



**COPPE/UFRJ**

**EFEITO DA TALIDOMIDA NA EXPRESSÃO GÊNICA DE CÉLULAS  
HUMANAS TRATADAS COM LPS**

Barbara Cattete Dias

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Biomédica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Biomédica.

Orientadores: Flávio Fonseca Nobre

Ulisses Gazos Lopes

Rio de Janeiro

Junho de 2009

EFEITO DA TALIDOMIDA NA EXPRESSÃO GÊNICA DE CÉLULAS HUMANAS  
TRATADAS COM LPS

Barbara Cattete Dias

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA BIOMÉDICA.

Aprovada por:

---

Prof. Flávio Fonseca Nobre, Ph.D.

---

Prof. Ulisses Gazos Lopes, D.Sc.

---

Prof. Jurandir Nadal, D.Sc.

---

Prof. Milton Ozório Moraes, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2009

Dias, Barbara Cattete

Efeito da talidomida na expressão gênica de células humanas tratadas com LPS. / Barbara Cattete Dias – Rio de Janeiro: UFRJ/COPPE, 2009.

VIII, 75 p.: il.; 29,7 cm.

Orientadores: Flávio Fonseca Nobre

Ulisses Gazos Lopes

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Biomédica, 2009.

Referências Bibliográficas: p. 69-75.

1. LPS. 2. Talidomida. 3. Microarranjo. 4. ANOVA I. Nobre, Flávio Fonseca, *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Biomédica. III. Título.

## **Agradecimentos**

Ao professor Flavio pela enorme sabedoria compartilhada, dedicação, tempo, paciência e amizade.

Ao professor Ulisses pela obtenção dos dados e valiosa colaboração para este trabalho.

Aos amigos do LESS, Renata, Marcelo, Claudia, Roberta, Robson, Sandro, etc. por toda ajuda, amizade, parceria e piadas.

Aos amigos da turma de 2006 do PEB, por todos os momentos que passamos juntos: horas de estudo, dificuldades, superações e, sobretudo realizações.

Ao Rafael, pela inspiração que desde o primeiro dia tornou essa conquista possível.

À minha família, por todo amor, credibilidade, apoio e compreensão.

Aos demais amigos por todo o resto.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

EFEITO DA TALIDOMIDA NA EXPRESSÃO GÊNICA DE CÉLULAS HUMANAS  
TRATADAS COM LPS

Barbara Cattete Dias

Junho/2009

Orientadores: Flavio Fonseca Nobre

Ulisses Gazos Lopes

Programa: Engenharia Biomédica

O fator de necrose tumoral (TNF) é uma citocina que participa da resposta aguda inflamatória do hospedeiro em caso de infecção, além de contribuir para a patogênese de doenças infecciosas e autoimunes. Os lipopolissacarídeos (LPS) da membrana externa de bactérias Gram-negativas são importantes estimuladores de monócitos e macrófagos para a síntese dos mediadores inflamatórios incluindo o TNF. A Talidomida é uma droga anti-inflamatória utilizada em diversas doenças tais como o eritema nodoso leproso, uma complicação da hanseníase. No presente estudo, os efeitos anti-inflamatórios da Talidomida foram investigados através de dados obtidos de experimentos de microarranjos de oligonucleotídeos de células mononucleares de sangue periférico (PBMC) tratadas com o Lipopolissacarídeo (LPS) de bactérias Gram-negativas. Genes diferencialmente expressos foram identificados usando os métodos de modelagem ANOVA e seleção de razões de expressão intensidade-dependentes (ISER). Os resultados apontaram genes relacionados com desenvolvimento neuronal e embrionário, oncogênese, sinalização, divisão e morte celular.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

EFFECT OF THALIDOMIDE IN GENE EXPRESSION OF HUMAN CELLS TREATED  
WITH LPS

Barbara Cattete Dias

June/2009

Advisors: Flavio Fonseca Nobre

Ulisses Gazos Lopes

Department: Biomedical Engineering

Tumor necrosis factor (TNF) is a cytokine that participates in the host immune inflammatory response in case of infection, also contributing to the pathogenesis of infections and autoimmune diseases. Gram-negative bacterial lipopolysaccharide are important macrophage and monocyte stimulators to the synthesis of inflammatory mediators including TNF. Thalidomide is an anti-inflammatory drug that is largely used in several diseases, such as erythema nodosum leprosum, a complication of leprosy. In the present study, the anti-inflammatory effects of Thalidomide were investigated through data obtained from DNA oligoarrays having peripheral blood mononuclear cells (PBMC) treated with Gram-negative lipopolysaccharide bacterial endotoxin (LPS). Differentially expressed genes were identified using ANOVA modeling and Intensity Dependent Selection of Expression Ratios (ISER) methods. Results showed genes mostly related to neuronal and embryonic development, oncogenesis, cell signaling, division and death.

# Índice

<b>1. Introdução</b>	<b>1</b>
<b>2. Fundamentos teóricos</b>	<b>4</b>
2.1. DNA e expressão gênica	4
2.1.1. Estrutura	4
2.1.2. Transcrição	6
2.1.3. Processamento do RNA	7
3.1.4. Regulação da expressão gênica	9
2.2. Microarranjos de DNA	11
<b>3. Revisão da literatura</b>	<b>16</b>
3.1. Bactérias Gram-negativas e Lipopolissacarídeos	16
3.2. TNF	18
3.2.1. Resposta imune	18
3.2.2. Produção de TNF	19
3.2.3. Sinalização	19
3.2.4. Efeitos celulares do TNF na resposta imune	21
3.2.5. Papéis fisiológicos do TNF	22
3.3 Talidomida	23
3.3.1. Histórico	23
3.3.2. Estrutura	24
3.3.3. Aplicações clínicas	25
3.3.4. Efeitos colaterais	28
3.4. Análise de dados de microarranjos	29
3.4.1. Aquisição da imagem e pré processamento	29
3.4.2. Análises estatísticas	34
<b>4. Bases Metodológicas</b>	<b>36</b>
4.1. Análise de variância de dados de microarranjo	36
4.2. ISER – Seleção de razões de expressão intensidade-dependentes	39
<b>5. Material e Métodos</b>	<b>42</b>
<b>6. Resultados</b>	<b>47</b>
<b>7. Discussão</b>	<b>61</b>
Genes Diferencialmente expressos	64
<b>8. Conclusão</b>	<b>67</b>
<b>9. Referências</b>	<b>69</b>

## Lista de Símbolos

A	Adenina
ANOVA	Análise de variância
C	Citosina
Cy3	Cianina 3
Cy5	Cianina 5
cDNA	DNA complementar
DNA	Ácido desoxirribonucléico
ENL	Eritema nodoso leproso
G	Guanina
IFN	Interferon
IL	Interleucina
ISER	Seleções de razões de expressão intensidade-dependentes
LPS	Lipopolissacarídeo
MAANOVA	Análise de variância de microarranjos
PBMC	Células mononucleares de sangue periférico
RNA	Ácido Ribonucléico
RNA <sub>m</sub>	RNA mensageiro
RNA <sub>r</sub>	RNA ribossomal
RNA <sub>t</sub>	RNA transportador
T	Timina
TNF	Fator de necrose tumoral
U	Uracila



# 1. Introdução

O fator de necrose tumoral (TNF) é uma citocina produzida principalmente por monócitos e macrófagos, e está envolvido na resposta imune do hospedeiro a infecções causadas por vírus, parasitas, fungos e bactérias (CORRAL e KAPLAN, 1999). A importância do TNF e do mecanismo de sinalização de seus receptores na resposta imunológica se tornou mais evidente em decorrência de inúmeros estudos realizados

Apesar do TNF ser crucial para a resposta imune, ele também está envolvido na patogênese de doenças infecciosas e autoimunes. Sabe-se que ele é um potente regulador de cascatas inflamatórias e representa um alvo terapêutico em doenças associadas com altas concentrações desta citocina em tecidos de pacientes. O TNF está relacionado em doenças como a hanseníase, doença inflamatória do intestino, doença de Crohn e diversos tipos de câncer (FRANKS *et al.*, 2004).

Sabe-se que um dos indutores eficazes de TNF são os lipopolissacarídeos (LPS) da membrana externa de bactérias Gram-negativas. Em função disso, o LPS é amplamente utilizado em diversos estudos para estimular mecanismos de inflamação que permitam uma melhor compreensão de infecções e seus tratamentos (SHANNON *et al.*, 2008).

A talidomida (α-N-ftalimidoglutarimide), apesar de suas propriedades teratogênicas e neurotóxicas, é uma droga imunomodulatória e anti-inflamatória que foi inicialmente utilizada como um sedativo e antiemético. O mecanismo de ação anti-inflamatório da talidomida ocorre pela inibição seletiva da produção de TNF a partir do aumento da degradação do RNA mensageiro, inibindo entre outros, a migração de células inflamatórias em direção às lesões e levando a uma diminuição da produção de outras citocinas da cascata inflamatória (SAMPAIO *et al.*, 2006; OLIVEIRA *et al.*, 1999).

Atualmente, existe um interesse crescente sobre o potencial uso clínico da Talidomida. Presentes aplicações incluem o tratamento de hanseníase, particularmente do eritema nodoso leproso (ENL), desnutrição associada ao HIV/AIDS; uma variedade de condições dermatológicas e algumas desordens imunológicas e angiogênicas (FERNANDEZ-MARTINEZ *et al.*, 2004). Novos análogos estruturais da talidomida que possuem efeitos imunomodulatórios aprimorados, além de maior estabilidade e menos efeitos adversos, têm sido desenvolvidos e avaliados.

O estudo da modulação da expressão de genes tratados com LPS e talidomida abre importantes perspectivas no entendimento de processos fisiopatogênicos envolvidos em doenças inflamatórias e seus tratamentos, além de auxiliar no desenvolvimento de novas drogas que sejam mais eficazes e possuam menos efeitos colaterais.

O desenvolvimento recente da genômica possibilitou a obtenção de novos conhecimentos sobre as doenças através da compreensão dos mecanismos básicos de funcionamento da célula normal, diagnóstico e desenvolvimento de novas drogas. Há alguns anos atrás, informações sobre expressão gênica eram somente obtidas com base em um gene de cada vez, tipicamente através da análise por *Northern Blotting* (REUE, 1998). Entretanto, a introdução de técnicas de hibridização em arranjos de DNA agora permite rapidamente rastrear a expressão de milhares de genes simultaneamente, gerando uma enorme quantidade de informação sobre interações e funções gênicas.

A tecnologia de microarranjo é uma ferramenta poderosa para a identificação de novos alvos moleculares e para a elucidação de mecanismos de ação de drogas. Adicionalmente, microarranjos permitem analisar o perfil global de expressão gênica em resposta a agentes farmacológicos específicos, fornecendo informações sobre eficácia e toxicidade de drogas (STOUGHTON, 2005).

Considerações estatísticas estão frequentemente em destaque na análise dos dados de microarranjo, uma vez que os pesquisadores têm de lidar com quantidades

massivas de dados e ajustá-los para várias fontes de variabilidade para identificar alguns genes importantes entre os milhares estudados. Recentemente, diversos métodos estatísticos têm sido usados como forma de se extrair informações de dados de microarranjos (KIM *et al.*, 2006). Entre estes métodos estão a análise de variância utilizando o software MAANOVA, e a seleção de razões de expressão intensidade-dependentes (ISER).

O objetivo deste trabalho foi identificar genes que obtiveram expressão diferencial após o tratamento com talidomida em células mononucleares de sangue periférico (PBMC) ativadas por LPS bacteriano empregando a modelagem ANOVA e o método ISER em dados de experimentos de microarranjos de DNA. A partir dos grupos de genes selecionados, espera-se encontrar genes potenciais para explicar os efeitos da Talidomida na resolução de processos inflamatórios.

## **2. Fundamentos teóricos**

### **2.1. DNA e expressão gênica**

As células dos organismos vivos são feitas de diferentes moléculas, tais como água, minerais, proteínas, açúcares, gorduras e ácido desoxirribonucléico (DNA). As proteínas são essenciais à estrutura e à função das células vivas e à regulação das funções celulares tais como transdução de sinal e metabolismo. Apesar da grande diversidade de proteínas que podem ser produzidas em organismos vivos, toda a informação necessária para a síntese dessas proteínas está codificada no DNA, um ácido ribonucléico que carrega instruções genéticas para o desenvolvimento biológico de todas as formas celulares de vida (LOO, 2004).

#### **2.1.1. Estrutura**

A estrutura do DNA se assemelha com uma longa escada retorcida em forma helicoidal ou de bobina. Cada metade é uma cadeia de oligonucleotídeos mantidos juntos por ligações fosfodiéster. As laterais da “escada” são formadas por uma estrutura constituída de desoxirriboses e fosfato e os “degraus” consistem nas bases nitrogenadas fracamente ligadas no meio por pontes de hidrogênio. O DNA é composto apenas de quatro moléculas básicas chamadas nucleotídeos que são idênticas exceto por conterem uma base nitrogenada diferente. . Um nucleotídeo de DNA é formado de uma molécula de desoxirribose, uma molécula de ácido fosfórico e uma das quatro bases nitrogenadas. Estas bases são descritas por letras (A, T, G e C) que correspondem à adenina, timina, guanina e citosina e a sua ordenação constitui o código genético. No pareamento dessas

bases, a adenina sempre se liga com a timina (A-T), e a guanina com a citosina (G-C) (GRIFFITHS *et al.*, 2002).

Embora tanto o ácido ribonucléico (RNA) quanto o DNA sejam ácidos nucléicos, o RNA difere de vários modos importantes. O RNA geralmente é unifilar e não uma dupla-hélice. Uma consequência de ser unifilar é que o RNA pode formar uma variedade muito maior de formas moleculares tridimensionais complexas que o DNA bifilar. O RNA tem a ribose em seus nucleotídeos, em vez de desoxirribose. As duas oses diferem na presença ou ausência de apenas um átomo de oxigênio. Analogamente aos filamentos individuais de DNA, o RNA possui uma estrutura de fosfato-ribose, com uma base covalentemente ligada a posição 1'. Os nucleotídeos do RNA possuem as bases adenina, guanina e citosina, mas a base pirimidínica uracila (U) é encontrada no lugar da timina. Entretanto, a uracila forma pontes de hidrogênio com a adenina, assim como a timina.

A sequência de bases de nucleotídeos no DNA funciona como um código que contém instruções genéticas para a formação de proteínas. Uma sequência de DNA contém milhares de genes, que são trechos específicos da sequência que determina os tipos de proteína que uma célula pode produzir. Um gene é tipicamente constituído de sequências de milhares de bases. Entretanto, o DNA não pode ser convertido em proteína diretamente. Ou seja, o gene envia uma informação, através do RNA mensageiro, que vai acarretar na síntese de proteínas (Figura 1).

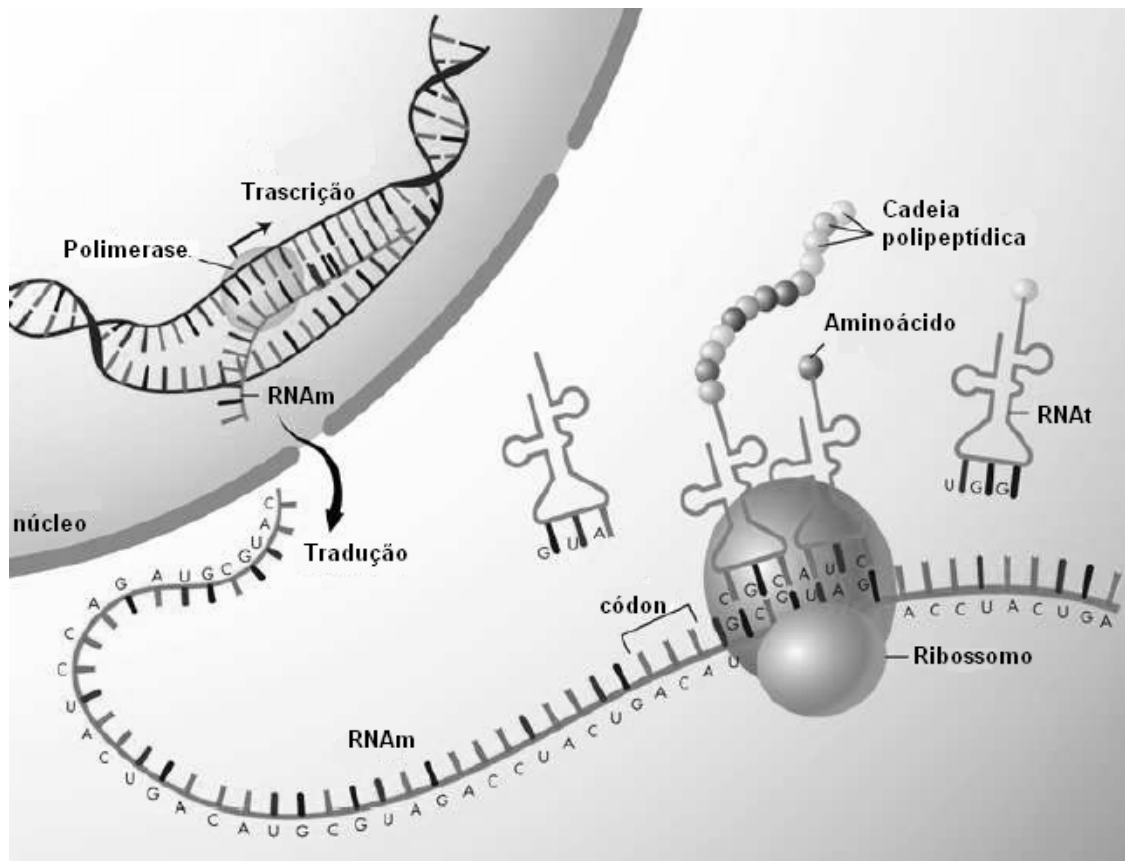


Figura 1: A síntese de proteínas: O RNAm é transcrito pela polimerase, transportado do núcleo, e traduzido pelo ribossomo para produzir proteínas (Imagem do National Human Genome Research Institute).

### 2.1.2. Transcrição

O RNAm é sintetizado no núcleo da célula a partir do DNA e então se desloca para o citoplasma, onde são sintetizadas as proteínas. Esse processo de síntese do RNAm é chamado de transcrição. Os dois filamentos da dupla hélice separam-se localmente, e um dos filamentos separados atua como um molde por meio de sua complementaridade de bases. Dessa forma, o ribonucleotídeo livre A alinha-se com T no DNA, G com C, C com G e U com A. O processo é catalisado pela enzima RNA polimerase, que se liga e se desloca ao longo do DNA adicionando ribonucleotídeos ao RNA crescente no sentido 5' ->

3'. O início da transcrição ocorre quando a RNA polimerase reconhece e se liga especificamente a uma região do DNA chamada de região promotora mais especificamente na sua porção 3'. Em células eucariotas, existem três tipos de RNA polimerase: A RNA polimerase I sintetiza o RNA ribossomal (RNAr), a RNA polimerase II sintetiza moléculas de RNAm que codificam uma proteína e a RNA polimerase III sintetiza RNA transportador (RNAt) (GRIFFITHS *et al.*, 2002; REECE, 2004).

### **2.1.3. Processamento do RNA**

O transcrito primário de RNA produzido no núcleo é geralmente processado de vários modos antes de ser transportado para o citoplasma onde é realizada a tradução. Primeiramente um *cap*, que consiste em uma 7-metilguanossina ligada à extremidade 5' do transcrito por uma ligação trifosfato é adicionado durante a transcrição. Então uma seqüência de adenosinas é adicionada à extremidade 3'. Estas caudas poli(A) têm de 150 a 200 unidades de comprimento. Após estas modificações e etapa de recomposição (*splicing*) remove partes internas do RNA transcrito. A recomposição remove os íntrons (seqüências intercalares que interrompem os segmentos de DNA que codificam a estrutura da proteína) e junta as regiões codificantes, chamadas éxons, para formar um RNAm (GRIFFITHS *et al.*, 2002).

Em seguida, o RNAm é transportado do núcleo da célula até o ribossomo. A função do ribossomo é “ler” a cópia do RNAm do gene e produzir a proteína apropriada (LOO, 2004). Na leitura de uma molécula de RNAm a partir de uma determinada extremidade cada três bases (códon) codificam um aminoácido. Esta seqüência é lida a partir de um ponto fixo e continua até o final da seqüência codificante. Cada aminoácido é trazido ao ribossomo para se ligar ao peptídeo em crescimento por meio de um ácido ribonucléico denominado RNAt . A molécula de RNAt possui vários sítios funcionais como

o sítio de ligação do aminoácido e o sítio que reconhece um códon no RNAm, chamado de anticódon, no qual suas bases são complementares e se pareiam com as bases do códon. Embora as moléculas de RNAt tenham muitas similaridades estruturais, cada uma tem uma forma tridimensional única, que permite o reconhecimento pela sintetase correta, a qual catalisa a união de um RNAt com seu aminoácido específico para formar um aminoacil-RNAt (um RNAt levando um único aminoácido) (GRIFFITHS *et al.*, 2002).

Os ribossomos contêm sítios específicos que possibilitam se ligar ao RNAm, RNAts e aos fatores protéicos específicos necessários à síntese de proteínas. O RNAm liga-se à subunidade 30s. Os RNAts ligam-se a dois sítios no ribossomo. O sítio A é o sítio de entrada de um aminoacil-RNAt. O peptidil-RNAt levando a cadeia polipeptídica em crescimento liga-se ao sítio P. Cada novo aminoácido é adicionado pela transferência da cadeia em crescimento para o novo aminoacil-RNAt, formando uma nova ligação peptídica e assim por diante (GRIFFITHS *et al.*, 2002).

O término da tradução ocorre quando o códon a ser lido é um códon finalizador ou terminal e, conseqüentemente, essa trinca não é reconhecida por nenhum RNAt e sim por fatores de liberação. O polipeptídeo é então liberado no sítio P e os ribossomos se dissociam em duas subunidades, finalizando a síntese da proteína (GRIFFITHS *et al.*, 2002). As proteínas citosólicas solúveis são simplesmente liberadas do ribossomo após o término da síntese do polipeptídeo, e já estão no local correto para assumir suas funções específicas. Porém, muitas proteínas eucariotas são destinadas para um determinado compartimento da célula e é o próprio polipeptídeo que codifica esta informação para o seu destino final. As proteínas que são destinadas ao núcleo contêm um sinal de localização nuclear. As proteínas que são destinadas a serem exportadas da célula, ou incorporadas em membranas celulares contêm um trecho aminoterminal de aminoácidos predominantemente hidrofóbicos chamado de sequência de sinal que é responsável pelo seu transporte para o retículo endoplasmático durante a tradução (REECE, 2004).



### **3.1.4. Regulação da expressão gênica**

O genoma é o conjunto de todo o DNA de um organismo vivo. Quase todas as células do organismo contêm a mesma cópia do genoma. O genoma humano codifica cerca de 30.000 genes, entretanto, a maioria dos genes está inativa na maior parte do tempo. Somente uma pequena porção dos genes está transcrevendo ativamente RNAm e a célula precisa ter a habilidade de ativar ou desativar certos genes em resposta a sinais externos. O nível total acumulado do RNAm de um gene corresponde ao nível de expressão deste (LOO, 2004; REECE, 2004).

Os níveis de proteínas podem ser regulados por mecanismos que operam na transcrição e pós-transcricionalmente. Para obter as taxas máximas de transcrição pela RNA polimerase II é necessária a cooperação de múltiplos elementos reguladores de ação *cis*. Próximos ao sítio de início da transcrição estão as sequências de ação *cis* proximais ao promotor que se ligam às proteínas que ajudam na ligação do promotor com a RNA polimerase II. Os sítios acentuadores são sequências de ação *cis* que ligadas a proteínas ativadoras podem aumentar muito as taxas de transcrição dos promotores na mesma molécula de DNA. Já os sítios silenciadores são sequências *cis* ligadas a proteínas repressoras, inibindo assim as ativadoras e reduzindo a transcrição. Os acentuadores e silenciadores podem regular a transcrição a grandes distâncias do promotor. Os elementos proximais ao promotor e os elementos independentes de distância são sítios reconhecidos por proteínas de ação *trans* que se ligam a sequências específicas do DNA, fazendo com que a RNA polimerase II inicie e atinja as taxas máximas de transcrição (GRIFFITHS *et al.*, 2002).

Alguns fatores de transcrição, como os acentuadores, são sintetizados apenas em tecidos específicos e as atividades dos fatores em si também são reguladas em tipos celulares diferentes. A atividade fisiológica está intimamente relacionada com a regulação transcricional, ou seja, as vias de diferenciação e padrão de desenvolvimento fornecem o contexto fisiológico para a ativação de fatores de transcrição específicos. Um sinal regulador que ativa os fatores de transcrição podem vir, por exemplo, de um órgão do sistema endócrino, para liberar hormônios no sistema circulatório. No caso, o sistema endócrino serve como um regulador para coordenar mudanças na transcrição em células de tecidos diferentes (GRIFFITHS *et al.*, 2002).

Além das proteínas e fatores de transcrição, a regulação da expressão gênica também inclui as mudanças reversíveis e herdáveis no genoma funcional, que não alteram a seqüência de nucleotídeos do DNA (ALBERTS *et al.*, 2004). As alterações epigenéticas explicam como os padrões de expressão são passados para os descendentes, como ocorre a mudança de expressão de genes durante a diferenciação de um tipo celular e como fatores ambientais podem mudar a maneira como os genes são expressos. Os três mecanismos principais de alterações epigenéticas são a metilação do DNA, modificações de histonas e ação de RNAs não codificadores (BIRD, 2002).

Teoricamente, todas as mudanças em um estágio ou tipo de célula podem ser correlacionadas com alterações dos níveis de RNAm dos genes. As variações de um ambiente celular podem afetar um grande número ou apenas um pequeno conjunto de genes. O conhecimento dos padrões de expressão de diversos genes ainda não caracterizados pode também fornecer pistas fundamentais de sua função. A análise das mudanças na expressão de todos os genes constituintes de um genoma é essencial para entender a interação entre os genes e seus produtos (REECE, 2004).

## 2.2. Microarranjos de DNA

Com a crescente quantidade de informação genômica disponível oriunda de sequenciamentos automatizados, novos métodos são requeridos para se extrair conhecimento dessa informação em alta velocidade. Os microarranjos de DNA foram desenvolvidos como um método de análise para milhares de genes simultaneamente e foram primeiro descritos em um estudo de expressão envolvendo 45 genes de *Arabidopsis spp.*, desenvolvido por SCHENA *et al.* (1995), onde a expressão diferencial foi medida por meio de fluorescência de hibridização de duas cores. No ano seguinte, DE RISI *et al.* (1996) aplicaram essa tecnologia no estudo de 1.661 genes para pesquisar diferenças na expressão gênica em melanoma.

Há décadas, a hibridização serve como base para técnicas tais como o *Southern Blotting* e o *Northern Blotting*. No *Southern Blotting*, pequenos fragmentos desnaturados de DNA, digeridos por enzimas de restrição e separados por tamanho por eletroforese em gel, são transferidos para uma membrana, marcados radioativamente e hibridizados (SOUTHERN, 1975). Já no *Northern Blotting*, um oligonucleotídeo marcado por radiação é utilizado na hibridização com RNA mensageiro (RNAm), também previamente separado por gel de eletroforese (REUE, 1998).

A técnica de microarranjos é basicamente o inverso das técnicas de *Southern* e *Northern Blotting*. Nela, é fornecido um filamento de DNA ao qual podem se ligar fragmentos marcados de DNA complementar (cDNA). Os fragmentos de DNA são fisicamente fixados em um suporte inerte, chamado de chip (REECE, 2004).

Diversas tecnologias diferentes são atualmente utilizadas para realizar experimentos de microarranjos. Estas diferem no comprimento da sequência de DNA e na forma que estas sequências são fixadas ao chip. Os dois sistemas mais comuns utilizados

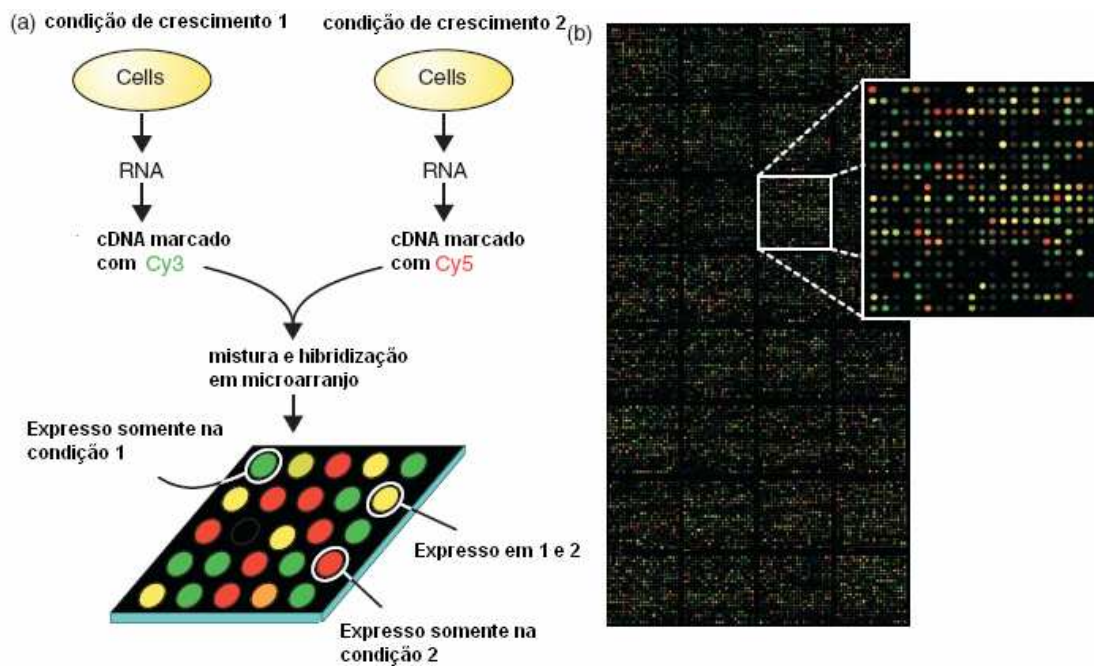
são o desenvolvido na Universidade de Stanford (CA, EUA) e o vendido comercialmente pela companhia *Affymetrix*.

Os chips *Affymetrix* contém oligonucleotídeos de filamento único de DNA quimicamente sintetizados (comprimento de cerca de 25 bases) que são fixados a uma superfície de quartzo usando agentes fotolábeis e técnicas fotolitográficas semelhantes às utilizadas na produção de microchips de silicone. Esse processo pode resultar em 500.000 moléculas individuais de DNA posicionadas precisamente em um chip de 1,28 cm<sup>2</sup>.

Os chips desenvolvidos por Stanford são desenvolvidos por deposição robótica de fragmentos de DNA, derivados de amplificação por PCR de genes inteiros, em locais precisos de uma lâmina de vidro medindo aproximadamente 2 cm<sup>2</sup>. Cada produto de PCR é depositado em um *spot* com localização precisa de modo que seja conhecida a sequência de DNA de qualquer *spot*. Após a deposição na lâmina os DNAs são secos por aquecimento a 100 °C por dois segundos, fixados por UV e desnaturados a 95 °C por dois minutos. Os chips preparados são então usados como moldes para a ligação de fragmentos cDNA marcados. São isoladas as amostras de RNA de dois grupos de células geralmente relacionadas, de modo que a maior parte dos genes que elas expressem seja semelhante. As células, no entanto, precisam ter crescido sob diferentes condições, ou derivadas de duas condições diferentes de mesmo tecido. Filamentos de cDNA são sintetizados usando a transcriptase reversa de um iniciador oligo (dT) na presença de três trifosfatos desoxinucleotídeos (dNTPs) normais e um único trifosfato desoxinucleotídeo marcado fluorescentemente. Uma das amostras de cDNA é produzida usando um marcador fluorescente excitado com laser, e emitido na faixa de intensidade visível verde (até 570 nm, nucleotídeo marcado com Cy3), enquanto a outra é produzida usando um marcador fluorescente excitado com laser, e emitido na faixa de intensidade visível vermelha (até 670 nm, nucleotídeo marcado com Cy5), de forma que o cDNA isolado de

cada condição ou tipo celular diferente vá possuir um marcador fluorescente diferente. As amostras de cDNA são então misturadas em quantidades iguais e deixadas hibridizar ao chip de microarranjo (KERR *et al.*, 2000; REECE, 2004). Após a hibridização, as moléculas que não estão pareadas ou que apresentam pareamento fraco são retiradas através de lavagens sucessivas que não afetam o material corretamente pareado com os fragmentos de DNA marcados (AFORNALI, 2006). A seguir, a leitura da lâmina por um scanner microscópico confocal revela para cada *spot* a intensidade de hibridização referente à quantidade relativa de RNA de um determinado gene no tecido estudado por meio da medida da fluorescência emitida de cada marcador excitado por laser separadamente, e essas medidas são usadas para se determinar a razão entre os canais (REECE, 2004).

Em um exemplo (Figura 2) em que as células crescidas sob a condição 1 fossem marcadas com Cy3 e expressassem um gene que não fosse expressado sob a condição 2, o cDNA misturado somente conteria a versão marcada com Cy3 do cDNA correspondente e o *spot* referente no microarranjo iria emitir fluorescência no comprimento de onda do verde. Similarmente, um gene somente expresso na condição de crescimento 2, no exemplo marcado com Cy5, iria gerar no arranjo um *spot* que emita fluorescência no comprimento de onda do vermelho. Ainda, os genes com mesmo nível de expressão em ambas as amostras, resultariam em *spots* contendo cDNAs marcados com Cy3 e Cy5 que poderiam se ligar à sequência complementar do arranjo. Nesse caso, seria gerado um *spot* emitindo fluorescência no comprimento de onda amarelo, mistura de verde e vermelho (REECE, 2004).



**Figura 2: Identificação de genes cuja expressão difere entre duas amostras de RNA usando microarranjos. (a) O RNA de células crescidas em uma condição 1 é isolado e convertido em cDNA usando um iniciador marcado com um corante que emite fluorescência no espectro visível verde (Cy3) enquanto o cDNA obtido a partir do RNA de células da condição 2 usa um iniciador marcado com outro corante que emite fluorescência no espectro visível vermelho (Cy5). As amostras de cDNA são misturadas em proporções iguais e hibridizadas no microarranjo. Visualizando o microarranjo com um microscópio de fluorescência são revelados uma série de *spots* coloridos. Um *spot* verde indica que o gene correspondente é predominantemente expresso na condição 1, enquanto um *spot* vermelho indica que um gene é predominantemente expresso na condição 2. Se quantidades iguais de cDNA derivado das células das condições 1 e 2 são hibridizadas no arranjo, então o *spot* correspondente será amarelo (Figura adaptada de REECE, 2004).**

Microarranjos de DNA permitiram que os pesquisadores conduzissem experimentos quantitativos em larga escala. Onde antes era possível realizar um par de *Northern Blots*, ou de *Southern Blots*, agora, de uma única vez, realizam-se, dezenas de milhares de hibridizações. O modo dominante na utilização da técnica de microarranjos é a sua aplicação na monitoração de mudanças na expressão de genes em determinado tecido ou tipo celular, em estágios de doenças, durante o desenvolvimento e em resposta a perturbações experimentais, como tratamentos com drogas. Os padrões de resposta

têm ajudado a esclarecer os mecanismos de doenças; identificar sub-fenótipos destas além de prever sua progressão; atribuir funções a determinados genes, agrupá-los e incluí-los em vias funcionais. Os microarranjos têm sido usados na identificação de novos genes, sítios de ligação de fatores de transcrição, mudanças no número de réplicas de DNA e variações em relação a uma sequência inicial, como mutações complexas causadas por doenças em genes humanos (STOUGHTON, 2005). Além disso, podem ser utilizados na construção de grupos hierárquicos de genes que compartilham padrões de expressão e que, portanto, devem ser ativados a partir dos mesmos estímulos. A correlação entre os genes de um mesmo grupo e determinadas vias metabólicas ou condições biológicas é uma forma de interpretar suas funções. Quando isto não é possível, a melhor alternativa é escolher alguns destes genes como alvos para investigação de sua estrutura e função (AFORNALI, 2006).

## **3. Revisão da literatura**

### **3.1. Bactérias Gram-negativas e Lipopolissacarídeos**

As bactérias podem ser classificadas em dois grupos principais através da coloração de Gram: Gram-positivas e Gram-negativas. A reação das bactérias à técnica de Gram expressa diferentes características, no que diz respeito à composição química, estrutura, permeabilidade da parede celular, fisiologia, metabolismo e patogenicidade.

A parede da célula Gram-negativa é constituída por estruturas de múltiplas camadas bastante complexas, que não retêm o corante quando submetidas a solventes nos quais o corante é solúvel, sendo descoradas e, quando acrescentado outro corante, adquirem a nova coloração. Já a parede da célula Gram-positiva consiste de única camada que retém o corante aplicado, não adquirindo a coloração do segundo corante.

Nas bactérias Gram-negativas, a parede celular está composta por uma camada de peptidoglicano e três outros componentes que a envolvem externamente: lipoproteína, membrana externa e lipopolissacarídeo.

O peptidoglicano, responsável pela forma das células e proteção do citoplasma frente às diferenças de pressão osmótica entre os meios externo e interno, confere rigidez ao corpo bacteriano. Está formado por dois sacarídeos aminados: o ácido N-acetil glicosamina e o ácido N-acetil murâmico, e por um tetrapeptídeo, sempre ligado ao resíduo de ácido N-acetil murâmico; as subunidades peptídicas de cadeias glicídicas adjacentes são unidas entre si por ligações diretas ou indiretas (pontes de ligação). O peptidoglicano situa-se no espaço periplásmico, localizado entre a membrana citoplasmática (interna) e a membrana externa. No espaço periplásmico também são encontradas enzimas hidrolíticas que facilitam a nutrição bacteriana (fosfatases,



nucleases, proteases e outras), proteínas de ligação, que participam da captação de açúcares e aminoácidos a partir do meio, e enzimas que inativam certos antibióticos.

A lipoproteína está ligada de modo covalente ao peptidoglicano e não covalente à membrana externa; sua função, inferida de estudos realizados com amostras mutantes, é estabilizar a membrana externa e ancorá-la à camada de peptidoglicano. A membrana externa é uma dupla camada, contendo fosfolípidos e proteínas e apresentando, em sua camada externa, o lipopolissacarídeo.

Entre suas funções, a membrana externa representa uma barreira molecular que previne ou dificulta a perda de proteínas periplasmáticas e o acesso de enzimas hidrolíticas e certos antibióticos ao peptidoglicano; possui receptores para bacteriófagos e bacteriocinas e participa da nutrição bacteriana. O lipopolissacarídeo consiste no lípido A (endotoxina), ao qual estão ligadas duas regiões de natureza polissacarídica, respectivamente, o “core” e as cadeias laterais. O lípido A é um glicofosfolípido cujo papel biológico consiste na participação nos mecanismos de patogenicidade da célula bacteriana.

O LPS é o maior fator de virulência das bactérias Gram-negativas, determinando efeitos biológicos que resultam na amplificação das reações inflamatórias. Esta endotoxina é um antígeno fraco não específico que é pobremente neutralizado por anticorpos, sendo capaz de ativar a cascata do sistema complemento além de mastócitos, basófilos e células endoteliais. O LPS induz os macrófagos a secretarem outras proteínas, as interleucinas (IL-1, IL-6 e IL-8), TNF, intermediários reativos de oxigênio, óxido nítrico (NO), interferon  $\alpha$ ,  $\beta$  e  $\gamma$ , fatores ativadores de plaquetas e prostaglandinas. Mesmo quantidades pequenas de endotoxinas são capazes de induzir a resposta inflamatória. Uma possível explicação para a multiplicidade de achados com endotoxinas é a variabilidade genética do LPS de diferentes microrganismos. As endotoxinas são encontradas em maior quantidade em doentes sintomáticos que naqueles assintomáticos.

## 3.2. TNF

### 3.2.1. Resposta imune

As citocinas participam em diversos processos fisiológicos incluindo a regulação das respostas imune e inflamatória. Essas moléculas controlam a amplitude e duração da resposta cujos intermediários são os linfócitos T *helper*. Estas células se proliferam após o contato com um antígeno e ativam outros tipos de células que agirão de maneira mais direta (FRANKS *et al.*, 2004).

Os linfócitos T *helper* são subdivididos em Th1 e Th2, com atividades imunorreguladoras específicas, que são mediadas por citocinas específicas. A subpopulação Th1 produz as citocinas IL-2 (interleucina 2), Interferon gama (IFN- $\gamma$ ) e TNF, responsáveis pela ativação e manutenção da resposta imune mediada por células, efetivas contra infecções intracelulares microbianas e virais. Os linfócitos Th2 produzem as citocinas IL-4, IL-5, IL-6, IL-8 e IL-10, responsáveis pela ativação e manutenção da resposta imune humoral, efetiva contra infecções extracelulares. Dependendo da subpopulação de células T *helper* estimuladas no início do processo inflamatório, haverá predominância de mecanismos de defesa ou de disseminação da infecção correlacionada (FOSS, 1997).

A IL-2, entre outros sinais específicos, ativa os linfócitos T, estimulando a formação de clones celulares, responsáveis pela manutenção da produção de citocinas como a IL-1 e IL-12 e, paralelamente, ativam células *Natural Killer* (NK), importante fonte inicial de IFN- $\gamma$  no sítio inflamatório. As IL-1, IL-12 e IFN- $\gamma$  agem sobre macrófagos, aumentando a capacidade de fagocitose e mecanismos de ativação celular como a

estresse oxidativo (FOSS, 1997), que levam à produção de intermediários reativos de oxigênio (ROI), tal como o peróxido de hidrogênio (H<sub>2</sub>O<sub>2</sub>) e do nitrogênio (RNI), por sua vez, óxido nítrico (NO), e outros que levam ao aumento da produção de TNF, que potencializa a ativação macrofágica atuando através de um mecanismo sinérgico cíclico.

O TNF foi identificado em 1975 como uma glicoproteína induzida por endotoxinas. Foi implicado em diversas condições inflamatórias, infecciosas e malignas, e sua importância na inflamação foi destacada pela eficácia de anticorpos anti-TNF ou pela administração de receptores solúveis de TNF no controle de doenças como artrite reumatóide e outras condições inflamatórias (BRADLEY, 2008).

### **3.2.2. Produção de TNF**

O TNF é produzido predominantemente por monócitos e macrófagos ativados, mas também por diversas células incluindo mastócitos, linfócitos T e B, células *Natural Killer*, neutrófilos, células endoteliais, células musculares lisas e cardíacas, fibroblastos e osteoclastos. É sintetizado como uma proteína pró-TNF de 26 kDa, a qual é expressa na membrana plasmática, e clivada no domínio extracelular, resultando na liberação da sua forma solúvel de 17 kDa. O TNF não é geralmente detectado em indivíduos saudáveis, porém são encontrados níveis séricos e teciduais elevados em condições inflamatórias e infecciosas e esses altos níveis são relacionados com a gravidade das infecções (BRADLEY, 2008).

### **3.2.3. Sinalização**

Todas as respostas celulares induzidas por TNF são mediadas pela sua ligação com um dos seus dois receptores TNFR1 (também conhecido como TNFRSF1A, CD120a

ou p55) e TNFR2 (também conhecido como TNFRSF1B, CD120b, p75), ambos pertencentes à superfamília de receptores de TNF e que são regulados de forma diferente em vários tipos celulares em tecidos normais e afetados. Quase todos os tipos celulares expressam um dos dois tipos de receptores de TNF (LIU e HAN, 2001). Apesar de não apresentarem homologia significativa em seus domínios intracelulares, a habilidade do TNFR1 e TNFR2 de interagir tanto com moléculas idênticas quanto não relacionadas entre si pode explicar suas funções compartilhadas e diversas. Com base em estudos com cultura de células e camundongos *knockout* para esses receptores, as vias pró-inflamatórias e de morte celular programada que são ativadas por TNF, e associadas com danos teciduais, são mediadas por TNFR1. As conseqüências da sinalização de TNFR2 não são tão bem caracterizadas, mas o TNFR2 tem mostrado mediar sinais que promovem reparo tecidual e angiogênese (BRADLEY, 2008). Há evidências de que a ligação do TNFR1 desencadeie a apoptose em diversas células diferentes, no entanto, o efeito pró-apoptótico de TNFR2 somente foi encontrado em algumas circunstâncias. Dependendo do tipo de célula alvo, a morte celular induzida por TNF poderia ser necrótica ou apoptótica (LIU e HAN, 2001).

As vias de transdução de sinal no TNF são complexas e ainda não foram completamente elucidadas, porém já se sabe que estão relacionadas com a regulação do fator de transcrição NF- $\kappa$ B (BRADLEY, 2008). O NF- $\kappa$ B é um dos fatores de transcrição chave que mediam muitas respostas celulares mediadas por TNF. Na maior parte das células, a atividade do NF- $\kappa$ B pode ser potencialmente elevada pelo tratamento com TNF (LIU e HAN, 2001).

Apesar do TNF ter sido nomeado dessa forma devido à sua habilidade de causar regressão tumoral, ele somente destrói seletivamente alguns tipos de células. Sabe-se hoje que uma das razões para essa ineficiência é a ativação de NF- $\kappa$ B em resposta ao

tratamento de TNF. Diversos estudos demonstraram que a ativação de NF-kB protege a célula contra apoptose induzida por TNF (LIU e HAN, 2001).

O efeito do TNF não é devido somente à sua citotoxicidade em certos tipos de células, mas também é uma consequência da indução de genes por esta citocina. O número de genes que podem ser regulados por estimulação de TNF é desconhecido, mas se sabe que quase todas as citocinas pró-inflamatórias são induzidas por sua estimulação. Outras moléculas, tais como proteases, que estão envolvidas em doenças inflamatórias são diretamente ou indiretamente induzidas por TNF *in vivo* (LIU e HAN, 2001).

#### **3.2.4. Efeitos celulares do TNF na resposta imune**

Apesar dos receptores do TNF apresentarem expressão diferencial em uma variedade de células e tecidos, diversos efeitos pró-inflamatórios do TNF podem ser explicados com base nos seus efeitos no endotélio e suas interações com leucócitos. Em resposta ao TNF, células endoteliais promovem inflamação por mostrarem diferentes combinações de moléculas de adesão para leucócitos, incluindo e-selectina, molécula de adesão intercelular 1 (ICAM-1) e molécula de adesão de célula vascular 1 (VCAM-1) (BRADLEY, 2008). Em combinação com a liberação de quimiocinas (incluindo IL-8, MCP-1 e IP-10), estas respostas levam ao recrutamento de diferentes populações de leucócitos independentemente do reconhecimento do antígeno. Além disso, diversas características de inflamação podem ser produzidas por efeitos locais do TNF nas células endoteliais. A expressão de cicloxigenase 2 induzida por TNF pode resultar em vasodilatação, causando rubor e calor através do aumento da corrente sanguínea. A formação de edemas pode ser resultante de um aumento na permeabilidade vascular mediada por TNF, permitindo a passagem transendotelial de fluido e macromoléculas.

Adicionalmente, a expressão de proteínas coagulantes induzida por TNF, tal como o fator tecidual e a redução de proteínas anticoagulantes, tais como a trombomodulina, podem causar trombose intravascular (BRADLEY, 2008, MARK *et al.*, 2001).

### **3.2.5. Papéis fisiológicos do TNF**

Um dos maiores papéis fisiológicos do TNF é a defesa do hospedeiro a infecções bacterianas, virais e por parasitas. Fisiologicamente, o TNF é importante para respostas normais a infecções, porém sua produção inapropriada ou em excesso pode ser prejudicial. O TNF foi originalmente identificado como um mediador humoral de endotoxina induzida por lipopolissacarídeo da caquexia, síndrome que compreende anorexia, perda de peso e proteínas, a qual que complica diversos tipos de câncer, inflamações e infecções crônicas. O TNF é o principal mediador do efeito letal de endotoxina derivada de *Escherichia coli* e está relacionado com a formação de granulomas, os quais limitam a disseminação de bactérias e outras infecções. Diversas infecções granulomatosas possuem associação com o uso de antagonistas de TNF para tratar doenças inflamatórias humanas (BRADLEY, 2008).

O TNF possui um papel chave em infecções por parasitas e vírus. Seus níveis séricos estão aumentados em pacientes de malária como uma resposta normal do hospedeiro ao *Plasmodium falciparum*, porém sua produção excessiva pode levar à forma cerebral da doença e levar a morte. Seu papel nas defesas do hospedeiro contra vírus foi comprovado pela existência de vírus codificantes de proteínas de ligação de TNF em seu genoma (BRADLEY, 2008).

O TNF pode em algumas circunstâncias contribuir para carcinogênese ao promover proliferação, invasão e metástase de células tumorais. No entanto, perfusões

isoladas com TNF em pacientes com metástases hepáticas, melanomas e sarcomas metastásicos vêm se mostrando eficazes (BRADLEY, 2008).

### **3.3 Talidomida**

#### **3.3.1. Histórico**

A molécula de talidomida foi sintetizada em 1953 na Alemanha. Como as propriedades sedativas da talidomida não apresentaram qualquer efeito colateral em pesquisas com animais, a droga foi então considerada “não-tóxica” e começou a ser distribuída em 1955 como sedativo com indicação na prevenção de crises em pacientes com epilepsia (SILVERMAN, 2002). A talidomida foi então utilizada em diversos países como sedativo e antiemético para enjoos matinais e proibida em 1961, quando seu efeito teratogênico em humanos foi comprovado (FRANKS *et al.*, 2004).

Em 1963, o médico israelense Jacob Sheskin, na tentativa de sedar um paciente crítico com eritema nodoso leproso administrou dois comprimidos de talidomida, fazendo finalmente o paciente dormir por horas e acordar com melhoras significativas de seus sintomas e sinais clínicos. Este resultado foi seguido de outras experiências favoráveis e de uma grande investigação que revolucionou o tratamento da hanseníase, (SILVERMAN, 2002, BORGES e FRÖEHLICH, 2003, TEO *et al.*, 2002). Após décadas de tratamento desta descoberta, nenhum outro medicamento foi tão eficaz quanto a talidomida, que ainda é a droga recomendada no tratamento do ENL (TEO *et al.*, 2002).

Em 1998 o FDA (*Federal Drug Administration*, EUA.) aprovou o pedido de distribuição e comercialização de talidomida na indicação do tratamento do ENL.(TEO *et al.*, 2002, BORGES e FRÖEHLICH, 2003). A utilização da talidomida no tratamento do mieloma múltiplo, por sua vez, foi aprovada em 2003. A talidomida, associada com a

dexametasona é atualmente o procedimento padrão no tratamento de mieloma múltiplo (GORDON e GOGGIN, 2003; MERCHIONNE *et al.*, 2007).

### 3.3.2. Estrutura

A molécula de talidomida (N-alfa-ftalimido-glutarimida -  $C_{13}H_{10}N_2O_4$  - Figura 3) é um derivado sintético do ácido glutâmico, cuja síntese é composta por duas etapas: Na primeira, o ácido glutâmico é condensado juntamente ao anidrido ftálico. Em seguida, este composto é condensado com amônia em temperatura elevada (BORGES E FRÖEHLICH, 2003). Sua estrutura é composta por um anel de ftalimida à esquerda e um anel de glutarimida à direita, apresentando um átomo de carbono assimétrico na posição 3' da estrutura glutarimida. Dessa forma, apresenta-se como uma mistura racêmica das duas formas opticamente ativas, a R(+) e a S(-)-talidomida (MUJAGIC *et al.*, 2002). Apesar de alguns trabalhos sugerirem que a forma S(-) inibe potencialmente a liberação de TNF, enquanto que a forma R(+) parece ser responsável pelo efeito sedativo, provavelmente mediado por receptores do sono, estes dados não podem ser confirmados, uma vez que a separação dos enantiômeros é pouco eficaz por ocorrer rápida interconversão espontânea entre estes (BORGES E FRÖEHLICH, 2003, FRANKS *et al.*, 2004).

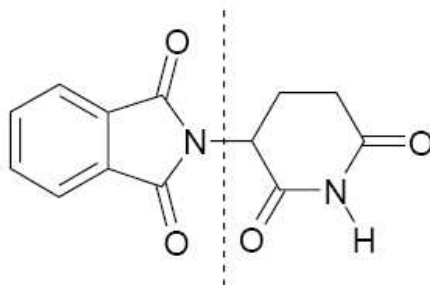


Figura 3: Estrutura da Talidomida. O anel de ftalimida está à esquerda e a glutarimida à direita.



### **3.3.3. Aplicações clínicas**

A talidomida possui diversos efeitos biológicos devido à sua habilidade de interferir com o sistema imune dependendo do tipo de célula e via de ativação. A inibição da síntese de TNF foi o primeiro efeito molecular descrito que poderia ser considerado para os muitos efeitos imunológicos da droga. Subsequentemente, descobertas adicionais das ações da talidomida levaram a várias novas aplicações em patologias (SAMPAIO, 2006).

O eritema nodoso leproso (ENL) é uma vasculite aguda vista em pacientes da forma multibacilar da hanseníase. É caracterizado por ulcerações nodulares dolorosas e sintomas sistêmicos como febre, atralgia, neurite e glomerulonefrite. Imunologicamente, envolve componentes tanto de resposta imune mediada por células com aumento de IFN- $\gamma$ , TNF, e IL-12, e de doença mediada pelo complexo imune. (GORDON e GOGGIN, 2003). Sua eficácia parece ser devido aos seus efeitos inibitórios de citocinas com resultados que mostram uma redução significativa nos níveis séricos de TNF (GORDON e GOGGIN, 2003). As lesões cutâneas granulomatosas causadas por sarcoidose, similares às do ENL, foram também controladas de maneira eficaz (FRANKS *et al.*, 2004).

A talidomida provou ser um agente antiprurítico eficaz em diversas desordens reumatológicas. Ela foi utilizada com sucesso no tratamento de prurido urêmico em pacientes submetidos à hemodiálise, e no tratamento do prurigo. Também tem sido usada com sucesso no tratamento de outras condições dermatológicas resistentes a terapias convencionais incluindo porfiria cutânea tardia, lúpus discóide, prurigo actínico, histiocitose cutânea X, e pioderma gangrenoso (GORDON e GOGGIN, 2003).

A talidomida mostrou ser eficaz no tratamento de ulceração aftosa oral resistente e ulceração aftosa associada ao HIV. Na doença de Behçet, diversos estudos mostraram que a talidomida é eficaz no tratamento e na prevenção da recorrência de ulceração

urogenital (GORDON e GOGGIN, 2003). Tipicamente, lesões orais são curadas em 3 a 4 semanas, porém recorrências são comuns após a interrupção da terapia (FRANKS *et al.*, 2004).

A doença enxerto versus hospedeiro (GvHD) geralmente afeta predominantemente a pele, trato gastrointestinal, fígado e sistema imune, e pode ser bastante debilitante após transplante de medula óssea. A talidomida parece ser mais útil como um adjunto à terapia imunossupressiva padrão do GvHD do que como tratamento único (FRANKS *et al.*, 2004).

O efeito da talidomida em dor articular relacionada com artrite reumatóide refratária parece ser eficaz, principalmente em um tratamento combinado com metotrexato e ou pentoxifilina (FRANKS *et al.*, 2004). Além disso, há diversos casos de pacientes com sarcoidose refratária, escleroderma e lúpus cutâneo que apresentaram resposta ao tratamento com talidomida (GORDON e GOGGIN, 2003).

O potencial da talidomida como tratamento de doenças inflamatórias intestinais foi primeiramente reportado em 1979 após sua utilização com sucesso em um paciente com colite ulcerativa severa. Após a descoberta do papel do TNF na patogênese da Doença de Crohn, novos tratamentos a partir de talidomida foram desenvolvidos. Estudos clínicos indicaram que pacientes de Doença de Crohn das formas resistentes ou refratárias a esteróides foram tratados com êxito com doses orais de talidomida (GORDON e GOGGIN, 2003). A talidomida também pode ser benéfica em pacientes refratários ao Infliximab, ou como um adjunto desta droga (FRANKS *et al.*, 2004).

Desnutrição e caquexia associadas com doença crônica são conhecidas por serem mediadas em parte através de uma resposta imune Th1 e produção acentuada de citocinas proinflamatórias tais como TNF, IL-1 $\beta$ , e IL-6. Essas condições são associadas com uma redução da expectativa de vida em pacientes de câncer. A talidomida mostrou ser capaz de retardar ou reverter a perda de peso em casos de desnutrição associada ao

HIV, tuberculose pulmonar ativa e desnutrição associada à malignidade esofageal (GORDON e GOGGIN, 2003).

A angiogênese é importante na progressão de malignidades hematológicas tais como leucemia, linfoma, mielodisplasia e mieloma. Em função disso, há um interesse considerável na utilização de drogas como a talidomida que possuem propriedades imunomodulatórias e antiangiogênicas. Os resultados mais promissores têm ocorrido no tratamento do mieloma múltiplo no qual a talidomida parece auxiliar na redução dos níveis plasmáticos e de medula óssea de fator de crescimento de fibroblastos, fator de crescimento vascular endotelial e TNF (GORDON e GOGGIN, 2003). A talidomida atualmente é recomendada como tratamento padrão para pacientes com mieloma múltiplo refratário em estado avançado em combinação com dexametasona ou quimioterapia citotóxica. Ela também pode também ser eficaz no tratamento de mielodisplasia de alto risco, leucemia mielóide crônica e outras desordens microproliferativas (FRANKS *et al.*, 2004).

Estudos relataram a eficácia do tratamento com talidomida para sarcoma de Kaposi, carcinoma de células renais e glioma de alto grau. A talidomida apresentou um pequeno efeito benéfico em pacientes com câncer de mama e melanoma avançado. No tratamento de câncer colorretal, a talidomida eliminou os efeitos colaterais do agente quimioterápico Irinotecan, apesar de ocorrência de diversos relatos de doenças tromboembólicas em pacientes avançados de câncer tratados com talidomida. Dessa forma, mesmo possuindo aparentemente alguma atividade contra tumores sólidos, a sua relação custo/benefício ainda é delicada e a talidomida ainda não é recomendada para este tipo de tratamento.

A talidomida sozinha ou em combinação com outros fármacos, possui um papel terapêutico em câncer de próstata de estágio avançado. Combinada com um agente citotóxico, a talidomida pôde ainda contribuir para a estabilização do crescimento do

câncer de próstata de baixa extensão, prolongando o controle da doença com potencial redução da toxicidade (FRANKS *et al.*, 2004).

#### **3.3.4. Efeitos colaterais**

A talidomida é normalmente administrada por cápsulas orais em doses de 50-400 mg/dia, embora até 1200 mg/dia possam ser permitidos. No entanto, a incidência de efeitos colaterais pode impor limites em relação à dose tolerada, como tem sido demonstrado no tratamento de pacientes com artrite reumatóide (MARRIOTT *et al.*, 1999). Além da indução de defeitos de nascimento, os principais efeitos colaterais da talidomida incluem fadiga e constipação. Ela também está associada a um aumento no risco de trombozes, especialmente quando combinada com a dexametasona. Altas doses de talidomida podem causar edema pulmonar, pneumonia, hipotensão entre outros. Outra questão é a sua baixa solubilidade e estabilidade aquática.

Há um grande interesse no desenvolvimento de novos compostos baseados na estrutura da talidomida, porém com aumento das atividades anti TNF, de estimulação de células T e redução ou ausência de toxicidade e efeitos colaterais. Um número considerável de análogos tem sido desenvolvido usando a talidomida como estrutura principal. Os análogos de talidomida, demonstraram ser até 50.000 vezes mais ativos que a talidomida e são altamente eficazes ao inibir a produção de TNF em células PBMC ativadas e em proteger camundongos de letalidades induzidas por LPS. A atividade imunomodulatória desses novos inibidores de TNF pode potencialmente permitir seu uso clínico no tratamento de uma variedade de desordens imunopatológicas de diferentes etiologias. (MARRIOTT *et al.*, 1999, CORRAL e KAPLAN, 1999).

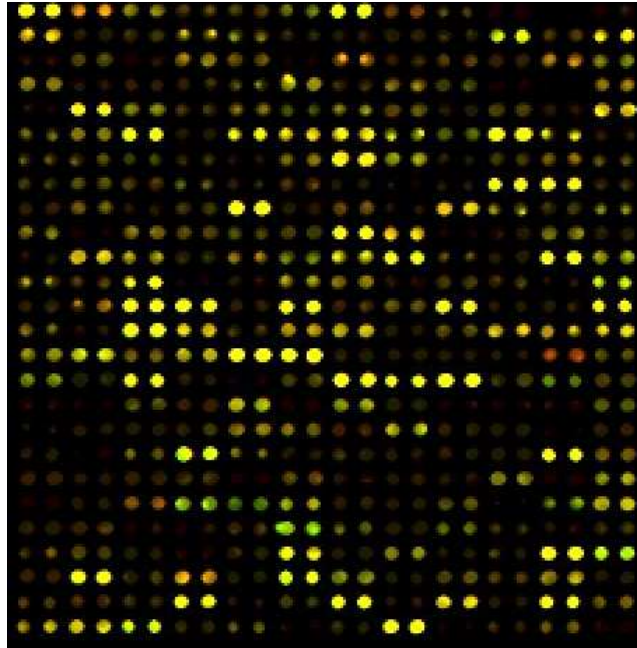
### **3.4. Análise de dados de microarranjos**

#### ***3.4.1. Aquisição da imagem e pré processamento.***

Os microarranjos de DNA proporcionam a análise da expressão de milhares de genes em um único experimento e, conseqüentemente, as análises são dispendiosas em equipamentos, material e tempo. Dessa forma, o planejamento é fundamental para que os resultados obtidos sejam informativos e compatíveis com os recursos empregados. Da mesma forma, a escolha dos métodos para análise das imagens e tratamento dos dados é imprescindível para o êxito deste tipo de análise (AFORNALI, 2006).

Os dados de microarranjos obtidos são geralmente organizados como uma matriz  $M$  ( $n \times m$ ), sendo  $n$  a representação dos genes (linhas) e  $m$  das amostras que foram estudadas (colunas). Conseqüentemente, a posição  $M_{ij}$  seria respectiva à expressão do gene  $i$  na amostra  $j$ .

A tecnologia de arranjos de DNA gera dados onde a expressão  $e_{ij}$ , do gene  $i$  na amostra  $j$ , é calculada como a razão  $r_{ij}/g_i$  do nível real de expressão do gene  $i$ ,  $r_{ij}$ , na amostra  $j$ , sobre o nível de expressão do gene  $i$  em uma amostra controle, ou ainda, outra amostra de interesse,  $g_i$ . Para a visualização dos dados, o valor de  $e_{ij}$  é codificado por cores em uma função da relação vermelho/verde (Figura 4).



**Figura 4: Composição de uma imagem de uma lâmina de microarranjo de DNA.**

A análise dos dados de microarranjo se inicia com os resultados das imagens. Neste ponto, as imagens são avaliadas, e os *spots* de má qualidade investigados. Os dados brutos são organizados em um banco de dados e verificados quanto à necessidade de transformação. A função da transformação é atingir propriedades estatísticas favoráveis, por exemplo, estabilizando a variabilidade e transformando fatores multiplicativos de um modelo linear de análise em fatores aditivos (CUI *et al.*, 2003).

O processo de remoção de efeitos sistemáticos é geralmente chamado de normalização. Este processo possui três etapas: correção de *background*, filtragem e transformação dos dados. A correção de *background* consiste na etapa de subtração do ruído de fundo da intensidade do sinal (*foreground* menos *background*) fluorescente de cada *spot*. A filtragem dos dados corresponde à eliminação ou marcação de genes cujos valores de expressão são insignificantes ou muito baixos em relação ao ruído da lâmina. Vários métodos de transformação dos dados já foram desenvolvidos, entre os mais

utilizados estão as razões, as transformações logarítmicas, as globais, as de ajuste de curvas e as de estabilização da variância.

A razão entre os dois sinais fluorescentes em cada *spot* é comumente usada na estimativa da razão das concentrações de RNA mensageiro nas duas amostras de RNA. O nome “Razão R/G” se deve ao uso do corante Cy5 (vermelho – *Red*) para a condição experimental e Cy3 (verde – *Green*) para a condição controle (Equação 1). Assim, a Razão R/G descreve o nível de expressão de um gene na condição experimental (R) em relação à expressão desse mesmo gene na condição controle (G).

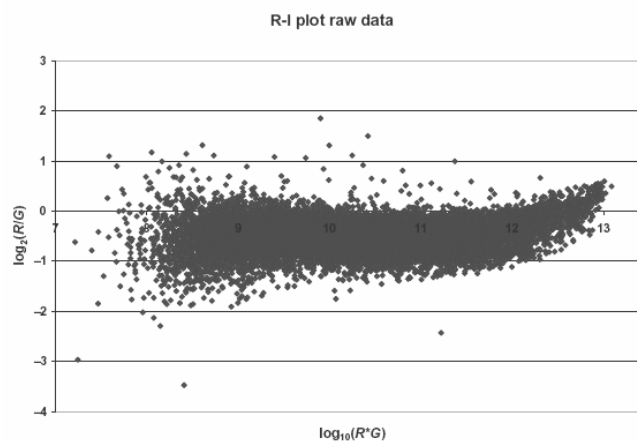
$$Y_i = \frac{R_i}{G_i} \quad (1)$$

A transformação logarítmica (geralmente de base 2) é geralmente aplicada aos dados de microarranjos no cálculo da variação dos sinais originais fluorescentes. Ela não só converte razões em diferenças entre os canais em cada *spot*, como também estabiliza a variância em *spots* de alta intensidade. Se os erros forem proporcionais à intensidade do sinal na escala original, eles serão constantes através da faixa de sinal de intensidade na escala logarítmica. Por outro lado, a presença de erros aditivos na escala original é problemática quando a transformação logarítmica é aplicada.

O logaritmo da razão (Equação 2) proporciona a atribuição do mesmo valor a genes super e subexpressos pelo mesmo fator, ocasionando apenas inversão no sinal. Assim,  $\log_2(2) = 1$  e  $\log_2(1/2) = -1$ . Logaritmos com diferentes bases podem ser utilizados, pois a base afeta apenas o valor obtido, não as propriedades desejadas.

$$Y_i = \log_2 \left( \frac{R_i}{G_i} \right) \quad (2)$$

Uma característica comum a dados de microarranjo é a dependência do logaritmo da razão em relação à intensidade fluorescente. Esta dependência é facilmente visualizada em um *Scatter-plot* (gráfico de espalhamento) do logaritmo da razão de intensidade, versus o logaritmo da intensidade (Figura 5), conhecido como *RI-plots* (DUDOIT *et al.*, 2000). Já o gráfico *MA-plot* serve para aumentar o espaço disponível à representação dos valores M, e facilita a visualização das relações não lineares entre as intensidades logaritmizadas (SMYTH *et al.*, 2002). O *MA-plot* é um *Scatter-plot* do logaritmo da razão de intensidade no eixo das ordenadas, versus o logaritmo da intensidade dividido por dois no eixo das abscissas. Sob a idéia de que a maior parte dos genes não é diferencialmente expressa, e que a frequência de genes sub e superexpressos é similar, a maioria dos pontos de um *RI-plot* e de um *MA-plot* devem se situar em torno de uma linha horizontal. Na prática, *RI-plots* e *MA-plots* podem mostrar vários tipos de curvatura, e existem diversas estratégias para transformar os dados de modo a remover esta curvatura.



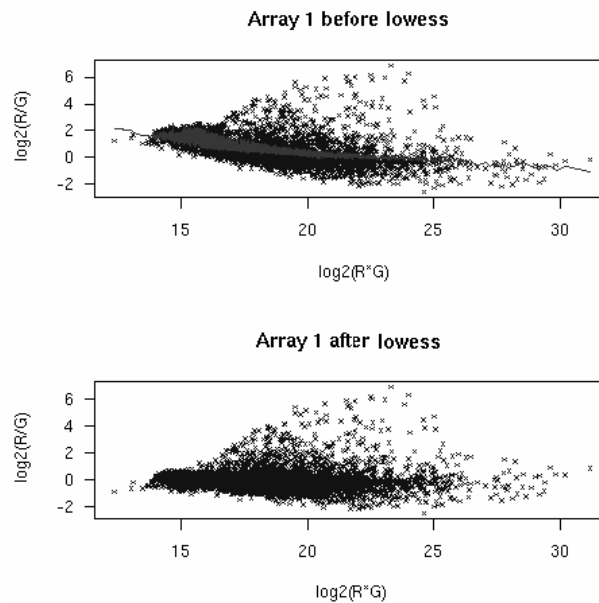
**Figura 5:** *RI-plot* que mostra o  $\log_2$  da razão ( $R_i/G_i$ ) para cada elemento no arranjo como sendo uma função do  $\log_{10}$  do produto ( $R_i \cdot G_i$ ) das intensidades e pode revelar efeitos sistemáticos intensidade-dependentes nos valores mensurados. (Retirado de QUACKENBUSH *et al.*, 2002).



A análise de regressão linear localmente ponderada (*lowess* para funções lineares e *loess* para funções quadráticas) é um método de regressão não paramétrico de grande robustez (CLEVELAND, 1979). Foi proposta como forma de normalização que pode remover efeitos intensidade-dependentes (CUI *et al.*, 2003) nos valores obtidos por logaritmo da razão. Quando dados originais são considerados, a variação aumenta conforme a intensidade do *spot* aumenta. Quando dados log-transformados são considerados, a variância é geralmente estável acima de certa intensidade, porém, *spots* de baixa intensidade de sinal podem ser altamente variáveis. A maneira mais simples de visualizar efeitos intensidade-dependentes, e o ponto de partida para a transformação *lowess* é através do *RI-plot* ou do *MA-plot*. O método *lowess* detecta desvios sistemáticos nestes gráficos e os corrige (Figura 6).

O ajuste pelo método *lowess* necessita da escolha de um valor que determine qual parcela dos dados é considerada central. A escolha desse valor é subjetiva, mas em teoria, o valor mais alto que remova a dependência da intensidade das razões logarítmicas é ideal. A maioria dos autores (CUI *et al.*, 2003, DUDOIT *et al.*, 2000, YANG *et al.*, 2002), normalmente escolhe um valor de 20 a 30%, ou seja, consideram 20 a 30% dos dados de intensidade como centrais.

O procedimento de ajuste pelo método *lowess* é bastante simples, mas perigoso. Primeiro, encorre-se o risco de ajustar os dados de forma excessiva e introduzir erros maiores do que os removidos. Segundo, deve-se ter a consciência de que se está forçando os dados ao encontro de expectativas subjetivas; quanto mais drástico for o ajuste, mais crítica se torna a validade dos resultados (CUI *et al.*, 2003).



**Figura 6: Transformação lowess, correção da curvatura do *RI-plot*. (Retirado de WU et al., 2006).**

### **3.4.2. Análises estatísticas**

A tecnologia de arranjos de DNA é a principal estratégia para o acesso dos perfis da expressão gênica numa larga escala. Entretanto, por causa da sua alta dimensionalidade, do ruído característico dos dados de arranjos e do limitado número de réplicas experimentais, um modelo estatístico para seleção de genes diferencialmente expressos não é facilmente determinado (DRUMMOND *et al.*, 2005).

A pesquisa de genes diferencialmente expressos (induzidos ou reprimidos) em um arranjo, ou seja, a análise do padrão de expressão gênico é realizada com experimentos de comparação entre duas amostras (controle x amostra de interesse). Para tal, as opções disponíveis dependem das características do arranjo, fundamentalmente da replicação do experimento. Em experimentos com replicação, há a possibilidade de utilização de métodos estatísticos mais confiáveis. Exemplos desses métodos incluem:

teste *t-student* (DUDOIT *et al.*, 2000), ANOVA (KERR *et al.*, 2000), testes não-paramétricos (GRANT *et al.*, 2001), Bayesianos (FRIEDMAN *et al.*, 2000) entre outros. Originalmente, na ausência de replicação, utiliza-se o método da razão (CUI *et al.*, 2003). Neste, utilizam-se os dados de expressão  $e_{ij}$ , selecionando-se genes que apresentam esse valor absoluto acima de um ponto de corte pré-determinado o que indicaria um aumento ou uma diminuição da expressão da amostra de interesse em relação ao controle ou à outra amostra de interesse. Este método foi logo substituído por outros modelos, embora ainda seja utilizado.

Para o caso de experimentos com pouca ou nenhuma replicação (a grande maioria dos dados de microarranjo), a estratégia frequentemente adotada é a seleção destes genes, cujos sinais de ambas as amostras mostra uma razão de expressão (*fold change*) de duas ou mais vezes (ou qualquer outro valor arbitrário fixo). Recentemente, foram desenvolvidas ferramentas estatísticas que permitem fazer uma análise precisa dos dados. Estes métodos podem ser univariados e multivariados. Dentre os multivariados, podemos citar os métodos de análise de variância de microarranjos, utilizando o aplicativo MAANOVA (WU *et al.*, 2003) e o de avaliação do erro global (*Global error assessment – MANSOURIAN et al.*, 2004). Como exemplos de modelos univariados, há o *ISER* (Seleção de razões de expressão intensidade-dependentes – DRUMMOND *et al.*, 2005) e o *EDGE* (Análise exploratória da expressão gênica diferencial – LOGUINOV *et al.*, 2004).

## 4. Bases Metodológicas

### 4.1. Análise de variância de dados de microarranjo

A estatística F, ou modelo ANOVA, é usada para problemas multiclasse, na detecção da diferença entre médias de duas ou mais “populações” (CHEN *et al.*, 2005). O teste F clássico (ANOVA) é uma generalização do teste t, que permite a comparação de mais de duas amostras. É usado para se detectar qualquer padrão de expressão diferencial entre as diversas condições, ao comparar a variação entre as amostras replicadas dentro e entre as condições (CUI *et al.*, 2003).

Um experimento de microarranjo pode envolver arranjos múltiplos para comparar amostras múltiplas. Cada medida em um experimento de microarranjo está associada com uma combinação particular de um arranjo, de um marcador (Cy5 ou vermelho; e, Cy3 ou verde), uma variedade e um gene (KERR *et al.*, 2000).

WU *et al.* (2003) desenvolveram um *software* que permite realizar a análise de variância de dados de microarranjo. O aplicativo *freeware* MAANOVA, versão 1.4.1 (THE JACKSON LABORATORY, 2002) fornece um conjunto completo de ferramentas para análise de dados de microarranjo, incluindo: (1) Visualização e inspeção da qualidade dos dados; (2) Transformação dos dados; (3) Ajuste do modelo ANOVA para efeitos fixos e mistos; (4) Análise de Cluster com *Bootstrapping*.

As funções do MAANOVA foram desenvolvidas para análise de experimentos de microarranjos de DNA de dois canais de qualquer escala e complexidade. Apesar de encorajar a utilização de desenhos balanceados eficientes, o *software* pode ser utilizado na análise de qualquer experimento que utilize mais de um microarranjo para cada conjunto de amostras.

Os dados originais de um experimento único de microarranjo consistem em um par de imagens representando as intensidades fluorescentes detectadas em que é feita a leitura da lâmina do microarranjo. As imagens são armazenadas como TIFF 16 bit e os dados extraídos por segmentação e quantificação da intensidade da imagem associada com cada *spot*. O arquivo de entrada em MAANOVA requer duas colunas com valores de intensidade de cada arranjo, e pode incluir colunas adicionais ou cabeçalhos se necessário. Por exemplo, em um experimento com quatro arranjos, as oito colunas de dados são dispostas como (R1; G1; R2; G2; R3; G3; R4; G4), sendo R os experimentos marcados em vermelho (Cy5) e G, os marcados em verde (Cy3).

Chamando a intensidade log-transformada (logaritmo de base 2) de  $Y$ , pode-se modelar essa intensidade, fatorando-a em fontes sistemáticas de variabilidade reconhecidas como o arranjo ( $A$ ); o marcador ( $D$ ), ou seja, verde (Cy3) ou vermelho (Cy5); a sequência gênica representante do gene ( $G$ ); e a variedade ( $V$ ), ou seja, se a amostra é de teste ou controle; assim como interações de primeira ordem como a do arranjo com o gene ( $AG$ ); a do marcador com o arranjo ( $AD$ ); a do marcador com o gene ( $DG$ ) e, de maior importância, a da variedade com o gene ( $VG$ ).

O modelo de ANOVA para microarranjos pode então ser descrito em duas etapas. A primeira é responsável pela normalização dos dados utilizando o modelo:

$$Y_{ijkgr} = \mu + A_i + D_j + AD_{ij} + r_{ijkgr} \quad (6),$$

onde  $\mu$ ,  $A$ ,  $D$  e  $AD$  englobam a média de intensidade global e as variações em relação a arranjos, marcadores e interação arranjo/marcador; Os resíduos desta normalização serão usados como dados de entrada para a segunda etapa, que realiza a modelagem dos efeitos gene-específicos:

$$r_{ijk} = G + AG_i + DG_j + VG_k + \varepsilon_{ijk} \quad (7),$$

onde, o termo  $G$  consiste na média de intensidade associada a um gene particular;  $AG$  a variação de cada gene em cada arranjo, o termo  $DG$  representa os efeitos gene-específicos dos marcadores e o termo  $VG$  são as variações dos níveis de expressão dos genes. O termo  $\varepsilon$  compreende os fatores não explicados e é assumido como sendo o erro com média zero (WU *et al.*, 2003).

Quando o modelo ANOVA é ajustado para os dados, obtêm-se estimativas para cada um dos termos individuais. De particular interesse neste contexto são os valores estimados de  $VG$  que são os valores relativos de expressão (nível de expressão de um gene  $g$  numa amostra  $k$  relativa à expressão ponderada média deste gene em todas as amostras do experimento) e seus valores estimados são denotados como  $\overline{VG}$ .

O teste  $F$  é designado para detectar qualquer padrão de expressão diferencial entre diversas condições através da comparação da variação entre amostras replicadas entre e dentre as condições. O MAANOVA fornece quatro tipos de teste  $F$  que podem ser usados individualmente ou em conjunto: O teste  $F$  gene-específico ( $F_1$ ), uma generalização do teste  $t$  de *student* gene-específico, é o teste  $F$  usual e é computado em uma base gene-a-gene. Assim como em testes  $t$ , também pode ser assumida uma variância comum para todos os genes, resultando no teste  $F$  da variância global ( $F_3$ ). Uma abordagem intermediária é obtida pelo teste  $F_2$ , análogo ao teste  $t$  regularizado. Este usa uma combinação ponderada das estimativas de variância global e gene-específica no denominador (CUI e CHURCHILL, 2003). O teste estatístico  $F_s$  usa um estimador da variância que não faz qualquer suposição prévia sobre distribuições da variância entre os genes. Este teste possui bons resultados tanto quando as variâncias são constantes como se elas variassem gene-a-gene.

Os resultados do teste  $F_s$  do MAANOVA podem ser visualizados por representações gráficas como *Volcano-plot*, em que o eixo vertical representa o logaritmo do p-valor da estatística  $F_s$  e o eixo horizontal é proporcional às razões de expressão. A linha horizontal representa o limite de significância para o teste  $F_s$ .

P-valores podem ser obtidos a partir de Tabelas padrão do teste F, mas as estatísticas  $F_2$ ,  $F_3$  e  $F_s$  não seguem a distribuição F tabulada e seus valores críticos devem ser determinados por análise de permutação.

A análise de permutação é uma abordagem não paramétrica para estabelecer a distribuição nula de um teste estatístico. A maneira de desenvolver uma estratégia de permutação é identificar unidades no experimento que sejam permutáveis sob a hipótese nula. Em experimentos de microarranjos, se é utilizada a heterogeneidade da variância gene-específica, então a unidade deverá ser o arranjo como um todo. Além disso, os arranjos a serem misturados, dependerão do desenho do experimento e dos fatores sendo testados. Arranjos de duas cores são um pouco mais complexos do que sistemas de uma cor, uma vez que o pareamento entre os dois canais do arranjo deve ser mantido nas unidades permutadas (STOREY e TIBSHIRANI, 2003).

#### **4.2. ISER – Seleção de razões de expressão intensidade-dependentes**

O método ISER (*intensity-dependent selection of expression ratios*) foi desenvolvido por DRUMMOND *et al.* (2005) com o objetivo de analisar dados de arranjos de DNA pela otimização da seleção de genes com variações mais significantes na expressão entre duas amostras de RNAm. Este algoritmo é desenhado para ser utilizado quando pouca ou nenhuma replicação de hibridização do arranjo é disponível.

Um dos diferenciais deste método é a transformação não-linear dos dados para normalidade por meio de uma janela de tamanho variável, permitindo o uso de

distribuições próximas à distribuição normal, ao invés de distribuições com “caudas pesadas” pelo número excessivo de *outliers*.

O algoritmo é baseado na hipótese de que a maior parte dos genes hibridizados em arranjos tem uma expressão constante nas amostras estudadas, de modo que genes diferencialmente expressos são identificados como aqueles que são *outliers* na distribuição global das razões de expressão. Outro aspecto positivo do ISER é o fato da variância das razões de expressão depender do logaritmo da intensidade média do gene.

A fundamentação inicial do ISER é encontrar uma transformação de dados que resulte numa distribuição normal das razões de expressão. Por exemplo, quando a distribuição das razões tem uma distribuição lognormal, a transformação dos dados efetuada pelo ISER é a distribuição logarítmica. Entretanto, se houver variações pequenas, os dados de entrada do ISER estão na forma de arquivo de texto contendo as intensidades do sinal (bruto ou normalizado) para ambas as amostras de RNA de cada gene. O algoritmo segue os passos abaixo:

(1) Para evitar trabalhar com valores zero, uma constante pequena ( $10^{-4}$  vezes o valor da expressão mínima diferente de zero) é adicionada aos dados. Para cada gene, a razão de expressão e a média geométrica dos sinais de ambas as amostras são calculados.

(2) É encontrada uma transformação dos dados que gere os valores transformados da razão mais próximos de uma distribuição normal. Isso é feito pela escolha do melhor expoente  $\lambda$  na família Box-Cox de transformações:

$$BC = \begin{cases} \log(x), & \text{se } \lambda = 0 \\ \frac{x^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \end{cases} \quad (8)$$



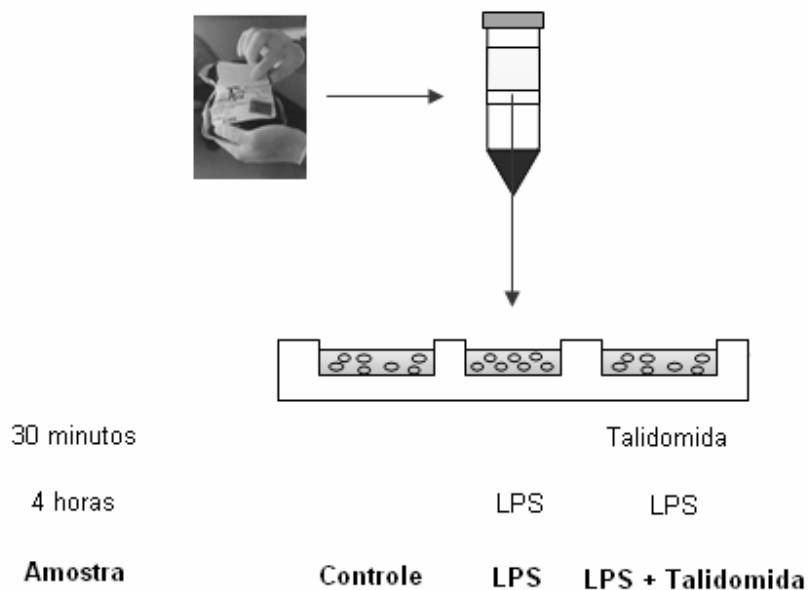
A transformação mais cabível é então aplicada às razões de expressão e às médias geométricas dos valores de expressão, para gerar o que é referido, respectivamente, como sendo as razões e as intensidades transformadas.

(3) Os genes são ordenados de modo crescente de acordo com suas intensidades transformadas. Para estimar a média e o desvio padrão das razões de expressão transformadas, é utilizada uma janela deslizante sobre estes valores ordenados de intensidade dos genes. O tamanho desta janela é determinado de modo a otimizar o valor do teste de *Kolmogorov-Smirnov* para normalidade, aplicado às razões transformadas dos genes dentro das janelas. Dentro de cada janela deslizante, o algoritmo calcula a média aparada (tradução livre de *trimmed mean*) e o desvio padrão (com respeito à média aparada) das razões transformadas, baseados nos valores 60% centrais. Estas estimativas e a determinação de um nível de significância são empregadas para delimitar os limites locais superiores e inferiores das razões transformadas, fora dos quais os genes são considerados como diferencialmente expressos.

(4) Um *spline* linear é usado para aproximar a relação entre estes limites e a média das intensidades transformadas em cada janela deslizante, estabelecendo os limiares intensidade-dependentes para as razões transformadas a serem selecionadas como diferencialmente expressas.

## 5. Material e Métodos

Os dados dos microarranjos foram obtidos através de uma parceria entre o Instituto de Biofísica Carlos Chagas Filho (UFRJ), o Laboratório de Hanseníase da FIOCRUZ e a Unidade de Microarranjos da Universidade Nacional Autônoma do México. Os experimentos consideram um pool de células PBMC de bolsas de sangue arterial coletadas de três indivíduos saudáveis como sendo o controle. Para a amostra de teste, células de mesma origem foram tratadas com lipopolissacarídeo por quatro horas, com e sem o tratamento prévio (30 minutos) de talidomida (Figura 7).



**Figura 7: Preparo das amostras controle e teste. A amostra controle consistiu em células mononucleares circulantes de doadores saudáveis extraídas sem que houvesse qualquer tratamento com elementos bacterianos ou drogas. Já a amostra de teste de LPS foi tratada durante 4 horas. A amostra de teste com LPS e Talidomida recebeu tratamento de 30 min com talidomida e de 4 horas com o LPS.**

Para os microarranjos, foram utilizados oligonucleotídeos (50 mer) representantes de 9984 genes humanos envolvidos em processos celulares como metabolismo, morte e crescimento celular, adquiridos da empresa MWG Biotech Company. Os oligonucleotídeos foram aplicados em duplicata sobre lâminas de vidro revestidas com superamina (Telechem International) utilizando o sistema Virtek ChipWriter, resultando em um total de 19968 *spots*. Cada experimento foi composto inicialmente por três lâminas.

Para se quantificar o *bias* de intensidade associado aos diferentes fluoróforos empregados nos microarranjos foi feita uma reversão de marcação (*reverse labeling*). Dessa forma, as amostras de controle são marcadas com o fluoróforo antes empregado nas amostras de teste, e vice-versa. Foi fabricada uma lâmina por tratamento para esse fim. Os fluoróforos utilizados em cada uma das lâminas dos dois experimentos estão esquematizados na Tabela 1. Cada experimento é composto por três lâminas no total.

**Tabela 1: Descrição das lâminas dos experimentos realizados e dos marcadores utilizados para controle e teste em cada uma delas.**

	<b>Lâminas</b>	<b>Cy3</b>	<b>Cy5</b>
Experimento 1	<b>H10KA-01-42</b>	Controle (PBMC)	Teste (LPS)
	<b>H10KA-02-05</b>	Teste (LPS)	Controle (PBMC)
	<b>H10KA-02-09</b>	Controle (PBMC)	Teste (LPS)
Experimento 2	<b>H10KA-01-41</b>	Controle (PBMC)	Teste (LPS + TAL)
	<b>H10KA-02-04</b>	Teste (LPS + TAL)	Controle (PBMC)
	<b>H10KA-02-08</b>	Controle (PBMC)	Teste (LPS + TAL)

Após a pré-hibridização (5x SSC, 0.1% SDS, 1% BSA) por uma hora a 42 °C, as amostras obtidas das células não tratadas (controle) e tratadas com LPS e LPS +

Talidomida (testes) foram adicionadas ao tampão Hybit 2 (Telechem International) e hibridizadas durante a noite em três arranjos para cada experimento (2 réplicas e uma de marcação reversa - *dye-swap*) a 42 °C em uma câmara umidificada (Corning). As lâminas hibridizadas foram então lavadas, secas e escaneadas usando o Scan Array 4000 (Packard Biochips). A intensidade de *spot* para cada arranjo foi obtida pelo software Array Pro Analyzer. As duas imagens de 16 bits geradas a partir de cada lâmina foram segmentadas e as intensidades foram corrigidas em função do *background*.

O pré-processamento dos dados consistiu na filtragem dos *spots* visando identificar aqueles com relação sinal-ruído (RSR) baixa. O critério escolhido para esta avaliação (JABLONKA *et al.*, 2006) seleciona mudanças significativas na razão entre a densidade média de todos os *spots* em relação ao *background* médio de todos os *spots* de um arranjo ( $D_{\text{médio}}/B_{\text{médio}}$ ). Foram considerados como *spots* de má qualidade aqueles em que os sinais de Cy3 ou Cy5 apresentassem razão menor do que o critério calculado ( $D_i/B_i < D_{\text{médio}}/B_{\text{médio}}$ ).

Na primeira abordagem, os *spots* que não passaram pelo critério foram assinalados, e a análise foi realizada com todos os genes (conjunto completo). Na segunda abordagem, os genes assinalados como tendo RSR inferior ao critério estabelecido foram excluídos da análise (conjunto reduzido). Para manter um desenho experimental balanceado em ambas as abordagens, foi descartada a lâmina que apresentou maior número de *spots* de má qualidade. Cada experimento passou a contar com duas lâminas, sendo uma de marcação direta e uma de marcação inversa. Para realização da análise, utilizou-se o desenho experimental de referência, onde as amostras de teste foram comparadas às amostras de controle.

Em cada lâmina foi aplicada aos dados de logaritmo de base 2 a transformação *lowess* que consiste numa regressão linear localmente ponderada, para melhor redução

da variabilidade dos dados (SPERANDEI, 2007). Em seguida, foi calculada a média dos *spots* de replicação interna para cada lâmina.

Usando o software MAANOVA, foi realizada a análise de variância dos dados transformados em experimentos de microarranjos em duas etapas. A primeira consistiu no ajuste dos dados utilizando um modelo que leva em conta efeitos de arranjo e corantes. Nesta etapa a qualidade do modelo ajustado foi avaliada por gráficos de dispersão e da normalidade dos resíduos<sup>1112</sup>. A segunda etapa compreendeu a utilização dos resíduos obtidos da etapa anterior na modelagem dos efeitos gene-específicos. Nesta etapa foi utilizado o teste  $F_s$  que deve ser mais eficiente quando há informação limitada para se estimar a variância gene-específica. O resultado do teste fornece p-valores e Tabelas padrão. Como o teste  $F_s$  não segue a distribuição F tabulada, seus valores críticos foram determinados por análise de 100 permutações (WU *et al.*, 2003).

Os resultados do teste  $F_s$  do MAANOVA foram visualizados em representações gráficas usando o *Volcano-plot* (Figuras 13 e 14). O valor relativo da razão de expressão (*fold change*) foi calculado para todos os genes como sendo a divisão do valor da amostra de teste pelo valor da amostra de controle. Se este número fosse menor do que 1, este gene possuiria expressão reduzida. Caso esse número fosse superior a 1, este gene apresentaria expressão aumentada.

Na análise utilizando o método ISER, os dados dos conjuntos completo e reduzido, após a transformação *lowess* e o cálculo da média entre as réplicas foram organizados em Tabelas da seguinte forma:

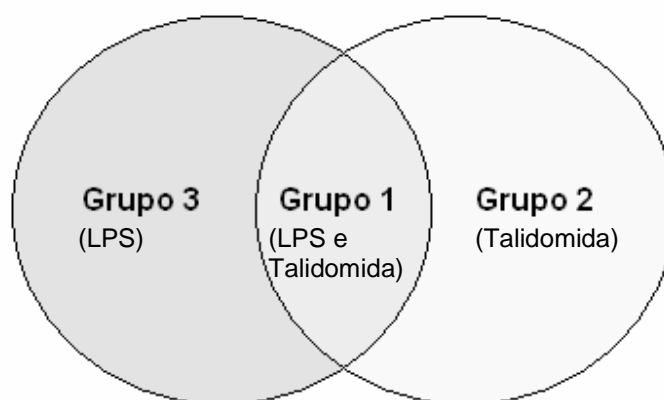
- 1ª análise: valores de expressão do controle (PBMC) x tratamento com LPS;
- 2ª análise: valores de expressão do controle (PBMC) x tratamento com LPS e

Talidomida.

O nível de significância foi ajustado para 0,001 e os resultados obtidos foram organizados em Tabelas que continham a informação do número do gene e a

classificação “0” para genes cuja alteração da expressão não foi considerada significativa, “1” para genes que tiveram aumento significativo de sua expressão e “-1” para genes que sofreram redução da expressão.

Em seguida foi feita a comparação entre as duas análises (Figura 8), na qual a interseção (Grupo 1) indicaria os genes cujos efeitos do tratamento com LPS se mantiveram mesmo no tratamento com a talidomida; os genes que somente mostraram expressão diferencial na condição de tratamento com talidomida (Grupo 2) e os genes que mostraram diferença nos níveis de expressão no tratamento com LPS, mas se expressavam normalmente na condição de tratamento com LPS e talidomida (Grupo 3).



**Figura 8: Esquema de interpretação dos resultados do método ISER de acordo com os genes selecionados nas duas análises (comparações entre lâminas) utilizadas.**

A função dos genes selecionados em ambos os métodos foi pesquisada na base de dados *Entrez Gene* que disponibiliza dados da estrutura, função, sequenciamento, níveis de expressão, citação e homologia a partir do número de acesso de cada gene (<http://www.ncbi.nlm.nih.gov/sites/entrez>).

## 6. Resultados

Usando a metodologia descrita para avaliar a qualidade dos *spots*, foram identificados 1388 genes apresentando razão entre sinal e ruído (RSR) igual ou superior ao critério utilizado em todas as lâminas do experimento. Os demais genes que não passaram pelo critério de qualidade foram assinalados e mantidos na primeira abordagem (conjunto completo) e excluídos na segunda abordagem (conjunto reduzido).

Os dados dos experimentos de microarranjo exibiam um padrão não linear, sugerindo a necessidade de uma transformação. As Figuras 9 e 10 mostram *MA-plots* para os conjuntos completo e reduzido, antes e depois da transformação *lowess*.

Após a transformação, utilizando o software MAANOVA, foi verificado o modelo que melhor se ajustou aos dados. No caso, o valor estimado de Y ( $\hat{Y}$ ) foi melhor ajustado pelo modelo que considerava os efeitos de Corante (D), Arranjo (A) e Amostra (S). As Figuras 11 e 12 mostram os gráficos contendo os resíduos em relação aos valores ajustados para cada arranjo em ambas as abordagens.

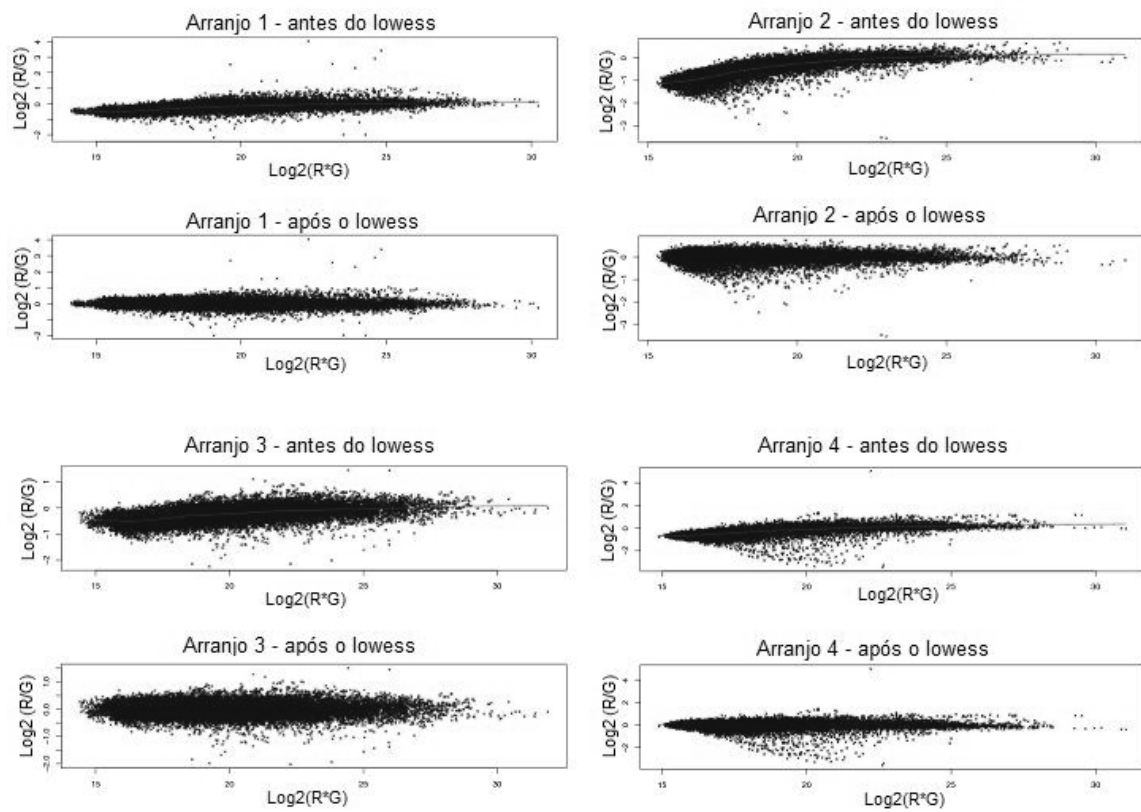
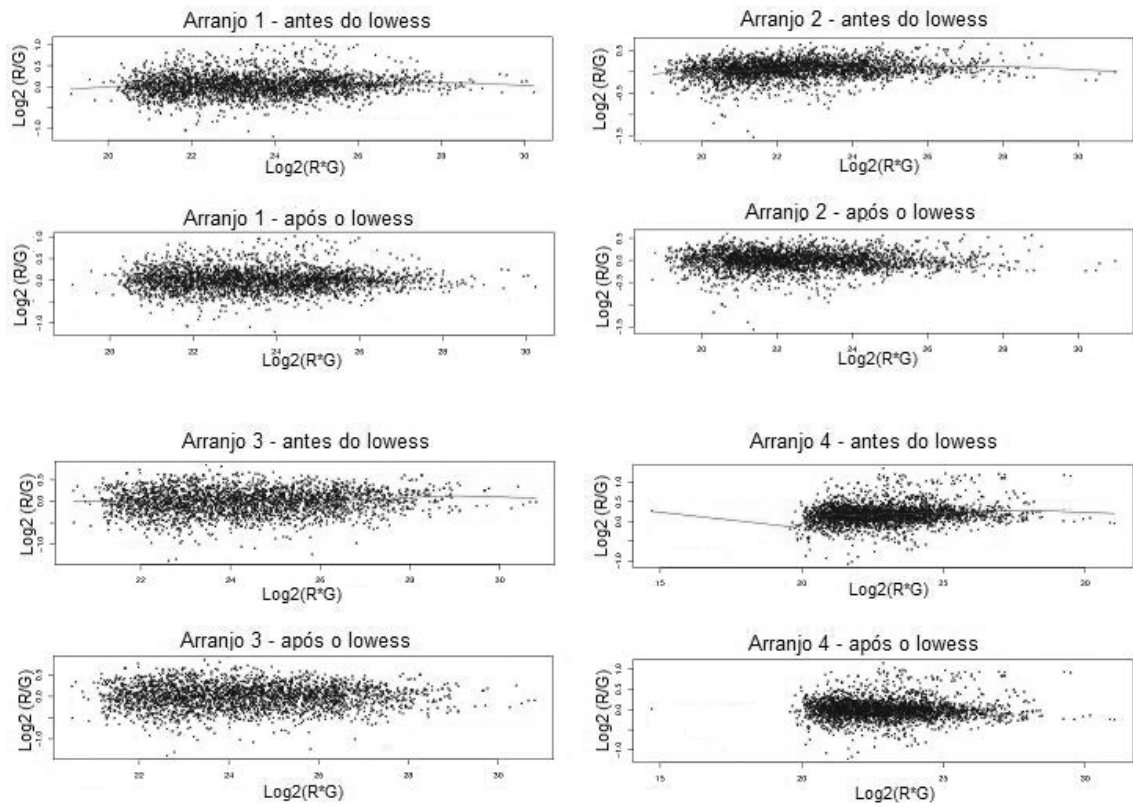


Figura 9: *MA-plot* contendo os dados dos quatro arranjos do conjunto completo antes e depois da transformação *lowess*.





**Figura 10: MA-plot contendo os dados dos quatro arranjos do conjunto reduzido antes e depois da transformação lowess.**

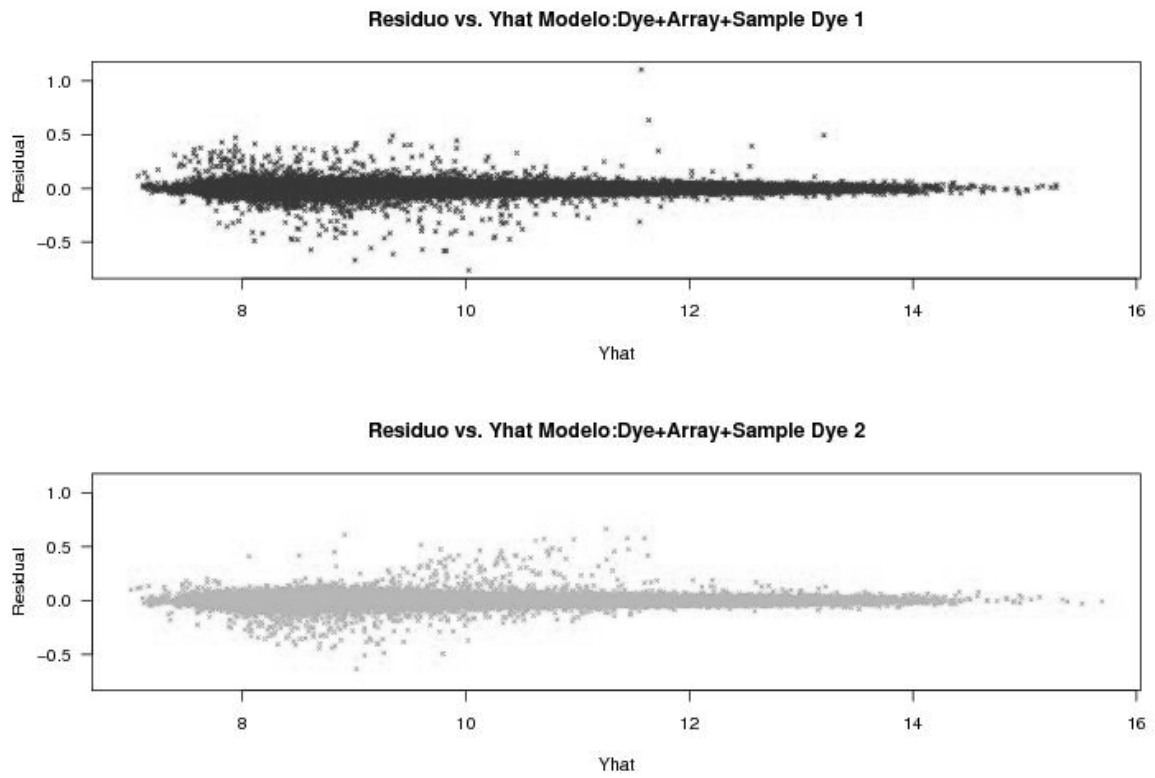
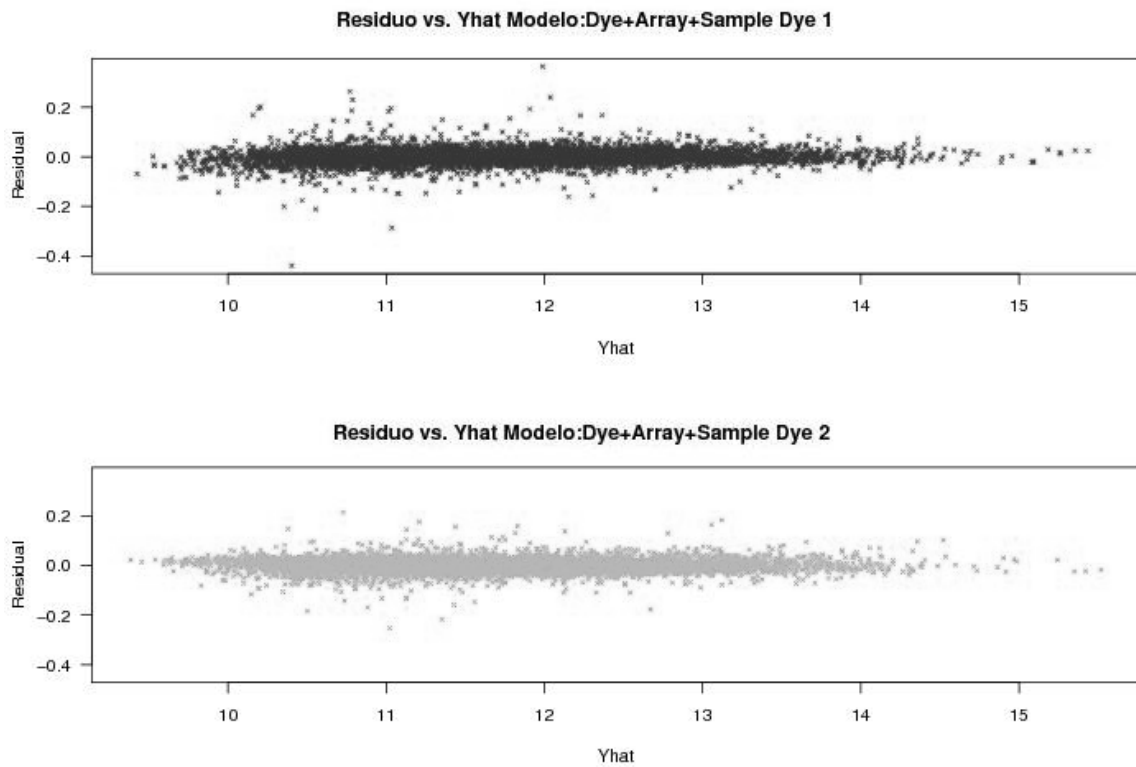


Figura 11: Gráfico com os resíduos em relação aos valores ajustados em um arranjo do conjunto completo.



**Figura 12: Gráfico com os resíduos em relação aos valores ajustados em um arranjo do conjunto reduzido.**

Os resíduos foram utilizados na modelagem ANOVA dos efeitos gene-específicos e os resultados obtidos pelo teste  $F_s$  foram visualizados usando o *Volcano-plot* (Figuras 13 e 14). Foram considerados diferencialmente expressos os genes que tiveram uma diferença significativa ( $p < 0.001$ ) para a estatística  $F_s$ . A seguir foi calculado o valor relativo da razão de expressão pela divisão do valor da amostra de teste pelo valor da amostra de controle.

Volcano plot para modelo de efeitos fixos (LPS/TAL)

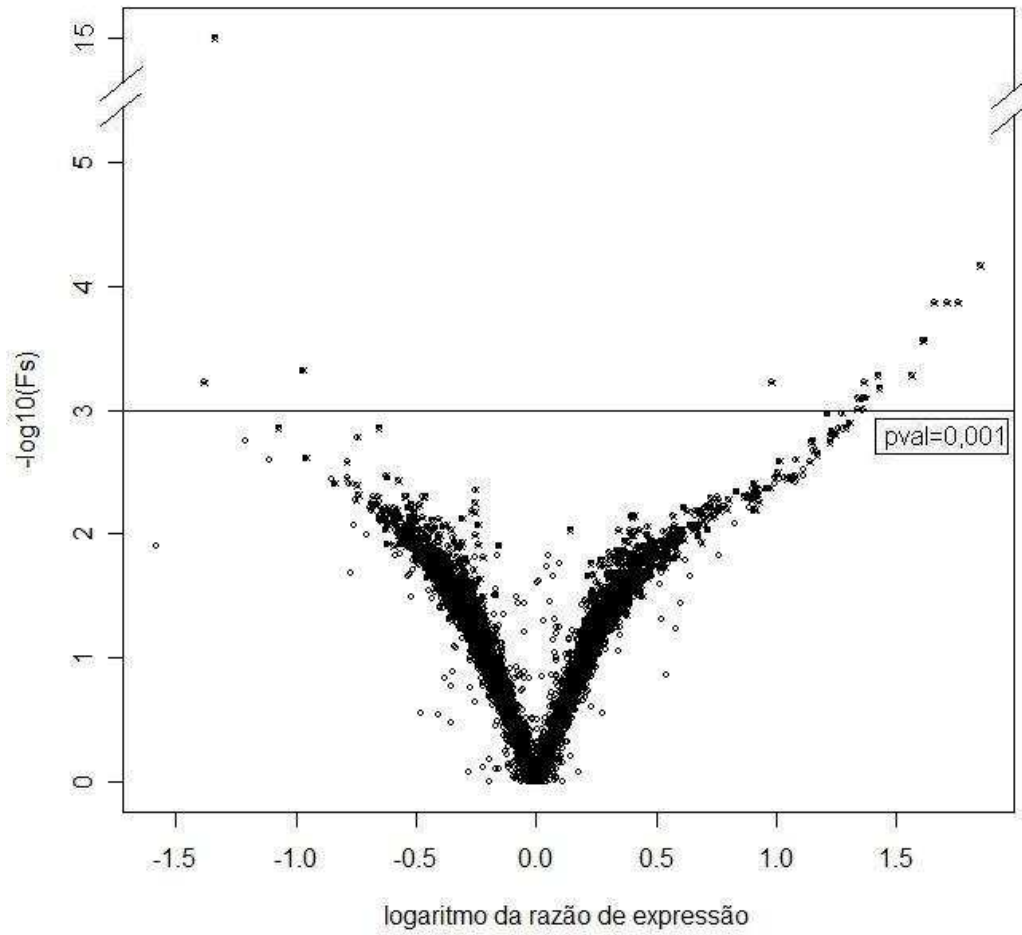
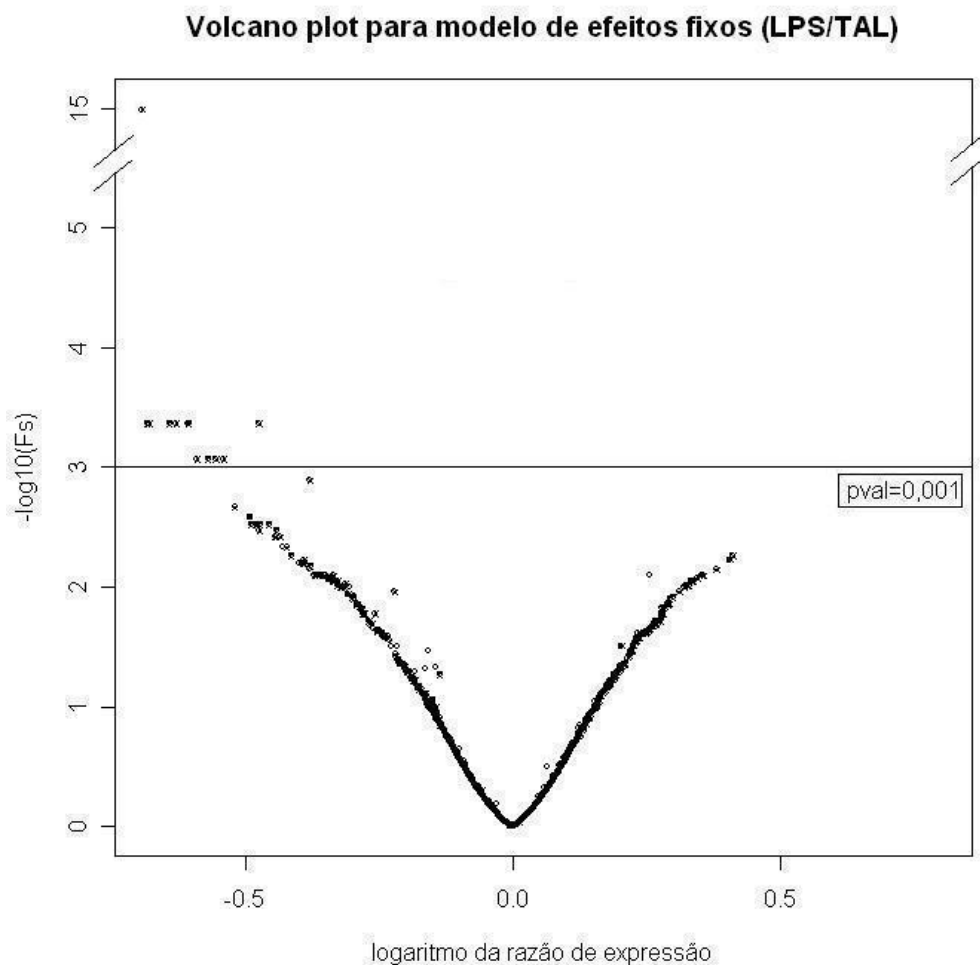


Figura 13: *Volcano-plot* com os resultados do teste  $F_s$  para o conjunto completo.



**Figura 14: Volcano-plot com os resultados do teste Fs para o conjunto reduzido.**

Na abordagem usando o conjunto completo de dados (Tabela 2), 22 genes foram identificados como diferencialmente expressos. No entanto, as amostras de todos os genes selecionados nesta estratégia haviam sido consideradas de qualidade insuficiente por sua razão entre o valor do sinal e o ruído (RSR) ser inferior ao critério estabelecido. Na outra abordagem (Tabela 3), ao eliminar da modelagem ANOVA os genes de qualidade inferior, 12 genes foram selecionados. Como nenhum dos genes selecionados pela análise feita com o conjunto completo sequer entrou na análise com o conjunto reduzido, não houve intersecções entre as listas de genes selecionados.

**Tabela 2: Genes selecionados do conjunto completo pelo teste Fs do software MAANOVA com p-valor inferior a 0,001, ordenados por p-valor.**

<b>Nº do Gene</b>	<b>p-valor</b>	<b>fold change</b>	<b>Símbolo</b>	<b>Descrição</b>
9972	0	0.396262	NM_005076	contactina 2 (axonal); cntn2
136	5,57E-05	3.608765	NM_014291	glicina c-acetiltransferase; gcat
148	0,000111	3.276312	NM_017660	proteína hipotética flj20085; gata zinc finger domain containing 2a
161	0,000111	3.155288	NM_004066	centrina 1; cetn1
149	0,000167	3.385615	XM_030485	gadd45 beta (proteína regulatória negativa do crescimento myd118) (proteína de resposta primária de diferenciação mielóide myd118) (h. sapiens); loc126563
65	0,000223	3.06469	NM_024979	proteína hipotética flj12122;
162	0,000223	3.057481	NM_006101	altamente expressa em cancer. rica em repetições de heptads de leucina; hec
9569	0,00039	0.510044	XM_006188	butirotetaina gamma 2-oxoglutarato dioxigenase gamma-butirotetaina hidroxilase bbox1
99	0,000453	2.95969	NM_018179	proteína hipotética flj10688
10	0,000509	2.684956	b-actin-r3	beta-actina-r3
32	0,000565	2.694345	NM_018264	proteína hipotética flj10900
54	0,000565	1.974208	NM_007368	proteína de ligação inositol 1,3,4,5-tetraquisfosfato;
151	0,000565	2.577821	NM_015937	cgi-06 protein; loc51604
2510	0,000565	0.38359	NM_003392	wingless-type mmtv família de sítios de integração precursor de membro 5a; wnt5a
17	0,000676	2.565199	NM_018013	proteína hipotética flj10159
41	0,000676	2.589936	NM_002571	proteína endometrial associada ao progestogênio (paep)
173	0,000676	2.533914	NM_005881	cadeia ramificada alfa-cetoácida desidrogenase kinase; bckdk
128	0,000739	2.592083	NM_000079	receptor colinérgico, nicotínico, alfa polipeptideo 1 (muscular) precursor; chrna1
43	0,000906	2.530575	NM_006244	proteína fosfatase 2, subunidade regulatória b (b56), isoforma beta
92	0,000906	2.315631	NM_023071	proteína hipotética flj13117
160	0,000906	2.580968	NM_003784	inibidor de proteinase serina (or cisteína), clade b (ovalbumina), membro 7
52	0,000962	2.421407	NM_024330	proteína hipotética mgc4365

**Tabela 3: Genes selecionados do conjunto reduzido pelo teste Fs do software MAANOVA com p-valor inferior a 0,001, ordenados por p-valor.**

<b>Nº do Gene</b>	<b>p-valor</b>	<b>fold change</b>	<b>Símbolo</b>	<b>Descrição</b>
8946	0	0.585509	NM_012175	proteína f-box 3, isoforma 1; fbxo3
6331	0.000436	0.719299	NM_004433	fator 3 tipo e74 (fator de transcrição de domínio epitélio-específico); elf3
6565	0.000436	0.623532	NM_006328	RNA binding motif protein 14; rbm14
7582	0.000436	0.656124	AF151045	antígeno associado ao melanoma (mutado) 1 hspc211
7930	0.000436	0.64048	NM_021927	proteína hipotética flj13220
7975	0.000436	0.646392	NM_014859	produto do gene kiaa0672
9020	0.000436	0.657025	NM_005717	complexo 2/3 proteína relativa a actina, subunidade 5 (16 kd); arpc5
9023	0.000436	0.622125	NM_018471	proteína de hipotálamo não-caracterizada ht010
4220	0.000873	0.687397	NM_006351	translocase of membrana interna de mitocôndria 44 homóloga (levedura); tim44
7748	0.000873	0.680372	NM_032991	pré-proteína caspase 3
7940	0.000873	0.66408	NM_000084	canal de cloro 5; clcn5
8688	0.000873	0.673267	NM_002310	precursor do receptor do fator inibitório de leucemia; lifr

As análises realizadas pelo método ISER, com os conjuntos de dados completo e reduzido, detectaram genes com expressão diferencial para cada lâmina das duas análises (Figuras 14 e 15). O nível de significância estipulado gerou limites superiores e inferiores, acima (ou abaixo) dos quais, o gene é considerado superexpresso (ou subexpresso). Os genes selecionados como diferencialmente expressos pelo método ISER na análise com o conjunto completo, juntamente com a classificação da qualidade dos *spots* correspondentes, estão indicados nas Tabelas 4, 5 e 6. A Tabela 4 indica os genes cujos efeitos da ativação com LPS se mantiveram mesmo no tratamento com a talidomida; a Tabela 5 identifica os genes que mostraram expressão diferencial somente no tratamento com Talidomida e a Tabela 6 mostra os genes que tiveram expressão diferencial somente na ativação com LPS.

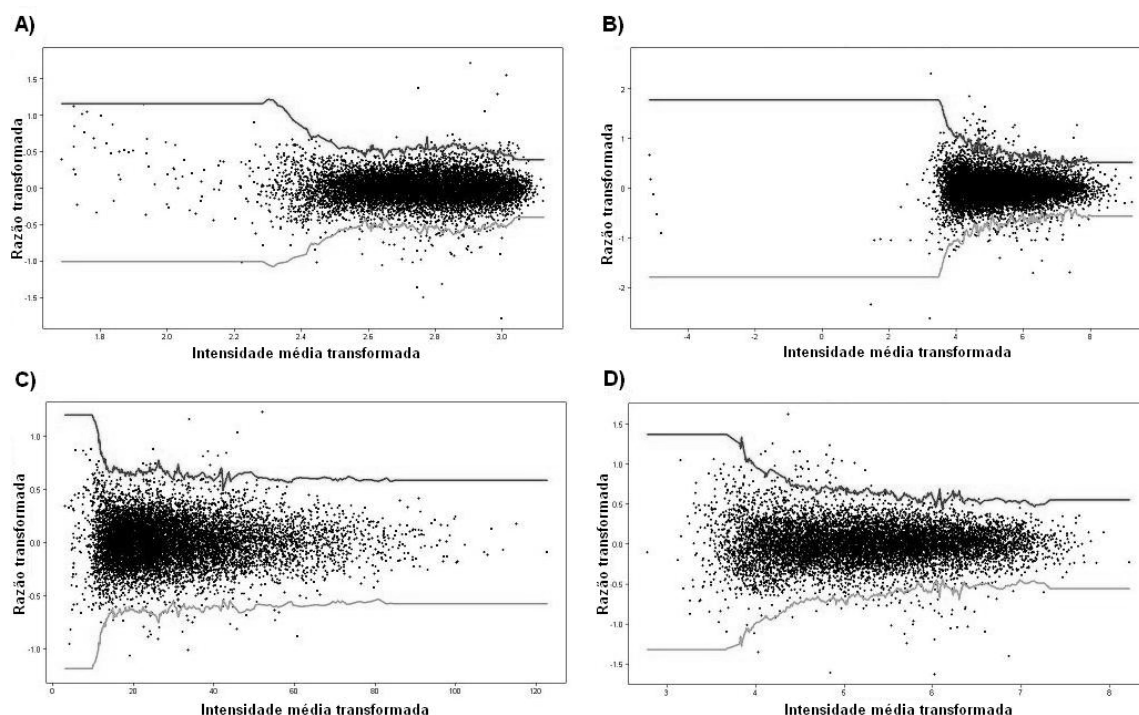


Figura 15: Gráfico gerado pelo método ISER dos valores dos dados do conjunto completo transformados da razão de expressão em função da Intensidade média para as lâminas 1 e 3 da análise com LPS (A e B) e da análise com LPS e Talidomida (C e D). As linhas indicam os limites de significância superior e inferior.

Tabela 4 – Genes selecionados pelas duas análises (interseção dos efeitos de LPS com Talidomida - Grupo 1), utilizando o conjunto completo.

Nº do gene	Símbolo	Descrição	Qualidade da lâmina
947	NM_004081	deletado in azoospermia; daz	Não
3299	NM_002663	fosfolipase d2; pld2	Não
3959	NM_001467	proteína de transporte glicose-6-fosfatase 1; g6pt1	Não
5103	NM_005617	proteína ribossomal s14; rps14	Não
5764	NM_033630	proteína contendo domínio scan; scand1	Não
6262	NM_001031	proteína ribossomal s28; rps28	Não
6699	NM_003333	precursor de proteína ubiquitina e ribossomal l40 precursor; uba52	Sim
7972	NM_012467	pré-proteína transmembrana triptase; tpsg1	Sim
8435	NM_000982	proteína ribossomal l21; rpl21	Não
8762	NM_001020	proteína ribossomal s16; rps16	Sim
9060	NM_002032	ferritina, polipeptídeo pesado 1; fth1	Sim
9365	NM_000990	proteína ribossomal l27a; rpl27a	Sim
9845	NM_000988	proteína ribossomal l27; rpl27	Não



**Tabela 5 – Genes selecionados somente na análise com Talidomida (sem efeito de LPS e com efeito de Talidomida - Grupo 2), utilizando o conjunto completo.**

<b>Nº do gene</b>	<b>Símbolo</b>	<b>Descrição</b>	<b>Qualidade da lâmina</b>
274	NM_007027	proteína de ligação topoisomerase (dna) ii ; topbp1	Não
474	NM_001943	pré-proteína desmogleina 2; dsg2	Sim
1197	NM_000524	receptor de 5-hidroxitriptamina (serotonina) 1a; htr1a	Não
1703	NM_001711	biglicana; bgn	Não
1707	NM_017712	proteína hipotética flj20208; flj20208	Não
1744	NM_014239	fator de iniciação de tradução eucariótica 2b, subunidade 2 (beta, 39kd)	Não
1972	NM_016454	proteína hipotética; loc51234	Não
1983	NM_000045	arginase, tipo 1; arg1	Não
2201	NM_018982	proteína hipotética; dj167a19.1	Não
2663	NM_013433	carioferina beta 2b, transportina; trn2	Não
2664	NM_015493	proteína dkfzp434n161; dkfzp434n161	Não
2704	NM_014758	produto do gene kiaa0254; kiaa0254	Não
3311	NM_005622	sa homóloga (associada à hipertensão do rato); sah	Não
3995	NM_003804	receptor de interação (tnfrsf) serina-treonina kinase 1; ripk1	Não
4283	NM_018549	proteína hipotética pro3073; pro3073	Não
4501	NM_012143	tuftelin-interacting protein; tip39	Não
4599	XM_029162	yotiao; akap9	Não
4669	NM_016079	proteína cgi-149; loc51652	Não
4775	NM_002486	proteína de ligação nuclear cap subunidade 1, 80kd; ncbp1	Não
4799	NM_002645	fosfoinositide-3-kinase, classe 2, alfa polipeptideo; pik3c2a	Não
4809	NM_002841	proteína tirosina fosfatase, tipo receptor, g; ptpg	Não
4908	NM_003635	n-deacetilase/n-sulfotransferase (heparan glicosaminil) 2; ndst2	Não
4950	NM_007217	morte celular programada 10; pdcd10	Não
5703	NM_024955	proteína hipotética flj23322	Não
6869	NM_003133	signal recognition particle 9kd; srp9	Não
7194	NM_006222	proteína (peptidil-prolil cis/trans isomerase) interação-nima tipo 1	Não
7389	NM_000391	tripeptidil peptidase i	Não
7406	NM_016118	ny-ren-18 antígeno; loc51667	Não
8958	NM_000606	polipeptídeo gamma componente de complemento 8; c8g	Não
9277	NM_001913	proteína de deslocamento tipo cut 1, ccaat (drosófila); cutl1	Não

**Tabela 6 – Genes selecionados na análise com LPS (com efeito de LPS e sem efeito de Talidomida - Grupo 3) , utilizando o conjunto completo.**

<b>Nº do gene</b>	<b>Símbolo</b>	<b>Descrição</b>	<b>Qualidade da lâmina</b>
174	NM_004765	célula b linfoma 7c; bcl7c	Não
475	NM_002002	antígeno de diferenciação do leucócito humano cd23	Sim
895	NM_018456	proteína de medula óssea não caracterizada 040	Não
1338	NM_021822	proteína forbolina-like mds019	Não
1826	NM_002500	diferenciação neurogênica 1; neurod1	Não
2203	XM_052761	proteína hipotética xp_052761; loc112451	Não
2269	NM_006332	proteína induzível de interferon gamma 30; ifi30	Não
2331	NM_001009	proteína ribossomal s5	Não
2449	NM_000998	proteína ribossomal L37a	Sim
2498	NM_000661	proteína ribossomal l9	Não
2868	NM_000994	proteína ribossomal l32	Não
3128	XM_017660	proteína hipotética xp_017660; loc95585	Não
3789	XM_006714	timeless drosophila homologncbi critical mRNA entry	Não
3909	NM_001015	proteína ribossomal s11	Não
4353	NM_014774	produto do gene kiaa0494	Não
4685	NM_003240	fator de crescimento transformante beta 4; ebf	Não
5558	NM_016483	proteína phd finger 7; hspc226	Sim
6565	NM_006328	rna binding motif protein 14; rbm14	Sim
7165	NM_021103	timosina beta 10; tmsb10	Sim
7785	NM_000975	proteína ribossomal l11; rpl11	Não
7930	NM_021927	proteína hipotética flj13220	Sim
8047	NM_002031	kinase fyn-relacionada; frk	Sim
8557	NM_017903	proteína hipotética flj20618	Sim
8725	NM_052985	domínio de repetição wd 10, isoforma 1; wdr10, homólogo ao transporte intraflagelar 122	Sim
8946	NM_012175	proteína f-box 3, isoforma 1; fbxo3	Sim
9198	NM_001958	fator de alongamento de tradução eucariótica 1 alpha 2; eef1a2	Não
9223	NM_006042	heparan sulfato d-glicosaminil 3-o-sulfotransferase 3a1	Não
9540	NM_000173	glicoproteína plaquetária ib alfa (polypeptide precursor); gp1ba	Sim

Os genes com expressão diferencial determinada pelo método ISER na análise com o conjunto reduzido estão indicados nas Tabelas 7 e 8. A Tabela 7 indica os genes cujos efeitos da ativação com LPS se mantiveram mesmo no tratamento com a talidomida e a Tabela 8 mostra os genes que tiveram expressão diferencial somente na ativação com

LPS. Nas duas lâminas analisadas com dados do conjunto reduzido, não foram encontradas intersecções entre os genes diferencialmente expressos somente no tratamento com Talidomida.

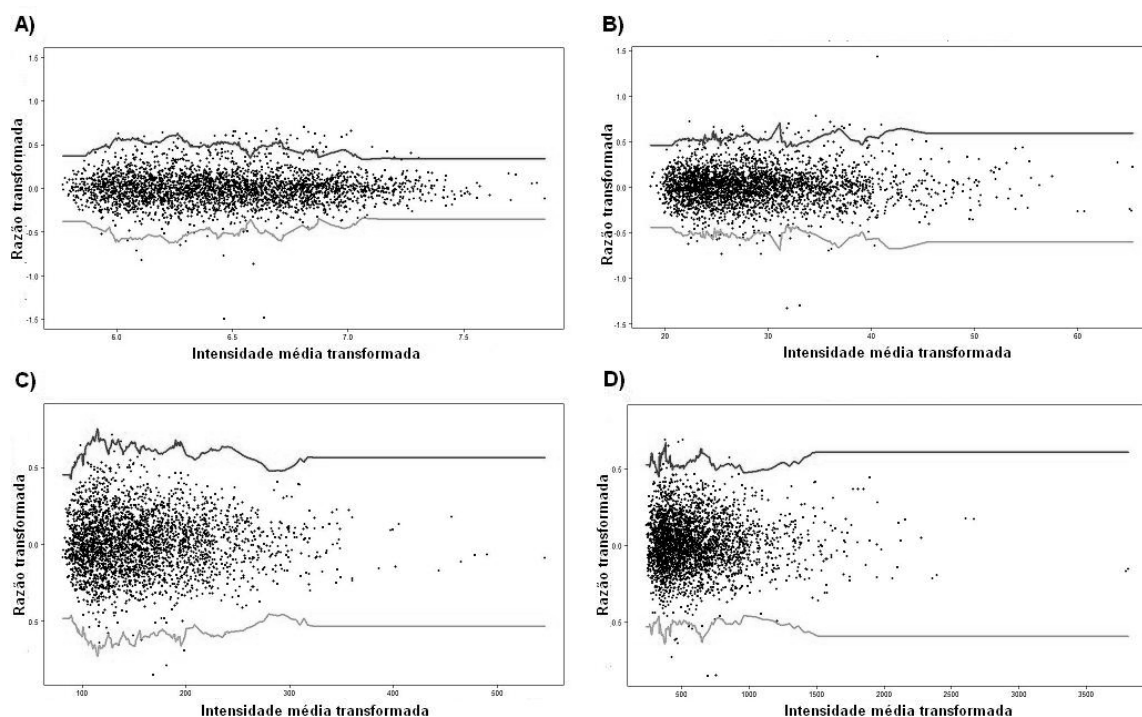


Figura 16: Gráfico do método ISER dos valores dos dados do conjunto reduzido transformados da razão de expressão em função da Intensidade média para as lâminas 1 e 3 da análise com LPS (A e B) e da análise com LPS e Talidomida (C e D). As linhas indicam os limites de significância superior e inferior.

Tabela 7 – Genes selecionados pelas duas análises (efeitos de LPS mantidos com Talidomida - Grupo 1), utilizando o conjunto reduzido.

Nº do gene	Símbolo	Descrição
8762	NM_001020	proteína ribossomal S16
9060	NM_002032	ferritina, polipeptídeo pesado 1; fth1

**Tabela 8 – Genes selecionados na análise com LPS (com efeito de LPS e sem efeito de Talidomida - Grupo 3), utilizando o conjunto reduzido. Os genes em negrito também foram selecionados pelo método MAANOVA.**

<b>Nº do gene</b>	<b>Símbolo</b>	<b>Descrição</b>
463	NM_001501	hormônio liberador de gonadotrofina 2
2449	NM_000998	proteína ribossomal L37a
3538	NM_018474	RNA não codificante de proteína 153
4353	NM_014774	produto do gene kiaa0494
5558	NM_016483	proteína phd finger 7; hspc226
<b>6565</b>	<b>NM_006328</b>	<b>rna binding motif protein 14; rbm14</b>
7165	NM_021103	timosina beta 10; tmsb10
<b>7930</b>	<b>NM_021927</b>	<b>GUF1 GTPase homóloga, proteína hipotética flj13220</b>
8047	NM_002031	kinase fyn-relacionada; frk
8557	NM_017903	proteína hipotética flj20618
<b>8688</b>	<b>NM_002310</b>	<b>precursor do receptor do fator inibitório de leucemia; lifr</b>
8725	NM_052985	domínio de repetição wd 10, isoforma 1; wdr10, homólogo ao transporte intraflagelar 122
<b>8946</b>	<b>NM_012175</b>	<b>proteína f-box 3, isoforma 1; fbxo3</b>
9353	AF_116656	pro1167 predicted protein of hq1167
9365	NM_000990	proteína ribossomal L27a

A maior parte dos genes selecionados pelo método ISER utilizando os dados do conjunto reduzido foi encontrada também na análise com o conjunto completo. Adicionalmente, ao utilizar os dados do conjunto reduzido, os métodos MAANOVA e ISER identificaram quatro genes em comum.

## 7. Discussão

No presente estudo, foram utilizados dados de expressão de 9.984 oligonucleotídeos representativos de genes humanos, obtidos a partir de experimentos de microarranjos de DNA. Usando dois métodos de análise de dados de microarranjo diferentes foi possível encontrar listas de genes cuja expressão foi considerada alterada. Os genes selecionados exibiram perfil de expressão gênica diferencial nas células PBMC ativadas por LPS após tratamento com talidomida. Isso possibilita uma análise dos mecanismos de ação desta droga em resposta a um quadro inflamatório, induzido nesse caso, pela presença do LPS.

O banco de dados recebido para análise apresentou uma grande quantidade de amostras de baixa qualidade nas lâminas hibridizadas. Isso foi constatado porque no emprego de um critério para verificar a relação entre o sinal e o ruído de cada *spot* na análise das 9984 genes, somente 1388 foram considerados satisfatórios. Utilizando o conjunto completo de dados na análise, os genes de qualidade inferior foram marcados. Esse enfoque permitiu que genes declarados diferencialmente expressos pudessem ser avaliados posteriormente quanto à qualidade do sinal original. Ao utilizar o conjunto reduzido, os genes de *spots* com qualidade do sinal inferior foram excluídos da análise.

A maior parte dos métodos de análise de dados de microarranjos conhecidos na literatura requer um maior número de réplicas para obter resultados estatísticos satisfatórios (LEE *et al*, 2000). No caso de número reduzido de amostras, a elaboração de um desenho de estudo para o experimento é uma etapa de extrema importância uma vez que é necessário extrair o máximo de informação a partir de um limitado número de lâminas. A aplicação desta tecnologia em experimentos de duas cores impõe alguns desafios na criação do desenho experimental, especialmente quando mais de dois grupos

são comparados. Nesse caso, os efeitos estimados dos tratamentos podem ser parcialmente confundidos com as diferenças entre as lâminas.

O experimento de microarranjo foi realizado com três lâminas para cada análise. No entanto, para trabalharmos com um desenho balanceado, a terceira lâmina, que era de marcação direta, foi excluída. Dessa forma, o desenho de estudo utilizado para a análise com o método ANOVA foi o desenho clássico de referência com uma lâmina de marcação direta e uma reversa para cada condição. O software MAANOVA ajusta o modelo de expressão gênica separando efeitos sistemáticos. No entanto, mesmo sendo amenizado pela inclusão da lâmina de marcação reversa, este desenho apresenta problemas aparentes porque em um experimento com poucas replicações os efeitos de corante se confundem com os efeitos de variação. Sendo assim, o efeito de interesse da variação dos níveis de expressão dos genes (VG) é confundido com os efeitos gene-específicos dos marcadores (DG).

Já a análise pelo método ISER apenas compara os valores obtidos nas duas condições comparadas, não separando efeitos sistemáticos. Consequentemente, não se justifica a utilização da lâmina de *Dye-Swap* com o objetivo de reduzir os efeitos de marcação neste método. Foram comparadas então as lâminas 1 e 3, ambas de marcação direta.

Os dois métodos de análise de dados de microarranjo utilizados possuem características diferenciadas no que diz respeito aos dados de entrada e aos fatores de análise considerados. O método ANOVA, por exemplo, requer um número maior de réplicas experimentais para obter um resultado mais robusto enquanto que o ISER afirma ter bom desempenho em experimentos com pouca ou nenhuma replicação. Adicionalmente, uma vez que na análise ANOVA foram utilizadas as lâminas 1 e 2 e no método ISER as lâminas 1 e 3, os resultados obtidos por cada um destes métodos não

podem ser comparados e nem se pode determinar a superioridade de um método em relação ao outro.

A utilização de dados brutos na análise de microarranjo para se estabelecer genes diferencialmente expressos é geralmente ineficiente (CUI *et al.*, 2003). Dados de microarranjos de duas cores frequentemente contêm variações sistemáticas que podem ser minimizadas pela transformação prévia dos dados.

Após a avaliação das amostras, eliminando ou marcando ou genes cujos *spots* apresentaram qualidade inferior, os dados de intensidade foram avaliados pelo *MA-plot* (Figuras 9 e 10) que revelou uma estrutura intensidade-dependente. Um padrão não linear no *MA-plot* indica a presença de efeitos sistemáticos e a necessidade de transformação nos dados. O método de transformação *lowess* centralizou a média dos experimentos em zero e foi capaz de diminuir a curvatura do *MA-plot* de cada lâmina.

A análise utilizando o software MAANOVA permitiu combinar a informação de diversos arranjos em uma única análise. Além disso, foi possível identificar fontes de variação nos dados atribuídos a outros fatores além da expressão diferencial dos genes. Quando o modelo ANOVA foi ajustado aos dados, foram obtidas estimativas para cada termo individual da análise e os valores de expressão relativa obtidos constituíram os dados normalizados uma vez que os efeitos relativos a arranjo, gene, *spot* e etc. foram removidos.

Uma consequência desta metodologia é lidar com um grande número de testes de hipótese, o que envolve uma relação entre especificidade e sensibilidade. Métodos de correção de p-valor fornecem um ajuste com prioridade na especificidade dos resultados. Neste trabalho, foram selecionados genes baseados nos p-valores não ajustados para comparações múltiplas. De acordo com BENDER e LANGE (2001), não há obrigatoriedade de realização de correções de testes múltiplos em estudos exploratórios. Adicionalmente, o número de réplicas do experimento era pequeno (um maior número de

réplicas aumentaria a precisão da análise) e o limite de p-valor considerado na metodologia foi muito baixo. Dessa forma, apesar de haver a possibilidade de selecionar um número elevado de genes falso-positivos, outros genes potencialmente importantes não foram indevidamente descartados. Uma vez que esses resultados serão confirmados posteriormente em laboratório, eliminando assim os genes falso-positivos, esta análise priorizou a sensibilidade em relação à especificidade.

Na análise utilizando o método ISER, os genes que mostraram expressão diferencial nas lâminas das células ativadas com LPS indicam relação com o mecanismo de ação desta toxina. Deste conjunto, os que também apareceram alterados nas lâminas de LPS tratadas com Talidomida (Grupo 1) sugerem que seus níveis de expressão não sofreram alteração pelos efeitos da droga, somente pelo LPS. Os genes que não se modificaram na presença do LPS e se mostraram alterados somente nas lâminas que sofreram tratamento com Talidomida, podem estar relacionados a outras propriedades além do mecanismo anti-inflamatório que a Talidomida pode possuir. Já os genes que apresentaram diferenças na expressão na ativação por LPS e este mesmo perfil não se confirmou na lâmina de LPS tratada com Talidomida (Grupo 3), podem sugerir que a Talidomida pode estar anulando a expressão diferencial correspondente ao efeito do LPS.

### **Genes Diferencialmente expressos**

O lipopolissacarídeo (LPS) possui um papel relacionado à ação indutiva de inflamação e a Talidomida um efeito anti-inflamatório. A análise de variância utilizando o aplicativo MAANOVA comparou a expressão gênica entre as duas condições, ou seja, supõe-se que os genes que foram selecionados como diferencialmente expressos teriam um papel na resolução do processo inflamatório (LPS/TAL em relação a LPS).



A proteína *Contactina 2* é uma proteína de membrana neuronal que tem a função de molécula de adesão celular. Ela pode ter um papel na formação das conexões dos axônios no sistema nervoso em desenvolvimento, ou ainda, estar envolvida na tumorigênese glial e prover um alvo potencial na intervenção terapêutica (NCBI Reference Sequences (RefSeq), 2009)..

A proteína codificada pelo gene *gadd45 beta* é induzida por TNF e está relacionada ao NFkB (fator nuclear kappa-B). O NFkB, induzido pelo TNFr (receptor de TNF), inativa a sinalização do gene JNK (c-Jun N-terminal kinase). A inibição de *gadd45* ocasiona uma ativação prolongada de JNK e citotoxicidade após a ativação de TNFr. O JNK é crucial para a morte celular programada induzida por TNFr e o mecanismo pelo qual o NFkB protege as células é pela sub-regulação da cascata de JNK através da ativação transcricional do *gadd45* (NCBI Reference Sequences (RefSeq), 2009).

A proteína de ligação Inositol 1,3,4,5-tetraquisfosfato aumenta a atividade da GTPase intrínseca das proteínas RAS permitindo o controle da proliferação e diferenciação celular (NCBI Reference Sequences (RefSeq), 2009).

A família do gene WNT compreende genes relacionados que codificam proteínas de sinalização relacionadas com a oncogênese e diversos processos de desenvolvimento, incluindo regulação de padronagem e morte celular durante a embriogênese (NCBI Reference Sequences (RefSeq), 2009).

O gene que codifica a *Proteína endometrial associada ao Progestogênio* possui três formas distintas que são encontrados no líquido amniótico, folicular e plasma seminal do sistema reprodutivo. Estas glicoproteínas têm papéis distintos e essenciais na regulação do ambiente uterino adequado para a gravidez e no controle da sequência de eventos no processo de fertilização (NCBI Reference Sequences (RefSeq), 2009).

O fator de transcrição de domínio epitélio-específico está relacionado à resposta inflamatória, desenvolvimento embrionário e diferenciação de células epiteliais. Seu gene precursor está envolvido em processos patológicos como carcinogênese e inflamações das vias aéreas (NCBI Reference Sequences (RefSeq), 2009).

O gene precursor do receptor do fator inibitório de leucemia codifica uma proteína que participa na formação de um complexo de receptor que regula a ação do fator inibitório de leucemia, uma citocina envolvida na diferenciação, proliferação e sobrevivência celular no adulto e no embrião. Mutações neste gene e seu promotor estão relacionadas com a síndrome Schwartz-Jampel tipo 2, doença pertencente do grupo de displasias ósseas e com o adenoma pleomórfico das glândulas celulares. Foram observadas diversas variantes deste gene codificando a mesma proteína (NCBI Reference Sequences (RefSeq), 2009).

O gene que codifica um membro da família de proteínas f-box constitui uma das quatro subunidades do complexo ligase da proteína ubiquitina chamado SCF (SKP1-cullin-F-box) que agem na ubiquitinação dependente de fosforilação (NCBI Reference Sequences (RefSeq), 2009)...

Apesar de não poder haver comparação entre os dois métodos de análise de dados de microarranjos empregados, conforme descrito previamente, os resultados obtidos pelo software MAANOVA e pelo ISER utilizando o conjunto reduzido ainda apresentaram quatro genes diferencialmente expressos em comum. Estes quatro genes, assim como outros genes selecionados que parecem estar relacionados com os mecanismos de inflamação do LPS e de ação da Talidomida, serão melhor investigados com relação a seus produtos em termos de processos biológicos, componentes celulares e funções moleculares.

## 8. Conclusão

A tecnologia de microarranjos de DNA se desenvolveu rapidamente para se tornar uma tecnologia de alto rendimento para a análise simultânea dos níveis relativos de expressão de milhares de genes individuais. Os dados originados desses ensaios frequentemente exibem variações grandes e inexplicáveis entre experimentos e plataformas, e a variedade de métodos disponíveis para analisar os dados, os quais incluem abordagens empíricas e estatísticas, podem resultar em interpretações drasticamente diferentes do mesmo conjunto de dados. A natureza complexa de um experimento de microarranjo introduz diversas fontes de variabilidade. Estas incluem extração da amostra, qualidade da amostra, desenho do arranjo, protocolo de hibridização, condições de hibridização e lavagem das amostras, *scanner*, processamento das imagens, normalização e análise dos dados, avaliação da qualidade dos dados e interpretação dos resultados. Todas essas variáveis contribuem para uma incerteza geral das conclusões tiradas de um experimento de microarranjo.

No presente estudo, visto que a maior parte dos *spots* foi considerada como de má qualidade, a probabilidade dos genes com variação mais significativa de expressão estarem nessas amostras não satisfatórias era grande. Ao utilizar a modelagem ANOVA, esses genes de qualidade inferior foram removidos da análise utilizando o conjunto reduzido, surgindo dessa forma novos resultados, que apesar de não terem significância estatística tão grande quanto os resultados do conjunto completo, são confiáveis devido à boa qualidade do sinal na lâmina.

Entre os genes selecionados estão precursores de proteínas relacionadas com desenvolvimento embrionário e neuronal, regulação de processos celulares como sinalização, divisão, morte celular, oncogênese entre outras. Grande parte dos genes

identificados neste estudo ainda não sofreram investigação suficiente quanto às suas propriedades e ainda são classificados como genes precursores de proteínas hipotéticas e ribossomais sem funções específicas atribuídas até o presente. Os genes que a princípio não parecem estar diretamente associados com essas funções necessitam de maior investigação quanto às vias intermediárias em que suas proteínas podem estar atuando.

Os resultados deste estudo necessitam de confirmação utilizando ferramentas moleculares como o PCR quantitativo em tempo real. Dessa forma, será possível a criação de um modelo de regulação da atividade da talidomida em células ativadas por LPS, baseando-se nas vias de regulação conhecidas dos genes diferencialmente expressos selecionados.

## 9. Referências Bibliográficas

AFORNALI, A., 2006, *Análise diferencial da expressão gênica de cultura primária de células musculares esqueléticas durante a infecção com taquizoítos de Toxoplasma gondii*". Dissertação de M. Sc., Instituto Oswaldo Cruz, Curitiba, PR, Brasil.

ALBERTS, B., JOHNSON, A., LEWIS, J. *et al.*, 2004, *Biologia Molecular da Célula*. 4ª ed. Porto Alegre, Artmed.

BENDER, R., LANGE, S., 2001, "Adjusting for multiple testing – when and how?" *Journal of Clinical Epidemiology*, v. 54, pp. 343-349.

BIRD, A., 2002, "DNA methylation patterns and epigenetic memory". *Genes Dev.* v. 16: pp. 6-21.

BORGES, L. G., FRÖEHLICH, P. E., 2003, "Talidomida – Novas perspectivas para utilização como antiinflamatório, imunossupressor e antiangiogênico". *Rev Assoc Med Bras.* v. 49(1), pp. 96-102.

BRADLEY J. R., 2008, "TNF-mediated inflammatory disease". *J Pathol* , v. 214, pp. 149–160.

CHEN, D., LIU, Z., MA, X., *et al.*, 2005, "Selecting genes by Test statistics". *Journal of Biomedicine and Biotechnology*, v. 2, pp. 132 – 138.

CLEVELAND, W. S., 1979, "Robust Locally Weighted Regression and Smoothing Scatterplots". *Journal of the American Statistical Association*, v. 74, n. 368, (Dec) Theory And methods section, pp. 829-836.

CORRAL, L. G. KAPLAN, G, 1999, "Immunomodulation by thalidomide and thalidomide analogues". *Ann Rheum Di*, v. 58 (Supl I) pp. I107–I113.

CUI, X., CHURCHILL, G. A., 2003, "Statistical tests for differential expression in cDNA microarray experiments". *Genome Biology*, v. 4, n. 4, article 210.

CUI, X., KERR, M. K., CHURCHILL, G. A., 2003, "Data Transformations for cDNA Microarray Data". *Stat. Appl. Genet. Mol. Biol.*, v. 2 (1), article 4 (Jun 18).

DE RISI, J., PENLAND, L., BROWN, P. O. *et al.*, 1996, "Use of a cDNA microarray to analyze gene expression patterns in human cancer". *Nature Genetics*, v.14, (Dec).

DRUMMOND, R. D., PINHEIRO, A., ROCHA, C. S., *et al.*, 2005, "ISER: selection of differentially expressed genes from DNA array data by non-linear data transformations and local fitting". *Bioinformatics*, v. 21 n. 24, pp. 4427–4429.

DUDOIT, S., YANG, Y., CALLOW, M. J., *et al.*, 2000, "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments". *Statist. Sinica*, v. 12, pp. 111-139.

FERNANDEZ-MARTINEZ, E., MORALES-RIOS, M. S., PEREZ-ALVAREZ, V., *et al.*, 2004, "Immunomodulatory effects of thalidomide analogs on LPS-induced plasma and hepatic cytokines in the rat". *Biochemical Pharmacology*, v. 68, pp. 1321–1329.

FOSS, N. T., 1997, "Aspectos imunológicos da hanseníase Medicina". In: *Simpósio: HANSENÍASE*, v. 30, Cap. II, pp. 335-339, Ribeirão Preto, Jul/Set.

FRANKS, M. E., MACPHERSON, G. R., FIGG, W. D., 2004, "Thalidomide". *The Lancet*, v. 363, pp. 1802-1811.

FRIEDMAN, N., LINIAL, M., NACHMAN, I., *et al.*, 2000: "Using Bayesian networks to analyze expression data". *J Comput Biol*, v. 7, n. 3-4, pp. 601–620.

GORDON, J. N., GOGGIN P. M., 2003, "Thalidomide and its derivatives: emerging from the Wilderness". *Postgrad Med J*. v. 79, n. 929, pp. 127–132.

GRIFFITHS, A. J. F., MILLER, J. H., SUZUKI, D. T., *et al.*, 2002, *Introdução à Genética*. 7 ed. Rio de Janeiro, Guanabara Koogan.

GRANT, G. R., MANDUCHI, E., STOECKERT, C. J., 2001, "Using nonparametric methods in the context of multiple testing to identify differentially expressed genes". In *Methods of Microarray Data Analysis*, Kluwer Academic Publishers, pp. 37–56, Boston.

HUGHES, T. R., MARTON, M. J., JONES, A. R., *et al.*, 2000, "Functional discovery via a compendium of expression profiles". *Cell*, v., 102, n. 19, pp. 109-126.

JABLONKA, W., GUZMÁN, S., RAMÍREZ, J., MONTERO-LOMELÍ, M., 2006, "Deviation of carbohydrate metabolism by the SIT4 phosphatase in *Saccharomyces cerevisiae*". *Biochimica et Biophysica Acta* v.1760, pp. 1281–1291.

KERR, M. K., MARTIN, M., CHURCHILL, G. A., 2000, "Analysis of Variance for Gene Expression Microarray Data". *J. Comput. Biol.*, v. 7, pp. 819-837.

KIM, S. Y., LEE J. W., SOHN, I. S., 2006, "Comparison of various statistical methods for identifying differential gene expression in replicated microarray data". *Statistical Methods in Medical Research*, v. 15, pp. 3–20.

LEE, M. T., KUO, F. C., WHITMORE, G. A. *et al.*, 2000, "Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations". *PNAS*, v. 97, n. 15, pp. 9834-9839.

LIU, Z-G., HAN, J., 2001, "Cellular Responses to Tumor Necrosis Factor". *Curr. Issues Mol. Biol.*, v. 3, n. 4, pp. 79-90.

LOGUINOV, A. V., MIAN, I. S., VULPE, C. D., 2004, "Exploratory differential gene expression analysis in microarray experiments with no or limited replication". *Genome Biology*, v. 5, n. 3, article R18.

LOO, L. H. 2004, *Identifying Differentially Expressed Genes in DNA Microarray Data*. Ph.D. dissertation, Drexel University, Philadelphia.



MANSOURIAN, R., MUTCH, D. M., ANTILLE, N., *et al.*, 2004, "The Global Error Assessment (GEA) model for the selection of differentially expressed genes in microarray data". *Bioinformatics*, v. 20, n. 16, pp. 2726 – 2737.

MARK, K. S., TRICKLER, W. J., MILLER, D. W., 2001, "Tumor Necrosis Factor-alpha Induces Cyclooxygenase-2 Expression and Prostaglandin Release in Brain Microvessel Endothelial Cells". *The Journal of Pharmacology and Experimental Therapeutics*, v. 297, n. 3, pp. 1051-1058.

MARRIOTT, J. B., MULLER, G., DALGLEISH, A. G., 1999, "Thalidomide as an emerging immunotherapeutic agent". *Trends Immunology Today*. v. 20, n. 12, pp. s38-s40.

MERCHIONNE, F., PEROSA, F., DAMMACCO, F., 2007, "New therapies in multiple myeloma", . *Clin Exp Med*. v. 3, pp. :83-97.

MUJAGIC, H., CHABNER, B. A., MUJAGIC, Z., 2002, "Mechanisms of Action and Potential Therapeutic Uses of Thalidomide". *Croatian Medical Journal*, v. 43, n. 3, pp. 274-285.

OLIVEIRA, R. B., MORAES, M. O., OLIVEIRA, E. B., *et al.*, 1999, "Neutrophils isolated from leprosy patients release TNF-a and exhibit accelerated apoptosis in vitro". *Journal of Leukocyte Biology*, v. 65, (Mar).

QUACKENBUSH, J., 2002, "Microarray data normalization and transformation". *Nature genetics supplement*, v. 32, pp. 496-501, (Dec).

REECE, R. J., 2004, *Analysis of Genes and Genomes*. 1 ed. West Sussex, John Wiley e Sons Ltd.

REUE, K., 1998, "mRNA Quantitation Techniques: Considerations for Experimental Design and Application", *J Nutr*, v. 128, n.11, pp.2038-2044.

SAMPAIO, E. P., CARVALHO, D. S., NERY, J. A. C., *et al.*, 2006, "Thalidomide: An Overview of its Pharmacological Mechanisms of Action". *Anti-Inflammatory e Anti-Allergy Agents in Medicinal Chemistry*, v. 5, n. 1, pp. 1-7.

SCHENA, M., SHALON, D., DAVIS, R. W., *et al.*, 1995, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray". *Science*, v. 270, n. 5235 (Oct), pp. 467–470.

SHANNON, E., NOVECK, R., SANDOVAL, F., KAMATH, B, 2008, "Thalidomide suppressed IL-1 beta while enhancing TNF and IL-10, whrn cells in hole blood were stimulated with lipopolysaccharide". *Immunopharmacol Immunotoxicol*, v.30, n. 3, pp. 447-457.

SILVERMAN, W. A., 2002, "The Schizophrenic Career of a "Monster Drug". *Pediatrics*, v. 110, pp. 404-406.

SMYTH, G. K., YANG Y. H. e SPEED, T., 2002, "Statistical Issues in cDNA Microarray Data Analysis". *Methods Mol Biol.*, v.224, pp. 111-136.

SOUTHERN, E. M., 1975, "Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis", *J Mol Biol*, v. 98, n.3, pp.503-517.

SPERANDEI, S. L. M., 2007, *Testes Estatísticos Aplicados a Microarranjos de DNA: Impacto dos métodos de Transformação e do Tamanho da Amostra*. Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

STOREY, J., TIBSHIRANI, R., 2003, "Sam thresholding and false discovery rates for detecting differential gene expression in DNA microarrays" In: *The Analysis of Gene Expression Data: Methods and Software*. New York, Springer.

STOUGHTON, R. B., 2005, "Applications of DNA microarrays in biology". *Annu. Rev. Biochem.*, v. 74, pp. 53–82.

TEO, S. K., RESZTAK, K. E., SCHEFFLER, M. A., *et al.*, 2002, "Thalidomide in the treatment of leprosy". *Microbes and Infection*, v. 4, pp. 1193–1202.

WU, H., KERR, M. K., CUI, X., CHURCHILL, G. A., 2003, "MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments". In: *The analysis of gene expression data: methods and software*, pp. 313 – 341, Springer.

YANG, Y.H., DUDOIT, S., LUU, P., *et al.*, 2002, "Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation". *Nucleic. Acids Res*, v. 30, n. 4, e15.