em diferentes estágios de desenvolvimento ou condições ambientais. Devido seu caráter prospectivo, essa tecnologia tem sido largamente utilizada em experimentos de genômica funcional projetados para estudar as funções e as interações dos genes dentro do contexto global do genoma de diversas espécies animais e vegetais. A tecnologia de *microarrays* tem sido também gradativamente incorporada na pesquisa com animais de interesse zootécnico e hoje é utilizada em diversas áreas, como por exemplo, crescimento e metabolismo, saúde e bem-estar animal, reprodução, e genética e melhoramento.

Entretanto, os experimentos com *microarrays* ainda são consideravelmente caros e trabalhosos e, como consequência, são geralmente conduzidos com tamanhos amostrais relativamente pequenos. Não obstante, tais experimentos envolvem uma série de procedimentos laboratoriais, desde a extração de RNA, transcrição reversa e marcação fluorescente, até a hibridização final, os quais invariavelmente introduzem diferentes níveis de variação adicional aos dados. Desta maneira, a condução de ensaios com *microarrays* requer cuidadoso delineamento experimental e análise estatística dos dados.

Assim sendo, este artigo oferece uma revisão da literatura existente e discute os principais aspectos relacionados ao planejamento e análise de experimentos com *microarrays*. Numa primeira seção é apresentada uma breve exposição em relação à tecnologia básica de *microarrays*, e de como tais experimentos funcionam. Esta seção também descreve os principais recursos existentes para a confecção de *microarrays* (tipos de lâminas e de sondas de DNA utilizadas), e discute os métodos de fixação das sondas e de marcação das amostras de RNA, bem como as variações existentes nos protocolos laboratoriais. Na seção seguinte, são apresentados os princípios básicos de delineamento de experimentos com *microarrays*, com maior ênfase nas diferentes estratégias para pareamento de amostras e marcação das mesmas em experimentos com

hibridizações competitivas. Nesta seção são contrastados os delineamentos com amostra referência e os delineamentos circulares, bem como algumas alternativas mais gerais. A seguir, tem-se uma seção relacionada à análise de dados de *microarrays*, a qual apresenta uma breve revisão sobre os três componentes geralmente envolvidos no tratamento estatístico dos dados: 1) obtenção dos valores de expressão, 2) normalização, e 3) inferência estatística propriamente dita. Um enfoque maior é dado aos diferentes procedimentos utilizados dentro de um contexto de modelos lineares, bem como ao problema relacionado à multiplicidade de testes em tais experimentos. Para finalizar, tem-se uma seção discutindo diferentes aplicações de experimentos com *microarrays* na pesquisa zootécnica, seguida pelas considerações finais sobre os diferentes tópicos abordados neste artigo.

## A tecnologia de *Microarrays*

A idéia de usar arranjos de ácidos nucleicos com o propósito de analisar simultaneamente o maior número possível de genes começou a ser aplicada no final da década de 70, com o advento da técnica conhecida como *Dot-Blot* (Kafatos *et al.*, 1979). Contudo, foi somente na metade da década de 90 que esta tecnologia adquiriu as características atuais (Schena *et al.*, 1995). Dois fatores foram fundamentais para essa evolução: 1) a disponibilidade de sistemas robotizados que permitiram a confecção de arranjos de alta densidade; e 2) a adoção do método de detecção óptica (fluorescência), que conferiu a sensibilidade ideal para se fazer as medidas de intensidade.

Os arranjos de DNA, também conhecidos como *chips* de DNA em alusão ao componente eletrônico miniaturizado que carrega milhões de transistores, são coleções de segmentos de DNA que se encontram ordenadamente distribuídos sobre uma superfície sólida. O pré-requisito fundamental para qualquer tipo de arranjo de DNA é a existência de um “endereço próprio” para cada componente da coleção, ou seja, uma posição individual para cada componente do arranjo (Chaudhuri, 2005).

Cada um desses endereços no arranjo é chamado de *spot* (ou ponto), e contém uma quantidade ínfima de DNA devidamente

imobilizada, denominada sonda (ou *probe*). Cada uma destas sondas tende a se ligar apenas a sua seqüência complementar de nucleotídeos, mediante processo chamado hibridização (Jaluria *et al.*, 2007). Essa seqüência complementar, normalmente um DNA complementar (cDNA) produzido à partir de um RNA mensageiro (mRNA), representa apenas um único gene do genoma e é chamada “alvo” ou *target*. A hibridização de cada sonda de DNA com o seu correspondente alvo (cDNA) é um processo baseado na complementaridade das cadeias de nucleotídeos, ou seja, na propriedade que duas cadeias homólogas têm de parear suas bases complementares (A com T e C com G), mediante a formação de pontes de hidrogênio.

A plataforma sólida mais utilizada na confecção dos arranjos é a lâmina de vidro, do tipo usado em microscopia, que depois de ter as sondas imobilizadas na sua superfície é normalmente referida como *slide* ou simplesmente *lâmina de microarray*. Basicamente existem duas formas de distribuir as sondas em um microarranjo de DNA. A primeira, mais simples, é feita por robôs de alta precisão que utilizam agulhas especiais para depositar as amostras de DNA na superfície de uma lâmina de vidro, processo que também é conhecido como “impressão do slide”. Normalmente essas amostras são constituídas de oligonucleotídeos pré-sintetizados, cDNAs produzidos em projetos de seqüenciamento, ou ainda produtos de amplificação por PCR (reação em cadeia da polimerase). A segunda, mais complexa, utiliza processos especiais (fotolitografia, por exemplo) para realizar a síntese química de oligonucleotídeos diretamente sobre a superfície da lâmina de vidro (Walsh & Henderson, 2004).

Desta maneira, a tecnologia de *microarrays* consiste na utilização de um *slide* (lâmina ou microarranjo) no qual as sondas (amostras de DNA) foram imobilizadas em quantidades e posições precisamente definidas (*spots*), para se fazer a hibridização com um *pool* de mRNAs extraídos de amostras biológicas (*targets*), que foram previamente marcados com fluoróforos (marcadores fluorescentes). Como as moléculas de mRNA são bastante instáveis quando manipuladas, a maioria dos protocolos laboratoriais utiliza o processo de transcrição reversa para convertê-las

nos correspondentes cDNAs durante o processo de marcação.

Após o processo de hibridização, cada lâmina é lavada para remoção dos “alvo” excedentes (que não se ligaram às sondas) e, em seguida, exposta à ação de raios laser que excitam os fluoróforos que foram incorporados aos “alvos”, fazendo com estes emitam luz (fluorescência). Em princípio, quanto maior for a expressão de um determinado gene, maior será a quantidade de “alvos” marcados com o fluoróforo e, conseqüentemente, maior será a intensidade da fluorescência do complexo alvo-sonda após a hibridização (Hiendleder *et al.*, 2005). Assim, a tecnologia de *microarrays* fornece uma medida indireta do nível de expressão gênica, mediante quantificação da abundância dos RNAs transcritos.

Entretanto, a tecnologia básica de *microarrays* descrita acima apresenta diversas variações (Rogojina *et al.*, 2003), dependendo do substrato (lâminas de vidro ou de sílica), tipo de sonda (oligonucleotídeos de cadeia curta, oligonucleotídeos de cadeia longa, cDNA ou produtos de PCR), do método de deposição das amostras (impressão por robôs ou síntese baseada na tecnologia de semicondutores), além de uma variedade de técnicas de extração e marcação do RNA, bem como de protocolos de hibridização. Desta maneira, diversos tipos de lâminas podem ser encontradas para diferentes espécies animais e vegetais, as quais são produzidas por empresas (como por exemplo: *Affymetrix*, *Agilent*, *Illumina*, etc.) ou por instituições de pesquisa e universidades.

Assim sendo, organizações como a *Microarray Gene Expression Data Society* - MGED (<http://www.mged.org/>) e o *European Bioinformatics Institute* – EBI (<http://www.ebi.ac.uk/>), têm estabelecido guias de orientação que auxiliam os pesquisadores a planejar e implementar seus experimentos com *microarrays*, cujo objetivo é padronizar as variantes dessa tecnologia, para que os resultados obtidos em experimentos diferentes possam ser comparados e utilizados como base para o planejamento de novas pesquisas sobre o mesmo tema. Um desses guias é o *Minimum Information About a Microarray Experiment* (MIAME), que contém diversas recomendações e padrões para coleta e análise de dados provenientes de experimentos com *microarrays*, para que estes

possam ser corretamente interpretados e reproduzidos (Jaluria *et al.*, 2007).

Outra tentativa interessante é o compartilhamento dos dados brutos obtidos em experimentos com *microarrays*, uma vez que não é possível incluir esse tipo de informação nas publicações. Dois bancos de dados criados com essa finalidade são o *Gene Expression Omnibus* (GEO), iniciativa do *National Center for Biotechnology Information* (NCBI) e o *ArrayExpress*, mantido pelo EBI.

Do ponto de vista de delineamento e análise estatística de experimentos de *microarray*, a distinção mais importante entre as diferentes tecnologias existentes refere-se ao número de amostras hibridizadas em cada lâmina. Neste sentido, os diversos tipos de tecnologias de *microarrays* podem ser divididos em dois grupos básicos: sistema de uma cor (*single-color* ou *single-channel microarray*) e sistema de duas cores (*two-color microarray*). Neste artigo estas duas tecnologias básicas serão referidas por lâminas de hibridizações independentes e lâminas de hibridizações competitivas, respectivamente. No sistema com hibridizações independentes, cada amostra de RNA é marcada e hibridizada individualmente numa lâmina. Uma vantagem deste tipo de tecnologia é que a condução dos experimentos é geralmente mais simples, mas por outro lado variações naturais entre as lâminas ficam de certa maneira confundidas com as diferenças entre as amostras. Este problema, no entanto, é minimizado se as diferenças entre as lâminas não forem importantes, e também a partir da utilização de procedimentos estatísticos específicos de ajuste dos dados para a correção dos mesmos, como discutido posteriormente neste artigo.

Já no sistema com hibridizações competitivas, as amostras de RNA são individualmente transcritas reversamente e marcadas com diferentes fluorocromos (geralmente Cyanine-3 e Cyanine-5, também referidos como Cy3 e Cy5), e então pares de amostras marcadas diferentemente são hibridizadas conjuntamente nas lâminas, permitindo uma comparação direta entre amostras (por exemplo, controle *versus* tratamento). A comparação entre amostras numa mesma lâmina é dada pela razão entre as intensidades de fluorescência relativas ao Cy3 e Cy5 (relativas a diferentes comprimentos de ondas, representadas pelas cores verde e vermelho) em cada *spot*, a qual

é uma medida da expressão relativa do gene correspondente nas duas amostras. Uma descrição mais detalhada deste processo pode ser encontrada, por exemplo, em Brown & Botstein (1999).

Acredita-se que a hibridização conjunta de pares de amostras (também chamada hibridização competitiva) minimiza o problema de variabilidade entre lâminas. Outra vantagem é que lâminas de hibridização competitiva, tanto com sondas de cDNA quanto de oligonucleotídeos longos, estão disponíveis para um maior número de espécies de interesse zootécnico. Ainda, como existem lâminas deste tipo produzidas por universidades e instituições de pesquisa sem fins lucrativos, elas podem ser adquiridas a um preço normal. No entanto, experimentos de *microarray* com hibridizações competitivas são um pouco mais complexos, especialmente quando mais do que dois grupos experimentais são comparados, como será discutido a seguir.

### Planejamento de experimentos com *Microarrays*

O planejamento de estudos de expressão gênica utilizando-se *microarrays* envolve uma série de aspectos, como a escolha do(s) tecido(s) a ser(em) estudado(s) – ou até mesmo a escolha do tipo de célula(s) de interesse –, e o(s) estágio(s) de desenvolvimento a serem avaliados. Como exemplo, suponha um estudo de genômica funcional para a comparação dos níveis de expressão gênica de diferentes raças bovinas em resposta à infestação com carrapatos. Neste caso, o pesquisador terá que escolher se investigará mudanças de expressão gênica na pele, em linfonodos, ou algum outro tecido, bem como em que estágio após a infestação, como por exemplo, algumas horas, após um dia ou uma semana, etc.

Mesmo após a escolha do tecido(s) de interesse, pesquisadores ainda devem refletir em relação à complexidade do(s) mesmo(s) para julgar se extrações de RNA devem ser conduzidas considerando-se cada tecido como um todo ou se as extrações devem ser específicas para determinado(s) tipo(s) de célula(s) de interesse em cada tecido. Quando o RNA total de determinado tecido é utilizado nos experimentos, mudanças na expressão gênica em específicos componentes deste tecido podem ser mascaradas pela

contribuição de transcritos de outros tipos de células e componentes que não apresentam variação nos níveis de expressão, ou até mesmo que apresentam mudança em direção oposta. Uma outra situação similar seria um determinado tecido no qual não ocorre qualquer alteração nos níveis de expressão gênica em nenhum dos seus componentes em resposta a diferentes estímulos ou tratamentos, mas sim uma modificação nas proporções dos diferentes tipos de células que compõem este tecido. Como alguns genes podem estar expressos diferencialmente nos diferentes tipos de células que compõem o tecido, uma alteração na proporção destes tipos de células em resposta aos diferentes tratamentos pode levar à falsa conclusão de que estes genes apresentam expressão alterada em decorrência dos tratamentos. Desta maneira, ainda que a complexidade e os custos dos experimentos sejam aumentados, muitas vezes a utilização de populações de células específicas (obtidas, por exemplo, utilizando-se a técnica de dissecação por laser) é necessária (Smith & Rosa, 2007).

Outro aspecto importante no planejamento de experimentos com *microarrays* refere-se à escolha do tipo de lâmina a ser utilizada, em termos de qual melhor se adequa aos específicos objetivos experimentais. Por exemplo, existem lâminas relativas a determinados grupos de genes (como genes relacionados à resposta imunológica, metabolismo, etc.), lâminas que possibilitam não só o estudo de variações nos níveis de expressão, mas também o estudo de diferentes variantes de transcritos de RNA produzidos por cada gene (ou do inglês, *alternative splicing*), lâminas com maior especificidade ou maior sensibilidade (geralmente relacionados ao comprimento das cadeias de DNA utilizados como sondas), e assim por diante.

A escolha do tipo de lâmina a ser utilizada, entretanto, é geralmente restringida pela disponibilidade de alternativas para o organismo em estudo. Por exemplo, lâminas de cDNA estão disponíveis apenas para espécies para as quais ESTs já foram obtidas de bibliotecas de cDNA. No caso de lâminas com sondas de oligonucleotídeos, as mesmas só podem ser geradas se seqüências de DNA estão disponíveis e, ainda, um conjunto adequado de sondas de oligonucleotídeo só pode ser obtido para aquelas espécies com os genomas completamente seqüenciados. Desta

maneira, tem-se uma diversidade maior de tipos de lâminas disponíveis para experimentos envolvendo humanos e organismos modelos (como animais de laboratório) do que para a maioria dos animais de interesse zootécnico ou animais selvagens.

Não obstante, diversos tipos de lâminas de hibridização competitiva, tanto com sondas de cDNA como de oligonucleotídeos longos, foram desenvolvidas por diferentes instituições de pesquisa para várias espécies de interesse zootécnico. Lâminas de hibridizações independentes, utilizando-se sondas de cadeias curtas de oligonucleotídeos, são comercializadas pela empresa *Affymetrix* para experimentos com bovinos, suínos ou aves. Espécies para as quais não existem ainda lâminas de *microarray*, ou estas não são de fácil obtenção (como por exemplo, ovinos e caprinos), experimentos com hibridização cruzada utilizando-se lâminas de alguma outra espécie próxima do ponto de vista evolutivo (especialmente lâminas de cDNA) é uma alternativa, ainda que não uma solução ideal.

Após a escolha do tipo de lâmina a ser utilizada, e do(s) tecido(s) e condições experimentais a serem estudados, o planejamento de experimentos com *microarrays* requer ainda a escolha do número de animais e de lâminas a serem utilizados e, no caso de experimentos com hibridizações competitivas, deve-se ainda definir como as amostras de RNA serão marcadas (em termos de Cy3 e Cy5), bem como as mesmas serão pareadas em cada hibridização. Como os custos dos experimentos de *microarray* são ainda relativamente altos, o número de lâminas utilizadas na maioria dos estudos (à exceção de algumas áreas, como por exemplo, na pesquisa médica) é geralmente determinado por restrições orçamentárias. Desta maneira, o estudo do tamanho amostral em experimentos com *microarrays* limita-se muitas vezes ao cálculo do poder de teste (i.e., poder de detecção de variações nos níveis de expressão gênica entre os grupos experimentais) para determinado número de lâminas, do que do número de repetições propriamente dito, para determinado poder estatístico desejado.

Uma vez estabelecidos os grupos experimentais a serem considerados, bem como o número de amostras em cada grupo e o número de lâminas disponíveis, outro passo fundamental no

delineamento experimental para ensaios com *microarrays* refere-se à estratégia de hibridização das amostras de RNA nas diferentes lâminas.

Para experimentos de *microarrays* com hibridizações independentes (como no caso de lâminas *Affymetrix*, por exemplo), este processo é relativamente simples. Em casos onde o número de amostras (animais) disponíveis é igual ao número de lâminas, cada amostra é hibridizada numa única lâmina. Em situações onde o número de animais é maior do que o número de lâminas disponíveis, como ocorre na maioria dos casos, uma subamostragem deve ser efetuada dentro de cada grupo experimental, e assim cada subamostra é hibridizada numa única lâmina. Uma outra alternativa para este tipo de situação é a extração de RNA de vários animais e a mistura de múltiplas amostras em cada grupo experimental, antes da realização das hibridizações. A mistura de amostras de RNA em *pools* apresenta vantagens e desvantagens. Ela pode ser uma boa estratégia para se diminuir a variabilidade biológica entre *pools*, mas apresenta o risco de que uma amostra com problemas pode ser mascarada após sua diluição no *pool*. Para uma discussão mais detalhada deste assunto, veja, por exemplo, Churchill (2002) e Kendzierski *et al.* (2005).

Existem ainda situações nas quais o número de amostras é um fator limitante, pelo custo ou dificuldade de obtenção das mesmas, como por exemplo, em estudos com clonagem ou doenças raras. Nestes casos, cada amostra biológica pode ser hibridizada em múltiplas lâminas para uma medida mais precisa dos níveis de expressão gênica em cada amostra. Experimentos desta natureza apresentam dois níveis de replicação, denominados replicação biológica (relativo às amostras independentes dentro de cada grupo experimental) e replicação técnica (relativa às medidas repetidas ou subamostras de cada amostra biológica).

Replicação técnica é bastante comum (ainda que nem sempre uma estratégia experimental eficiente) em experimentos de *microarray* com hibridizações competitivas. Para este tipo de experimentos, o delineamento experimental exige um cuidado maior em relação à marcação das amostras e o pareamento das mesmas para as hibridizações em cada lâmina (Churchill, 2002; Kerr, 2003; Rosa *et al.*, 2005). Se pensarmos em

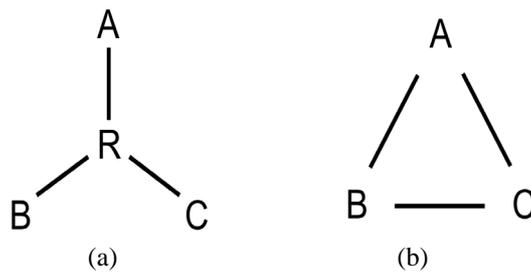
cada lâmina como um bloco, e dado que somente duas amostras são hibridizadas em cada lâmina, experimentos envolvendo mais do que dois grupos experimentais são essencialmente ensaios em blocos incompletos. Como tem-se ainda que as duas amostras numa mesma lâmina são marcadas diferentemente (introduzindo-se assim uma fonte adicional de variação no experimento, uma vez que o Cy3 e o Cy5 apresentam diferenças naturais em termos de potencial de incorporação), tais experimentos são mais especificamente estruturados como delineamentos em linhas-colunas (*row-column designs*) de dimensões  $2 \times s$ , onde  $s$  é o número de lâminas.

Dentro deste contexto, os delineamentos experimentais mais amplamente utilizados para experimentos de *microarray* com hibridizações competitivas referem-se aos delineamentos circulares (ou em *loop*) e aos delineamentos com amostra referência (Kerr & Churchill, 2001; Yang & Speed, 2002). Estas duas estruturas básicas estão representadas na Figura 1, onde cada tratamento é indicado por uma letra maiúscula e cada lâmina é representada por um segmento, o qual conecta as duas amostras hibridizadas na mesma.

No delineamento com amostra referência (Figura 1a), uma única amostra (denominada amostra referência, e representada aqui pela letra R) é hibridizada com cada uma das amostras de cada tratamento ou grupo experimental (A, B e C), semelhantemente a experimentos agrônômicos em blocos incompletos com tratamentos comuns. Como comparação, em experimentos com *microarrays* cada bloco (lâmina) tem tamanho dois, e existe um único tratamento comum (a amostra referência). A comparação entre tratamentos se dá, então, de maneira indireta. Por exemplo, a comparação entre os tratamentos A e B é estimada usando-se informação referente à diferença entre os contrastes entre o tratamento A e a referência, e o tratamento B e a referência, isto é:  $(A - R) - (B - R)$ .

Já em delineamentos circulares, ao invés da utilização de amostra referência, as amostras são pareadas alternando-se tratamentos (Figura 1b) de maneira que diferenças entre tratamentos são estimadas por combinações entre comparações diretas e indiretas. Por exemplo, na Figura 1b os tratamentos A e B são comparados diretamente em uma das lâminas, além da comparação indireta a

partir da conexão entre eles pelo tratamento C nas outras duas lâminas.



**Figura 1** - Estruturas básicas de experimentos de *microarrays* com hibridizações competitivas: (a) Delineamento com amostra referência e (b) Delineamento circular. Cada letra representa um grupo experimental (A, B e C) ou a amostra referência (R), e pares de amostras conectadas por um segmento representam cada uma das hibridizações realizadas.

*Figure 1. Basic layouts for microarray experiments with competitive hybridizations: (a) Reference design, and (b) Loop design. Each letter represents an experimental group (A, B and C) or the reference sample (R), and each segment represents a slide, which connects the two samples co-hybridized on it.*

A principal vantagem de delineamentos com amostra referência é que eles são mais simples de serem conduzidos no laboratório. A desvantagem, entretanto, relaciona-se ao fato de que metade das observações refere-se à amostra referência, a qual não é de interesse direto no experimento (ela é utilizada simplesmente para corrigir os dados para o efeito de lâminas). Já o delineamento circular refere-se a uma estrutura mais complexa, na qual o pareamento e a marcação das amostras exigem um cuidado adicional. Entretanto, estruturas circulares são geralmente mais eficientes do que delineamentos com amostra referência (Kerr & Churchill, 2001; Yang & Speed, 2002; Tempelman, 2005).

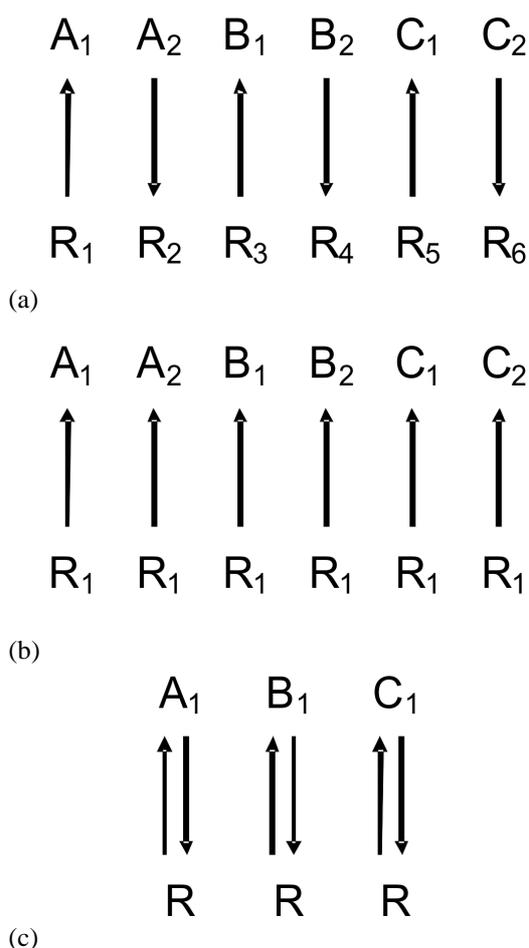
A Figura 1 apresentada, entretanto, é uma simplificação destes dois tipos básicos de delineamentos para experimentos de *microarray* com hibridizações competitivas. Evidentemente, múltiplas amostras são necessárias para cada tratamento e, em cada lâmina, tem-se ainda que designar qual marcação será usada para cada amostra. A utilização de diferentes níveis de replicação (as chamadas replicação biológica e

replicação técnica, como discutido anteriormente), bem como as diferentes maneiras para se marcarem as amostras em cada lâmina, gera uma imensa variedade de possíveis experimentos dentro destas duas estruturas básicas de delineamentos.

Como exemplos, na Figura 2 são representadas algumas alternativas de delineamentos com amostra referência para um experimento com três tratamentos (A, B e C) e seis lâminas. No primeiro caso (Figura 2a), duas replicações biológicas de cada grupo experimental (isto é, amostras de RNA de duas unidades experimentais de cada tratamento) são hibridizadas com a amostra referência com inversão da marcação, ou seja, numa lâmina a amostra referência é marcada com Cy3 e na outra com Cy5. É importante ressaltar que no experimento ilustrado na Figura 2a, a amostra referência é na verdade constituída de amostragens independentes de um grupo controle, ou grupo referência. Esta situação é semelhante ao que ocorre em experimentos agrônômicos com tratamentos comuns, nos quais o tratamento comum pode ser uma determinada variedade, e amostras (plantas) desta variedade específica são colocadas em cada bloco.

Em experimentos com *microarrays*, entretanto, uma mesma amostra de RNA pode ser fracionada e hibridizada em múltiplas lâminas (inclusive com diferente marcação). Desta maneira, um experimento com amostra referência pode ser conduzido como ilustrado na Figura 2b, no qual a amostra referência é constituída de uma única extração de RNA replicada em cada uma das lâminas. Uma outra diferença deste delineamento em relação ao experimento da Figura 2a é que a amostra referência está sempre com a mesma marcação (no caso, Cy3) e os tratamentos com a outra (Cy5). Nestas circunstâncias, existe um confundimento entre a diferença entre cada tratamento e a referência, e a diferença entre as duas marcações. O efeito de marcação, no entanto, se cancela quando efetuam-se as comparações indiretas entre tratamentos, isto é, em lâminas com o tratamento A tem-se comparações do tipo  $[(A + \text{Cy5}) - (R + \text{Cy3})]$  e em lâminas com o tratamento B tem-se  $[(B + \text{Cy5}) - (R + \text{Cy3})]$ , mas quando se comparam os tratamentos A e B tem-se  $[(A + \text{Cy5}) - (R + \text{Cy3})] - [(B + \text{Cy5}) - (R + \text{Cy3})] = [A - B]$ .

Como um terceiro exemplo, tem-se na Figura



**Figura 2** - Três delineamentos alternativos para experimentos de *microarray* com hibridizações competitivas utilizando-se amostra referência. Cada letra representa um grupo experimental (A, B e C) ou a amostra referência (R), e pares de amostras conectadas por uma seta representam cada uma das hibridizações realizadas. Ainda, a ponta da seta indica a amostra marcada com Cy5, sendo a outra amostra marcada com Cy3, e os índices indicam replicações biológicas.

*Figure 2 - Three alternative microarray experimental layouts for reference designs. Each letter represents an experimental group (A, B and C) or a reference sample (R), and each arrow represents a slide, which connects the two samples co-hybridized on it. Furthermore, the head and tail of each arrow indicate the samples labeled with Cy5 and Cy3, respectively, and the indexes represent biological replications.*

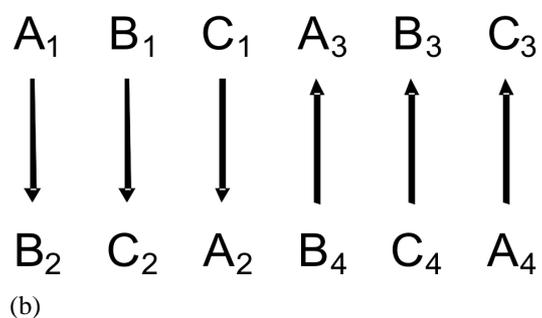
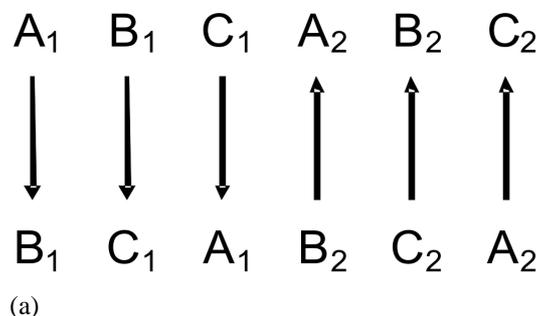
2c uma outra alternativa de delineamento bastante usual na literatura, na qual cada replicação biológica é hibridizada duas vezes com marcação alternada (o chamado *dye-swap*, em inglês). Como este delineamento envolve replicação técnica de cada amostra, o dobro de lâminas é necessário para

um mesmo tamanho amostral (em termos de replicação biológica) relativamente aos experimentos das Figuras 2a e 2b. Usuários deste tipo de delineamento experimental defendem que somente com *dye-swap* é possível se corrigir eficientemente os dados para o efeito de marcação. Outros pesquisadores, no entanto, incluindo os autores deste artigo, defendem que para um número fixo de lâminas (e em situações nas quais a disponibilidade de amostras biológicas não é um fator limitante), o aumento do número de replicações biológicas é geralmente mais vantajoso do ponto de vista estatístico do que a utilização de replicação técnica. Para uma detalhada discussão entre replicação biológica *versus* replicação técnica veja, por exemplo, Churchill (2002), Rosa *et al.* (2005) e Tempelman (2005). Especificamente com relação às diferentes estratégias para delineamentos com amostra referência, bem como a modelagem estatística de dados em experimentos desta natureza, consulte Steibel & Rosa (2005).

Assim como com o delineamento com amostra referência, inúmeras variações da estrutura circular também são possíveis, considerando-se diferentes níveis de replicação (técnica e biológica) e a maneira com a qual as amostras são marcadas. Por exemplo, na Figura 3 tem-se duas destas variações, as quais são comumente encontradas na literatura. Similarmente aos delineamentos ilustrados na Figura 2, tem-se aqui também um experimento com 3 tratamentos (A, B e C) e seis lâminas.

Na Figura 3a tem-se o chamado delineamento circular conectado. O termo ‘conectado’ refere-se ao fato de que cada amostra de RNA é hibridizada em múltiplas lâminas (neste caso específico, em duas lâminas). Veja que o delineamento representado nesta figura apresenta dois “*loops*”, um com as amostras  $A_1$ ,  $B_1$  e  $C_1$ , e outro com as amostras  $A_2$ ,  $B_2$  e  $C_2$ . No primeiro *loop*, a amostra  $A_1$  é marcada com Cy3 e amostra  $B_1$  é marcada com Cy5 numa das lâminas, depois a mesma amostra  $B_1$  é marcada com Cy3 e hibridizada com a amostra  $C_1$ , a qual é marcada com Cy5, e finalmente o *loop* se completa com a amostra  $C_1$  sendo marcada com a Cy3 e hibridizada com a amostra  $A_1$ , a qual é agora marcada com Cy5. O segundo *loop* é elaborado de maneira similar, mas com marcações em ordem invertida. Por exemplo, a lâmina envolvendo os tratamentos A e B tem a amostra  $A_2$  marcada com Cy5 e amostra  $B_2$

marcada com Cy3, e assim por diante. É importante notar que neste delineamento da Figura 3a tem-se duas replicações biológicas para cada tratamento, com duas replicações técnicas de cada amostra.



**Figura 3** - Duas alternativas de delineamentos circulares para experimentos de *microarray* com hibridizações competitivas. Cada letra representa um grupo experimental (A, B e C), e pares de amostras conectadas por uma seta representam cada uma das hibridizações realizadas. Ainda, a ponta da seta indica a amostra marcada com Cy5, sendo a outra amostra marcada com Cy3, e os índices indicam replicações biológicas.

*Figure 3 - Two alternative microarray experimental layouts for loop designs. Each letter represents an experimental group (A, B and C), and each arrow represents a slide, which connects the two samples co-hybridized on it. Furthermore, the head and tail of each arrow indicate the samples labeled with Cy5 and Cy3, respectively, and the indexes represent biological replications.*

A utilização de conexões entre lâminas utilizando-se replicações técnicas, ainda que bastante comum em experimentos encontrados na literatura, não é de maneira alguma necessária do ponto de vista estatístico, para a comparação entre médias de tratamentos. Neste sentido, na Figura 3b é apresentada uma alternativa de delineamento na qual há ainda certa circularidade na maneira

pela qual as amostras são pareadas e hibridizadas nas lâminas, mas aqui é dada preferência para replicação biológica ao invés de replicação técnica. Veja que neste delineamento numa primeira lâmina tem-se a comparação de uma amostra do tratamento A com uma amostra do tratamento B (as quais são marcadas com Cy3 e Cy5, respectivamente, de maneira similar ao experimento da Figura 3a), mas numa segunda lâmina uma outra amostra do tratamento B é hibridizada com uma amostra do tratamento C. Neste caso, ainda com somente seis lâminas no experimento, tem-se agora quatro amostras independentes de cada grupo experimental, mas sem replicação técnica de qualquer amostra.

Tanto o delineamento da Figura 3a quanto o da Figura 3b são, na medida do possível, balanceados em termos do número de replicações para cada tratamento, lâminas e marcações, e até mesmo para certas combinações destes fatores, como por exemplo, tratamento  $\times$  marcação. Existem vantagens e desvantagens para cada um destes delineamentos, mas do ponto de vista estatístico mais uma vez é geralmente recomendado priorizar-se replicação biológica. Para uma comparação mais detalhada em termos de eficiência e poder estatístico destas alternativas de delineamento, bem como a robustez destes delineamentos em relação à perda de lâminas (por exemplo, por problemas técnicos na condução das hibridizações no laboratório), veja, por exemplo, Rosa *et al.* (2005), Steibel & Rosa (2005) e Tempelman (2005).

Como discutido acima, existem sempre várias maneiras de se parearem e marcarem as amostras em experimentos com *microarrays* com hibridizações competitivas. Foram apresentadas aqui algumas alternativas em termos de delineamentos com amostra referência e delineamentos circulares, utilizando-se diferentes níveis de replicação biológica e técnica. Entretanto, existem obviamente outras estruturas possíveis de delineamentos para experimentos com *microarrays*, incluindo estruturas gerais de delineamentos em linhas-colunas. Por exemplo, dadas restrições em termos de material biológico e número de lâminas disponíveis, pode-se utilizar algoritmos de busca para a obtenção de delineamentos ótimos (ou quase ótimos) para determinados objetivos específicos dos experimentos (Wit *et al.*, 2005; Bueno *et al.*,

2006).

Outra importante questão em experimentos com *microarrays*, a qual não é abordada neste trabalho, refere-se ao cálculo do poder estatístico e tamanho de amostras. Como mencionado anteriormente, muito freqüentemente o tamanho dos experimentos de *microarray* são definidos por restrições orçamentárias em termos do número de lâminas disponíveis. Nestes casos, busca-se uma otimização dos experimentos em termos do delineamento experimental para a maximização do poder estatístico e eficiência de tais ensaios. Exemplos de cálculos neste contexto são apresentados, por exemplo, por Rosa *et al.* (2005) e Tempelman (2005) num contexto de modelos lineares de efeitos mistos. Cui *et al.* (2005) apresentam também cálculos dos custos de experimentos em termos de replicações técnicas e biológicas. Dada a multiplicidade de testes que são efetuados em cada experimento de *microarray*, os quais geralmente envolvem milhares de genes, cálculos relativos a poder de teste e tamanho amostral para tais experimentos devem ser baseados, por exemplo, no conceito de taxa de falsos positivos (FDR, do termo em inglês *False Discovery Rate*), como apresentado por Gadbury *et al.* (2004), Muller *et al.* (2004), Dobbin & Simon (2005), Hu *et al.* (2005) e Jung (2005).

### Análise de dados de *Microarrays*

A análise estatística de dados de *microarray* geralmente envolve três componentes: 1) obtenção dos dados e eliminação de valores espúrios; 2) pré-ajuste dos dados para efeitos sistemáticos (este ajuste é conhecido por normalização dos dados); e 3) análise estatística propriamente dita, a qual geralmente utiliza ferramentas metodológicas relativas a testes de significância, análise discriminante ou análise de agrupamento.

A obtenção dos dados refere-se à análise de imagem de cada lâmina para a extração dos valores de intensidade fluorescente em cada *spot*, os quais são medidas indiretas da abundância de transcritos de RNA dos genes representados pelas sondas. Existe uma série de procedimentos e programas computacionais disponíveis para a leitura das imagens para diferentes tipos de lâminas. Para lâminas de cDNA ou oligos longos utilizando o sistema de hibridizações competitivas, os valores

de intensidade são geralmente obtidos a partir da classificação de cada pixel da imagem como pertencente a determinado *spot* (i.e., sonda) ou a espaços vazios entre *spots*, denominados *foreground* e *background*, respectivamente. Posteriormente, para cada lâmina, os pixels de *foreground* relativos a cada *spot* são combinados em alguma medida resumo, como média, mediana ou intensidade total, as quais muitas vezes são ajustadas para valores de *background*. Este mesmo procedimento é efetuado tanto para as intensidades relativas à Cy3 quanto aquelas relativas à Cy5, de maneira que para cada *spot* têm-se duas medidas de intensidade, das quais obtém-se uma medida da expressão relativa de cada gene nas duas amostras hibridizadas em cada lâmina.

Dependendo do tipo de lâmina e tecnologia utilizada, entretanto, outros métodos são utilizados para a obtenção dos valores de intensidade. Em lâminas *Affymetrix*, por exemplo, cada gene é representado por um grupo (geralmente de 11 a 20 pares) de sondas de cadeias curtas de oligonucleotídeos (de 25 bases). Cada par inclui uma sonda com seqüência nucleotídea idêntica ao gene (chamada *perfect match*, PM), e outra sonda com uma mudança nucleotídea na 13ª base (chamada *mismatch*, MM). De maneira similar às lâminas de hibridização competitiva, os valores de intensidade observados para cada gene são geralmente combinados numa única medida resumo para expressar o nível de abundância de transcritos de RNA, como por exemplo, utilizando-se a média das diferenças entre PM e MM para cada gene, dada por:

$$\text{AvDiff}_g = \frac{1}{K} \sum_{i=1}^K (\text{PM}_{gi} - \text{MM}_{gi}),$$

Onde  $\text{AvDiff}_g$  é a medida de expressão relativa ao gene  $g$ ,  $\text{PM}_{gi}$  e  $\text{MM}_{gi}$  são as intensidades PM e MM relativas ao  $j$ -ésimo par de probes ( $j=1,2,\dots,K$ ) do gene  $g$ .

A medida resumo  $\text{AvDiff}$  foi inicialmente proposta pela *Affymetrix*, mas hoje já existem outras metodologias alternativas, supostamente melhores, para a sumarização das intensidades observadas para cada gene, como por exemplo o procedimento MAS5.0 da própria *Affymetrix*, o

MBEI (*Multiplicative Model-Based Expression Index*; proposto por Li & Wong, 2001) e o RMA (*Robust Multi-array Average*; proposto por Irizarry *et al.*, 2003).

Após a obtenção dos dados e eliminação de valores espúrios em decorrência de possíveis problemas na fixação das sondas, marcação das amostras, hibridização etc., um ajuste geral dos dados é geralmente necessário antes de uma análise estatística mais formal dos mesmos. Este processo de correção dos dados é geralmente denominado normalização, e considera ajustes para diferenças entre lâminas (em termos de média ou mediana e variância), efeito de marcação, etc. Alguns procedimentos de normalização dos dados baseiam-se em somente alguns genes presentes nas lâminas (como genes controles ou genes com expressão supostamente constante nos diversos grupos experimentais), outros baseiam-se em todos os genes e utilizam procedimentos estatísticos robustos, com a suposição de que a maioria dos genes é não diferencialmente expressa entre os grupos experimentais.

Um procedimento comumente utilizado para a normalização de dados de *microarray* em hibridização competitiva utiliza uma metodologia de regressão não-paramétrica robusta (denominada LOWESS) para estabilizar a relação entre o logaritmo da razão de intensidades e a média do logaritmo das intensidades em cada lâmina (Yang *et al.*, 2002). Esta metodologia é também utilizada para dados de *Affymetrix*, mas neste caso o procedimento LOWESS é aplicado sucessivamente para cada par de lâminas, e a normalização final para cada lâmina é dada pela média geral dos resultados de cada um de seus pareamentos. O procedimento é repetido até convergência, i.e., até que pareamentos adicionais não alterem a normalização dos dados, e é denominado LOWESS cíclico. Outro procedimento bastante comum de normalização é denominado normalização quantílica (Bolstad *et al.*, 2003). Este procedimento faz com que todas as lâminas apresentem mesma distribuição empírica dos valores de intensidade, de maneira que elas são coincidentes em termos de locação (incluindo medidas de centralidade e percentis) e escala ou variabilidade.

Após a análise das imagens e normalização dos dados (as quais são por si próprios procedimentos

bastante interessantes e importantes em qualquer estudo de *microarray*), procedimentos estatísticos mais formais são utilizados, dependendo do objetivo dos experimentos. Por exemplo, análise de agrupamento é bastante utilizada tanto para se agruparem genes quanto amostras, no intuito de se descobrirem grupos de genes ou grupos de amostras com padrões de expressão similares. Análise discriminante é também bastante comum em estudos médicos utilizando-se amostras de pacientes sadios e doentes, para o desenvolvimento de modelos de classificação para uso em testes de diagnóstico. Outro procedimento bastante comum com dados de *microarray* refere-se a testes de significância para a detecção de genes diferencialmente expressos em amostras provenientes de diferentes grupos experimentais. Neste contexto, metodologias relativas a modelos lineares (como modelos de ANOVA e modelos mistos) são as mais frequentemente utilizadas, e são então foco da discussão que se segue.

A utilização da técnica da análise de variância (ANOVA) para dados de *microarrays* foi sugerida por Kerr *et al.* (2000), os quais propuseram o seguinte modelo para dados relativos a hibridizações competitivas:

$$y_{\text{smgr}} = \mu + S_s + M_m + (SM)_{\text{sm}} + G_g + (SG)_{\text{sg}} + (MG)_{\text{mg}} + (TG)_{\text{tg}} + \epsilon_{\text{smgr}}, \quad (1)$$

Onde  $y_{\text{smgr}}$  representa as intensidades de expressão, na escala logarítmica;  $\mu$  é uma constante geral;  $S_s$ ,  $M_m$  e  $(SM)_{\text{sm}}$  são os efeitos de lâminas, marcação (Cy3 e Cy5), e a interação entre lâminas e marcação, os quais são genericamente denominados como *fatores globais*;  $G_g$  refere-se aos efeitos dos genes;  $(SG)_{\text{sg}}$  representa a interação entre genes e lâminas;  $(MG)_{\text{mg}}$  refere-se a efeitos de marcação específicos para cada gene;  $(TG)_{\text{tg}}$  representa os efeitos de tratamentos na expressão de cada gene (e são portanto os efeitos de interesse); e  $\epsilon_{\text{smgr}}$  é um termo residual com variância  $\sigma_{\epsilon}^2$ .

O trabalho de Kerr *et al.* (2000) foi uma importante contribuição para a literatura em análise de experimentos com *microarrays*, pois nele foi proposto que múltiplos fatores (como marcação, lâminas, tratamentos, etc.) deveriam ser considerados simultaneamente nas análises dos dados. O modelo por eles proposto, entretanto,

apresenta algumas limitações com relação às suposições consideradas, as quais são extremamente fortes e pouco plausíveis. Primeiro, uma variância residual comum é considerada para todos os genes; hoje já é consenso que esta suposição não é adequada, mesmo na escala logarítmica. E em segundo e mais importante, um modelo de efeitos fixos foi adotado, ignorando-se as múltiplas fontes de variação aleatória, e considerando-se cada *spot* nas lâminas como unidade experimental (Kerr & Churchill, 2001).

Wolfinger *et al.* (2001) estenderam o modelo de ANOVA de Kerr *et al.* (2000), incluindo efeitos aleatórios para modelar a dependência entre observações relativas aos mesmos *spots* (i.e., intensidades Cy3 e Cy5) ou lâminas, e permitindo componentes de variância específicos para gene. Por conveniência computacional, estes autores propuseram um procedimento de ajuste do modelo em duas etapas. O primeiro estágio, denominado “normalização global”, considera todos os dados conjuntamente e ajusta um modelo expresso por:

$$y_{\text{smgr}} = \mu + S_s + M_m + (SM)_{\text{sm}} + e_{\text{smgr}}, \quad (2)$$

Onde os termos são como definidos no modelo (1). A partir deste modelo de normalização global são obtidos os resíduos estimados,  $\hat{e}_{\text{smgr}}$ . O segundo estágio compreende uma série de modelos específicos para cada gene, os quais são expressos como:

$$\hat{e}_{\text{smgr}} = \mu_g + S_{\text{sg}} + M_{\text{mg}} + T_{\text{tg}} + \varepsilon_{\text{smgr}}, \quad (3)$$

Onde  $\mu_g$ ,  $S_{\text{sg}}$ ,  $M_{\text{mg}}$  e  $T_{\text{tg}}$  são constante geral, efeito de lâmina, efeito de marcação e efeito de tratamento, específicos para cada gene  $g$ . Componentes de variância, também específicos para cada gene, são estimados para os efeitos aleatórios de lâmina e resíduo. Outros fatores devem ainda ser incluídos no modelo (veja Rosa *et al.* (2005) para uma discussão neste sentido), dependendo da estrutura de tratamentos (por exemplo, experimentos fatoriais) e do delineamento experimental.

Modelos lineares têm sido também bastante utilizados na análise de dados de *microarray* com hibridizações independentes. Por exemplo, com experimentos utilizando-se lâminas *Affymetrix*, modelos lineares podem ser utilizados tanto para

a análise das medidas resumo (i.e., índices de expressão tipo MAS5.0 ou RMA) como para a análise dos dados no nível de sondas, na qual a variável resposta refere-se à intensidade observada para cada sonda PM, na escala logarítmica. Alternativamente, a análise pode ser efetuada considerando-se o logaritmo da diferença entre as sondas PM e MM, ou ainda considerando-se os valores de MM como covariável no modelo para PM (Chu *et al.*, 2002).

Um problema bastante comum na análise de dados de *microarray* refere-se ao fato de que o número de observações por amostra (i.e., o número de genes em cada lâmina) é muitíssimo maior do que o número de amostras (ou lâminas) no experimento. Esta condição impossibilita a utilização de métodos multivariados clássicos, por exemplo, para se modelar covariâncias entre os níveis de expressão dos diferentes genes. Desta maneira, muito frequentemente as análises são conduzidas independentemente para cada gene. Entretanto, dado o pequeno tamanho amostral para cada gene, estimativas de componentes de variância são instáveis e o poder estatístico para se detectarem diferenças entre os grupos experimentais é geralmente pequeno. Uma estratégia bastante comum para se contornar este problema é a utilização de estimadores de “encolhimento” (*shrinkage estimators*) para os componentes de variância (veja, por exemplo, Smyth, 2004; Cui *et al.*, 2005; e Feng *et al.*, 2006).

Um outro problema bastante interessante na análise de dados de *microarray* refere-se ao fato de que múltiplos testes são efetuados em cada experimento (um para cada gene), de maneira que um grande número de falsos positivos é esperado mesmo em situações nas quais não há qualquer expressão diferenciada entre os grupos experimentais. Para se entender melhor este problema, basta lembrar que para determinado nível de significância  $\alpha$  adotado, a probabilidade de rejeição errônea de uma hipótese nula (erro tipo I) é  $\alpha$ . Da mesma maneira, se forem testados múltiplos genes, e assumindo-se que estes são expressos independentemente, tem-se probabilidade  $[1-(1-\alpha)^m]$  de pelo menos uma falsa rejeição, onde  $m$  é o número de testes efetuados. Por exemplo, se for adotado  $\alpha = 0,05$  para determinado estudo de expressão gênica tem-se 5% de chance de se detectar um específico gene

como diferencialmente expresso em dois grupos experimentais quando na verdade não existe diferença entre os grupos. Se 10 genes (independentes) forem considerados no estudo, tem-se aproximadamente 40% de chance de pelo menos um falso positivo. Esta probabilidade sobe para algo em torno de 64% no caso de 20 genes, e ultrapassa 90% com apenas 45 genes.

Outra maneira de se pensar este problema de testes múltiplos em experimentos com *microarrays* é que o número esperado de falsos positivos é igual a  $\alpha \times m$ . Por exemplo, suponha um experimento com *microarrays* no qual cada amostra de RNA é subdividida e hibridizada em duas lâminas, as quais são então divididas aleatoriamente em dois grupos (A e B), de maneira que cada animal tem uma lâmina no grupo A e outra no grupo B. Nestas circunstâncias tem-se um experimento no qual não existe qualquer diferença real na expressão gênica entre os dois grupos experimentais. Assim, para um experimento considerando-se  $m = 10.000$  genes, por exemplo, espera-se que 500 deles sejam detectados como diferencialmente expressos se  $\alpha = 0,05$ . Isso significa que a probabilidade de obtenção de p-valores menores do que 0,05 é 5%, como consequência do fato de que p-valores obtidos a partir de múltiplos testes independentes, todos sob a hipótese de nulidade, apresentam distribuição uniforme entre 0 e 1 (Allison *et al.*, 2002).

Existem vários procedimentos disponíveis para o ajuste do nível de significância quando múltiplos testes são efetuados num experimento, os quais variam em relação às suposições consideradas (por exemplo, independência entre os testes) e o tipo de controle desejado do erro tipo I. O método mais simples e utilizado para a correção do nível de significância em ensaios com testes múltiplos é denominado correção de Bonferroni, o qual controla a probabilidade global de qualquer falso positivo (i.e., um ou mais falsos positivos) no experimento. Para se aplicar a correção de Bonferroni, simplesmente especifica-se o nível de significância geral  $\alpha$  desejado, o qual é então dividido pelo número de testes a serem realizados. Este nível de significância corrigido  $\alpha^* = \alpha/m$  é então utilizado como limite de significância para cada um dos testes efetuados.

No caso de *microarrays*, entretanto, dada a grande quantidade de genes estudados (da ordem

de milhares), a correção de Bonferroni passa a ser muitíssimo conservativa, diminuindo drasticamente o poder de teste para a detecção de genes diferencialmente expressos. Além disso, num experimento com tantos genes e testes considerados, dificilmente é este o tipo de controle desejado de erro tipo I e, em geral, os pesquisadores aceitam maiores riscos em termos de falsos positivos para diminuir a chance de que importantes diferenças em expressão gênica não sejam detectadas (erro tipo II, relativo a falsos negativos).

Um critério menos restritivo para múltiplos testes refere-se à taxa de falsos positivos (FDR), definida como a proporção esperada de falsos positivos entre todos os testes significativos (Benjamini & Hochberg, 1995), i.e.,  $E[V/(V+S)] = E[V/R]$ , onde os valores V, S e R são definidos na Tabela 1. Desta maneira, ao invés de se controlar a probabilidade de  $V > 0$ , como no caso da correção de Bonferroni, o FDR controla a proporção de hipóteses nulas erroneamente rejeitadas entre todas as hipóteses rejeitadas para um dado experimento. Um procedimento de FDR cada vez mais comum na literatura refere-se ao conceito de q-valor, definido como o menor valor de FDR para o qual uma determinada hipótese seria rejeitada (Storey, 2003). Para uma extensiva revisão dos diferentes métodos para controle do erro tipo I com testes múltiplos veja, por exemplo, Dudoit *et al.* (2003). Programas computacionais disponíveis gratuitamente para análises de experimentos de *microarray*, incluindo testes de significância utilizando-se procedimentos baseados em estimadores de encolhimento (como os programas LIMMA e R/MAANOVA) e métodos de FDR podem ser encontrados na página de *web* do *Bioconductor* (<http://www.bioconductor.org/>).

A comparação dos padrões de expressão gênica entre grupos experimentais não se limita à geração de uma lista de genes supostamente diferencialmente expressos, obtidos a partir de testes de significância. Procedimentos complementares de análise estatística e bioinformática são geralmente utilizados, como a análise de agrupamento para o estudo de genes co-regulados, e o estudo de listas de genes super-representadas (Hosack *et al.*, 2003). Por exemplo, os genes podem ser classificados em termos de processos metabólicos ou

**Tabela 1-** Número de hipóteses nulas falsas e verdadeiras, as quais são rejeitadas ou não utilizando-se determinado teste estatístico, para um total de  $m$  testes.

*Table 1 - Number of false and true null hypotheses, which are rejected or not using a specific statistical test, for a total of  $m$  tests.*

Hipótese nula <i>Null Hypothesis</i>	Resultado do teste <i>Test Result</i>		Total <i>Total</i>
	Não rejeita-se $H_0$ <i>Do not reject <math>H_0</math></i>	Rejeita-se $H_0$ <i>Reject <math>H_0</math></i>	
Verdadeira / <i>True</i>	U	V	$m_0^{(1)}$
Falsa / <i>False</i>	T	S	$m - m_0$
Total / <i>Total</i>	$m - R$	R	$m$

<sup>(1)</sup> O número de hipóteses nulas verdadeiras  $m_0$  é obviamente desconhecido, de maneira que as únicas quantidades observáveis na Tabela 1 são o número total de testes ( $m$ ) e o número de hipóteses nulas rejeitadas ( $R$ ) ou não ( $m - R$ ).

<sup>(1)</sup>*The number of true null hypotheses  $m_0$  is obviously unknown, such that the only observable quantities in Table 1 are the total number of tests ( $m$ ) and the number of null hypotheses rejected ( $R$ ) or not ( $m - R$ ).*

grupos ontológicos (*gene ontology*) gerando tabelas de contingência  $2 \times 2$ , uma para cada processo metabólico ou grupo ontológico. No exemplo ilustrado na Tabela 2, tem-se  $R$  genes detectados como diferencialmente expressos e  $(m - R)$  genes tidos como similarmente expressos entre os grupos experimentais, de um total de  $m$  genes considerados. Estes mesmos  $m$  genes são então classificados como pertencentes ou não a determinado grupo de genes (baseado, por exemplo, em ontologia ou informação relativa a processos metabólicos), sendo que no exemplo  $m_G$  e  $m_{\bar{G}}$  genes foram classificados como pertencentes ou não ao grupo de interesse, respectivamente. A seguir, as frequências  $m_{GD} / m_G$  e  $m_{\bar{G}D} / m_{\bar{G}}$  são comparadas, para cada grupo ontológico, utilizando-se por exemplo testes de qui-quadrado e/ou teste exato de Fisher, para a identificação de grupos de genes com proporções significativamente maiores de genes detectados como diferencialmente expressos (Beissbarth &

Speed, 2004).

### Exemplos de aplicação na pesquisa zootécnica

Devido sua característica prospectiva e capacidade de analisar milhares de genes simultaneamente, a tecnologia de *microarrays* está sendo cada vez mais utilizada em pesquisas da área animal, notadamente nos estudos que envolvem características de herança complexa e/ou de baixa herdabilidade.

Aplicações dessa tecnologia podem ser encontradas em diferentes áreas da zootecnia, incluindo crescimento e metabolismo, reprodução, resposta imune a doenças e parasitoses, e resposta a fatores de estresse não-infecciosos como restrição alimentar, exposição a elementos tóxicos e outras condições ambientais desfavoráveis. Nesta secção são apresentados alguns exemplos dessas aplicações.

**Tabela 2** - Tabela de contingência  $2 \times 2$  para genes classificados de acordo com o resultado do teste estatístico (significativo ou não) para expressão diferenciada, e como pertencentes ou não a determinado grupo metabólico ou ontologia.

*Table 2.  $2 \times 2$  contingency table with genes classified according to the result of the statistical test (significant or not) for differential expression, and as pertaining or not to a determined metabolic group or gene ontology.*

Grupo Ontológico <i>Ontology group</i>	Resultado do teste / <i>Test Result</i>		Total <i>Total</i>
	Expressão diferenciada <i>Differential Expression</i>	Expressão similar <i>Similar expression</i>	
Pertencente / <i>Pertaining</i>	$m_{GR}$	$m_{GS}$	$m_G$
Não-pertencente / <i>Non-pertaining</i>	$m_{\bar{G}R}$	$m_{\bar{G}S}$	$m_{\bar{G}}$
Total / <i>Total</i>	R	$m - R$	$m$

### **Resposta imune a doenças e parasitoses**

O estudo de doenças patogênicas a nível molecular em animais de produção tem sido limitado pela escassez de ferramentas que permitam analisar conjuntamente as respostas do hospedeiro e do patógeno. Estas limitações estão desaparecendo com o advento da tecnologia de *microarrays*, que facilita uma rápida caracterização da expressão gênica global a um nível individual de células e tecidos (Wilson *et al.*, 2005).

Neste sentido, com o objetivo de estudar as respostas celulares à colonização com *E. coli* O157:H7, Li & Hovde (2007) realizaram um experimento para caracterizar os padrões de expressão gênica na mucosa da junção reto-anal de bezerros desafiados com diferentes cepas de *E. coli* e, para tanto, utilizaram *microarrays* de cDNA que continham 4.608 transcritos oriundos de tecidos gastrintestinais e das glândulas mamária e pituitária de bovinos. Os autores identificaram 49 genes diferencialmente expressos, cujas funções evidenciaram o envolvimento de múltiplos processos na resposta celular a *E. coli* O157:H7 e sugeriram que mudanças dinâmicas na estrutura celular e nas vias de transdução de sinais cálcio-dependente podem ter papel importante no processo de colonização.

Na tentativa de entender os mecanismos moleculares envolvidos na resposta hepática aos lipopolissacarídeos (LPS) produzidos por bactérias gram-negativas, Cao *et al.* (2006) utilizaram um *microarray* de cDNA bovino contendo 109 seqüências de genes associados com o processo inflamatório, sistema endócrino e resposta imune para analisar a expressão gênica no fígado de ovinos desafiados com diferentes dosagens de LPS de *E. coli* 0111:B4. Os autores não apenas identificaram diversos genes que desempenham funções importantes no processo inflamatório, mas sugeriram também que a resposta aos LPS é dependente de vários fatores, dentre os quais o intervalo de amostragem do tecido, a dosagem e o tipo de LPS utilizado. Abordagem semelhante foi utilizada por Gu & Bertone (2004), que produziram *microarrays* contendo sondas de oligonucleotídeos representando 3.098 genes de eqüinos e avaliaram o padrão de expressão gênica em sinoviócitos de cavalos adultos desafiados com LPS de *E. coli* 055:B5<sup>w</sup>. Além de identificar um conjunto de 102 genes cuja expressão foi

modulada pela endotoxina, os autores descobriram indícios de novos padrões de sinalização e adesão celular que abrem perspectivas para novas pesquisas.

A paratuberculose bovina é uma doença infecciosa crônica causada pelo *Mycobacterium paratuberculosis*, cuja incidência provoca perdas econômicas significativas devido à redução na produtividade, maior susceptibilidade a outras doenças, aumento dos custos sanitários e maior taxa de descarte precoce dos animais. A falta de um método eficiente para o diagnóstico precoce dessa enfermidade tem sido um dos principais entraves para o controle efetivo da sua propagação. Na tentativa de identificar uma *assinatura genética* que possa ser utilizada no diagnóstico precoce da paratuberculose, Skovgaard *et al.* (2006) usaram *microarrays* de cDNA para medir simultaneamente a expressão de mais de 1.300 genes em leucócitos de bovinos portadores subclínicos dessa enfermidade, conseguindo identificar um conjunto de genes diferentemente expressos nos animais infectados com potencial de serem utilizados para essa finalidade.

Os produtos avícolas constituem-se num importante veículo de transmissão da *Salmonella* entérica e uma maneira efetiva de reduzir a contaminação dos alimentos seria a criação aves mais resistentes a este patógeno. No entanto, a complexidade dos fatores envolvidos na manifestação dessa característica dificulta muito a compreensão das suas bases genéticas. Assim sendo, van Hemert *et al.* (2006) utilizaram *microarrays* de cDNA contendo seqüências identificadas no intestino e baço de galinhas, para comparar o perfil da expressão gênica no intestino de duas linhagens de frangos (crescimento rápido e lento) infectadas com *Salmonella*. Os resultados obtidos evidenciaram que as duas linhagens apresentam diferenças significativas nos padrões de expressão gênica em resposta à infecção com *Salmonella*, indicando que os genes de ativação das células T têm a expressão induzida na linhagem de crescimento rápido, enquanto que na de crescimento lento predomina a modulação dos genes envolvidos na ativação dos macrófagos. Os autores também observaram a existência de diferenças nos padrões de expressão gênica dentro grupo controle (aves não infectadas de ambas as linhagens), sugerindo que a suscetibilidade à

*Salmonella* pode estar associada a diferenças no desenvolvimento do intestino.

Já Sarson *et al.* (2006), produziram um *microarray* contendo 84 seqüências gênicas relacionadas com a resposta imune e o desenvolvimento de processos inflamatórios nas aves, para investigar os efeitos de outro tipo de patógeno, o vírus causador da doença de Marek (MDV), enfermidade que está presente em praticamente todas as criações avícolas do mundo e caracteriza-se pelo aparecimento de tumores na pele, nos órgãos internos, no globo ocular e no sistema nervoso central e periférico, provocando imunossupressão, mortalidade e condenações na indústria de processamento. Os resultados obtidos permitiram identificar diversos mecanismos genéticos que são induzidos pela infecção *in vivo* com esse vírus, expandindo os conhecimentos sobre a resposta imune do hospedeiro.

A síndrome respiratória e reprodutiva suína (PRRS) foi descrita inicialmente nos Estados Unidos no final da década de 1980 e na Europa e Ásia no início da década de 1990, se caracterizando principalmente pela inapetência e problemas respiratórios em suínos de todas as idades e baixa taxa de concepção em rebanhos de reprodutores. Assim, com o objetivo de melhor entender os mecanismos envolvidos na manifestação dessa enfermidade, Miller & Fox (2004) usaram um *microarray* comercial (*LifeArray V2.34, Incyte Genomics*) contendo mais de 9.200 sondas para examinar a expressão de genes relacionados à apoptose em células MARC-145 infectadas com o vírus da PRRS e não observaram variações nos padrões de expressão desse grupo específico de genes durante as primeiras 24 horas após a infecção.

O vírus da síndrome da mancha branca (WSSV) é um patógeno que infecta crustáceos e se constitui no maior obstáculo para a sustentabilidade do cultivo de camarões no mundo todo. Para avaliar a eficácia da ciclohexamida, um inibidor da síntese protéica, no bloqueio da produção de proteínas virais de WSSV, Liu et al. (2005) construíram um *microarray* com produtos de PCR que representavam as 532 ORFs deste vírus (prováveis genes, identificados apenas por análise computacional). Utilizando o mRNA extraído da guelra de camarões tratados ou não com ciclohexamida antes da inoculação com o

WSSV, os autores foram capazes de identificar seis genes desse vírus que são transcritos independentemente da presença de proteínas virais, avançando nos conhecimentos sobre a interação hospedeiro-patógeno. Por outro lado, Robalino *et al.* (2007) produziram *microarrays* de cDNA englobando 2.469 genes expressos na guelra, hemócitos circulantes e hepatopâncreas de camarão (*Litopenaeus vannamei*), que foram utilizados para avaliar a expressão gênica em animais desafiados com o WSSV. Os resultados evidenciaram que a infecção viral estimula a expressão de diversos genes relacionados com a resposta imune e, ao mesmo tempo, inibe a expressão de genes que têm conexão com a proteção contra o estresse oxidativo.

Os nematóides gastrintestinais são uma importante causa de morbidade e mortalidade em ruminantes sob sistema de pastejo. Assim, Keanel *et al.* (2006) estudaram a expressão gênica no duodeno de ovelhas pertencentes a duas linhagens, resistente e susceptível a estes nematóides, utilizando *microarrays* de cDNA representando 9.238 genes de ovinos. Foram identificados 41 genes com padrão de expressão diferenciado entre as duas linhagens, sendo que os animais susceptíveis apresentaram maior número de genes relacionados ao estresse. Com base nestas evidências, os autores sugeriram que a susceptibilidade inata desses indivíduos pode estar associada a um comprometimento do trato gastrintestinal.

Localizadas no intestino delgado, as placas de Peyer são constituídas por tecidos linfóides organizados e estão relacionadas com a resistência às doenças. Contudo, pouco se sabe sobre seus mecanismos de ação, notadamente no que se refere aos processos envolvidos na identificação dos patógenos. Para melhor entender as funções dessas estruturas, Machado *et al.* (2005) investigaram os padrões normais de expressão gênica nas placas de Peyer presentes no jejuno de suínos jovens e adultos, criados sob condições convencionais. Utilizando *microarrays* de cDNA com 2.604 seqüências gênicas, estes autores verificaram que tanto nos animais jovens, quanto nos adultos, as placas de Peyer apresentavam diversos genes relacionados com a imunidade inata estimulados, enquanto que um grupo de genes relacionados com o crescimento celular e com a apoptose tiveram a

expressão alterada somente nos animais jovens.

#### ***Exposição a elementos tóxicos e outras condições ambientais desfavoráveis***

O microclima é um fator que está diretamente relacionado com o conceito de conforto térmico e, portanto, exerce influência no desempenho produtivo e reprodutivo dos animais criados de maneira intensiva e industrial. Nesse contexto, Collier *et al.* (2006), utilizaram *microarrays* de cDNA contendo cerca de 18.000 seqüências de genes bovinos para avaliar o perfil de expressão gênica em culturas de células do epitélio mamário que foram submetidas ao estresse térmico (42°C) por 24 horas. Os autores verificaram que os genes com funções relacionadas à biossíntese celular, metabolismo e morfogênese se encontravam reprimidos, enquanto aqueles associados com a resposta ao estresse e ao reparo de proteínas tinham a expressão estimulada, sugerindo que o efeito do estresse térmico na redução da produção de leite pode estar relacionado com a diminuição do crescimento do epitélio mamário.

A temperatura da água é um dos fatores ambientais mais importantes para a criação de peixes e o estresse provocado pela sua flutuação favorece a incidência de doenças e aumenta a taxa de mortalidade. Os ectotermos aquáticos adaptam-se a grandes variações de temperatura, porém os mecanismos moleculares que estão por trás dessa adaptabilidade ainda não são bem entendidos. Assim sendo, Ju *et al.* (2002) utilizaram um *microarray* de cDNA contendo 660 genes para estudar os efeitos da baixa temperatura sobre a expressão gênica no cérebro de *catfish* (*Ictalurus punctatus*), e foram capazes de identificar 61 genes diferencialmente expressos, a maior parte dos quais codificando chaperoninas ou proteínas envolvidas na transdução de sinais, sugerindo que, tanto a modificação de proteínas existentes, como a síntese *de novo* de proteínas induzidas pelo frio, são mecanismos importantes para a adaptação e tolerância dessa espécie aos ambientes de baixa temperatura.

A contaminação ambiental com hormônios e outros poluentes causadores de desbalanços no metabolismo hormonal é um problema sério, especialmente para os organismos aquáticos. A presença de níveis elevados de estradiol e outros xenoestrogênios na água têm provocado profundas

alterações histológicas em peixes, produzindo inclusive o hermafroditismo e induzindo a produção de vitelogenina (característica de fêmeas sexualmente maduras) em indivíduos machos. Assim, para caracterizar mudanças na expressão gênica que ocorrem quando linguados europeus (*Platichthys flesus*) são expostos a altas doses de 17-beta estradiol, Williams *et al.* (2007) usaram *microarrays* de cDNA (*GENIPOL European flounder microarray*) com sondas representando 3.336 transcritos dessa espécie e identificaram 175 transcritos diferencialmente expressos, dentre os quais aqueles que codificam vitelogeninas e outros indicadores da exposição aos estrogênios.

Ainda em relação à adaptação ambiental, Moens *et al.* (2007) construíram um *microarray* de cDNA com 960 genes de carpa para avaliar os efeitos da toxicidade crônica em peixes submetidos ao fluxo contínuo de efluentes por 21 dias e observaram que as funções da maior parte dos genes diferencialmente expressos estavam associadas com o balanço de energia no organismo (vias de metabolismo dos nutrientes e enzimas digestivas).

#### ***Resposta a fatores de estresse não-infecciosos***

O estresse nutricional é resultante da limitação qualitativa ou quantitativa de nutrientes fornecidos pela dieta, impedindo que o animal expresse seu potencial de crescimento. Neste sentido, Lehnert *et al.* (2006) utilizaram *microarrays* de cDNA (8.856 sondas em duplicatas) para avaliar as mudanças no perfil da expressão gênica no músculo *Longissimus dorsi* de bovinos da raça *Belmont Red* submetidos à restrição alimentar e, posteriormente, a um período de realimentação. Os resultados mostraram que após um período prolongado de restrição alimentar ocorre uma forte redução nos níveis de expressão de genes relacionados com a produção de proteínas estruturais, enzimas metabólicas e constituintes da matriz extracelular, os quais retornam aos níveis normais após o período de realimentação. Esses autores também observaram que as funções da maioria dos genes que tiveram a expressão reduzida em decorrência da restrição alimentar estão relacionadas com atividades específicas das fibras musculares de contração rápida, que têm papel fundamental no processo de adaptação à restrição calórica.

De forma similar, da Costa *et al.* (2004) utilizaram *microarrays* de cDNA com 5.500 seqüências provenientes de músculo esquelético de suínos para caracterizar a expressão gênica diferencial nos músculos *Longissimus dorsi* e *psaos maior* de suínos alimentados de forma convencional ou submetidos a uma restrição alimentar moderada, verificando que os genes relacionados com o *turnover* de substratos metabólicos (proteínas, carboidratos e lipídeos) tiveram a expressão altamente estimulada pela restrição alimentar. Ademais, os autores foram capazes de identificar novos genes que atuam como reguladores do crescimento.

Com o objetivo de estudar a influência da nutrição sobre a produção de leite, Ollier *et al.* (2007) utilizaram um *microarray* de oligonucleotídeos de cadeia longa (*Bovine Genome Oligo Set VI.1, Operon Biotechnologies*) para examinar os efeitos da restrição alimentar sobre a expressão gênica na glândula mamária de caprinos, encontrando 161 genes diferencialmente expressos, 88% dos quais com a expressão diminuída. Dentre estes, diversos genes com funções relacionadas com o metabolismo de substratos (proteínas, carboidratos e lipídeos), sem contar um grupo específico de 14 genes provavelmente envolvidos na inibição da proliferação e diferenciação das células do epitélio mamário.

Os neutrófilos são células sanguíneas leucocitárias de ação fagocítica, que estão presentes em fenômenos agudos de inflamação devido ao seu alto potencial de diapedese e rápida velocidade de migração. No entanto, possuem um curto tempo de vida (seis horas). Desse modo, Madsen *et al.* (2004) usaram *microarrays* de cDNA com 1.056 genes provenientes de leucócitos (<http://cafg.msu.edu>) para estudar as mudanças induzidas pelo parto na expressão gênica de neutrófilos obtidos antes, durante e depois do parto de vacas Holandesas e identificaram diversos genes diferencialmente expressos cujas funções estão relacionadas com a regulação da apoptose. Baseados nestes e também em outros resultados de experimentos biológicos, os autores sugeriram que o parto em bovinos pode prolongar a sobrevivência de neutrófilos.

### **Reprodução**

A competência dos oócitos, ou seja, a habilidade de serem fertilizados e se desenvol-

verem até o estágio de blastocisto é um fator de grande importância para a produção de embriões com elevado potencial de desenvolvimento, mas pouco se sabe sobre as características moleculares (em nível de expressão gênica) dos oócitos competentes. Assim sendo, Patel *et al.* (2007) utilizaram *microarrays* de cDNA (representando cerca de 15.200 genes), para comparar o perfil de transcritos em oócitos e embriões em estágio inicial do desenvolvimento, utilizando dois modelos de estudo, bezerras pré-púberes e fêmeas adultas. Os resultados evidenciaram um número significativo de genes (193 nos animais adultos e 233 nos pré-púberes) cuja expressão se encontrava estimulada, com predominância daqueles relacionados com a regulação da secreção de hormônios. Estes autores sugeriram que existe uma associação positiva entre a abundância de mRNAs da folistatina e a competência dos oócitos nos dois modelos estudados.

Nesta mesma linha de pesquisa, Bonnet *et al.* (2007) utilizaram um *microarray* de cDNA com 9.152 clones espécie-específicos para investigar os efeitos da indução da ovulação sobre o transcriptoma dos ovos de trutas (*Oncorhynchus mykiss*) e verificaram que o controle do processo ovulatório, tanto por indução hormonal, como pela manipulação do fotoperíodo, provocam alterações significativas na abundância dos transcritos presentes nos ovos de truta, que podem afetar seu potencial de desenvolvimento.

Na tentativa de ampliar os conhecimentos sobre a reprogramação do genoma que ocorre durante o estágio de pré-implantação do embrião, Misirlioglu *et al.* (2006) avaliaram os padrões de expressão gênica em dois estágios fundamentais do desenvolvimento embrionário: oócitos maduros e embriões no estágio de 8 células. Utilizando *microarrays* de oligonucleotídeos (*Affymetrix GeneChip Bovine Genome Array*), que representam mais de 23.000 transcritos de bovinos, os autores identificaram um grande número de genes relacionados com os processos de regulação da transcrição gênica, manutenção da estrutura da cromatina e com as vias de sinalização celular, cuja expressão é estimulada nos estágios iniciais da embriogênese.

Abordagem semelhante também foi utilizada por Whitworth *et al.* (2005), que avaliaram os padrões de expressão gênica nos estágios iniciais

do desenvolvimento de embriões suínos, produzidos *in vitro* e *in vivo*. Utilizando *microarrays* de cDNA que representavam aproximadamente 15.000 genes de suínos, os autores obtiveram um grande volume de dados que deverão auxiliar o entendimento do processo de embriogênese, assim como a identificação de marcadores para avaliação do potencial de desenvolvimento do embrião.

Com o objetivo de estudar a relação existente entre o perfil transcricional do embrião e o sucesso da prenhez em bovinos, El-Sayed *et al.* (2006) compararam a expressão gênica em biópsias de blastocistos realizadas antes da transferência para as receptoras, as quais foram posteriormente agrupadas em três conjuntos: a) biópsias de embriões que não resultaram em prenhez; b) biópsias de embriões que resultaram em prenhez, mas foram reabsorvidos; c) biópsias de embriões que resultaram em nascimentos. Utilizando dois tipos diferentes de *microarrays* de cDNA, um dos quais com seqüências exclusivas de embriões na fase de implantação, estes autores foram capazes de identificar grupos de genes diferencialmente expressos em cada uma das situações estudadas, que podem ser importantes para a determinação do destino do embrião após a transferência.

Já Caetano *et al.* (2004) utilizaram *microarrays* de cDNA (4.608 sondas específicas) para estudar o perfil da expressão gênica nos ovários e nos folículos de marrãs provenientes de linhagem que apresenta superioridade para algumas características reprodutivas (taxa de ovulação, sobrevivência embrionária e tamanho da leitegada). Os autores identificaram 71 genes nos ovários e 59 nos folículos que apresentavam expressão diferenciada, sugerindo que a seleção para características reprodutivas resultou em mudanças significativas no controle fisiológico da maturação folicular.

#### **Crescimento e metabolismo**

A adeno-hipófise ou glândula pituitária anterior exerce função preponderante na regulação do sistema endócrino, sintetizando e secretando os hormônios tróficos que modulam as secreções de outras glândulas endócrinas e, portanto, está diretamente relacionada com o controle de diversos processos fisiológicos, dentre os quais o crescimento e o metabolismo.

Com o objetivo de caracterizar a diferenciação (início da atividade) dos tireotrófos, somatotrófos e lactotrófos, Ellestad *et al.* (2006) produziram um *microarray* contendo 5.128 cDNAs expressos no sistema neuro-endócrino para avaliar as mudanças na expressão gênica na adeno-hipófise de frangos durante o desenvolvimento embrionário e identificaram vários fatores de transcrição e outras moléculas de sinalização que ainda não haviam sido associados com o desenvolvimento da pituitária.

A seleção para eficiência na produção e qualidade da carne são prioridades na produção moderna de suínos e o músculo esquelético é o tecido-alvo para a identificação de genes candidatos para estas características. Deste modo, Lin & Hsu (2005) estudaram os perfis de transcrição gênica no músculo *Longissimus dorsi* de suínos das raças Duroc e Taoyuan, que apresentam dramáticas diferenças no desenvolvimento muscular pós-natal. Empregando um *microarray* de cDNA humano (*Incyte Genomics*) contendo 9.182 sondas, os autores identificaram 73 genes induzidos e 44 genes reprimidos, a maioria relacionada com proteínas e enzimas que desempenham funções importantes no crescimento e metabolismo de miócitos esqueléticos e, dentre esses, um potencial marcador molecular para crescimento muscular em suínos.

Abordagem similar foi utilizada por Cagnazzo *et al.* (2006) que investigaram as diferenças nos perfis de transcritos músculo-específicos durante o desenvolvimento pré-natal de suínos das raças Duroc e Pietrain, que diferem na taxa de crescimento e na composição da carcaça na fase adulta, verificando que nos estágios iniciais do desenvolvimento (14 a 49 dias de gestação), os embriões da raça Duroc apresentaram maior expressão de genes relacionados à miogênese, enquanto que nos estágios mais avançados (63 a 91 dias de gestação), a expressão desse grupo de genes foi maior nos embriões da raça Pietrain.

Devido suas particularidades genéticas e fisiológicas, o fenótipo *callipyge*, que se manifesta em ovinos, é um dos modelos mais interessantes para se estudar a miogênese, notadamente os aspectos relacionados com a hipertrofia das fibras musculares. De maneira que Vuocolo *et al.* (2007) utilizaram o *Bovine Affymetrix GeneChip microarray* (que representa aproximadamente

19.000 genes) para comparar a expressão gênica em amostras de músculos provenientes de ovinos portadores e não portadores do genótipo *callipyge*, coletadas logo após o nascimento (antes da manifestação desse fenótipo) e as 12 semanas de idade. Estes autores puderam identificar um número significativo de genes cuja expressão é influenciada pelo genótipo *callipyge*, verificando também que estas alterações se iniciam antes mesmo dos animais manifestarem o fenótipo. Com base nestes resultados, propuseram um modelo descrevendo a rede regulatória principal e as modificações epigenéticas que provavelmente explicam as mudanças nos tipos de fibra e a hipertrofia que caracterizam esse fenótipo em ovinos.

A tecnologia de *microarrays* também tem sido útil para decifrar a complexa interação existente entre os componentes químicos das dietas (nutrientes) e o material genético dos indivíduos (genótipo), ou seja, dos mecanismos moleculares envolvidos na interação nutrientes-genótipo e a maneira como estas afetam o fenótipo (Kaput & Rodriguez, 2004).

O fígado tem uma função central na regulação do *status* metabólico, partição de nutrientes e gasto de energia. Assim, com o objetivo de melhorar os conhecimentos sobre a regulação das vias metabólicas que ocorrem neste órgão, Ponsuksili *et al.* (2007) comparam os perfis de expressão gênica em amostras do fígado de suínos pertencentes a raças que possuem características de carcaça divergente (obesa e magra), coletadas em diferentes estágios do desenvolvimento. Utilizando *microarrays* de oligonucleotídeos de cadeia longa (representando 13.297 genes específicos de suínos), estes autores evidenciaram diferenças marcantes nos padrões de expressão gênica, tanto na comparação entre as raças, como entre os estágios de desenvolvimento, identificando diversos genes candidatos para estudos funcionais relacionados com características de carcaça.

A adiposidade excessiva em frangos de corte é a consequência de muitos anos de seleção para elevadas taxas de ganho de peso, que hoje está sendo um dos maiores problemas desse setor da avicultura. Entretanto, pouco se sabe sobre os mecanismos genéticos que regulam as vias metabólicas e fisiológicas envolvidas na

manifestação dessa característica. Sendo assim, Bourneuf *et al.* (2006) utilizaram *microarrays* de cDNA contendo sondas tecido-específicas para estudar a expressão gênica no fígado de frangos pertencentes a linhagens que apresentam diferentes teores de gordura na carcaça (alta e baixa adiposidade), identificando diversos genes que podem estar relacionados com a ontologia da obesidade em frangos de corte.

O desenvolvimento dos depósitos de gordura corporal está estreitamente associado ao aumento no número de adipócitos que ocorre ao longo da vida e, no caso dos animais que têm predisposição genética para desenvolver o *marmoreio*, a proliferação dessas células nos depósitos intramusculares ocorre numa taxa sensivelmente maior do que nos demais depósitos do organismo. Com o objetivo de melhor entender as bases genéticas dessa característica, Wang *et al.* (2005) estudaram a expressão gênica no músculo *Longissimus dorsi* de bovinos da raça *Japanese Black*, que se caracteriza pela notável propensão de depositar grandes quantidades de gordura de baixo ponto de fusão nos depósitos intramusculares. Utilizando *microarrays* de cDNA contendo seqüências tecido-específicas, detectaram diferenças nos padrões de expressão gênica que forneceram pistas importantes sobre a regulação de algumas características particulares da raça *Japanese Black*, especialmente no que se refere ao início do desenvolvimento e taxa de deposição do tecido adiposo, vias metabólicas preferenciais e mecanismos de sinalização envolvidos na conversão de carboidratos em lipídeos, que deverão contribuir para o avanço nas pesquisas nesta área.

Uma importante prioridade na indústria avícola é desenvolver um rápido e acurado método para selecionar linhagens de aves que produzam elevadas taxas de ovos de alta qualidade. Com o objetivo de identificar transcritos diferencialmente expressos na glândula da casca em duas linhagens de galinhas (alta e baixa produção de ovos), Yang *et al.* (2006) desenvolveram um *microarray* de cDNA representando 2.743 transcritos, com o qual identificaram 85 genes que são expressos de forma diferente nestas duas linhagens, alguns deles com potencial para serem utilizados como marcadores moleculares nos programas de

seleção para produção de ovos.

### ***Estudos com insetos utilizados em sistemas de produção comercial***

De forma similar ao que está ocorrendo com as demais espécies exploradas pela zootecnia, a tecnologia de *microarrays* começa a ser utilizada em estudos com insetos utilizados em sistemas de produção comercial, como o são a sericicultura e a apicultura. Entretanto, essas pesquisas iniciais não estão focadas diretamente nas questões relativas à produção, mas sim em outros aspectos biológicos, uma vez que tais insetos também são utilizados como modelo de estudo em diversas áreas da biologia.

O bicho-da-seda, *Bombyx mori*, é um importante modelo biológico utilizado para estudar a fisiologia de insetos devido à disponibilidade de mutantes de linhagens homogêneas. Neste sentido, Kawasaki *et al.* (2004), construíram *microarrays* de cDNA representando 5.760 seqüências espécie-específicas, que foram utilizados para investigar as mudanças no perfil da expressão gênica durante o processo de metamorfose de *Bombyx mori*. Os resultados obtidos permitiram que os autores caracterizassem os perfis genéticos associados com as mudanças morfológicas que ocorrem durante esse processo. Posteriormente, Hong *et al.* (2006) construíram *microarrays* de cDNA contendo 2.445 seqüências expressas em ovos de bicho-da-seda (estágio de formação da banda germinativa), que foram utilizados para estudar o perfil da expressão gênica durante vários estágios do desenvolvimento embrionário. Estes autores foram capazes de identificar 241 transcritos diferencialmente expressos que estão associados com o estágio de formação da banda germinativa, dentre os quais, diversos codificando proteínas envolvidas no desenvolvimento e comunicação celular.

A abelha (*Apis mellifera*) é um importante modelo para estudos da plasticidade neural e comportamental, particularmente com respeito ao comportamento social, aprendizado e memória. Assim sendo, com o objetivo de estudar as bases moleculares do comportamento social de abelhas, Takeuchi *et al.* (2002) produziram um *microarray* de cDNA contendo 480 seqüências expressas no cérebro desses insetos, verificando que genes relacionados com via de sinalização cálcio-

dependente se apresentavam diferencialmente expressos em regiões do cérebro que estão ligadas a modulação do comportamento. Mais recentemente, Whitfield *et al.* (2006) construíram e validaram um *microarray* de cDNA contendo 8.872 seqüências expressas no cérebro de abelhas, com o objetivo de criar recursos para acelerar a análise dos mecanismos moleculares envolvidos no comportamento desses insetos.

### **Considerações finais**

Como pode ser observado nos exemplos que foram apresentados, a tecnologia de *microarrays* está reformulando a maneira tradicional de se fazer pesquisa na área de zootecnia, pois permite que questões relevantes sejam abordadas sob uma perspectiva na qual genética e fisiologia estão totalmente integradas, favorecendo o surgimento de novos paradigmas científicos que passam a nortear os rumos das pesquisas dessa área.

Neste sentido, é preciso destacar o caráter prospectivo da tecnologia de *microarrays* que, ao estabelecer associações entre grupos de genes e características fisiológicas, também define a ligação que existe entre o genoma e os processos biológicos envolvidos na manifestação do fenótipo, gerando novas hipóteses para serem testadas em estudos posteriores, mediante uso de abordagens mais direcionadas.

Outro paradigma que começa a ser adotado é a caracterização das chamadas *assinaturas moleculares*, termo utilizado para descrever o conjunto de genes que são característicos de um processo biológico ou de um fenótipo específico, ao invés de se manter o foco unicamente sobre o pequeno número de genes que apresentam expressão altamente diferenciada. Esta tendência terá grande impacto nos estudos de características fenotípicas complexas, que envolvem a participação de um grande número de genes, porém com efeito individual pequeno. Assim, a representatividade do genoma passa a ser uma característica de fundamental importância na confecção dos *microarrays* para estudos de expressão gênica com as espécies zootécnicas.

Entretanto, existem duas limitações para a confecção de *arrays* contendo grande número de genes. A primeira é a disponibilidade de informações sobre o genoma da espécie, ou seja, é preciso

que a maior parte dos seus genes já tenha sido seqüenciada e identificada, uma vez que as tecnologias para produção de *microarrays* de alta densidade são baseadas no uso de oligonucleotídeos. Já a segunda, diz respeito à demanda de utilização desses *microarrays*, pois, em função dos elevados custos de produção, os mesmos têm que ser confeccionados em escala industrial. Desta maneira a disponibilidade de *microarrays* com elevado número de genes deverá ficar restrita por algum tempo às espécies de maior importância econômica, para as quais existe maior volume de recursos para pesquisa.

Além da parte tecnológica, o planejamento, a análise e a interpretação dos dados em experimentos com *microarray* também apresentam seus obstáculos e desafios. Como discutido anteriormente, os estudos com *microarray* ainda são relativamente pequenos, especialmente na área zootécnica. Tais experimentos, no entanto, geram uma enormidade de dados, de dimensões e complexidade sem precedentes. Desta maneira, um cuidadoso planejamento de tais experimentos é crucial para o sucesso das pesquisas envolvendo *microarrays*. Um planejamento ineficiente geralmente incorre em situações experimentais nas quais há confundimento entre fatores, os quais não são possíveis de serem separados por nenhuma metodologia estatística.

Uma vez conduzidos os experimentos com *microarray*, o desafio seguinte refere-se à análise dos dados obtidos. Primeiro, dadas as dimensões e complexidade de tais conjuntos de dados, metodologias computacionais apropriadas são necessárias para a eficiente armazenagem e manuseio dos arquivos de dados, principalmente para experimentos com grande número de lâminas ou para a análise conjunta de múltiplos experimentos independentes. Segundo, métodos estatísticos específicos são necessários para uma análise apropriada dos mesmos, principalmente em relação à multiplicidade de testes efetuados em cada experimento, a qual pode gerar uma taxa de falsos positivos inadvertidamente alta se não for modelada objetivamente.

Diversos métodos estatísticos têm sido sugeridos para a análise de dados de *microarray* para diferentes objetivos experimentais. Inicialmente, a análise de dados de tais experimentos eram conduzidos de maneira bastante

rudimentar, por exemplo estudando-se somente a mudança (em valores absolutos, ou *fold change*) nos níveis de expressão gênica entre os grupos experimentais, sem nem mesmo levar-se em conta a variabilidade dos dados. Posteriormente os experimentos com *microarrays* passaram a ser analisados de maneira mais objetiva e formal do ponto de vista estatístico, com a utilização de testes de hipóteses para o estudo das mudanças nos níveis de expressão gênica. Os métodos estatísticos inicialmente utilizados, no entanto, eram adaptações de procedimentos metodológicos clássicos. Como a maioria destas metodologias foi desenvolvida em outros contextos experimentais, estes procedimentos clássicos nem sempre são adequados ou eficientes para a análise de dados com estruturas tão complexas e envolvendo tantas variáveis respostas simultaneamente.

Felizmente, mais recentemente diversas metodologias estatísticas têm sido desenvolvidas especificamente para a análise de dados de *microarrays*, como por exemplo, modelos mais robustos ou com suposições mais adequadas para experimentos desta natureza, modelos que combinam informação dos vários genes estudados em cada lâmina, procedimentos alternativos para o controle de FDR, etc. Uma outra área promissora na análise de dados de *microarray* refere-se a procedimentos de bioinformática que incorporam aos modelos estatísticos informação relativa a similaridades de seqüências de DNA, posição dos genes no genoma, ontologias, etc.

Outros resultados de experimentos de *microarrays* usando animais de produção estarão disponíveis em breve e o próximo desafio será explorar os dados gerados a fim de examinar minuciosamente genes específicos e redes regulatórias que possam ajudar no entendimento dos processos biológicos relacionados às características de interesse. Outra área de grande potencial neste aspecto refere-se a estudos que buscam integrar estudos de expressão gênica com outros tipos de informação, como por exemplo marcadores moleculares e características fenotípicas (veja, por exemplo, Rosa (2007) neste mesmo exemplar da *Revista Brasileira de Zootecnia*, para uma discussão neste sentido). É possível ainda que num futuro próximo novos usos e outras aplicações de experimentos com *microarrays* serão encontrados. O contínuo

desenvolvimento e melhoramento da tecnologia de *microarrays*, bem como dos conceitos e metodologias aqui apresentados, são cruciais para o contínuo sucesso do uso dessa ferramenta em estudos de genômica funcional em animais.

### Agradecimento

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela concessão da bolsa de doutorado sanduíche no exterior ao aluno Leonardo B. Rocha.

### Literatura citada

- ALLISON, D.B.; GADBURY, G.L.; HEO, M.S. *et al.* A mixture model approach for the analysis of microarray gene expression data. **Computational Statistics & Data Analysis**, v.39, n.1, p.1-20, 2002.
- BEISSBARTH, T.; SPEED, T.P. Gostat: find statistically overrepresented Gene Ontologies within a group of genes. **Bioinformatics**, v.20, n.9, p.1464-1465, 2004.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society B**, v.57, p.289-300, 1995.
- BONNET, E.; FOSTIER, A.; BOBE, J. Microarray-based analysis of fish egg quality after natural or controlled ovulation. **BMC Genomics**, v.8, n.55, 2007.
- BOURNEUF, E.; HÉRAULT, F.; CHICAULT, C. *et al.* Microarray analysis of differential gene expression in the liver of lean and fat chickens. **Gene**, v.372, p.162-170, 2006.
- BOLSTAD, B.M.; IRIZARRY, R.A.; ASTRAND, M. *et al.* A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. **Bioinformatics**, v.19, n.2, p.185-193, 2003.
- BROWN, P.O.; BOTSTEIN, D. Exploring the new world of the genome with DNA microarrays. **Nature Genetics**, v.21, p.33-37, 1999.
- BUENO, J.S.D.; GILMOUR, S.G.; ROSA, G.J.M. Design of microarray experiments for genetical genomics studies. **Genetics**, v.174, n.2, p.945-957, 2006.
- CAETANO, A.R.; JOHNSON, R.K.; FORD, J.J. *et al.* Microarray Profiling for Differential Gene Expression in Ovaries and Ovarian Follicles of Pigs Selected for Increased Ovulation Rate. **Genetics**, v.168, p.1529-1537, 2004.
- CAGNAZZO, M.; TE PAS, M.F.W.; PRIEM, J. *et al.* Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. **Journal of Animal Science**, v.84, p.1-10, 2006.
- CAO, H.; KABAROFF, L.C.; YOU, Q. *et al.* Characterization of ovine hepatic gene expression profiles in response to *Escherichia coli* lipopolysaccharide using a bovine cDNA microarray. **BMC Veterinary Research**, v.2, n.34, 2006.
- CHAUDHURI, J.D. Genes arrayed out for you: the amazing world of microarrays. **Medical Science Monitor**, v.11, n.2, p.52-62, 2005.
- CHU, T.M.; WEIR, B.; WOLFINGER, R. A systematic statistical linear modeling approach to oligonucleotide array experiments. **Mathematical Biosciences**, v.176, n.1, p.35-51, 2002.
- CHURCHILL, G.A. Fundamentals of experimental design for cDNA microarrays. **Nature Genetics**, v.32, p.490-495, 2002.
- COLLIER, R.J.; STIENING, C.M.; POLLARD, B.C. *et al.* Use of gene expression microarrays for evaluating environmental stress tolerance at the cellular level in cattle. **Journal of Animal Science**, v.84, n.E. Suppl., p.E1-E13, 2006.
- CUI, X.; HWANG, J.T.G.; QIU, J. *et al.* Improved statistical tests for differential gene expression by shrinking variance components estimates. **Biostatistics**, v.6, n.1, p.59-75, 2005.
- DA COSTA, N.; MCGILLIVRAY, C.; BAI, Q. *et al.* Restriction of dietary energy and protein induces molecular changes in young porcine skeletal muscles. **Journal of Nutrition**, v.134, p.2191-2199, 2004.
- DOBBIN, K.; SIMON, R. Sample size determination in microarray experiments for class comparison and prognostic classification. **Biostatistics**, v.6, n.1, p.27-38, 2005.
- DUDOIT, S.; SHAFFER, J.P.; BOLDRICK, J.C. Multiple hypothesis testing in microarray experiments. **Statistical Science**, v.18, n.1, p.71-103, 2003.
- ELLESTAD, L.E.; CARRE, W.; MUCHOW, M. *et al.* Gene expression profiling during cellular differentiation in the embryonic pituitary gland using cDNA microarrays. **Physiological Genomics**, v.25, p.414-425, 2006.
- EL-SAYED, A.; HOELKER, M.; RINGS, F. *et al.* Large-scale transcriptional analysis of bovine embryo biopsies in relation to pregnancy success after transfer to recipients. **Physiological Genomics**, v.28, p.84-96, 2006.
- FENG, S.; WOLFINGER, R.D.; CHU, T.M. *et al.* Empirical Bayes analysis of variance component models for microarray data. **Journal of Agricultural Biological and Environmental Statistics**, v.11, n.2, p.197-209, 2006.
- GADBURY, G.L.; PAGE, G.P.; EDWARDS, J. *et al.* Power and sample size estimation in high dimensional biology. **Statistical Methods in Medical Research**, v.13, n.4, p.325-338, 2004.
- GU, W.; BERTONE, A.L. Generation and performance of an equine-specific large-scale gene expression microarray. **American Journal of Veterinary Research**, v.65, p.1664-1673, 2004.
- HIENDLEDER, S.; BAUERSACHS, S.; BOULESTEIX, A. *et al.* Functional genomics: tools for improving farm animal health and welfare. **Revue scientifique et technique (International Office of Epizootics)**, v.24, n.1, p.354-377, 2005.
- HONG, S.M.; NHO, S.K.; KIM, N.S. *et al.* Gene expression profiling in the silkworm, *Bombyx mori*, during early embryonic development. **Zoological Science**, v.23, n.6, p.517-28, 2006.
- HOSACK, D.A.; DENNIS, G.; SHERMAN, B.T. *et al.* Identifying biological themes within lists of genes with EASE. **Genome Biology**, v.4, n.10, article R70, 2003.
- HU, J.H.; ZOU, F.; WRIGHT, F.A. Practical FDR-based sample size calculations in microarray experiments. **Bioinformatics**, v.21, n.15, p.3264-3272, 2005.
- IRIZARRY, R.A.; HOBBS, B.; COLLIN, F. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. **Biostatistics**, v.4, n.2, p.249-264, 2003.
- JALURIA, P.; KONSTANTOPOULOS, K.; BETENBAUGH, M. *et al.* A perspective on microarrays: current applications, pitfalls, and potential uses. **Microbial Cell Factories**, v.13, n.6, Article 4, 2007.
- JUNG, S.H. Sample size for FDR-control in microarray data analysis. **Bioinformatics**, v.21, n.14, p.3097-3104, 2005.
- JU, Z.; DUNHAM, R.A.; LIU, Z. Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation. **Molecular Genetics and Genomics**, v.268, p.87-95, 2002.
- KAFATOS, F.C.; JONES, C.W.; EFSTRATIADIST, A. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. **Nucleic Acids Research**, v.7, n.6, p.1541-1552, 1979.
- KAPUT, J.; RODRIGUEZ, R. L. Nutritional genomics: the next frontier in the postgenomic era. **Physiological Genomics**,

- v.16, p.166-177, 2004.
- KAWASAKI, H.; OTE, M.; OKANO, K. *et al.* Change in the expressed gene patterns of the wing disc during the metamorphosis of *Bombyx mori*. **Gene**, v.343, p.133-142, 2004.
- KEANEL, O.M.; ZADISSA, A.; WILSON, T. *et al.* Gene expression profiling of Naïve sheep genetically resistant and susceptible to gastrointestinal nematodes. **BMC Genomics**, v.7, n.42, 2006.
- KENDZIORSKI, C.; IRIZARRY, R.A.; CHEN, K. *et al.* On the utility of pooling biological samples in microarray experiments. **Proceedings of the National Academy of Sciences USA**, v.102, n.12, p.4252-4257, 2005.
- KERR, M.K. Design considerations for efficient and effective microarray studies. **Biometrics**, v.59, p.822-828, 2003.
- KERR, M.K.; CHURCHILL, G.A. Experimental design for gene expression microarrays. **Biostatistics**, v.2, p.183-201, 2001.
- KERR, M.K.; MARTIN, M.; CHURCHILL, G.A. Analysis of variance for gene expression microarray data. **Journal of Computational Biology**, v.7, n.6, p.819-837, 2000.
- LEHNERT, S.A.; BYRNE, K.A.; Reverter, A. *et al.* Gene expression profiling of bovine skeletal muscle in response to and during recovery from chronic and severe undernutrition. **Journal of Animal Science**, v.84, p.3239-3250, 2006.
- LI, C.; WONG, W.H. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. **Proceedings of the National Academy of Sciences USA**, v.98, n.1, p.31-36, 2001.
- LI, J.; HOVDE, C.J. Bovine gene expression profiles at the recto-anal junction mucosa 4 during colonization with *E. coli* O157:H7. **Applied and Environmental Microbiology**, v.73, n.7, p.2380-2385, 2007.
- LIN, C.S.; HSU, C.W. Differentially transcribed genes in skeletal muscle of Duroc and Taoyuan pigs. **Journal of Animal Science**, v.83, p.2075-2086, 2005.
- LIU, W.J.; CHANG, Y.S.; WANG, C.H. *et al.* Microarray and RT-PCR screening for white spot syndrome virus immediate-early genes in cycloheximide-treated shrimp. **Virology**, v.334, p.327-341, 2005.
- MACHADO, J.G.; HYLAND, K.A.; DVORAK, C.M.T. *et al.* Gene expression profiling of jejunal peyer's patches in juvenile and adult pigs. **Mammalian Genome**, v.16, p.599-612, 2005.
- MADSEN, S.A.; CHANG, L.C.; HICKEY, M.C. *et al.* Microarray analysis of gene expression in blood neutrophils of parturient cows. **Physiological Genomics**, v.16, p.212-221, 2004.
- MILLER, L.C.; FOX, J.M. Apoptosis and porcine reproductive and respiratory syndrome virus. **Veterinary Immunology and Immunopathology**, v.102, p.131-142, 2004.
- MISIRLIOGLU, M.; PAGE, G.P.; SAGIRKAYA, H. *et al.* Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. **Proceedings of the National Academy of Sciences USA**, v.103, p.18905-18910, 2006.
- MOENS, L.N.; SMOLDERS, R.; VAN DER VEN, K. *et al.* Effluent impact assessment using microarray-based analysis in common carp: A systems toxicology approach. **Chemosphere**, v.67, n.11, p.2293-2304, 2007.
- MULLER, P.; PARMIGIANI, G.; ROBERT, C. *et al.* Optimal sample size for multiple testing: The case of gene expression microarrays. **Journal of the American Statistical Association**, v.99, n.468, p.990-1001, 2004.
- OLLIER, S.; ROBERT-GRANIÉ, C.; BERNARD, L. *et al.* Mammary transcriptome analysis of food-deprived lactating goats highlights genes involved in milk secretion and programmed cell death. **Journal of Nutrition**, v.137, p.560-567, 2007.
- PATEL, O.V.; BETTEGOWDA, A.; IRELAND, J.J. *et al.* Functional genomics studies of oocyte competence: evidence that reduced transcript abundance for follistatin is associated with poor developmental competence of bovine oocytes. **Reproduction**, v.133, p.95-106, 2007.
- PONSUKSILI, S.; MURANI, E.; WALZ, C. *et al.* Pre- and postnatal hepatic gene expression profiles of two pig breeds differing in body composition: insight into pathways of metabolic regulation. **Physiological Genomics**, v.29, p.267-279, 2007.
- ROBALINO, J.; ALMEIDA, J.S.; MCKILLEN, D. *et al.* Insights into the immune transcriptome of the shrimp *Litopenaeus vannamei*: tissue specific expression profiles and transcriptomic responses to immune challenge. **Physiological Genomics**, v.29, p.44-56, 2007.
- ROGOJINA, A.T.; ORR, W.E.; SONG, B.K. *et al.* Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. **Molecular Vision**, v.9, p.482-496, 2003.
- ROSA, G.J.M. Delineamento de experimentos em genética genômica. **Revista Brasileira de Zootecnia**, v.36, Suplemento Especial, p.211-218, 2007.
- ROSA, G.J.M.; STEIBEL, J.P.; TEMPELMAN, R.J. Reassessing design and analysis of two-colour microarray experiments using mixed effects models. **Comparative and Functional Genomics**, v.6, n.3, p.123-131, 2005.
- SARSON, A.J.; ABDUL-CAREEM, M.F.; ZHOU, H. *et al.* Transcriptional analysis of host responses to Marek's disease viral infection. **Viral Immunology**, v.19, n.4, p.747-758, 2006.
- SCHENA, M.; SHALON, D.; DAVIS, R.W. *et al.* Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, v.270, n.5235, p.467-470, 1995.
- SKOVGAARD, K.; GRELL, S.N.; HEEGAARD, P.M.H. *et al.* Differential expression of genes encoding CD30L and P-selectin in cattle with Johne's disease: Progress toward a diagnostic gene expression signature. **Veterinary Immunology and Immunopathology**, v.112, p.210-224, 2006.
- SMITH, G.W.; ROSA, G.J.M. Interpretation of microarray data: trudging out of the abyss towards elucidation of biological significance. **Journal of Animal Science**, v.85, n.13(Suppl., 1), p.E20-E23, 2007.
- SMYTH, G.K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. **Statistical Applications in Genetics and Molecular Biology**, v.3, n.1, Article 3, 2004.
- STEIBEL, J.P.; ROSA, G.J.M. On reference designs for microarray experiments. **Statistical Applications in Genetics and Molecular Biology**, v.4, n.1, Article 36, 2005.
- STOREY, J.D. The positive false discovery rate: A Bayesian interpretation and the q-value. **Annals of Statistics**, v.31, n.6, p.2013-2035, 2003.
- TAKEUCHI, H.; FUJIYUKI, T.; SHIRAI, K. *et al.* Identification of genes expressed preferentially in the honeybee mushroom bodies by combination of differential display and cDNA microarray. **FEBS Letters**, v.513, p.230-234, 2002.
- TEMPELMAN, R.J. Assessing statistical precision, power, and robustness of alternative experimental designs for two color microarray platforms based on mixed effects models. **Veterinary Immunology and Immunopathology**, v.105, p.175-186, 2005.
- VAN HEMERT, S.; HOEKMAN, A.J.W.; SMITS, M.A. *et al.* Gene expression responses to a *Salmonella* infection in the chicken intestine differ between lines. **Veterinary Immunology and Immunopathology**, v.114, p.247-258, 2006.
- VUOCOLO, T.; BYRNE, K.; WHITE, J. *et al.* Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. **Physiological Genomics**, v.28, p.253-272, 2007.
- WALSH, B.; HENDERSON, D. Microarrays and beyond: what potential do current and future genomics tools have for

- breeders?. **Journal of Animal Science**, v.82, n. E. Suppl., p.E292-E299, 2004.
- WANG, Y.H.; REVERTER, A.; MANNEN, H. *et al.* Transcriptional profiling of muscle tissue in growing Japanese Black cattle to identify genes involved with the development of intramuscular fat. **Australian Journal of Experimental Agriculture**, v.45, n.8, p.809-820, 2005.
- WHITWORTH, K.M.; AGCA, C.; KIM, J.G. *et al.* Transcriptional profiling of pig embryogenesis by using a 15-K member unigene set specific for pig reproductive tissues and embryos. **Biology of Reproduction**, v.72, p.1437-1451, 2005.
- WHITFIELD, C.W.; BEN-SHAHAR, Y.; BRILLET, C. *et al.* Genomic dissection of behavioral maturation in the honey bee. **Proceedings of the National Academy of Sciences USA**, v.103, p. 16068-16075, 2006.
- WILLIAMS, T.D.; DIAB, A.M.; GEORGE, S.G. *et al.* Gene expression responses of European flounder (*Platichthys flesus*) to 17- $\beta$  estradiol. **Toxicology Letters**, v.168, p.236-248, 2007.
- WILSON, H.L.; AICH, P.; ROCHE, F.M. *et al.* Molecular analyses of disease pathogenesis: Application of bovine microarrays. **Veterinary Immunology and Immunopathology**, v.105, p.277-287, 2005.
- WIT, E.; NOBILE, A.; KHANIN, R. Near-optimal designs for dual-channel microarrays studies. **Applied Statistics**, v.54, n.5, p.817-830, 2005.
- WOLFINGER, R.D.; GIBSON, G.; WOLDINGER, E.D. *et al.* Assessing gene significance from cDNA microarray expression data via mixed models. **Journal of Computational Biology**, v.8, n.6, p. 625-637, 2001.
- YANG, K.T.; LIN, C.Y.; LIOU, J.S. *et al.* Differentially expressed transcripts in shell glands from low and high egg production strains of chickens using cDNA microarrays. **Animal Reproduction Science**, In Press, 2006, doi:10.1016/j.anireprosci.2006.09.004.
- YANG, Y.H.; DUDOIT, S.; LUU, P. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. **Nucleic Acids Research**, v.30, n.4, article e15, 2002.
- YANG, Y.H.; SPEED, T. Design issues for cDNA microarray experiments. **Nature Reviews Genetics**, v.3, p.579-588, 2002.