

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

*Loci de Suscetibilidade Genética à Artrite
Reumatoide Estabelecem Redes de Regulação
Transcricional com Outros Genes*

Guilherme Liberato da Silva

Ribeirão Preto
2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

GUILHERME LIBERATO DA SILVA

*Loci de Suscetibilidade Genética à Artrite
Reumatoide Estabelecem Redes de Regulação
Transcricional com Outros Genes*

Dissertação apresentada ao Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências, área de concentração: Genética

Orientador: Prof. Dr. Geraldo A. S. Passos

**Ribeirão Preto
2010**

Silva, Guilherme Liberato

Loci de suscetibilidade genética à artrite reumatoide estabelecem redes de regulação transcricional com outros genes

Ribeirão Preto, 2010

159 p. 30 cm

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto, USP,
Área de Concentração: Genética

Orientador: Passos, Geraldo Aleixo da Silva

1. Artrite reumatoide 2. Loci de suscetibilidade 3. cDNA microarrays 4. Redes gênicas

Apoio e Suporte Financeiro

Este trabalho foi realizado no Laboratório de Imunogenética Molecular do Departamento de Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo com apoio e suporte financeiro das seguintes entidades e instituições:

- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)
 - Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
 - Faculdade de Medicina de Ribeirão Preto (FMRP)
-

Dedico este trabalho

*Aos meus queridos pais, José e Maria Helena,
pela única certeza que carrego no peito!*

Agradecimentos

Agradeço especialmente ao Prof. Dr. Geraldo Aleixo da Silva Passos, pela oportunidade, confiança, incentivo e pelos ensinamentos passados ao longo de tantos anos de orientação.

A todas amigas do Laboratório de Imunogenética Molecular, pelo aprendizado e agradável convivência, principalmente à Dra. Cristina Moraes Junta, por ter cedido os dados de cDNA microarrays.

Aos docentes, alunos e funcionários do Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto, principalmente ao coordenador Prof. Dr. Ademilson Espencer Egea Soares, pelo apoio recebido.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro.

O grilo procura
no escuro
o mais puro diamante perdido.

O grilo
com as suas frágeis britadeiras de vidro
perfura

as implacáveis solidões noturnas.

E se o que tanto busca só existe
em tua límpida loucura

- que importa? -

isso
exatamente isso
é o teu diamante mais puro!

Mário Quintana

SUMÁRIO

ÍNDICE DE FIGURAS	iii
ÍNDICE DE TABELAS	v
RESUMO	vii
ABSTRACT	ix
1 INTRODUÇÃO	2
1.1 ARTRITE REUMATOIDE	2
1.1.1 <i>Definição e Epidemiologia</i>	3
1.1.2 <i>Etiologia</i>	3
1.1.3 <i>Fisiopatologia</i>	10
1.1.4 <i>Apresentação Clínica</i>	11
1.1.5 <i>Diagnóstico</i>	13
1.1.6 <i>Tratamento</i>	16
1.2 TECNOLOGIA DE cDNA MICROARRAYS	21
1.3 MOTIVAÇÃO	25
1.4 OBJETIVOS.....	28
1.4.1 <i>Objetivo Geral</i>	28
1.4.2 <i>Objetivos Específicos</i>	28
2 MATERIAL E MÉTODOS	31
2.1 CONSIDERAÇÕES INICIAIS	31
2.2 SELEÇÃO DE PACIENTES E CONTROLES.....	31
2.3 ASPECTOS ÉTICOS.....	32
2.4 NA BANCADA	32
2.4.1 <i>Tipificação do Alelo HLA-DRB1</i>	32
2.4.2 <i>Detecção de Anticorpos anti-CCP</i>	32
2.4.3 <i>Preparação das Amostras de RNA</i>	33
2.4.4 <i>Hibridização dos cDNA Microarrays</i>	33
2.5 NO COMPUTADOR	34
2.5.1 <i>Pipeline de Bioinformática</i>	34
2.5.2 <i>Quantificação das Imagens dos Microarrays</i>	36
2.5.3 <i>Normalização dos Dados</i>	36
2.5.4 <i>Análise Estatística</i>	39
2.5.5 <i>Inferência das Redes Gênicas</i>	42
2.5.6 <i>Funcionamento do GeneNetwork</i>	47
2.5.7 <i>Interpretação Textual das Redes Gênicas</i>	50
3 DELINEAMENTO EXPERIMENTAL	52
4 RESULTADOS	57
4.1 CARACTERIZAÇÃO DOS PACIENTES COM AR	57

4.2 DETERMINAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS.....	58
4.3 INFERÊNCIA DAS REDES GÊNICAS REGULATÓRIAS.....	61
4.4 DESCRIÇÃO DO PROGRAMA MANDALA.....	62
4.5 DETERMINAÇÃO DAS INTERAÇÕES REGULATÓRIAS CARACTERÍSTICAS DE CADA GRUPO DE PACIENTES...	71
5 DISCUSSÃO.....	87
5.1 CONSIDERAÇÕES INICIAIS.....	87
5.2 SOBRE AS REDES DE REGULAÇÃO TRANSCRICIONAL.....	90
5.3 SOBRE O PROGRAMA MANDALA.....	98
6 CONCLUSÕES.....	102
7 REFERÊNCIAS BIBLIOGRÁFICAS.....	104
8 APÊNDICE A.....	129
9 APÊNDICE B.....	143

ÍNDICE DE FIGURAS

FIGURA 1 - PADRÕES DE EVOLUÇÃO CLÍNICA DA AR	13
FIGURA 2 - METODOLOGIAS DE UM EXPERIMENTO TÍPICO COM CDNA MICROARRAYS	24
FIGURA 3 - DAS 4608 SEQUÊNCIAS DEPOSITADAS NA LÂMINA DE CDNA MICROARRAYS, 229 ESTÃO LOCALIZADAS EM <i>LOCI</i> DE SUSCETIBILIDADE À ARTRITE REUMATOIDE, O QUE CORRESPONDE A APROXIMADAMENTE 5% DO TOTAL.....	26
FIGURA 4 - <i>PIPELINE</i> MOSTRANDO AS TAREFAS ENVOLVIDAS NO PROCESSAMENTO DAS INFORMAÇÕES PRODUZIDAS PELOS EXPERIMENTOS DE MICROARRAYS.....	35
FIGURA 5 - GRÁFICO MA DOS DADOS PROVENIENTES DE UMA ÚNICA LÂMINA. AS LINHAS COLORIDAS SÃO LINHAS DE TENDÊNCIA DA EXPRESSÃO DE <i>SPOTS</i> DEPOSITADOS POR CADA UMA DAS AGULHAS DO ROBÔ	38
FIGURA 6 - GRÁFICO <i>BOX-PLOT</i> POR LÂMINA	39
FIGURA 7 - DIAGRAMA DE VENN DEMONSTRANDO O NÚMERO DOS GENES DIFERENCIALMENTE EXPRESSOS QUE SÃO EXCLUSIVOS E O DAQUELES COMPARTILHADOS ENTRE OS PACIENTES REUNIDOS NOS SEGUINTE GRUPOS: DAS28 > 5.1 (N=9), DMARDS + ANTI-TNF (N=5), HLA-DRB1 SE (N=10) E ANTI-CCP (N=9)	61
FIGURA 8 - TELA INICIAL DO PROGRAMA MANDALA.....	63
FIGURA 9 - TELA REFERENTE À PRIMEIRA OPÇÃO DO <i>MENU</i> PRINCIPAL DO PROGRAMA MANDALA	64
FIGURA 10 - GRÁFICO HIPOTÉTICO CONSTITUÍDO POR SEIS GENES, CUJOS NOMES SÃO A, B, C, X, Y E Z, E POR VINTE INTERAÇÕES REGULATÓRIAS ENTRE OS MESMOS	65
FIGURA 11 - ESTRUTURAS DE DADOS REFERENTES À LISTA COM OS NOMES DOS GENES E À MATRIZ REGULATÓRIA, CAPAZ DE DETERMINAR COMO ELES INTERAGEM UNS COM OS OUTROS.....	65
FIGURA 12 - TELA REFERENTE À SEGUNDA OPÇÃO DO <i>MENU</i> PRINCIPAL DO PROGRAMA MANDALA	68
FIGURA 13 - REDE DE REGULAÇÃO TRANSCRICIONAL INFERIDA A PARTIR DOS DADOS DE EXPRESSÃO GÊNICA DOS PACIENTES REUNIDOS NO GRUPO HLA-DRB1 SE	73
FIGURA 14 - REDE DE REGULAÇÃO TRANSCRICIONAL INFERIDA A PARTIR DOS DADOS DE EXPRESSÃO GÊNICA DOS PACIENTES REUNIDOS NO GRUPO ANTI-CCP	74
FIGURA 15 - REDE DE REGULAÇÃO TRANSCRICIONAL INFERIDA A PARTIR DOS DADOS DE EXPRESSÃO GÊNICA DOS PACIENTES REUNIDOS NO GRUPO DAS28 > 5.1	75

FIGURA 16 - REDE DE REGULAÇÃO TRANSCRICIONAL INFERIDA A PARTIR DOS DADOS DE EXPRESSÃO GÊNICA DOS PACIENTES REUNIDOS NO GRUPO DMARDS + ANTI-TNF	76
FIGURA 17 - GRÁFICO DEMONSTRATIVO DO NÚMERO DE GENES REPRESENTADOS NAS REDES DE REGULAÇÃO TRANSCRICIONAL QUE PARTICIPAM DE CADA PROCESSO BIOLÓGICO CELULAR (IDENTIFICADO NO BANCO DE DADOS GENE ONTOLOGY POR MEIO DE UMA SÉRIE NUMÉRICA	78
FIGURA 18 - GRÁFICO DEMONSTRATIVO DO NÚMERO DE GENES REPRESENTADOS NAS REDES DE REGULAÇÃO TRANSCRICIONAL QUE SÃO RESPONSÁVEIS POR CADA FUNÇÃO MOLECULAR (IDENTIFICADA NO BANCO DE DADOS GENE ONTOLOGY POR MEIO DE UMA SÉRIE NUMÉRICA	79
FIGURA 19 - GRÁFICO DEMONSTRATIVO DO NÚMERO DE GENES REPRESENTADOS NAS REDES DE REGULAÇÃO TRANSCRICIONAL QUE TERÃO SEU PRODUTO LOCALIZADO NUMA REGIÃO CELULAR CARACTERÍSTICA (IDENTIFICADA NO BANCO DE DADOS GENE ONTOLOGY POR MEIO DE UMA SÉRIE NUMÉRICA	80

ÍNDICE DE TABELAS

TABELA 1: CRITÉRIOS PARA O DIAGNÓSTICO DA ARTRITE REUMATOIDE ESTABELECIDOS PELO ACR EM 1987 ..	15
TABELA 2: NÚMERO DE SITUAÇÕES EXPERIMENTAIS NECESSÁRIAS PARA INFERIR UMA REDE REGULATÓRIA COM N GENES.....	48
TABELA 3: NÚMERO DE GENES IDENTIFICADOS DURANTE A ANÁLISE ESTATÍSTICA COMO SENDO DIFERENCIALMENTE EXPRESSOS PARA CADA UM DOS PACIENTES COM AR	59
TABELA 4: GENES CUJOS mRNAs PARTICIPAM DAS REDES DE REGULAÇÃO TRANSCRICIONAL APRESENTADAS NAS FIGURAS 13-16	81

Resumo

SILVA, G. L. ***Loci de suscetibilidade genética à artrite reumatoide estabelecem redes de regulação transcricional com outros genes***, 159p. Dissertação de Mestrado – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2010.

RESUMO

Estudos recentes de *linkage* permitiram identificar as regiões cromossômicas potencialmente associadas à ocorrência da artrite reumatoide (AR), tendo destaque o antígeno leucocitário humano (HLA)-DRB1, localizado no *locus* de suscetibilidade (LS) 6p21.3. No entanto, estima-se que sua contribuição seja parcial, de maneira que outros genes fora da região HLA desempenhem um importante papel na suscetibilidade à AR. Para testar esta hipótese, a tecnologia de cDNA microarrays foi utilizada para acessar o perfil transcricional de células mononucleares do sangue periférico de 15 pacientes com AR, os quais foram agrupados de acordo com seus padrões clínicos e moleculares, além do tipo de medicamento recebido. Por meio de programas de Bioinformática, foi possível identificar quais são os genes diferencialmente expressos, bem como reconstruir as interações regulatórias realizadas pelos respectivos mRNAs. Para cada grupo de pacientes, as redes de regulação transcricional resultantes demonstram *loci* de suscetibilidade interagindo com outros genes, como é o caso do FNDC3A (*fibronectin type III domain containing 3A*), cujos níveis de expressão gênica parecem depender da modulação de HLA-DRB1. Dado que fragmentos de fibronectina podem estimular mediadores da destruição de cartilagem e matriz extracelular, a evidência desta interação pode ajudar a elucidar a influência do HLA-DRB1 sobre a patogênese da AR.

Abstract

SILVA, G. L. **Genetic susceptibility loci in rheumatoid arthritis establish transcriptional regulatory networks with others genes**, 159p. Master's Dissertation – Medical School at Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2010.

ABSTRACT

Linkage studies have identified the human leukocyte antigen (*HLA*)-*DRB1* as a putative rheumatoid arthritis (RA) susceptibility locus (SL). Nevertheless, it was estimated that its contribution was partial, suggesting that other non-HLA genes may play a role in RA susceptibility. To test this hypothesis, we conducted microarray transcription profiling of peripheral blood mononuclear cells in 15 RA patients and analyzed the data, using bioinformatics programs (significance analysis of microarrays method and GeneNetwork), which allowed us to determine the differentially expressed genes and to reconstruct transcriptional networks. The patients were grouped according to disease features or treatment with tumor necrosis factor blocker. Transcriptional networks that were reconstructed allowed us to identify the interactions occurring between RA SL and other genes, for example, *HLA-DRB1* interacting with *FNDC3A* (fibronectin type III domain containing 3A). Given that fibronectin fragments can stimulate mediators of matrix and cartilage destruction in RA, this interaction is of special interest and may contribute to a clearer understanding of the functional role of *HLA-DRB1* in RA pathogenesis.

Introdução

1 INTRODUÇÃO

1.1 Artrite Reumatoide

“Era uma dor frequente. Às vezes, as mãos de Analícia Durval, então com 17 anos, amanheciam retesadas e com uma coloração estranha. Dos médicos, recebeu diversos diagnósticos, menos o certo. ‘Alguns diziam que era porque eu ficava muito no computador, outros falavam que era síndrome do túnel do carpo. Fiz muitos exames, mas nunca aparecia nada. Minha mãe chegou a pensar que eu estava inventando para chamar a atenção’, conta a estudante de música, hoje com 22 anos. Até que a professora de flauta transversal da jovem deu a dica: ‘Ela tem artrite reumatoide e me falou que eu podia ter a mesma coisa’. Um exame clínico e outro de sangue confirmaram a suspeita.

O diagnóstico pegou a família de surpresa. ‘Para nós, era uma doença de idosos’, diz Analícia. Embora seja mais comum em pessoas com mais de 35 anos, principalmente mulheres, a incidência da doença vem aumentando na faixa etária mais jovem, chegando a afetar até crianças. ‘Não se sabe exatamente o motivo. Mas deve-se considerar que as pessoas circulam cada vez mais e os fatores ambientais podem influenciar. Todos estão vivendo mais, se expondo mais. Várias questões podem contribuir’, diz a reumatologista Elisa Hacbarth Freire, do Hospital Professor Edmundo Vasconcelos, em São Paulo. A doença não tem causas definidas. Daí a dificuldade de se entender os mecanismos que levam ao avanço do problema sobre os mais jovens.

Para os pacientes, não é fácil aceitar a artrite reumatoide. Além dos sintomas físicos, há também os danos sociais e psicológicos, como aponta uma pesquisa internacional feita pelo instituto de pesquisa Opinion Health com 474 pacientes. Quarenta e dois por cento dos entrevistados disseram que se sentem sozinhos, e 40% relataram que a enfermidade afeta o relacionamento sexual. Um terço das pessoas ouvidas afirmou que a família não compreende como a doença atinge suas vidas.

‘É muito ruim. Para mim, que sou instrumentista, então, foi bastante traumático’, revela Analícia. Com os estudos prejudicados, ela teve de adiar os planos de ingressar no curso de música da Universidade Federal da Bahia, onde mora. Porém, com a dica da professora, a jovem conseguiu detectar a doença menos de um ano depois que os sintomas apareceram. ‘No início, não reagi bem ao tratamento. Mas minha médica mudou o remédio e hoje convivo muito bem com a doença. As dores quase sumiram’.

Especialista em psicologia médica, área que oferece assistência a pacientes e familiares que enfrentam situações difíceis por causa de doenças, Raimundo Azevedo explica que a dor contínua pode levar até mesmo à depressão, exigindo tratamento psiquiátrico. ‘Os pacientes sentem-se inutilizados. Há casos em que atividades simples, como abrir torneiras ou garrafas, tornam-se impossíveis devido à dor. Nesses momentos, é muito importante o apoio da família e dos amigos’, diz”.

Transcrição de matéria veiculada no jornal Correio
Braziliense em 27 de Janeiro de 2009.

1.1.1 Definição e Epidemiologia

Artrite reumatoide (AR) pode ser definida como uma doença autoimune crônica e sistêmica, caracterizada principalmente pela inflamação da membrana sinovial com destruição progressiva da articulação, causando deformação e perda de função (VENABLES & MAINI, 2006) (OMIM # 180300).

Trata-se do tipo mais comum de artrite inflamatória (FIRESTEIN, 2003) e da segunda doença reumática com maior prevalência, atrás apenas da osteoartrite, (IGNATAVICIUS, 2001), atingindo cerca de 0,5 - 1% da população mundial (ALAMANOS & DROSOS, 2005). O pico de incidência ocorre entre a quarta e quinta décadas de vida, mas pode desenvolver-se na infância e na velhice. Além disso, as mulheres têm de duas a quatro vezes mais chances de desenvolvê-la (GROSSMAN & BRAHN, 1997).

1.1.2 Etiologia

Embora grandes avanços alcançados pela comunidade científica tenham promovido um melhor entendimento sobre a artrite reumatoide, sua etiologia ainda permanece inconclusiva. Na realidade, os pesquisadores sugerem que o desenvolvimento da AR envolva uma complexa interação entre componentes ambientais e autólogos. Além disso, propõem a existência tanto de fatores disparadores como de outros que favoreçam a progressão da doença.

Dentre os componentes ambientais, os agentes infecciosos são os mais frequentemente citados como causadores de doenças autoimunes. Os mecanismos pelas quais uma infecção conduziria o estabelecimento de uma condição autoimunológica são suportados por cinco hipóteses principais (WUCHERPFENNIG, 2001), a saber:

- Mimetismo molecular: explicado pela possível identidade entre as sequências de aminoácidos de proteínas autólogas e invasoras. Neste caso, poderia acontecer o fenômeno conhecido como reação cruzada, no qual

antígenos próprios estruturalmente semelhantes aos antígenos patogênicos desencadeariam a produção de anticorpos autorreativos (ZHAO *et al.*, 1998; BACHMAIER *et al.*, 1999).

- Superantígenos: normalmente de origem viral ou bacteriana, seriam capazes de se ligar aos receptores de células T exclusivamente por meio do domínio variável da cadeia β . Esta forma diferenciada de ativação (policlonal) pelos superantígenos, aliada a sua habilidade de se ligar a uma grande variedade de moléculas MHC de classe II, induziria uma hiperexpansão de linfócitos T (até 25% da população celular original), com subsequente liberação de citocinas (SCHERER *et al.*, 1993).
- Expansão de epítomos: devido ao aumento característico no processamento e apresentação de antígenos resultante de episódios inflamatórios, a resposta imune celular acabaria estendendo sua especificidade para epítomos não relacionados com aqueles que causaram a infecção. Desta forma, ocorreria migração do alvo dos autoanticorpos para regiões da mesma proteína ou de proteínas distintas (intermoleculares) (TOPFER *et al.*, 1995).
- Ativação Bystander: infecções virais ocasionariam um aumento na produção local de citocinas, que, por sua vez, estimulariam a expansão clonal de linfócitos T autorreativos previamente ativados (e que até então seriam inofensivos devido à baixa titulação) (MURALI-KRISHNA *et al.*, 1998).
- Ativação policlonal: justificada pela infecção de células B por vírus linfotrópicos, como o da hepatite C (HCV), por exemplo. A proliferação de linfócitos B causaria uma produção aumentada de imunoglobulinas e o acúmulo de complexos imunes circulantes, o que poderia acarretar danos ao próprio tecido (FERRI & ZIGNEGO, 2000).

É importante observar que alguns mecanismos disparadores da condição autoimune não são diretamente dependentes dos agentes patogênicos, mas sim do processo inflamatório estabelecido com a infecção. Além disso, eles não são

mutuamente exclusivos e podem ser particularmente relevantes em diferentes estágios do desenvolvimento da doença (WUCHERPFENNIG, 2001).

No caso específico da AR, estudos como o de Von Bodungen e colaboradores demonstram que cabras infectadas com clones de lentivírus da artrite encefalite (*caprine arthritis encephalitis virus*, CAE) apresentaram, após 33 dias, lesões nas membranas sinoviais histologicamente muito semelhantes às daquelas de animais com história de artrite crônica de dois anos. Seus resultados apoiam a ideia de que a resposta imune humoral pode contribuir para a patogênese da artrite encefalite em caprinos, semelhante à AR humana (VON BODUNGEN *et al.*, 1998).

Ainda com relação aos componentes externos, várias referências na literatura apontam a prática do tabagismo como um importante fator de risco para o desenvolvimento da AR (HUTCHINSON *et al.*, 2001; CRISWELL *et al.*, 2002). Num estudo de coorte retrospectivo com 377.481 profissionais da saúde, das quais 7.697 eram pacientes com AR, Karlson e colegas concluíram que mulheres que tinham fumado por mais de vinte anos apresentavam um risco de 24% a 39% maior em desenvolver AR, além de um risco de 30% a 49% maior em desenvolver a doença com fator reumatoide positivo (FR+), sendo que tal variação era dependente do número de cigarros fumados ao dia (KARLSON *et al.*, 1999).

Stolt e colaboradores, por sua vez, desenvolveram um estudo do tipo caso-controle com 1576 indivíduos no qual constataram que fumantes e ex-fumantes possuíam um risco maior em desenvolver AR com FR+ (e não com FR-), quando comparados aos indivíduos que nunca tinham fumado. Entre aqueles que declararam sempre ter fumado, este risco chegava a ser 70% maior para as mulheres e 90% maior entre os homens. Em resumo, o risco relativo aumentado foi evidente em indivíduos que tinham fumado por mais de vinte anos, sendo proporcional à intensidade da dose (número de cigarros/dia). Além disso, o risco aumentado permanecia por até 19 anos após a interrupção do hábito tabagista (STOLT *et al.*, 2003).

Observe que ambos os trabalhos mencionados são concordantes quanto ao maior risco de soro-conversão do fator reumatoide (considerado um agravante no diagnóstico da AR) entre os indivíduos que fumavam por mais de duas décadas. Realmente, a maior agressividade na apresentação clínica da doença pode estar associada à prática do tabagismo, que, portanto, poderia ser utilizada para prever a progressão da AR (SAAG *et al.*, 1997).

Por outro lado, como bem suscitado por Criswell, o desenvolvimento de doenças complexas como a AR não deve ser justificado exclusivamente por fatores ambientais, mas sim os considerando juntamente com características intrínsecas ao organismo, como, por exemplo, a predisposição genética (CRISWELL *et al.*, 2006).

O envolvimento de fatores genéticos na patogênese da artrite reumatoide é amplamente fundamentado na literatura. Num estudo populacional finlandês, Aho e colaboradores demonstraram que a porcentagem de concordância para AR era de 12,3% para gêmeos monozigóticos (MZ), contra 3,5% obtida para gêmeos dizigóticos (DZ). Além disso, o risco relativo ajustado por idade e sexo foi calculado em 8,6 para gêmeos MZ e 3,4 entre pares DZ (AHO *et al.*, 1986).

Já no Reino Unido, 15,4% dos gêmeos MZ e 3,6% dos gêmeos DZ apresentavam concordância com relação à AR. Os autores ainda concluíram que o risco do segundo irmão MZ em desenvolver AR era diretamente proporcional ao tempo de apresentação clínica da doença no irmão inicialmente afetado (SILMAN *et al.*, 1993).

Estudos populacionais considerando gêmeos consistem numa abordagem clássica para determinar a influência genética no desenvolvimento de certa doença. No entanto, eles possuem o seguinte viés: quanto maior a prevalência da doença naquela população, maior as taxas absolutas de concordância entre gêmeos, independentemente da contribuição genética investigada.

Uma estimativa mais robusta é fornecida pelo cálculo da chamada *heritability* (SMITH, 1970), que quantifica a contribuição de fatores genéticos para o

desenvolvimento de uma doença, sem levar em consideração sua prevalência na população.

A partir do conjunto de dados de ambos os trabalhos mencionados acima, Alexander e colegas concluíram que aproximadamente 60% da responsabilidade sobre o desenvolvimento da AR pode ser atribuída a fatores genéticos, independentemente de sexo, idade, idade quando do surgimento da doença, ou severidade da mesma (McGregor *et al.*, 2000).

Mais recentemente, a meta-análise do conjunto de dados de quatro diferentes estudos de mapeamento genômico, totalizando 767 famílias e 964 pares de irmãos, permitiu identificar as regiões cromossômicas provavelmente associadas com a suscetibilidade à artrite reumatoide. Os loci de suscetibilidade são: 1p31.1-p13.3, 2q34-q35, 6q15-q23.2, 6p21.1-q15, 6p22.3-p21.1, 8pter-p22, 12p11.21-q15, 12p12.1-p11.21, 16p13-q12.2 e 18q12.3-q22.1 (CHOI *et al.*, 2006).

A principal região cromossômica de suscetibilidade à AR localiza-se no braço curto do cromossomo 6 (mais especificamente 6p21), no chamado Complexo Principal de Histocompatibilidade (*Major Histocompatibility Complex* - MHC), região que se estende por 4Mb e abriga um grande número de genes altamente polimórficos.

Variações do gene HLA-DRB1 têm sido particularmente associadas à ocorrência de AR em diferentes populações. Localizados na região conhecida como MHC de classe II, estes alelos determinam os subtipos HLA-DR4 (HLA-DRB1*0401/0404/0405/0408), HLA-DR1 (HLA-DRB1*0101/0102) e também HLA-DRB1*1001, *1402 e *1406.

Todos os alelos DRB1 compartilham uma região codificando uma sequência comum de aminoácidos entre as posições 70-74 da terceira região supervariada (HVR3), constituinte da cadeia β no heterodímero HLA-DR. Este padrão conservado de aminoácidos, QKRAA/QRRAA/RRRAA, geralmente é referido como epítopo compartilhado (*shared epitope* - SE).

O estudo de uma amostra populacional do leste da França identificou que a prevalência do alelo HLA-DRB1*0401 era significativamente maior em pacientes com AR (46,6%) do que em indivíduos controle (17,3%). Os subtipos mais comumente encontrados entre aqueles pacientes foram o HLA-DRB1*0101 (33,3% dos casos) e o HLA-DRB1*0401 (28,3%). Além disso, os autores concluíram que a ocorrência de dupla dose do SE estava associada a um risco 4,83 vezes maior em desenvolver a doença (TOUSSIROT *et al.*, 1999).

No Brasil, a análise de uma amostra populacional com 140 pacientes com AR e 161 indivíduos sadios revelou que a frequência dos alelos HLA-DRB1*0101, *0401, *0404, *0408, *0405 e *1001 era significativamente maior em pacientes com AR do que em controles. Considerando todos os alelos SE, sua frequência acumulada foi de 39,6% no conjunto de pacientes, contra 12,5% característica do grupo controle (OR=6,02). Por outro lado, considerando todos os alelos codificadores da sequência de aminoácidos DERAA (DRB1*0103, *0402, *1102, *1103, *1301, *1302 e *1304), os autores encontraram uma frequência acumulada de 19,3% no conjunto de pacientes e 37,2% no grupo controle (OR=0,4), reafirmando seu efeito protetor contra o desenvolvimento da AR. Os autores ainda concluíram que anticorpos anti-CCP (e FR+) estavam presentes em 64,5% dos pacientes SE-positivo, contra 44,7% de pacientes SE-negativo, sugerindo que anticorpos anti-CCP sejam responsáveis por intermediar a associação entre alelos HLA-DRB1 e AR (LOUZADA-JÚNIOR *et al.*, 2008).

Adicionalmente, vários trabalhos têm identificado uma especial associação entre a ocorrência de dupla dose do alelo HLA-DRB1*04 e uma maior severidade na apresentação clínica da AR, como manifestações extra-articulares, vasculite e quadro erosivo (GORMAN *et al.*, 2004; GORONZY *et al.*, 2004; TURESSON *et al.*, 2005).

Entretanto, de acordo com Deighton e colegas, os alelos HLA-DRB1 contribuem com apenas um terço do risco genético para AR (DEIGHTON *et al.*, 1989),

evidenciando que outros genes fora da região MHC de classe II desempenham um papel importante na suscetibilidade à doença.

Verpoort e colegas, por exemplo, encontraram que a frequência do alelo HLA-DR3 era significativamente maior entre pacientes com AR que não apresentavam anticorpos anti-CCP do que entre controles (OR=1,84), o que não foi observado na comparação de pacientes AR anti-CCP+ versus controles (OR=0,92). Situação semelhante foi verificada no subgrupo de pacientes com artrite indiferenciada. Segundo os autores, os resultados sugerem que o desenvolvimento de AR sem anticorpos anti-CCP esteja associado à ocorrência do alelo HLA-DR3, localizado na região MHC de classe III (VERPOORT *et al.*, 2005).

Tokuhiro e colaboradores, por sua vez, identificaram uma associação significativa entre AR e o polimorfismo de um único nucleotídeo (*single nucleotide polymorphism* - SNP) do SLC22A4, pertencente a um grupo de genes de citocinas localizado na posição 5q31. Codificador de um transportador de cátions, o gene SLC22A4 encontra-se superexpresso na artrite induzida por colágeno. O SNP diminui sua eficiência de transcrição devido à alteração na afinidade com o fator de transcrição RUNX1 (*Runt-related transcription factor 1*) (citobanda 21q22). Também foi observada uma forte associação entre AR e um SNP do RUNX1. Os autores sugerem que a regulação da expressão do gene SLC22A4 pelo RUNX1 esteja relacionada com a suscetibilidade à AR, representando um exemplo do efeito epistático entre dois genes sobre a doença (TOKUHIRO *et al.*, 2003).

Os parágrafos anteriores discorreram sobre a importância dos fatores genéticos como um dos componentes autólogos responsáveis por contribuir para o desenvolvimento da artrite reumatoide. Outra característica intrínseca possivelmente associada a este risco é a condição dos níveis hormonais do indivíduo. Algumas evidências que suportam esta ideia são a diferença na prevalência da AR entre homens e mulheres e a redução dos sintomas durante a gravidez.

1.1.3 Fisiopatologia

Nos pacientes com AR, o exame da membrana sinovial revela um intenso processo inflamatório caracterizado por hipertrofia do tecido e transformação funcional de suas células constituintes, os sinoviócitos. A membrana sinovial inflamada, denominada *pannus*, possui uma constituição marcante de populações de células dendríticas, fibroblastos sinoviais, macrófagos e linfócitos T, o que evidencia o envolvimento da imunidade celular na fisiopatologia da AR (FIRESTEIN *et al.*, 1987; WOOLLEY & TETLOW, 1997).

Sucintamente, a reação autoimune tem início com a ativação antígeno-específica de linfócitos T CD4⁺ por peptídeos ligados à superfície de células apresentadoras de antígenos. Quando o heterodímero composto por molécula MHC de classe II + peptídeo encontra-se com o receptor de célula T (TCR), ocorre sua diferenciação em linfócitos T auxiliares de tipo 1 (*T helper type 1* - Th1) e tipo 2 (Th2).

Responsáveis por eliminar patógenos intercelulares, os linfócitos Th1 também são capazes de acarretar lesões tissulares devido à indução da atividade de macrófagos e linfócitos T citotóxicos. Por outro lado, os linfócitos Th2 conduzem a secreção de autoanticorpos por células B autorreativas. De acordo com Panayi, os próprios linfócitos B apresentariam papel fundamental na secreção de citocinas pró-inflamatórias, produção do chamado fator reumatoide (FR) e ativação de células T (PANAYI, 2005).

Além disso, a maturação de linfócitos Th provoca a liberação de uma grande quantidade de citocinas pró-inflamatórias, como interferon- γ , interleucina-1 (IL-1), interleucina-6 (IL-6) e fator de necrose tumoral (TNF- α). Estas estimulam as células sinoviais a produzir enzimas hidrolíticas (incluindo colagenases, agreganases - proteínas que degradam agreganos -, catepsinas L e D derivadas de neutrófilos e metaloproteinases do tipo I, III e VIII), as quais levam à destruição progressiva da articulação. De fato, os níveis de IL-1 são maiores nos pacientes com AR que apresentam quadro de doença erosiva (FONG *et al.*, 1994).

1.1.4 Apresentação Clínica

Embora as manifestações clínicas da artrite reumatoide sejam muito variáveis, as mais frequentes são dor, edema articular simétrico e rigidez matinal de duração variável (sendo mais prolongada em processos inflamatórios mais intensos). No estágio inicial, as articulações acometidas são preferencialmente pequenas e periféricas, como radiocárpicas (punhos), metacarpofalangeanas e interfalangeanas proximais (mãos), e metatarsofalangeanas (pés). Já em estágios mais tardios, ocorre o comprometimento de grandes articulações, como ombros, cotovelos, quadris, joelhos e tornozelos (GRASSI *et al.*, 1998).

O envolvimento de tecidos periarticulares pode acontecer sob as formas de rupturas tendíneas, síndromes neurológicas compressivas (especialmente a síndrome do túnel do carpo), atrofia e fraqueza muscular, por exemplo.

As manifestações extra-articulares podem incluir nódulos reumatoides (presentes em metade dos pacientes com AR, estando relacionados a um pior prognóstico), vasculites (frequentemente nas extremidades inferiores, representam o tipo de manifestação mais grave, também indicando um pior prognóstico), neuropatias, pericardite, pneumonite intersticial, pleurite, glomerulonefrite, esclerite, ceratoconjuntivite seca (associada à síndrome de Sjögren), síndrome de Felty e anemia (GORDON *et al.*, 1973; TURESSON *et al.*, 1999; YOUNG & KODURI, 2007). Os pacientes ainda podem apresentar febre, mialgia, astenia, inapetência, perda de peso e suor nas extremidades (BERTOLO & MAGALHÃES, 2006).

A evolução clínica da doença também é bastante heterogênea. Existem casos que evoluem espontaneamente para a cura; outros acompanham progressivas deformidades do sistema locomotor, acarretando uma importante debilidade funcional. Alguns casos implicam o comprometimento de outros órgãos e sistemas, podendo levar ao óbito.

De maneira geral, a AR pode se estabelecer de acordo com os seguintes padrões: monocíclico, policíclico e progressivo. O primeiro caracteriza-se pelo curso autolimitado, no qual sinais e sintomas (principalmente dor e rigidez articular)

regridem após um período de 1-2 anos, com pouca ou nenhuma medicação. O padrão monocíclico ocorre em aproximadamente 20% dos pacientes, sendo mais frequente em homens, em indivíduos soronegativos para o fator reumatoide e entre aqueles com o menor número de articulações afetadas (MASI, 1983) (veja a figura 1 na próxima página).

O padrão policíclico, por sua vez, caracteriza-se pela evolução lenta e progressiva, marcada por períodos de repentina e intensa atividade inflamatória e outros de remissão clínica. Este tipo de curso intermitente ocorre em aproximadamente 70% dos pacientes. Já o padrão progressivo corresponde à forma mais agressiva da doença, caracterizada pelo comprometimento articular aditivo sem períodos de remissão, causando sérias deformidades (como anquiloses, desvio ulnar, dedos em forma de pescoço de cisne ou martelo) e morte prematura se não tratada a tempo. É o tipo de curso menos comum, encontrado em pacientes com o maior número de articulações afetadas quando do diagnóstico (MASI, 1983).

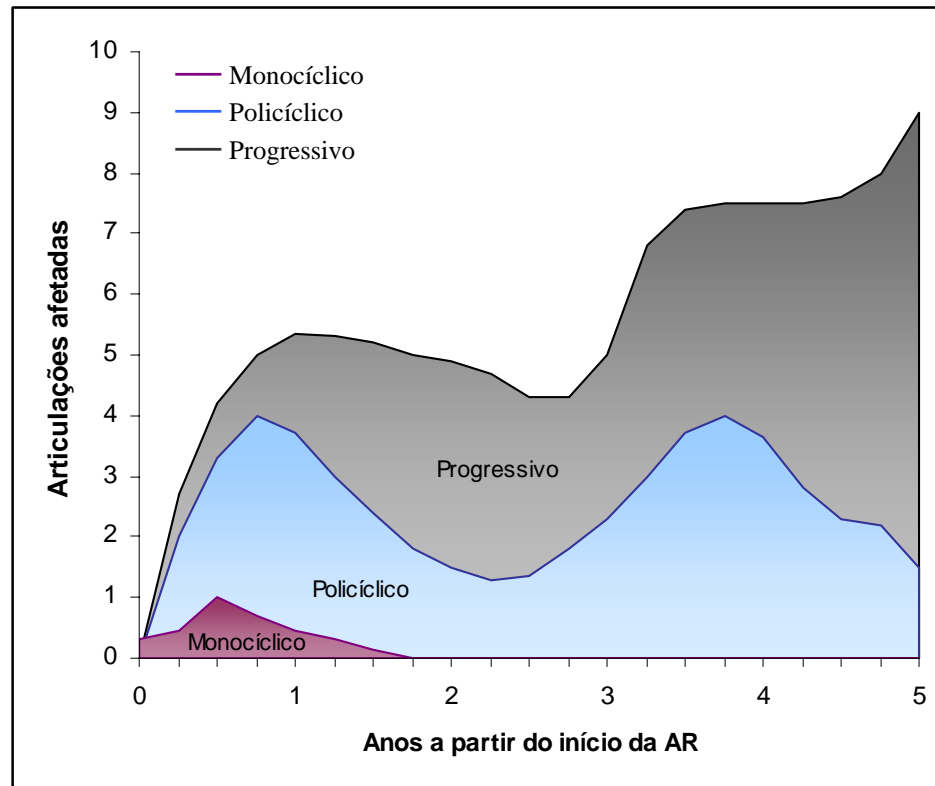


Figura 1 – Padrões de evolução clínica da AR. O gráfico indica que o número de articulações comprometidas ao longo dos anos varia de acordo com o tipo de curso da doença. No padrão monocíclico, ocorre envolvimento monoarticular de forma insidiosa, com remissão clínica espontânea. No padrão policíclico, a instalação da doença geralmente é insidiosa e poliarticular, evoluindo de maneira intermitente, com períodos de crise e melhora. O padrão progressivo é marcado pelo envolvimento poliarticular agudo, podendo resultar em deformidades incapacitantes (adaptado de HADLER & GILLINGS, 1985).

1.1.5 Diagnóstico

Embora as manifestações clínicas da AR sejam bem caracterizadas nos períodos mais avançados da doença, as manifestações clínicas observadas no início da AR podem dificultar sua identificação, abrangendo vários diagnósticos diferenciais. Na realidade, ainda não existe uma definição consensual com relação à AR inicial. Seu diagnóstico deve ser feito com base na presença de sinovite persistente afetando pelo menos três articulações, dolorimento à compressão das

metacarpofalangeanas ou metatarsofalangeanas e rigidez matinal de pelo menos trinta minutos (BERTOLO & MAGALHÃES, 2006).

Além da investigação das manifestações clínicas, achados laboratoriais e exames por imagem podem representar ferramentas úteis tanto no diagnóstico diferencial da AR quanto no prognóstico e acompanhamento desses pacientes.

O hemograma pode revelar alterações compatíveis com anemia normocrômica/normocítica, característica do início da doença, ou anemia microcítica/hipocrômica caso o processo inflamatório não seja controlado. As taxas de alfa-1-glicoproteína ácida, proteína C reativa (PCR) e velocidade de hemossedimentação (VHS), embora inespecíficas, apresentam-se elevadas durante o processo inflamatório da doença.

O fator reumatoide (FR), correspondente à imunoglobulina IgM anti-IgG, está presente em pelo menos 80% dos pacientes com AR. Apesar de sua baixa sensibilidade, 69% de acordo com (SHMERLING & DELBANCO, 1992) ou 66% segundo (GOLDBACH-MANSKY *et al.*, 2000), o FR pode auxiliar na confirmação do diagnóstico de AR quando levado em consideração junto com os demais aspectos clínicos do paciente. Além disso, em títulos elevados, pode ser indicativo de doença erosiva e de um maior número de manifestações extra-articulares (AGRAWAL *et al.*, 2007). Ainda de acordo com Goldbach-Mansky e colaboradores, o valor preditivo positivo do FR corresponde a 80%, semelhantemente àquele encontrado por Wolfe e colegas (WOLFE *et al.*, 1991).

Já os autoanticorpos contra peptídeos citrulinados cíclicos (*cyclic citrullinated peptides* - anti-CCP) possuem valor preditivo positivo maior que o FR. De fato, a curva ROC característica de anticorpos anti-CCP aproxima-se da extremidade superior esquerda, definindo uma área equivalente a 0.775; valor significativamente maior que a área de 0.577 obtida com a curva ROC do FR (SHOVMAN *et al.*, 2005).

Embora o exame radiográfico represente a abordagem clássica de investigação por imageamento na AR, as alterações evidentes aos raios-X (como osteopenia,

estreitamento do espaço articular, erosões ósseas marginais e deformidades) geralmente acontecem em estágios mais avançados da doença.

Desta forma, a ultrassonografia tem sido preferencialmente utilizada para detectar alterações articulares decorrentes da AR inicial, como derrame articular, proliferação sinovial e erosões corticais, anormalidades que não seriam visualizadas por meio de radiografias convencionais. Já a ressonância nuclear magnética (RNM), apesar do alto custo, é capaz de identificar precocemente o surgimento de erosões ósseas e processos inflamatórios de partes moles, como sinovites e tenossinovites.

Baseado em parâmetros clínicos, laboratoriais e radiológicos, o diagnóstico da AR obedece a sete critérios estabelecidos pelo *American College of Rheumatology* (ACR) em 1987 (ARNETT *et al.*, 1988), dos quais o paciente deve preencher pelo menos quatro. Veja a tabela a seguir:

Tabela 1: Critérios para o diagnóstico da artrite reumatoide estabelecidos pelo ACR em 1987.

1. Rigidez matinal articular de pelo menos uma hora
2. Artrite de três ou mais áreas articulares: mãos, punhos, cotovelos, joelhos, tornozelos e pés
3. Artrite das articulações das mãos
4. Artrite simétrica
5. Nódulos reumatoides
6. Fator reumatoide positivo
7. Alterações radiológicas: osteopenia periarticular e/ou erosões
<i>Obs.: os quatro primeiros critérios devem estar presentes há pelo menos seis semanas</i>

Outra importante ferramenta comumente utilizada para o diagnóstico da AR e avaliação da resposta ao tratamento é o chamado Índice de Atividade da Doença Baseado em 28 Articulações (*Disease Activity Score 28 - DAS28*), proposto por (PREVOO *et al.*, 1995) de maneira a simplificar a fórmula do DAS original, desenvolvida anteriormente pela *European League Against Rheumatism* (EULAR) (VAN DER HEIDJE *et al.*, 1993).

O cálculo do índice DAS28 leva em consideração vinte e oito articulações: ombros, cotovelos, punhos, metacarpofalangeanas, interfalangeanas proximais e joelhos, bilateralmente. Contam também a velocidade de hemossedimentação (VHS) e uma variável denominada Avaliação Global de Saúde (AGS), que é medida numa escala visual analógica variando de 0 a 100. A fórmula é definida do seguinte modo:

$$DAS28 = 56 * \sqrt{NAD} + 0,28 * \sqrt{NAE} + 0,70 * \ln(VHS) + 0,014 * AGS \quad (1.1)$$

na qual NAD é o número de articulações doloridas e NAE, o número de articulações edemaciadas.

Entende-se por dolorosa uma articulação que apresente algum grau de desconforto, não necessariamente dor intensa, quando avaliada pela dígito-pressão de sua interlinha ou por sua mobilização passiva (PINHEIRO, 2007).

Assim, o índice DAS28 apresenta uma escala contínua variando de 0 a 9.4 e geralmente possui uma distribuição gaussiana em populações de indivíduos com AR. O nível de atividade da doença pode ser classificado como baixo quando se obtém $DAS28 \leq 3.2$, moderado no intervalo $3.2 < DAS28 \leq 5.1$, e alto quando se encontra $DAS28 > 5.1$ (VAN GESTEL *et al.*, 1998). Além disso, quase todos os pacientes com doença em remissão apresentam um $DAS28 < 2.3$, o que não quer dizer que todos os pacientes com um índice desses estão em remissão. (FRANSEN & VAN RIEL, 2005). Sua correlação com o DAS original foi calculada em 0.97, sendo muito mais prático de ser aplicado (ALETAHA & SMOLEN, 2005).

1.1.6 Tratamento

O tratamento adequado da AR requer uma abordagem multidisciplinar que envolva, além da terapia medicamentosa, o estabelecimento de fisioterapia,

suporte psicossocial e mesmo a intervenção cirúrgica. O paciente e seus familiares devem ser orientados sobre os possíveis efeitos adversos da medicação estabelecida, bem como a possibilidade de reativação da doença. Também devem ser buscadas soluções para as dificuldades e limitações enfrentadas pelo paciente no seu cotidiano.

Drogas Sintomáticas

Dentre as drogas sintomáticas atualmente disponíveis, destacam-se os anti-inflamatórios não hormonais (AINHs) e os glicocorticoides, cujos principais objetivos são o alívio da dor e a redução do processo inflamatório já instalado na articulação.

O mecanismo de ação dos AINHs consiste na inibição da síntese de cicloxigenases (COX), enzimas que são responsáveis pela produção de prostaglandinas tanto nos processos fisiológicos (via COX-1) quanto em reações inflamatórias (caso da COX-2) (SMITH, 1992). Desta forma, embora não apresentem eficácia superior aos demais AINHs, os inibidores seletivos de COX-2 possuem a vantagem de causar menos efeitos colaterais (principalmente problemas gastrointestinais), pois não interferem na cadeia das prostaglandinas participantes de processos fisiológicos (WOLF *et al.*, 1993). Celecoxibe, Etoricoxibe e Valdecoxibe são os medicamentos pertencentes a este grupo. O Rofecoxibe foi retirado do mercado norte-americano por comprovadamente aumentar o risco de doenças cardíacas.

Os glicocorticoides, apesar de pertencerem à classe mais potente de supressores inflamatórios, devem ser utilizados com precaução devido aos vários efeitos adversos associados à terapêutica. Seus mecanismos de ação incluem a supressão da imunidade humoral e celular, a alteração na distribuição e função de leucócitos e a intervenção na cadeia de síntese de prostaglandinas (WILDER, 1996).

Pequenas doses de prednisona são indicadas principalmente no início do tratamento (período no qual são aguardados os efeitos de drogas antirreumáticas

modificadoras da doença), devendo ser reduzidas ou suspensas assim que possível. Doses maiores são reservadas aos casos mais graves, com manifestações sistêmicas. Injeções intra-articulares de glicocorticoides representam um importante adjuvante na redução da inflamação de juntas recalcitrantes.

Drogas Antirreumáticas Modificadoras da Doença

As chamadas DMARDs (do acrônimo em inglês *Disease-Modifying Antirheumatic Drugs*) correspondem aos medicamentos mais importantes no tratamento da AR. Diferentemente das drogas sintomáticas, estas promovem a remissão da atividade da doença, o que pode ser mensurado pelo retardo da progressão radiológica.

Dentre as DMARDs sintéticas, destacam-se:

- Metotrexato (MTX): é considerado o tratamento padrão-ouro devido à rápida ação, boa relação eficácia/toxicidade e ao baixo custo. Age bloqueando a síntese de DNA por meio da inibição da diidrofolato-reductase. Seu efeito terapêutico é obtido suprimindo a proliferação de linfócitos e a expressão de citocinas pró-inflamatórias como IL-1, IL-8 e TNF. Em contrapartida, possui implicações relacionadas ao comprometimento pulmonar, hepático e renal, além do potencial teratogênico;
- Leflunomida: apesar do custo mais elevado, apresenta eficácia comparável ao MTX (STRAND *et al.*, 1999). Seu mecanismo de ação consiste na inibição da diidrato-desidrogenase, levando à diminuição na síntese de pirimidinas. O efeito terapêutico baseia-se na supressão da proliferação e da ativação linfocítica, com subsequente modulação sobre IL-2, TGF- α e a produção de imunoglobulinas (MLADENOVIC *et al.*, 1995). A principal contraindicação refere-se aos pacientes em fase de procriação devido a sua propriedade teratogênica;
- Antimaláricos: são preferencialmente prescritos sob a forma de difosfato de cloroquina (DFC) e sulfato de hidroxicloroquina (HCQ). Embora os mecanismos de ação não estejam totalmente elucidados, são

reconhecidamente bastante eficazes, bem toleradas, seguras e de baixo custo. O efeito colateral mais importante envolve a toxicidade ocular, exigindo acompanhamento oftalmológico regular;

- Sulfassalazina (SSZ): resultante da ligação entre o antibiótico sulfapiridina e o anti-inflamatório 5-ácido-aminossalicílico, a SSZ age principalmente impedindo a expansão clonal de células B e reduzindo a concentração plasmática de linfócitos ativados. Normalmente é administrada em combinação com MTX e drogas antimaláricas. A adesão ao tratamento é diminuída em decorrência dos frequentes problemas gastrointestinais;
- Ciclosporina: devido à toxicidade característica, é reservada aos pacientes refratários (ou seja, não responsivos) aos demais tratamentos. Seu potencial terapêutico está relacionado à inibição da síntese de receptores de IL-1, IL-2 e TGF, além de restringir a interação entre macrófagos e linfócitos T, interferindo negativamente na ativação destes. Possui implicações quanto à hipertensão arterial e comprometimento hepático e renal.

Ciclofosfamida, Azatioprina, D-penicilamina, Clorambucil, Micofenolato Mofetil (MMF), Tacrolimus (FK506) e sais de ouro integram a lista de DMARDs sintéticas.

Mais recentemente, importantes avanços no campo da biotecnologia permitiram o desenvolvimento das DMARDs biológicas, também chamadas de agentes biológicos, que basicamente consistem de anticorpos monoclonais com alvos bastante específicos, além de proteínas solúveis com estruturas semelhantes a moléculas de superfície e receptores envolvidos na fisiopatologia da AR.

As principais características dos agentes biológicos disponíveis são apresentadas a seguir:

- Infliximabe: anticorpo monoclonal quimérico com ação especificamente antagonista ao TNF- α . Aproximadamente 25% de sua estrutura proteica tem origem murina (KNIGHT *et al.*, 1993), o que pode ocasionar efeito imunogênico quando administrado como monoterapia (WOLBINK *et al.*,

2006). Vários estudos têm demonstrado sua eficácia superior ao placebo quando administrados juntamente com MTX, além dos maiores efeitos terapêuticos obtidos por meio da associação Infiximabe + MTX em comparação aos produzidos por este último sozinho (MAINI *et al.*, 1999; LIPSKY *et al.*, 2000);

- Adalimumabe: anticorpo monoclonal recombinante, totalmente humano, com alta afinidade para o TNF- α , impedindo sua ligação aos receptores celulares p55 (tipo I) e p75 (tipo II) (SALFELD *et al.*, 1998). De acordo com Breedveld e colegas, os pacientes com AR grave (diagnosticados há menos de três anos) tratados com Adalimumabe + MTX apresentaram melhores efeitos terapêuticos (redução de sintomas, inibição na progressão radiográfica e mesmo remissão da doença) do que aqueles tratados exclusivamente com uma destas drogas (BREEDVELD *et al.*, 2006);
- Etanercepte: proteína solúvel resultante da fusão entre a porção extracelular do receptor p75 e a região constante da imunoglobulina IgG1 humana, sendo, portanto, capaz de se ligar tanto às citocinas TNF- α circulantes quanto àquelas ligadas aos receptores de células-alvo (MOHLER *et al.*, 1993). Vários trabalhos têm concluído que a combinação Etanercepte + MTX ou Etanercepte + SSZ produz uma resposta terapêutica mais eficaz do que a monoterapia com tais DMARDs sintéticas (WEINBLATT *et al.*, 1999; COMBE *et al.*, 2006);
- Abatacepte: proteína recombinante obtida por meio da fusão entre o domínio extracelular da molécula CTLA4 e a porção constante da IgG1 humana. É capaz de se ligar com alta afinidade aos receptores CD-80/86 das APCs (LINSLEY *et al.*, 1991), bloqueando a coestimulação positiva de CD28, suprimindo a ativação e até 95% da proliferação de linfócitos T (ALEGRE *et al.*, 2001; DAVIS *et al.*, 2007). Em estudos comparativos, mostrou-se mais eficaz que placebo em pacientes não responsivos ao MTX (KREMER *et al.*, 2006) ou a antagonistas de TNF- α (WESTHOVENS *et al.*, 2006);

- Rituximabe: anticorpo monoclonal quimérico responsável pela depleção de linfócitos B por meio da ligação aos antígenos CD-20 (EDWARDS & CAMBRIDGE *et al.*, 2001). É indicado em associação com MTX para pacientes com resposta inadequada aos inibidores de TNF- α (COHEN *et al.*, 2006);
- Anakinra: possui ação antagonista ao receptor tipo I de IL-1, impedindo, por concorrência, que esta citocina pró-inflamatória efetue transdução de sinais (MCINTYRE *et al.*, 1991). Ainda não é comercializado no Brasil.

1.2 Tecnologia de cDNA Microarrays

Os recentes avanços no campo da Genética Molecular têm aberto perspectivas grandiosas sobre a investigação dos complexos processos moleculares subjacentes aos organismos vivos. Questões que tradicionalmente eram consideradas no singular – por exemplo, quais as características estruturais e funcionais de um determinado gene? –, podem agora ser propostas no plural (D’HAESELEER *et al.*, 2000). Começando pelas novas técnicas de sequenciamento do DNA, entre as quais se destaca o pirosequenciamento (RONAGHI, 2001), é possível identificar as características estruturais de milhares de genes, fornecendo importantes indícios quanto as suas funções.

Dos métodos de mensuração da expressão gênica em larga escala, tais como a hibridização com cDNA microarrays (SCHENA *et al.*, 1995) ou o RTPCR quantitativo (BUSTIN, 2000), passando pela eletroforese bidimensional em gel (GYGI *et al.*, 2000) ou microarrays de anticorpos (GLÖKLER & ANGENENDT, 2003), existem inúmeras tecnologias especialmente desenvolvidas para avaliar as concentrações das moléculas de RNA e proteínas em diferentes níveis de profundidade e precisão, sendo, inclusive, capazes de determinar o perfil do transcriptoma ou do proteoma de um organismo inteiro (ZHU *et al.*, 2001; MOCKLER *et al.*, 2005).

Conceitualmente, o transcriptoma pode ser definido como o conjunto total de moléculas de RNA encontradas numa célula ou população delas num determinado instante (BERNOT, 2004). Essas moléculas incluem os chamados *coding* RNAs, conhecidos como mensageiros (mRNAs), e os *non-coding* RNAs, das quais fazem parte o transportador (tRNA), o ribossômico (rRNA), e também *small nucleolar* (sno)RNAs, microRNAs, *small interfering* (si)RNAs.

Porém, as metodologias destinadas à mensuração da concentração de transcritos em larga escala normalmente levam em consideração apenas as moléculas de mRNAs, pois são elas as responsáveis por carregar as instruções da síntese proteica, representando o nível mais elevado em termos da expressão do genoma. Dentre as principais técnicas com este propósito estão a chamada *Serial Analysis of Gene Expression* - SAGE e os cDNA microarrays.

Sucintamente, a metodologia do SAGE consiste na extração das moléculas de RNA de uma amostra celular, as quais são clivadas em pontos específicos de sua sequência para, depois de unidas numa única fita, serem clonadas com o auxílio de um vetor. A próxima etapa corresponde ao sequenciamento da fita amplificada, de maneira que a concentração de determinado transcrito é avaliada pela frequência na qual a sequência característica é identificada (VEUCULESCU *et al.*, 1995).

Desta forma, o SAGE baseia-se na amostragem dependente de sequências de mRNA, que, na verdade, não precisam ser conhecidas *a priori*. Esta é uma de suas principais vantagens, pois podem ser descobertos genes ou variantes de genes ainda desconhecidos. Por outro lado, seu custo é relativamente bastante elevado, o que inviabiliza sua utilização em experimentos de larga-escala.

A tecnologia de cDNA microarrays, por sua vez, é baseada no princípio de que dois fragmentos de cDNA (molécula de DNA produzida a partir de um mRNA, e, portanto, livre de íntrons) com sequências complementares tendem a hibridizar quando unidas num mesmo experimento (DUGGANS *et al.*, 1999).

A premissa dos experimentos utilizando cDNA microarrays é que as medidas de luminosidade para cada gene arranjado no substrato representam a concentração

das moléculas de mRNA correspondente. Os padrões de expressão biologicamente relevantes são identificados por meio da comparação, gene a gene, entre os níveis de concentração de seu mRNA nas diferentes condições experimentais investigadas (MELLO, 2004).

Os experimentos geralmente são conduzidos da seguinte forma: inicialmente, moléculas de cDNA correspondentes aos genes cuja expressão se deseja analisar são amplificadas por meio da reação em cadeia da polimerase (*polymerase chain reaction* - PCR). Os produtos resultantes são purificados e depositados ordenadamente sobre a superfície de um suporte sólido com o auxílio de um robô de alta precisão. Este substrato pode ser uma membrana de nylon ou uma lâmina de vidro quimicamente revestida, que ao final do processo possui milhares de pontos, cada um contendo concentrações picomolares de sequências de cDNA (alvo) que irão hibridizar com o mRNA complementar.

Para medir a quantidade relativa dos transcritos correspondentes às duas condições experimentais, as amostras celulares de mRNA são convertidas em cDNA por meio da transcriptase reversa, e marcadas com fluorocromos Cy3 (verde) e Cy5 (vermelho). No caso dos experimentos cujo suporte é constituído por membranas de nylon, utiliza-se marcação radioativa com ^{33}P .

Após a marcação, estas moléculas de cDNA (denominadas sondas) são depositadas na lâmina de modo a criar as condições necessárias para que o processo de hibridização aconteça. Passado certo tempo, a lâmina é lavada para retirar as moléculas de cDNA que não hibridizaram. A emissão de fluorescência é captada por um scanner de alta resolução que utiliza dois comprimentos de onda diferentes, produzindo uma imagem para cada canal verde e vermelho. Quando se recorre à marcação radioativa, um procedimento semelhante identifica a emissão de radiação específica de cada *spot*.

A figura a seguir representa as principais etapas envolvidas na condução de um experimento típico com cDNA microarrays.

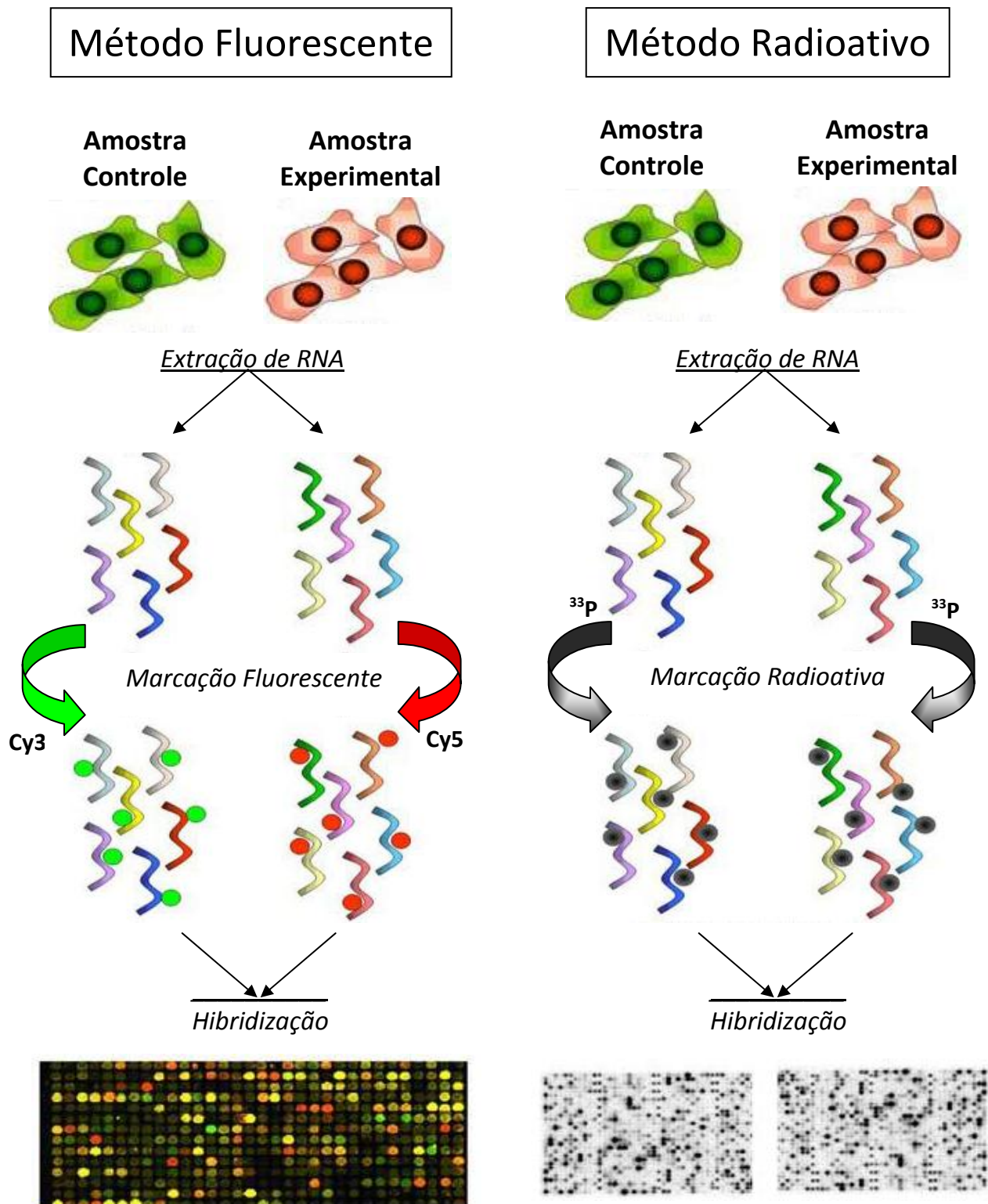


Figura 2 – Metodologias de um experimento típico com cDNA microarrays utilizando: A) marcação fluorescente com fluorocromos Cy3 (verde) e Cy5 (vermelho) e suporte constituído por lâmina de vidro; B) marcação radioativa com ^{33}P e suporte constituído por membrana de nylon (adaptado de SAKAMOTO-HOJO *et al.*, 2003).

1.3 Motivação

Conforme apresentado anteriormente, a meta-análise realizada por Choi e colaboradores permitiu a identificação das regiões cromossômicas que potencialmente conferem suscetibilidade à AR. Recapitulando, são elas: 1p31.1-p13.3, 2q34-q35, 6q15-q23.2, 6p21.1-q15, 6p22.3-p21.1, 8pter-p22, 12p11.21-q15, 12p12.1-p11.21, 16p13-q12.2 e 18q12.3-q22.1.

Considerando estas informações, é possível afirmar que do total de 4608 sequências alvo depositadas nas lâminas de cDNA microarrays projetadas no Laboratório de Imunogenética Molecular (cujas confecção consta da seção 2.4 de Material e Métodos), exatamente 229 estão localizadas em *loci* de suscetibilidade à AR.

Ou seja, aproximadamente 5% dos genes representados nesses cDNA microarrays pode ser encontrado ao longo de alguma das regiões cromossômicas associadas à ocorrência da AR. De maneira mais detalhada, pode-se afirmar que a maioria (33%) das 229 sequências pertence ao cromossomo 6, que é conhecido por abrigar a região MHC de classe II e, conseqüentemente, os alelos HLA-DRB1, os quais contribuem com um terço do risco genético à AR de acordo com Deighton e colegas.

Além disso, 55 destas 229 sequências podem ser encontradas no cromossomo 12, o que lhe atribui uma representatividade de 24% dentre a parcela de genes localizados em *loci* de suscetibilidade à AR. Já o cromossomo 1 fica com a terceira melhor posição, representado por 17% das sequências em *loci* de suscetibilidade à AR. A próxima figura sumariza estas observações e acrescenta outros detalhes.

- Genes em outras regiões cromossômicas
- Genes em LS à AR

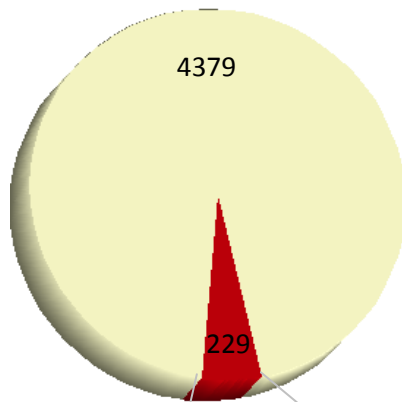
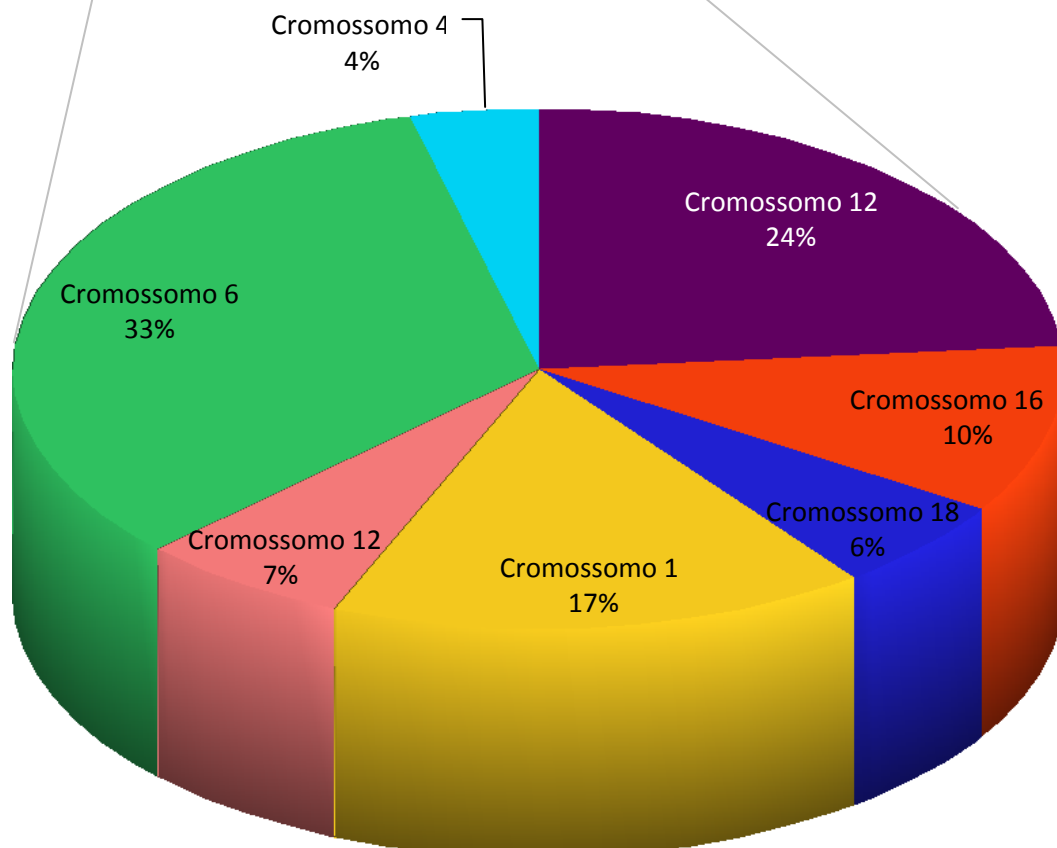


Figura 3 – Das 4608 sequências depositadas na lâmina de cDNA microarrays, 229 estão localizadas em *loci* de suscetibilidade à artrite reumatoide, o que corresponde a aproximadamente 5% do total. Esta parcela de genes, por sua vez, encontra-se distribuída ao longo de todos os sete cromossomos contendo regiões de suscetibilidade à AR, sendo que a maior concentração, 33%, é observada no cromossomo 6, o qual é conhecido por abrigar a região MHC de classe II e os alelos HLA-DRB1.



Tomadas as seguintes considerações:

- Contando com aproximadamente 5% do total de sequências depositadas na lâmina de cDNA microarrays, os *loci* de suscetibilidade à AR possuem uma representatividade suficiente para que se proponha a investigação de suas interações regulatórias com outras regiões cromossômicas;
- Já existem vários dados acerca da expressão gênica característica de pacientes com AR, obtidos por meio da utilização dos mesmos cDNA microarrays durante experimentos realizados pela Dra. Cristina Moraes Junta em seu trabalho de pós-doutoramento, os quais foram gentilmente disponibilizados para este projeto;
- Também é possível ter acesso a um conjunto de ferramentas computacionais (detalhadas na seção 2.5.1 *Pipeline* de Bioinformática) destinadas a manipular um grande volume de dados de expressão gênica a fim de identificar seus padrões diferenciais e realizar a modelagem de redes de regulação transcricional;
- Apesar de observada uma forte tendência de investigação sobre redes de regulação transcricional características de várias condições experimentais, principalmente de doenças humanas, ainda não há trabalhos divulgados na literatura científica acerca de redes gênicas relacionadas com AR;

Considerando ainda a relevância do tema e a possibilidade de contribuir com informações que ajudem a elucidar as vias de sinalização molecular e regulação transcricional da artrite reumatoide, a partir das quais poderia se direcionar o desenvolvimento de novas estratégias terapêuticas, inclusive para outras doenças autoimunes, puderam-se estabelecer, neste projeto, os seguintes objetivos:

1.4 Objetivos

1.4.1 Objetivo Geral

❖ *Inferir interações regulatórias transcricionais características de genes localizados em loci de suscetibilidade à artrite reumatoide utilizando-se programas para modelagem de dados de expressão gênica obtidos por meio da tecnologia de cDNA microarrays.*

Considerando que os níveis de expressão gênica e, conseqüentemente, as interações regulatórias poderiam variar de acordo com o nível de atividade da doença, com a presença de marcadores moleculares tais como o epítipo compartilhado HLA-DRB1 ou anticorpos contra peptídeos citrulinados cíclicos, e mesmo devido à administração medicamentosa, foram definidos os seguintes objetivos específicos:

1.4.2 Objetivos Específicos

❖ *Inferir uma rede de interações regulatórias transcricionais envolvendo todos os genes localizados em loci de suscetibilidade à artrite reumatoide que são diferencialmente expressos no grupo de pacientes com grande atividade da doença, ou seja, demonstrando o parâmetro clínico DAS28 > 5.1;*

❖ *Inferir uma rede de interações regulatórias transcricionais envolvendo todos os genes localizados em loci de suscetibilidade à artrite reumatoide que são diferencialmente expressos no grupo de pacientes com o epítipo compartilhado HLA-DRB1 em dose única ou dupla;*

❖ *Inferir uma rede de interações regulatórias transcricionais envolvendo todos os genes localizados em loci de suscetibilidade à artrite reumatoide que são*

diferencialmente expressos no grupo de pacientes apresentando anticorpos contra peptídeos citrulinados cíclicos;

❖ Inferir uma rede de interações regulatórias transcricionais envolvendo todos os genes localizados em loci de suscetibilidade à artrite reumatoide que são diferencialmente expressos no grupo de pacientes tratados com o agente biológico Infiximabe associado à droga antirreumática modificadora da doença Metotrexato.

Material e Métodos

2 MATERIAL E MÉTODOS

2.1 Considerações Iniciais

O trabalho aqui apresentado integra as atividades do programa de Pós-Graduação em Genética com ênfase em Bioinformática, oferecido pela Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP). O projeto foi realizado com a orientação do Prof. Dr. Geraldo Aleixo da Silva Passos, chefe do Laboratório de Imunogenética Molecular, cujas linhas de pesquisa incluem o estudo da expressão gênica diferencial em doenças autoimunes por meio da tecnologia de cDNA-microarrays.

É importante ressaltar que neste caso foram utilizados os resultados provenientes do trabalho de pós-doutoramento desenvolvido pela Dra. Cristina Moraes Junta sob a supervisão do Prof. Dr. Geraldo A. S. Passos, cujo título é: “Análise da expressão gênica em pacientes com artrite reumatoide do adulto estratificados o padrão de suscetibilidade genética” (FAPESP 01/09519-1).

2.2 Seleção de Pacientes e Controles

Foram coletadas amostras de sangue de quinze pacientes com artrite reumatoide provenientes do ambulatório do Serviço de Imunologia do Hospital das Clínicas de Ribeirão Preto (HC-FMRP), por meio da colaboração do Prof. Dr. Eduardo A. Donadi. Todos os pacientes satisfaziam os critérios estabelecidos pelo *American College of Rheumatology* para o diagnóstico da AR. Suas características e informações clínicas, obtidas a partir dos respectivos prontuários médicos, são apresentadas na seção 3.1 do capítulo de Resultados.

A elaboração do conjunto de controles, por sua vez, envolveu a coleta de amostras de sangue de onze indivíduos saudáveis, sem vínculo parentesco, doadores do banco de transplante de medula óssea da Fundação Hemocentro de Ribeirão Preto. As características apresentadas pelos indivíduos controle tornavam pertinente sua comparação com o grupo de pacientes.

2.3 Aspectos Éticos

Os aspectos éticos referentes à experimentação com seres humanos foram aprovados junto à Comissão de Ética em Pesquisa do Hospital das Clínicas de Ribeirão Preto (protocolos #8090/2001), e todos os indivíduos se propuseram a participar da pesquisa sob consentimento esclarecido.

2.4 Na Bancada

2.4.1 Tipificação do Alelo HLA-DRB1

As células mononucleares do sangue periférico (*peripheral blood mononuclear cells* – PBMC) dos pacientes com AR foram separadas por gradiente de densidade Ficoll-Hypaque (Sigma, St. Louis – MO, EUA), a partir das quais o DNA foi extraído e amplificado utilizando-se a reação em cadeia da polimerase (*polymerase chain reaction* – PCR). A tipificação do HLA-DRB1 foi realizada por meio da hibridização com *primers* específicos, disponíveis em kits comerciais (One-Lambda, Canoga Park – CA, EUA).

2.4.2 Detecção de Anticorpos anti-CCP

A detecção de anticorpos IgG anti-CCP2 foi realizada por meio de kits comerciais ELISA contendo peptídeos sintéticos de acordo com as instruções do fabricante

(Quanta Lite anti-CCP2 Inova, San Diego - CA, EUA). As amostras de soro que apresentavam titulação maior que 20U/ml foram consideradas positivas.

2.4.3 Preparação das Amostras de RNA

Amostras de RNA total das PBMC foram extraídas por meio do reagente Trizol[®] (Invitrogen, Carlsbad - CA, EUA) seguindo as normas do fabricante. A qualidade e integridade das amostras foi avaliada por eletroforese em gel de agarose sob condições padrão, e por análise Northern-blot utilizando-se sondas de oligonucleotídeos marcados com ³²P reconhecendo a fração 28S do rRNA.

As sondas de cDNA foram preparadas por transcrição reversa utilizando-se 10µg de RNA total, sendo posteriormente marcadas com fluorocromo Cy3 por meio do kit de marcação CyScribe Post-Labeling (GE Healthcare Life Sciences).

Como referência para o procedimento de hibridização, foram usadas quantidades equimolares de cDNA obtido a partir do RNA total proveniente de um *pool* de quatro tipos distintos de linhagens celulares humanas: astrocitoma cerebral (U343), útero (Hela), laringe (HEp-2) e linhagem linfocítica (Jurkat). Esta estratégia permite estimar a quantidade relativa de cDNA alvo em cada *spot* do microarray, além de possibilitar a normalização e a comparação entre experimentos independentes.

2.4.4 Hibridização dos cDNA Microarrays

Os microarrays foram construídos utilizando-se 4.608 sequências de cDNA depositadas sobre lâminas de vidro revestidas Ultra GAPS (# 40015, Corning[®], New York - NY, EUA) por meio do sistema Generation III Array Spotter (Amersham Molecular Dynamics, Sunnyvale - CA, EUA).

As 4.608 sequências foram selecionadas a partir de uma biblioteca de cDNA de cerca de 9.000 clones humanos (*expressed sequence tags* - ESTs) provenientes do Consórcio IMAGE (LENNON *et al.*, 1996). A biblioteca foi gentilmente cedida pela

Dra. Catherine Nguyen, do Centre d'Immunologie de Marseille-Luminy (Marseille, França).

A reação de hibridização foi realizada num sistema automatizado (ASP, Amersham Biosciences) por um período de 15 horas, do qual se seguiu a aquisição das imagens dos microarrays pelo scanner confocal a laser Generation III (Amersham Biosciences).

2.5 No Computador

2.5.1 *Pipeline* de Bioinformática

Partir de imagens em tons de cinza referentes aos experimentos de microarrays e chegar a imagens representativas de redes gênicas regulatórias requer considerável transformação de dados. Todo o percurso em busca de informações biológicas relevantes envolve um grande processamento computacional, que, basicamente, pode ser estruturado nas seguintes etapas: 1º) quantificação das imagens dos microarrays, 2º) normalização dos dados, 3º) análise estatística para determinar os genes diferencialmente expressos e 4º) inferência das redes gênicas.

A definição e organização do conjunto de tarefas que devem ser executadas em cada etapa do processamento permitem a construção de um *pipeline* de Bioinformática, como o apresentado na figura 4 a seguir.

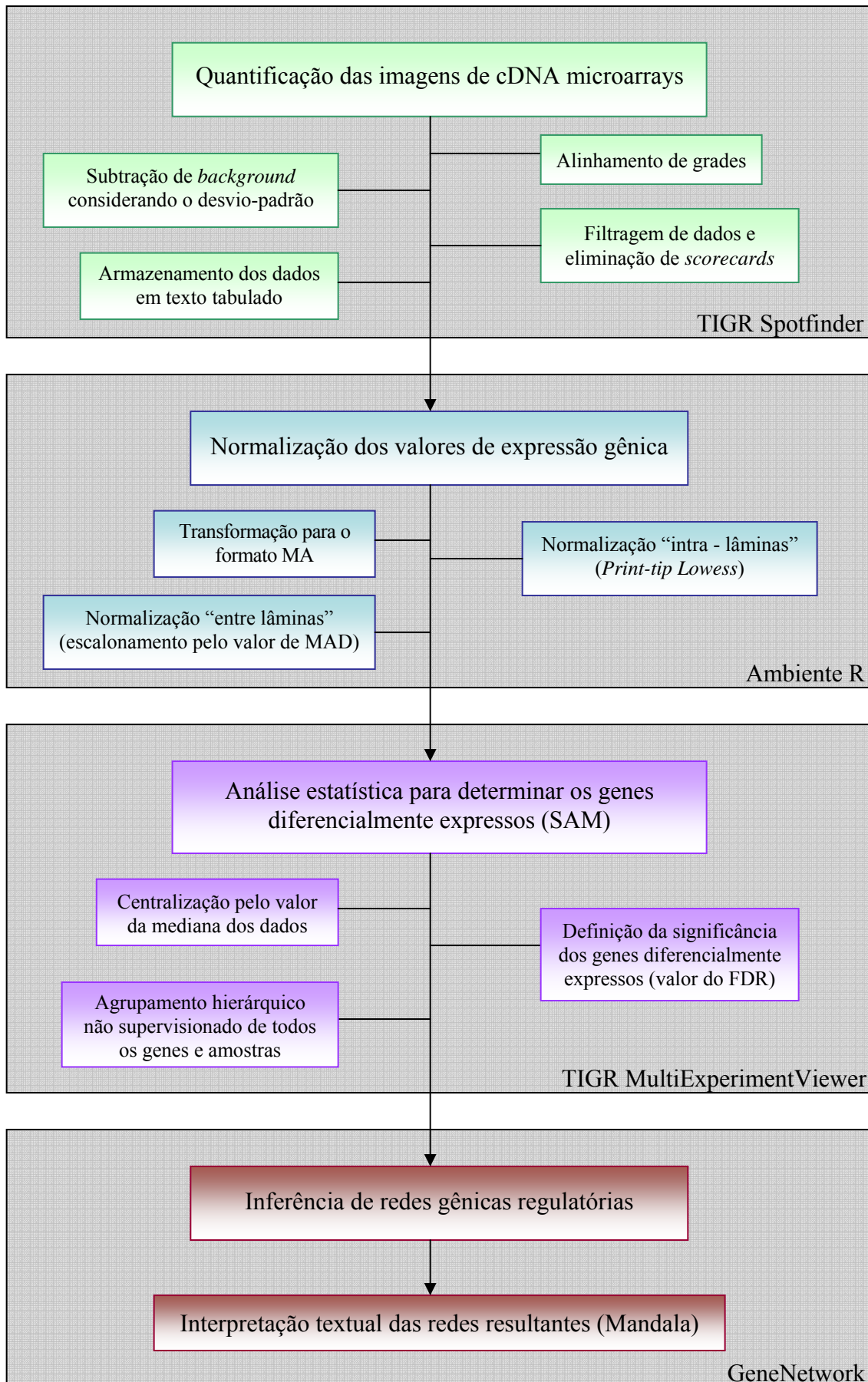


Figura 4 – Pipeline mostrando as tarefas envolvidas no processamento das informações produzidas pelos experimentos de microarrays. 35

2.5.2 Quantificação das Imagens dos Microarrays

O primeiro passo na análise dos microarrays refere-se à identificação dos níveis de expressão gênica característicos das condições experimentais em questão. O procedimento de quantificação, portanto, corresponde à atribuição de valores numéricos quantitativos da intensidade de fluorescência de cada ponto na lâmina. Ele foi realizado com auxílio do software TIGR Spotfinder v. 2.2.3 (SAEED *et al.*, 2003), disponível para download na página <http://www.tm4.org/spotfinder.html>.

A quantificação envolveu a confecção e alinhamento de grades, correção do sinal do background considerando o valor do desvio-padrão, eliminação de *scorecards* e filtragem dos dados levando-se em consideração a saturação, formato e tamanho dos pontos. Deve-se destacar que a quantificação foi responsável por garantir a qualidade dos dados repassados para as etapas posteriores.

2.5.3 Normalização dos Dados

O procedimento de normalização consiste na aplicação de métodos desenvolvidos com o objetivo de corrigir erros sistemáticos inerentes aos experimentos com cDNA microarrays. Dentre as possíveis fontes de erro dessa tecnologia, estão: alteração na eficiência de deposição das agulhas ao longo do *array*, variação na eficiência de incorporação por diferentes fluorocromos, variação na emissão de fluorescência quando da excitação pelo *laser* do scanner, divergência nas propriedades de escaneamento devido a diferenças no ajuste dos canais, alteração na medição da intensidade de fluorescência em áreas distintas do *array* (talvez devido a irregularidades na superfície da placa), ou até divergência nas condições ambientais do robô em diferentes rodadas de deposição (SMYTH & SPEED, 2003).

Portanto, a normalização é essencial para extrair medidas confiáveis das intensidades de fluorescência e para garantir que as diferenças observadas de fato reflitam uma expressão gênica diferencial e não um artefato experimental (YANG & THORNE, 2003).

O processo de normalização foi realizado por meio do ambiente R, que consiste numa linguagem de programação e também um ambiente integrado para a realização de cálculos estatísticos e a produção de gráficos (IHAKA & GENTLEMAN, 1996). Disponibilizado no endereço <http://cran.r-project.org/>, o software R provê um ambiente destinado à execução de um amplo conjunto de funções estatísticas, sendo flexível para suportar pacotes de funções desenvolvidos para aplicações específicas, como o Bioconductor (GENTLEMAN, 2004) para tratar dados genômicos e o AROMA (BENGTSSON, 2004) para análise de dados de microarrays.

Os métodos de normalização utilizados neste projeto foram o *Print-tip Lowess* (*Lowess* por agulhas) e o reescalonamento dos valores M pelo valor de MAD (*Median Absolute Deviation*), ambos contidos no pacote AROMA.

A correção *Lowess* (*Locally Weigthed Non-Linear Regression*) é aplicada para a detecção de desvios sistemáticos do *plot* R-I, corrigindo-os por uma série de regressões locais em várias janelas de dados que se sobrepõem ao longo do *array*, juntando em seguida essas regressões de modo a formar uma curva suave.

A regressão *Lowess* é um teste avançado, sendo geralmente realizado em quatro passos: a) cálculo do *log* da intensidade média e do *log* da razão para cada elemento; b) produção do gráfico MA; c) aplicação da regressão *Lowess* aos dados e d) cálculo do *log* da razão normalizada para cada elemento, sendo que este é obtido pela subtração do valor de correção (calculado para cada elemento) do valor da razão em *log* do valor original.

A regressão *Print-tip Lowess* é muito similar ao método acima descrito, com a diferença de não ser aplicada ao *array* como um todo, e sim em cada *grid* do *array* separadamente. A normalização “intra-slides” é muito útil quando existem erros introduzidos por diferentes agulhas de deposição do robô (erros espaciais), bem como quando os dados provenientes de diferentes fluorocromos possuem características diferentes ao serem depositadas por agulhas diferentes (STEKEL,

2003). A figura 5 mostra o efeito da normalização *Print-tip Lowess* sobre o gráfico MA.

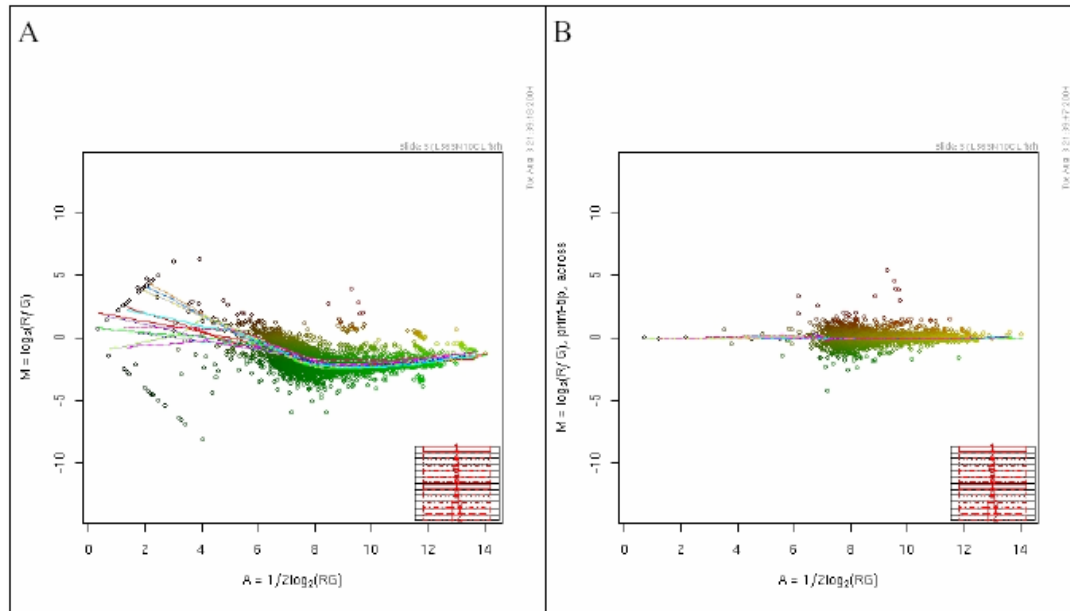


Figura 5 – Gráfico MA dos dados provenientes de uma única lâmina. As linhas coloridas são linhas de tendência da expressão de *spots* depositados por cada uma das agulhas do robô. A: Gráfico MA dos dados antes dos procedimentos de filtragem e normalização. B: Gráfico MA após o processo de normalização *Print tip Lowess*, no qual erros espaciais inseridos pela diferença na deposição do produto de PCR realizada por cada agulha são corrigidos (adaptado de MELLO, 2004).

Uma vez corrigidas as possíveis fontes de erro dentro de um mesmo array, ainda é necessário aplicar métodos de normalização que tornem diferentes experimentos comparáveis entre si. A normalização “entre slides” foi realizada pelo método de escalonamento, visando reescalonar os valores (M) dos vários microarrays de maneira que apresentem o mesmo desvio absoluto da mediana (*Median Absolut Deviation* - MAD) (MELLO, 2004). A figura 6 a seguir ilustra por meio de um *box-plot* os efeitos deste tipo de normalização.

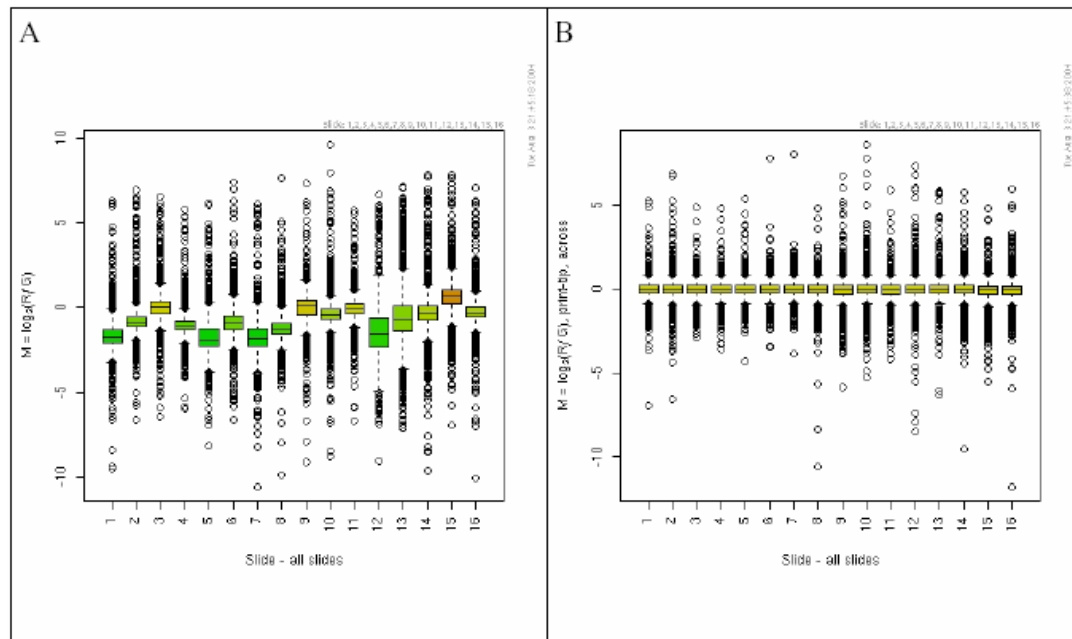


Figura 6 – Gráfico *box-plot* por lâmina. A: dados não normalizados, demonstrando a variabilidade dos dados nas diferentes lâminas. B: Após a normalização por meio do rescalonamento pelo MAD, esta distribuição torna-se homogênea, permitindo a comparação dos dados provenientes das diferentes lâminas (adaptado de MELLO, 2004).

Para os propósitos deste projeto, os dados normalizados ainda passaram por um filtro *Cutoff* com valor de 0.9, de modo a desconsiderar os genes cujos valores de expressão eram nulos em mais de 10% das amostras. Além disso, foram centrados com relação à mediana dos genes e das amostras.

2.5.4 Análise Estatística

O procedimento de análise estatística desempenha um papel fundamental na descoberta de informações biológicas relevantes resultantes da experimentação com cDNA microarrays, já que é responsável, dentre outras tarefas, pela identificação do conjunto de genes diferencialmente expressos nas situações investigadas.

A análise estatística foi realizada por meio do software TIGR MultiExperiment Viewer – Mev v. 3.1, (SAEED *et al.*, 2003), que pode ser obtido gratuitamente na

página <http://www.tm4.org/mev.html>. Este aplicativo Java, desenvolvido com o objetivo de favorecer o processo de *data mining* de experimentos de microarrays, reúne um conjunto de ferramentas para:

- Visualizar e ajustar os dados normalizados;
- Agrupar genes e amostras de acordo com a similaridade dos perfis de expressão gênica, como por exemplo: Hierarchical Clustering - HCL, Tree-EASE, Support Trees;
- Realizar testes estatísticos para determinar os genes diferencialmente expressos, como: Significance Analysis of Microarrays - SAM, Analysis of Variance - ANOVA, Two-factor ANOVA - TFA, T-tests - TTEST;
- Aplicar técnicas de classificação e aprendizado de máquina: Support Vector Machines - SVM, K-Nearest-Neighbor Classification - KNNC, Discriminant Analysis Module - DAM;

A ferramenta escolhida para realizar o procedimento de análise estatística foi o SAM (TUSHER *et al.*, 2001), devido ao seu desempenho reconhecido em vários estudos comparativos e por ser o exemplo típico de uma classe de técnicas abrangendo conceitos similares (THOMAS *et al.*, 2001; STOREY & TIBSHIRANI, 2003). Na verdade, o SAM possui uma abordagem matemática similar àquela adotada pelo Test-T: ele avalia cada gene levando em consideração a variação no seu nível de expressão com relação ao desvio-padrão de repetidas observações (a ideia básica do Test-T consiste em calcular a diferença entre médias, que então é dividida por uma estimativa do desvio-padrão).

Uma importante característica do SAM diz respeito ao cálculo do chamado *False Discovery Rate* (FDR), que fornece a proporção de genes erroneamente identificados como diferencialmente expressos. Esta estimativa é obtida por meio dos seguintes passos:

1. É fixado um valor limite para considerar os genes como diferencialmente expressos;

2. O número de genes tidos como sendo diferencialmente expressos é assinalado em cada permutação (falso-positivos);
3. Calcula-se o número mediano de falso-positivos ao longo de todas as permutações;
4. O FDR corresponde ao número de falso-positivos dividido pelo número de genes nos dados originais (adaptado de DRAGHICI, 2003).

Em todas as análises, foi considerado um FDR menor que 1%, o que confere aos resultados uma confiança estatística maior que 99%. Além disso, os genes selecionados não passaram por nenhum tipo de seleção baseada no valor do *fold-change*.

Finalmente, os resultados do teste estatístico são apresentados sob a forma de uma matriz colorida conhecida como *heat map*, no qual as linhas correspondem aos genes e as colunas indicam as amostras. Ele possui diferentes gradações de cores que evidenciam os níveis de expressão gênica ao longo de todas as amostras: normalmente, tons de vermelho indicam indução, tons de verde denotam repressão e tons de preto demonstram ausência de modulação.

O *heat map* também apresenta dendrogramas (dendro = árvore) nos quais fica evidente a disposição de genes e amostras de acordo com seus perfis de similaridade. A técnica utilizada para a criação destes dendrogramas foi o chamado Agrupamento Hierárquico. Além disso, a métrica de distância empregada denomina-se correlação de Pearson, que é explicada por meio do seguinte exemplo:

Dados dois vetores com dimensão n , $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$, sua distância calculada pela correlação de Pearson corresponde a:

$$d_R(\mathbf{x}, \mathbf{y}) = 1 - r_{xy} \quad (2.1)$$

no qual r_{xy} é o coeficiente de correlação de Pearson dos vetores \mathbf{x} e \mathbf{y} :

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x} \sqrt{s_y}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.2)$$

Agora, considere cinco situações experimentais nos quais o gene g_1 possui níveis de expressão iguais a 1, 2, 3, 4 e 5, respectivamente. Este gene pode ser representado como $g_1 = (1,2,3,4,5)$. Considere ainda dois outros genes cujos níveis de expressão correspondem a $g_2 = (100,200,300,400,500)$ e $g_3 = (5,4,3,2,1)$.

Aplicando estes valores nas equações acima, conclui-se que g_1 e g_2 possuem máxima correlação, pois a distância entre $(1,2,3,4,5)$ e $(100,200,300,400,500)$ é $d_R(g_1, g_2) = 1 - r = 1 - 1 = 0$. Por outro lado, g_1 e g_3 são inversamente correlacionados, já que a distância entre $(1,2,3,4,5)$ e $(5,4,3,2,1)$ é $d_R(g_1, g_3) = 1 - (-1) = 2$. Assim, g_1 estaria mais próximo de g_2 e mais distante de g_3 na construção do *cluster* (adaptado de DRAGHICI, 2003).

A correlação de Pearson, portanto, estabelece que duas variáveis sejam próximas caso variem de maneira semelhante, independentemente da magnitude de seus valores. Esta propriedade é de particular importância no caso dos experimentos com microarrays, uma vez que genes com perfis de expressão parecidos são reunidos num mesmo grupo, fornecendo indícios para esclarecer sua função ou tipo de regulação.

2.5.5 Inferência das Redes Gênicas

Dentre os algoritmos propostos na literatura para a inferência de redes gênicas regulatórias, existem aqueles que se baseiam na chamada lógica Booleana, como o programa REVEAL (LIANG *et al.*, 1998); os que exploram conceitos da Teoria da Informação, como o ARACNE (BASSO *et al.*, 2005; MARGOLIN *et al.*, 2006); aqueles que consideram equações diferenciais ordinárias, como os programas NIR,

NMI e TSNI (GARDNER *et al.*, 2003; DI BERNARDO *et al.*, 2005; BANSAL *et al.*, 2006); e também os que solucionam o problema de inferência por meio de redes Bayesianas, como o Banjo (YU *et al.*, 2004). A estratégia de cada um desses algoritmos é apresentada a seguir.

A lógica Booleana proporciona a forma mais simples e computacionalmente efetiva de modelagem das redes gênicas regulatórias (KAUFFMAN, 1993). Sua aplicação é fundamentada nas seguintes suposições (SOMOGYI & SNIEGOSKI, 1996): o estado de cada gene é caracterizado como 1 (ON) ou 0 (OFF); o controle regulatório da expressão gênica pode ser determinado por regras da lógica Booleana, tais como “OR”, “AND”, “NOT” e “EXOR”; e todos os genes são atualizados sincronicamente com relação ao seu estado (ON ou OFF) de acordo com a condição anterior do sistema.

Como mencionado anteriormente, o programa REVEAL utiliza o método Booleano para realizar a engenharia-reversa das redes gênicas. Porém, Peters Spirtes e colaboradores (SPIRITES *et al.*, 2000) advertem sobre sua inadequação, pois o sistema como descrito não é robusto na ocorrência do fenômeno de agregação, e omitiu fatores como erros de medição, células não sincronizadas, ou autorregulação gênica. Portanto, decidiu-se por não empregar o REVEAL nas análises deste projeto.

A inferência de redes gênicas a partir da Teoria da Informação, por sua vez, vale-se do conceito de Informação Mútua (*Mutual Information* – MI), função que mede o grau de dependência estatística entre os níveis de expressão gênica. A variável MI é calculada para cada par de genes, e tal interação recebe um valor normalizado dentro do intervalo [0;1], de maneira que MI_{ij} torna-se zero caso os genes i e j sejam estatisticamente independentes. Por outro lado, $MI=1$ indica fortemente que essa interação não é aleatória.

É importante observar que essa abordagem não permite estabelecer uma relação de causalidade às interações, mas apenas probabilidades condicionais entre os genes. Além disso, pode ser demonstrado que MI é simétrico, ou seja, $MI_{ij} = MI_{ji}$, e

por isso não é possível identificar qual a variável efetora da interação. A rede resultante acaba sendo apresentada como um grafo não direcionado (no qual a interação fica evidenciada por uma linha), justificando a inadequação do software ARACNE para o tipo de análise proposto neste projeto.

Outra estratégia de inferência toma como premissa que a concentração de um produto gênico é influenciada pela expressão dos demais genes e por componentes externos que perturbem o sistema. Trata-se, portanto, de uma abordagem determinística, no qual o comportamento de cada gene pode ser estimado por uma equação diferencial ordinária.

Neste caso, o desafio consiste em definir o conjunto de parâmetros que compõem a função. No modelo diferencial linear, por exemplo, ela é definida como uma simples soma ponderada das concentrações de produtos gênicos. Outros modelos mais complexos, como o chamado S-System (SAVAGEAU, 1969ab, 1976; VOIT, 1991, 2000; TORRES & VOIT, 2003), levam em consideração as taxas com que esses produtos são produzidos e degradados.

As equações diferenciais ordinárias são capazes de extrair relações de causalidade entre os genes, sendo que as interações agora ficam evidenciadas por setas positivas ou negativas, indicando atividades de indução e repressão, respectivamente. Outra vantagem é que uma vez conhecidos os seus parâmetros, a função pode ser utilizada para prever o comportamento do sistema biológico em diferentes condições.

Implementadas pelos algoritmos do NIR (*Network Identification by Multiple Regression*), MNI (*Microarray Network Identification*) e TSNI (*Time-series Network Identification*), as equações diferenciais ordinárias são úteis para identificar os genes alvos que interagem diretamente com um composto de interesse, situação bastante comum no processo de descoberta de novas drogas por exemplo. No entanto, como a estratégia de inferir redes gênicas a partir de estímulos externos ao sistema foge do escopo deste projeto, aqueles softwares não puderam ser utilizados.

Por fim, o problema de inferência de redes gênicas também pode ser solucionado por meio do formalismo matemático das redes Bayesianas. Conceitualmente, elas consistem num modelo baseado em grafos para a chamada distribuição de probabilidades de ligação (ou união), que define as propriedades de dependência condicional entre variáveis aleatórias, neste caso os níveis de expressão gênica.

Uma rede Bayesiana é constituída por dois componentes: o primeiro é um grafo direcionado denominado G , no qual os vértices (ou nodos) representam as variáveis aleatórias e as setas entre eles indicam suas interações, entendidas no presente contexto como atividade de regulação. O segundo componente, \square , corresponde aos parâmetros da função que define a distribuição de probabilidades de ligação de cada variável.

Assim, a topologia do grafo G (forma de organização estrutural dos vértices e suas ligações) fornece as suposições de dependência condicional que permitem estabelecer os parâmetros da distribuição de ligação, geralmente representada pela seguinte função:

$$P(X) = \prod_{i=1}^n P(X_i | P(a_i)) \quad (2.3)$$

no qual a_i denota os vértices que regulam X_i .

A estratégia de engenharia reversa de uma rede gênica regulatória pelo método Bayesiano consiste em encontrar o grafo G que melhor descreve o conjunto de dados de expressão. Esta tarefa é alcançada por meio de uma função de pontuação, responsável por avaliar as possíveis topologias para o grafo G e eleger a estrutura que maximiza tal pontuação. Normalmente, essa função tem a seguir:

$$\begin{aligned} \text{Pontuação}(G : D) &= \log P(G | D) \\ &= \log P(D | G) + \log P(G) + C \end{aligned} \quad (2.4)$$

no qual $P(G | D)$ é a probabilidade de ocorrência da topologia G dado o conjunto de dados de expressão D , e

$$P(D | G) = \int P(D | G, \theta) P(\theta | G) d\theta \quad (2.5)$$

refere-se à chamada probabilidade marginal, que calcula a probabilidade de ocorrência do conjunto de dados D com relação a todos os parâmetros designados pelo grafo G .

Na realidade, o que a função de pontuação acaba fazendo é averiguar se o conjunto de dados de expressão gênica pode ser explicado de maneira plausível a partir de determinada estrutura dentre as várias possíveis para o grafo G .

Dentre as vantagens de utilizar os conceitos matemáticos das redes Bayesianas para a modelagem de redes gênicas regulatórias, é possível enumerar as seguintes: 1) existem vários algoritmos bem definidos para implementar os fundamentos estatísticos das redes Bayesianas a partir de dados experimentais, e que têm alcançado sucesso em diferentes aplicações; 2) esta metodologia é especialmente indicada para descrever domínios esparsos, devido a sua habilidade de aprendizagem de processos estocásticos complexos a partir de observações incompletas ou ruidosas; 3) as redes Bayesianas tornam possível reproduzir explicitamente os processos pelos quais os dados foram obtidos; 4) elas são capazes de apresentar uma hipótese estatística com base no grafo de relações causais entre os níveis de expressão gênica.

Um dos softwares que utilizam as redes Bayesianas é o chamado Banjo, capaz de inferir redes gênicas partindo tanto de dados de expressão temporais (*time-series*) quanto de uma única condição experimental (*steady-state*). Os testes realizados por Di Bernardo e colaboradores, no entanto, demonstram que o Banjo não possui um desempenho adequado aos propósitos deste projeto.

Aqueles autores produziram dados de expressão *in silico* juntamente com as devidas redes regulatórias, o que resultou num conjunto de vinte grafos contendo

10, 100 e 1000 genes com uma variação média no nível de expressão de 2, 10 e 100, respectivamente. As redes construídas englobavam dados de expressão *steady-state* com perturbação global (todos os genes da rede eram perturbados simultaneamente em cada experimento) e perturbação local (um único gene era perturbado em cada experimento); bem como dados de expressão temporais.

Os modelos inferidos pelo algoritmo do Banjo utilizando-se os mesmos dados de expressão gênica foram então comparados com as redes originais. Os resultados obtidos foram os seguintes: considerando o conjunto de dados *steady-state* com perturbação global e local, o Banjo apresentou uma baixa precisão com relação às redes de 10 genes, mas alcançou uma precisão muito alta com as redes de 100 genes, embora caracterizada por uma sensibilidade muito baixa.

Já considerando os dados de expressão temporais, houve a necessidade de um número muito maior de experimentos (pontos de tempo) comparado com o número de genes. Além disso, o algoritmo do Banjo não obteve êxito ao manipular o conjunto de dados contendo 1000 genes, devido às limitações de memória do programa. Tendo em vista que os experimentos analisados neste projeto normalmente ultrapassam o número de 1.500 genes, concluiu-se sobre a inviabilidade de utilização do software Banjo.

Outro software que abrange o formalismo das redes Bayesianas é o chamado GeneNetwork (WU *et al.*, 2004), que, diferentemente do Banjo, possui recursos para manipular aproximadamente 2000 genes. Devido a esta vantagem, dentre outras importantes características (apresentadas na próxima seção), o programa GeneNetwork foi eleito o responsável por inferir as redes gênicas regulatórias deste projeto.

2.5.6 Funcionamento do GeneNetwork

Escrito em linguagem C++, a última versão do software pode ser encontrada no sítio <http://genenetwork.sbl.bc.sinica.edu.tw/Download/download.html>.

Inicialmente, os dados de expressão gênica devem ser fornecidos sob a forma de uma planilha, na qual as linhas indicam os genes analisados, as três primeiras colunas contêm informações sobre esses genes (nome, identificação do clone e descrição sumária, respectivamente), e as colunas restantes representam as situações experimentais investigadas.

De acordo com D'Haeseleer (D'HAESELEER, 2000), existe um problema comum a todas as técnicas de inferência, conhecido como Maldição da Dimensionalidade: o tamanho (ou complexidade) do espaço de busca aumenta exponencialmente com o número de variáveis do modelo matemático. No caso específico das redes gênicas, quanto maior o número de genes, maior a quantidade de situações experimentais necessárias para se obter um modelo plausível de interações regulatórias.

A tabela a seguir, desenvolvida por aquele autor, estabelece o número de situações experimentais exigidas para a engenharia-reversa de uma rede com N genes:

Tabela 2: Número de situações experimentais necessárias para inferir uma rede regulatória com N genes. Conectividade total significa que cada gene pode interagir com todos os outros genes; K corresponde ao número máximo de interações que podem ser realizadas por um gene; aditivo e linearmente separável indicam que a regulação pode ser calculada por uma soma ponderada; na correlação aos pares, o nível de significância atribuído às dependências estatísticas entre os genes deve ser inversamente proporcional ao número de variáveis.

Modelo	Número de Situações Experimentais
Booleano, conectividade total	2^N
Booleano, K conexões	$2^K (K + \log(N))$
Booleano, K conexões (linearmente separável)	$K \log(N/K)$
Contínuo, conectividade total, aditivo	$N + 1$
Contínuo, K conexões, aditivo	$K \log(N/K)$
Correlação aos pares	$\log(N)$

Entretanto, o alto custo imposto pelos experimentos com cDNA microarrays, e a própria onerosidade temporal em sucedê-los, acabam por inviabilizar o cumprimento das exigências por quantidades suficientes de condições amostrais.

O programa GeneNetwork tenta contornar esta dificuldade propondo a interpolação dos dados de expressão gênica disponíveis. Por meio das técnicas de Interpolação Linear, Interpolação Polinomial *Lagrange* (CONSTANTINIDES & MOSTOUFI, 1999) ou Interpolação Cúbica *Spline* (CONSTANTINIDES & MOSTOUFI, 1999), é possível aumentar a quantidade dos dados de situações experimentais, adequando-os, portanto, para o processo de inferência. Neste caso, foi utilizada a Interpolação Linear.

Outro poderoso recurso deste software é o emprego de um algoritmo genético com o objetivo de aperfeiçoar a busca pelo grafo G que maximiza a função de pontuação. Os Algoritmos Genéticos, originalmente sugeridos por Alex Fraser (FRASER, 1957; FRASER & BURNELL, 1970) e aprimorados por John Holland (HOLLAND, 1975), correspondem a uma classe particular de Algoritmos Evolucionários, assim denominados por abordarem conceitos como hereditariedade, mutação e seleção (computação bioinspirada).

A codificação do algoritmo genético e os parâmetros estabelecidos de acordo com os propósitos deste projeto são apresentados a seguir:

- Cada topologia possível para o grafo G representava um indivíduo, ou seja, a forma como as interações gênicas estão dispostas no grafo correspondia a um fenótipo;
- Foi estabelecida uma população inicial de 200 indivíduos. Assim, foram produzidas aleatoriamente 200 topologias a partir das quais foi feita a escolha do grafo G;
- A taxa de mutação determinada empiricamente foi de 0,02. Dado o tamanho inicial da população, a cada geração quatro cromossomos aleatórios foram modificados, ou seja, quatro grafos sofreram alterações na forma como suas conexões estavam organizadas;

- Todas as topologias eram pontuadas obedecendo ao critério de quão melhor elas representavam as interações regulatórias de acordo com os dados de expressão gênica. Os grafos melhor pontuados (ou “mais plausíveis”) eram então selecionados pela função de avaliação para a próxima geração;
- Foi estabelecido que a função de avaliação elege o grafo G com maior pontuação após 2000 gerações terem sido transcorridas.

Como tradicionalmente é feito pelos algoritmos genéticos, a solução encontrada é disponibilizada sob a forma de uma cadeia de caracteres binários. Mais especificamente, o programa GeneNetwork apresenta o grafo G escolhido sob a forma de uma matriz regulatória RM (quadrada, $N \times N$, sendo N o número de genes participantes da rede), na qual $RM[i][j] = 1$ indica que o gene da linha i modula o gene da coluna j . Em contrapartida, $RM[i][j] = 0$ assinala que não existe interação entre aqueles genes.

Além disso, o software oferece algumas funções gráficas que possibilitam a visualização e manipulação do grafo resultante. Este recurso é particularmente útil quando se obtém uma rede com algumas dezenas de genes. No entanto, grafos mais complexos, compostos por centenas de genes e milhares de conexões (em questão neste trabalho), constituem um tipo de representação biológica que inviabiliza a interpretação puramente visual por parte do usuário.

2.5.7 Interpretação Textual das Redes Gênicas

Sem o auxílio de alguma ferramenta computacional que favoreça a análise das redes gênicas, a busca por informações relevantes acerca das interações regulatórias características da situação investigada torna-se definitivamente impraticável. Este obstáculo motivou o desenvolvimento de um algoritmo capaz de representar os grafos num formato textual, de modo a facilitar sua interpretação. A solução encontrada culminou com a criação do programa intitulado Mandala, que será apresentado na seção 4.4 do capítulo de Resultados.

Delincamento Experimental

3 DELINEAMENTO EXPERIMENTAL

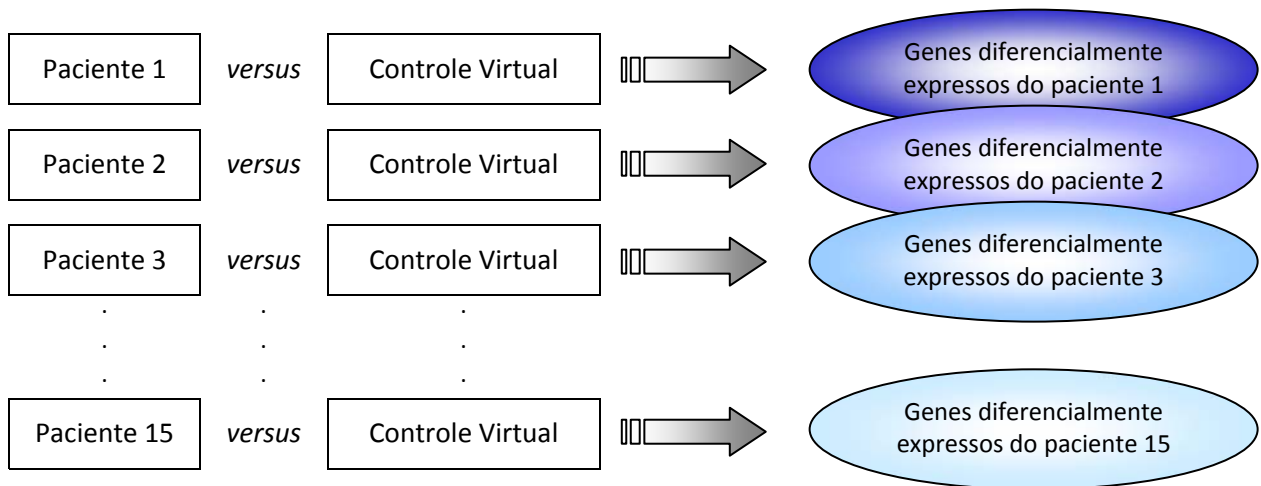
O delineamento experimental deste projeto pode ser estruturado de acordo com as seguintes etapas:

1. Simular o perfil de expressão gênica característico de um indivíduo saudável da população, obtido a partir dos dados normalizados de todas as amostras controle. Para cada gene na lâmina de microarrays, considerou-se a mediana dos valores correspondentes ao seu nível de expressão ao longo dos onze integrantes do grupo controle, produzindo, assim, uma espécie de “controle virtual” único. Veja a ilustração abaixo:

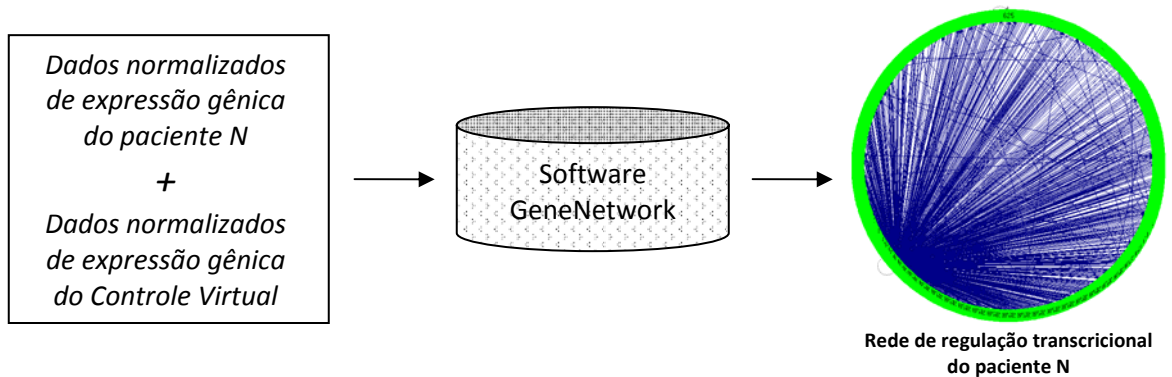
		Contr. 1	Contr. 2	Contr. 3	Contr. 4	Contr. 5	Contr. 6	Contr. 7	Contr. 8	Contr. 9	Contr. 10	Contr. 11	CONTROLE VIRTUAL	
V A L O R E S E X P R E S S Ã O	gene 1	3	1,5	5	5	7	7,5	3,5	9	2,5	4	6	4	
	gene 2	8	2,5	6	6,5	2	1	9,5	7	10	3	5,5	5,5	
	gene 3	1	3	10	8,5	2	7	5	6,5	4	2,5	10	5	
	gene 4	9	7,5	3	1,5	6	8,5	0,5	2	5	1	2,5	3	

	gene 4607	7	12	5	3	3	2,5	8	9,5	10	4	9	7	
	gene 4608	2	3,5	1	7	9	6,5	8,5	3	10	11,5	8	6,5	

2. Descobrir quais são os genes diferencialmente expressos em cada paciente com AR, que podem ser identificados, durante a análise estatística, ao se contrastar os níveis de expressão encontrados no paciente e no controle virtual. Confira a ilustração na sequência:



3. Conhecidos os genes com expressão diferencial em cada paciente com AR, optou-se por substituir seu valor numérico resultante pelo respectivo valor de expressão normalizado. Este procedimento é necessário porque os métodos estatísticos acabam por alterar os valores dos níveis de expressão gênica do paciente em decorrência de sua comparação com o controle virtual.
4. Proceder com a inferência de quinze redes de regulação transcricional, uma para cada paciente com AR. Para isso, o software responsável pela modelagem das redes foi alimentado com diferentes planilhas, que continham, nesta ordem, os dados normalizados de expressão gênica do controle virtual e do paciente em questão. Observe a ilustração na próxima página.

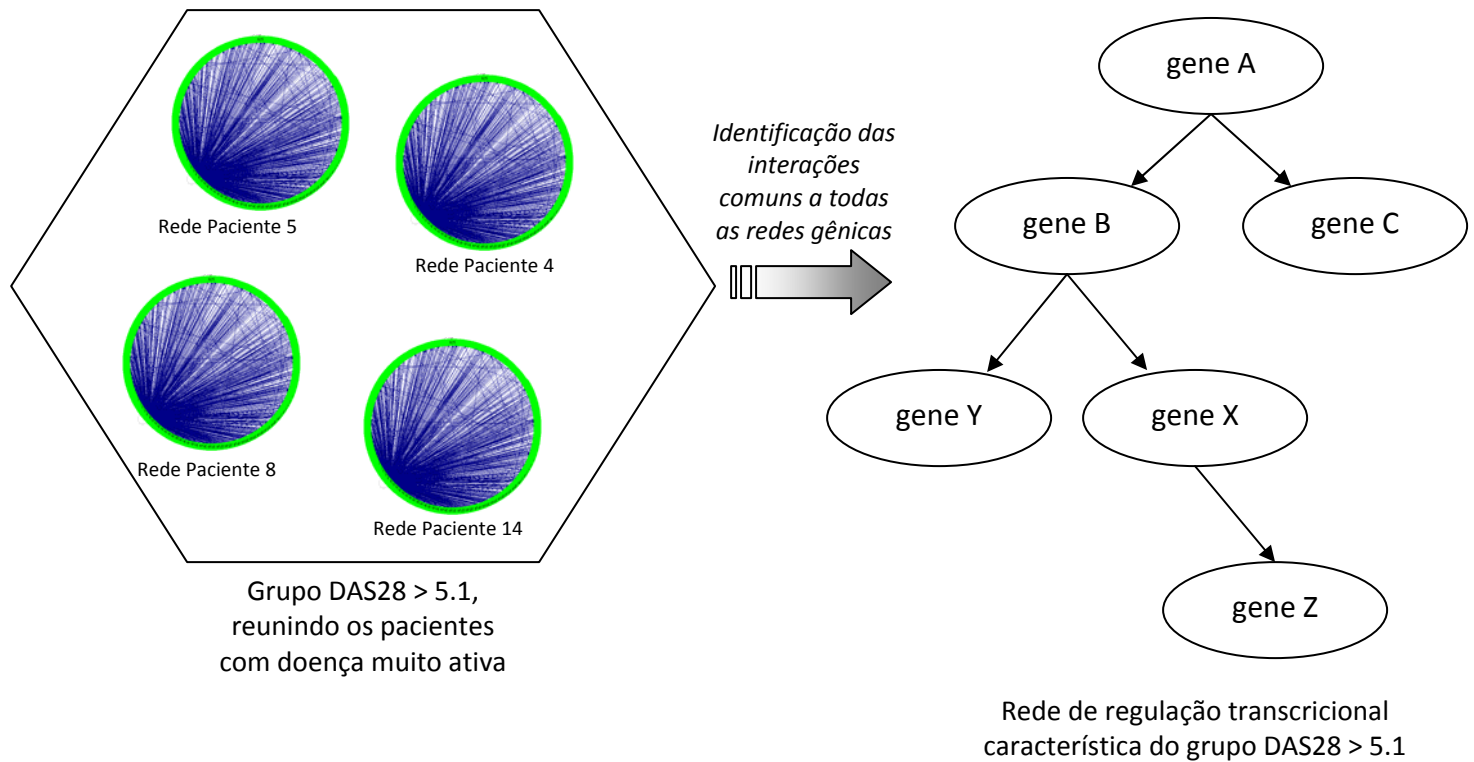


5. Agrupar os pacientes com AR de acordo com os parâmetros clínicos e moleculares, bem como com o tipo de medicação administrada a cada um.

Foram propostas quatro grupos de interesse, a saber:

- a. Conjunto de pacientes cujo parâmetro DAS28 possui valor maior que 5.1, ou seja, aqueles apresentando doença muito ativa;
- b. Grupo dos pacientes com autoanticorpos contra peptídeos citrulinados cíclicos (anti-CCP), os quais são averiguados durante o diagnóstico clínico devido ao seu reconhecido valor preditivo;
- c. Grupo dos pacientes demonstrando o epítipo compartilhado (SE) em dose única ou dupla, haja vista sua importância para a suscetibilidade à artrite reumatoide;
- d. Grupo de pacientes tratados com agentes biológicos bloqueadores do fator de necrose tumoral (TNF- α), neste caso o Infliximabe, em combinação com Metotrexato (MTX) mais prednisona. Enquanto a primeira consiste numa droga modificadora da atividade da doença reumática (DMARD), a segunda é basicamente sintomática;

6. Para cada grupo, comparar todas as redes gênicas características dos pacientes nele reunidos, de maneira a identificar suas interações regulatórias comuns e estabelecer, assim, uma rede de regulação transcricional que represente a variável em questão. Por exemplo:



Resultados

4 RESULTADOS

4.1 Caracterização dos Pacientes com AR

Por meio do acesso aos registros dos prontuários médicos, foi possível determinar que os pacientes com AR (n=15) apresentavam as seguintes características: quatorze eram do sexo feminino, com faixa etária variando de 32 a 72 anos (média de 50,6 anos) e etnicamente subdivididos em doze caucasianos e três negroides. Sete indivíduos eram fumantes.

Com relação aos parâmetros clínicos, nove pacientes possuíam índice DAS28 maior que 5.1, exibindo, portanto, doença muito ativa; quatro possuíam índice DAS28 entre 3.2 e 5.1, demonstrando doença moderadamente ativa; e apenas dois enquadravam-se no índice DAS28 menor que 3.2, o qual indica doença pouco ativa. Além disso, cinco indivíduos apresentavam manifestações extra-articulares.

Já com relação à administração medicamentosa, todos os pacientes utilizavam o DMARD Metotrexato, além de pequenas doses de Prednisona (menos que 10mg/dia). Cinco deles também faziam uso de medicação com ação bloqueadora de TNF (Infliximabe, 3-5mg/Kg/dose) associada ao MTX.

Os procedimentos de tipificação do alelo HLA-DRB1 e detecção de anticorpos anti-CCP, detalhados na seção 2.4, permitiram distinguir os seguintes parâmetros moleculares: seis pacientes com AR possuíam o epítipo compartilhado em dupla dose (SE/SE), quatro o possuíam em dose única (SE/X) e cinco não apresentavam epítipo compartilhado (X/X). Por fim, nove pacientes eram reagentes para anticorpos anti-CCP.

De acordo com o delineamento experimental descrito no terceiro capítulo, podem ser estabelecidos então quatro conjuntos de pacientes obedecendo aos critérios clínicos e moleculares de interesse: 1) o grupo DAS28 > 5.1 reúne nove

indivíduos com doença muito ativa; 2) o grupo DMARDs + anti-TNF distingue cinco pacientes que, além de MTX e Prednisona, recebiam terapia à base do agente biológico Infliximabe; 3) dez indivíduos que apresentavam epítipo compartilhado em dupla dose ou dose única compõem o grupo HLA-DRB1 SE; 4) já o grupo anti-CCP inclui os nove pacientes reagentes àquele anticorpo.

Conforme explicado anteriormente, cada grupo detém uma rede de interações gênicas distinta, originada a partir da identificação das interações regulatórias comuns às redes gênicas próprias dos pacientes nele reunidos. Para isso, foi necessário primeiro determinar o conjunto de genes diferencialmente expressos característico de cada paciente com AR, o que será discutido a seguir.

4.2 Determinação dos Genes Diferencialmente Expressos

De acordo com os procedimentos definidos no *pipeline* de Bioinformática (seção 2.5.1), a quantificação das imagens dos microarrays e a normalização dos dados de expressão gênica foram realizadas com sucesso.

A partir dos dados normalizados de todas as amostras de indivíduos normais, foi criado um controle “virtual” que simulasse o perfil de expressão gênica encontrado num indivíduo saudável da população. Para cada gene da lâmina, considerou-se seu valor de expressão mediano ao longo das onze amostras pertencentes ao grupo controle.

Além disso, a aplicação de um filtro *Cutoff* permitiu selecionar um conjunto de 3.876 genes cujos níveis de expressão tinham valores válidos em pelo menos 90% daquelas amostras, o que representa mais de 84% do total de 4.608 sequências depositadas na lâmina de microarrays.

Durante o procedimento de análise estatística, realizado por meio da ferramenta Significance Analysis of Microarrays - SAM, este conjunto seletivo de genes serviu como referência única para se chegar aos genes diferencialmente expressos

característicos de cada paciente com AR, cujos números, em ordem crescente, são apresentados na tabela a seguir:

Tabela 3: Número de genes identificados durante a análise estatística como sendo diferencialmente expressos para cada um dos pacientes com AR.

Paciente #	Número de genes diferencialmente expressos
1	634
2	714
3	831
4	894
5	934
6	1017
7	1585
8	1594
9	1595
10	1607
11	1608
12	1621
13	1647
14	1670
15	1685

Retomando os quatro grupos definidos na seção anterior (DAS28 > 5.1, DMARDs + anti-TNF, HLA-DRB1 SE e anti-CCP), dados os pacientes que integram cada grupo e conhecendo-se quais genes são diferencialmente expressos em cada paciente, é possível afirmar que:

- 1) O grupo DAS28 > 5.1 reúne um *pool* de 3493 genes distintos (ou seja, sem repetições) identificados como diferencialmente expressos no conjunto dos nove pacientes que o compõem;

- 2) O grupo DMARDs + anti-TNF contém um *pool* de 3134 genes distintos reconhecidos como diferencialmente expressos na totalidade de seus cinco pacientes;
- 3) Um *pool* de 3675 genes distintos é encontrado no grupo HLA-DRB1 SE, resultado da seleção dos genes diferencialmente expressos característicos de seus dez pacientes;
- 4) O grupo anti-CCP possui um *pool* de 3645 genes distintos e diferencialmente expressos nos nove pacientes nele organizados;
- 5) 2.841 é o número de genes diferencialmente expressos identificados quando da intersecção desses quatro grupos, ou seja, são genes que podem ser encontrados em qualquer um dos grupos;
- 6) Vinte e um genes diferencialmente expressos são exclusivos de pacientes com doença muito ativa ($DAS28 > 5.1$);
- 7) Quinze genes diferencialmente expressos são reconhecidos como pertencentes apenas ao grupo DMARDs + anti-TNF;
- 8) Os pacientes com dupla dose ou dose única do epítipo compartilhado HLA-DRB1 não apresentam nenhum gene diferencialmente expresso em particular quando comparados aos integrantes dos outros grupos;
- 9) Setenta e um genes diferencialmente expressos são exclusivos dos pacientes reunidos no grupo anti-CCP;

Todas as informações acima estão sumarizadas na figura 7 a seguir, que representa um diagrama de Venn formado pelos conjuntos $DAS28 > 5.1$, DMARDs + anti-TNF, HLA-DRB1 SE e anti-CCP, de modo a organizar a quantidade de genes diferencialmente expressos exclusivos e comuns aos vários grupos.

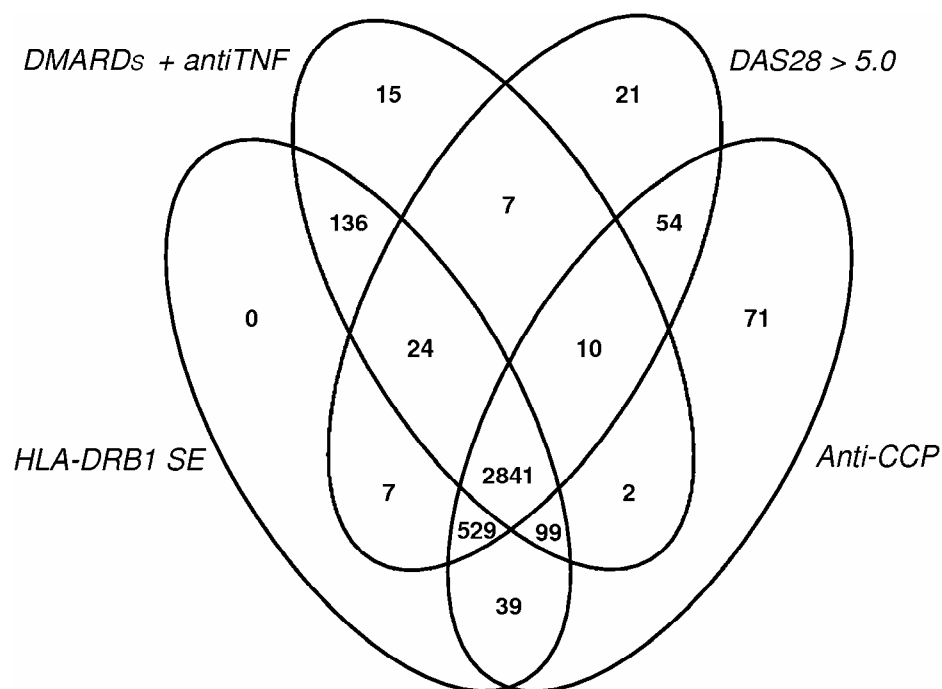


Figura 7 – Diagrama de Venn demonstrando o número dos genes diferencialmente expressos que são exclusivos e o daqueles compartilhados entre os pacientes reunidos nos seguintes grupos: DAS28 > 5.1 (N=9), DMARDs + anti-TNF (N=5), HLA-DRB1 SE (N=10) e anti-CCP (N=9).

4.3 Inferência das Redes Gênicas Regulatórias

Conhecidos quais genes eram diferencialmente expressos para cada paciente, decidiu-se por substituir os valores numéricos resultantes da análise estatística pelos respectivos valores de expressão normalizados. Conforme já foi mencionado no Delineamento Experimental, este procedimento é necessário porque os métodos estatísticos acabam por alterar os valores originais de expressão gênica em função das amostras experimentais comparadas (paciente *versus* “controle virtual”).

Com todos os dados adequadamente ajustados, procedeu-se com inferência de uma rede de regulação transcricional para cada um dos quinze pacientes. Este procedimento exigiu a dedicação de grande quantidade de recursos computacionais, já que o emprego de algoritmos genéticos pelo programa GeneNetwork faz com que o tempo de processamento aumente consideravelmente

quanto maior for a quantidade de dados de expressão gênica. Para se ter uma ideia, dispendo de memória exclusiva, o tempo de processamento dual necessário para a obtenção de uma rede composta por cerca de mil genes é de aproximadamente seis dias!

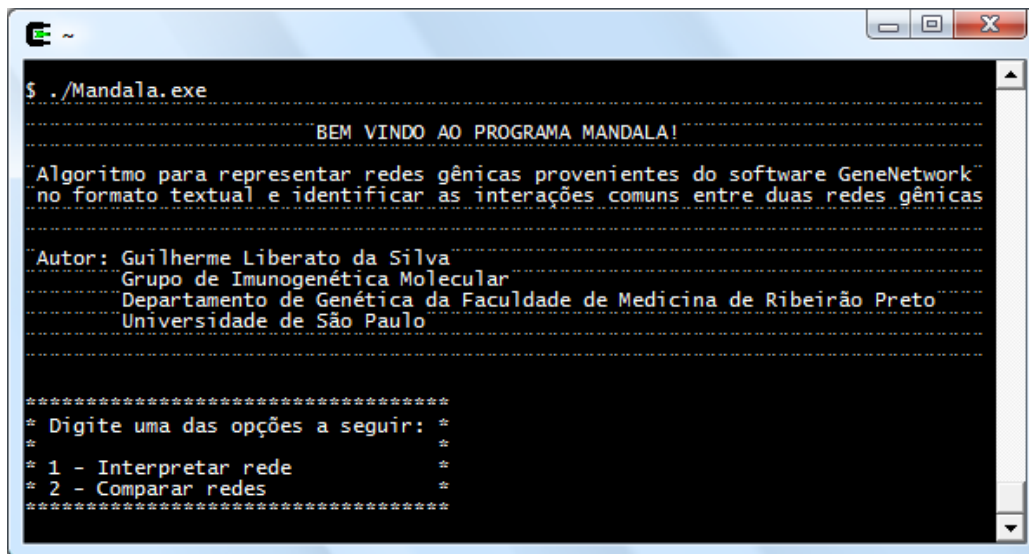
A modelagem de todas as redes regulatórias obteve êxito, mas as imagens resultantes serão omitidas devido ao número e complexidade das mesmas. A lista de genes e a matriz regulatória, que juntas codificam o grafo de interações obtido para cada paciente, foram então armazenadas em arquivos com o intuito de serem fornecidas ao programa Mandala.

4.4 Descrição do Programa Mandala

Basicamente, o programa Mandala foi concebido para atender a dois propósitos: o primeiro consiste em “interpretar” as redes gênicas provenientes do programa GeneNetwork, representando-as num formato textual; o segundo é de identificar quais interações gênicas são comumente encontradas em duas redes regulatórias. Estas funcionalidades serão apresentadas de forma detalhada mais adiante.

Seu algoritmo foi desenvolvido em linguagem C e possui aproximadamente seiscentas linhas de código-fonte (transcritas no Apêndice A). Todas as funções foram organizadas em módulos com o intuito de facilitar possíveis alterações. Pelo fato do programa ser executado em linhas de comando, usuários Windows necessitam de um emulador do sistema operacional UNIX/Linux, como o Cygwin, que é disponibilizado gratuitamente em <http://www.cygwin.com/>.

Ao executar o programa no terminal, o usuário é recepcionado com uma mensagem de apresentação (contendo título, propósitos da ferramenta, autoria e instituição), sendo instruído a digitar um número de acordo com o que deseja fazer: 1 para interpretar uma rede gênica ou 2 para comparar duas redes regulatórias. Observe a figura 8 a seguir:



```
$ ./Mandala.exe
.....
BEM VINDO AO PROGRAMA MANDALA!
.....
Algoritmo para representar redes gênicas provenientes do software GeneNetwork
no formato textual e identificar as interações comuns entre duas redes gênicas
.....
Autor: Guilherme Liberato da Silva
      Grupo de Imunogenética Molecular
      Departamento de Genética da Faculdade de Medicina de Ribeirão Preto
      Universidade de São Paulo
.....
*****
* Digite uma das opções a seguir: *
*                               *
* 1 - Interpretar rede           *
* 2 - Comparar redes            *
*                               *
*****
```

Figura 8 – Tela inicial do programa Mandala. São apresentadas informações gerais: título, propósitos, autoria e instituição, bem como as duas funções, que o usuário poderá acessar digitando os números correspondentes. Optou-se por utilizar o recurso gráfico de *menu* com o intuito de criar uma interface mais amigável.

Caso escolha a primeira opção, o usuário deve informar qual modelo estatístico-matemático, dentre os disponíveis no programa GeneNetwork, foi utilizado para inferir a rede gênica em questão. Este tipo de informação garante que o programa Mandala apresente o resultado textual de forma coerente com a representação gráfica característica de cada modelo.

A seguir, o usuário é orientado a digitar o nome do arquivo contendo a lista de genes participantes da rede e o nome do arquivo correspondente à matriz regulatória, ambos previamente formatados e armazenados no diretório C:\cygwin\home. É necessário ainda atribuir um nome ao arquivo que receberá o resultado do procedimento de interpretação. Veja a figura 9 a seguir:


```

1
*****
* Entre com o número correspondente ao módulo de inferência *
* utilizado no programa GeneNetwork: *
* *
* 1 - Boolean Network *
* 2 - Linear Model *
* 3 - S-Systems *
* 4 - Bayesian Network *
*****
4
Módulo utilizado: Bayesian Network

Digite o nome do arquivo correspondente à lista de genes:
lista_de_genes.txt

O número de genes nesta lista é: 1685

Digite o nome do arquivo correspondente à matriz regulatória:
matriz_regulatoria.txt

Digite o nome do arquivo que receberá o resultado da análise.
O arquivo deve ter extensão '.txt'.
resultado_da_interpretacao.txt

*****
* Deseja realizar a pesquisa a partir dos: *
* *
* 1 - Genes que modulam *
* 2 - Genes que são modulados *
*****

```

Figura 9 – Tela referente à primeira opção do *menu* principal do programa Mandala. São apresentadas instruções para que o usuário indique o modelo estatístico-matemático utilizado durante a modelagem da rede gênica no software GeneNetwork, e forneça os nomes dos arquivos referentes à lista de genes, à matriz regulatória e ao resultado do procedimento de interpretação. É preciso também selecionar qual a orientação na qual o resultado será apresentado: a partir dos genes que modulam ou levando em consideração os que são modulados.

Por fim, é possível escolher se a descrição textual da rede será organizada a partir dos genes que modulam ou levando em consideração os genes que são modulados (figura 9). Ao determinar a orientação na qual as interações regulatórias serão analisadas, o usuário pode adequar o formato de apresentação do resultado de acordo com os possíveis interesses da investigação biológica.

Por exemplo, considere um grafo hipotético obtido a partir do software GeneNetwork, formado por seis nodos (ou vértices) representando os genes denominados A, B, C, X, Y e Z, e vinte setas indicando as interações regulatórias realizadas entre eles. Observe a figura 10 a seguir:

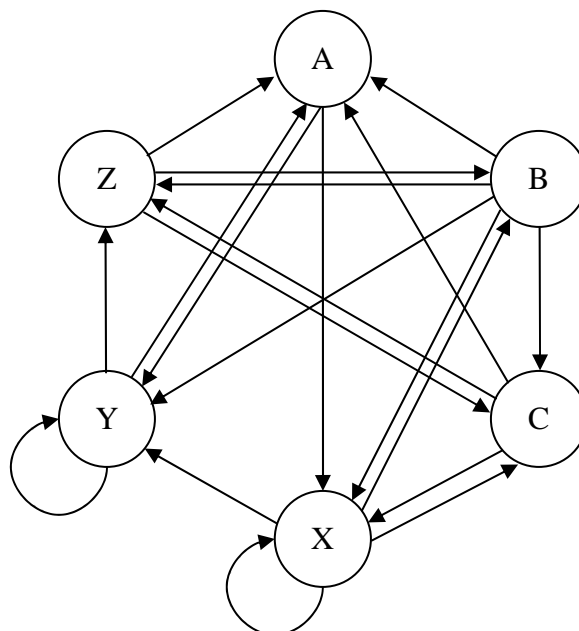


Figura 10 – Grafo hipotético constituído por seis genes, cujos nomes são A, B, C, X, Y e Z, e por vinte interações regulatórias entre os mesmos.

Este grafo será codificado, ainda no programa GeneNetwork, por meio de uma lista contendo os nomes dos seis genes, organizados em linhas, além de uma matriz binária (ou seja, composta por dígitos 0 e 1) indicativa da ocorrência das vinte interações regulatórias (veja a figura 11). Ambas as estruturas devem ser armazenadas em arquivos distintos com o intuito de serem fornecidas ao programa Mandala (conforme foi demonstrado na figura 9 acima), o qual procederá com a manipulação das informações neles contidas de modo a reproduzir o grafo original no formato textual.

A	0	0	0	1	1	0
B	1	0	1	1	1	1
C	1	0	0	1	0	1
X	0	1	1	1	1	0
Y	1	0	0	0	1	1
Z	1	1	1	0	0	0

Figura 11 – Estruturas de dados referentes à lista com os nomes dos genes e à matriz binária, capaz de determinar como eles interagem uns com os outros.

Se o usuário decidir que a descrição textual deverá ser disponibilizada a partir dos genes que modulam, o Mandala realizará a leitura da matriz regulatória sequencialmente de cima para baixo e da esquerda para a direita.

Começando pela primeira linha da matriz, que se refere ao primeiro registro da lista de genes (no exemplo ilustrativo: A), o programa pode identificar a ocorrência do dígito 1 (que denota a existência de interação) na quarta e quinta colunas, as quais correspondem ao quarto e quinto registros da lista de genes. Fica assim estabelecido que o gene A modula os genes X e Y.

Seguindo para a segunda linha da matriz (neste caso: gene B), é possível constatar a ocorrência de interações com todos os outros genes da lista, uma vez que todas as colunas, menos a segunda, estão preenchidas com o número 1. O programa então reporta que o gene B modula os genes A, C, X, Y e Z. Desta forma, ao final do processo aquela rede gênica será representada da seguinte maneira:

GENES QUE MODULAM

O gene A modula os genes:

X
Y

O gene B modula os genes:

A
C
X
Y
Z

O gene C modula os genes:

A
X
Z

O gene X modula os genes:

B
C
X
Y

O gene Y modula os genes:

A
Y
Z

O gene Z modula os genes:

A
B
C

Caso o usuário prefira que a descrição textual da rede seja organizada a partir dos genes que são modulados, o programa, ao contrário, realizará a leitura da matriz regulatória sequencialmente da esquerda para a direita e de cima para baixo, e o resultado ficará assim:

GENES QUE SÃO MODULADOS

O gene A é modulado pelos genes:

B
C
Y
Z

O gene B é modulado pelos genes:

X
Z

O gene C é modulado pelos genes:

B
X
Z

O gene X é modulado pelos genes:

A
B
C
X

O gene Y é modulado pelos genes:

A
B

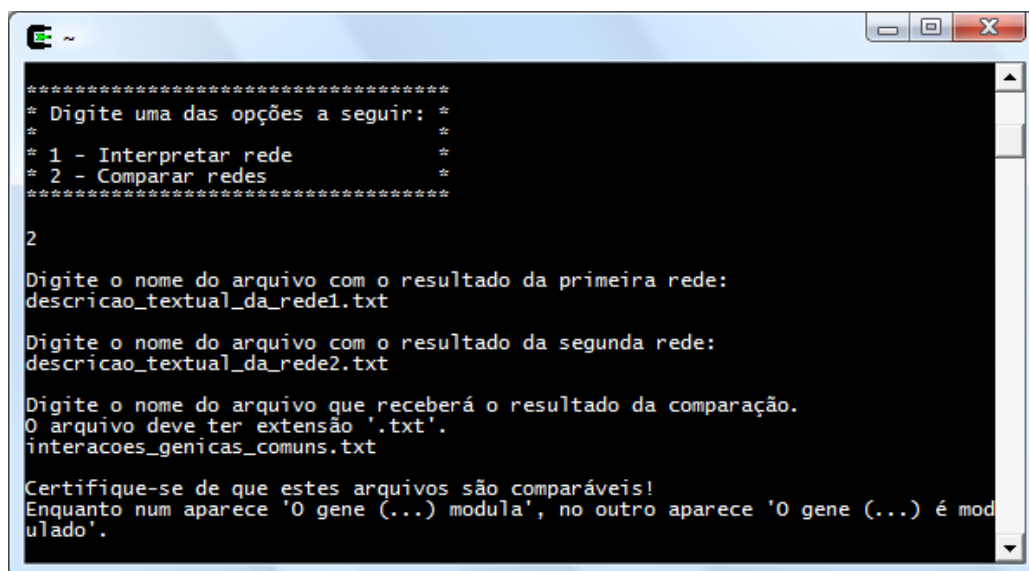
X
Y

O gene Z é modulado pelos genes:

B
C
Y

Vale a pena destacar que, enquanto a rede gênica na forma gráfica permanece disponível somente durante o funcionamento do programa GeneNetwork, sua descrição textual poderá ser acessada imediatamente e analisada pelo tempo que for necessário. Tais características acabam sendo exigências do processo de investigação biológica das redes gênicas de alta complexidade.

O usuário pode também estar interessado em comparar duas redes regulatórias (segunda opção do *menu* principal). Para isto, é necessário informar os nomes dos respectivos arquivos contendo os resultados do procedimento de interpretação realizado previamente. Além disso, o usuário deve dar um nome ao arquivo que receberá o resultado do procedimento de comparação. Confira a figura 12:



```
*****
* Digite uma das opções a seguir: *
*                                     *
* 1 - Interpretar rede                *
* 2 - Comparar redes                  *
*****

2

Digite o nome do arquivo com o resultado da primeira rede:
descricao_textual_da_rede1.txt

Digite o nome do arquivo com o resultado da segunda rede:
descricao_textual_da_rede2.txt

Digite o nome do arquivo que receberá o resultado da comparação.
O arquivo deve ter extensão '.txt'.
interacoes_genicas_comuns.txt

Certifique-se de que estes arquivos são comparáveis!
Enquanto num aparece 'O gene (...) modula', no outro aparece 'O gene (...) é modulado'.
```

Figura 12 – Tela referente à segunda opção do *menu* principal do programa Mandala. São fornecidas instruções para que o usuário digite os nomes dos arquivos contendo a descrição textual das redes que se deseja comparar. Repare que uma mensagem de erro é emitida caso os arquivos contenham interações gênicas organizadas de maneira diferente.

Desta vez, suponha duas redes gênicas cujas descrições no formato textual tenham sido obtidas por meio do programa Mandala:

Arquivo nº 1

O gene ZNF236 modula os genes:

HMGCLL1

GPR172A

LASS6

RBM23

CUL3

CRY1

IMAGE:25061

O gene LIN28B modula os genes:

CCDC22

SYN2

SCOC

RAPGEF1

NAGA

O gene HLA-DRB1 modula os genes:

ADC

C5orf42

C7orf43

CRB1

FNDC3A

EFHC1

Arquivo nº 2

O gene USP45 modula os genes:

FNDC3A

MOBP

LOC100129345

SERPINH1

UBR5

O gene ZNF236 modula os genes:

PLCE1

RPL24

C5orf42

RBM23

GPR172A

O gene SLC38A1 modula os genes:

RBJ

EXPH5

RAPGEF1

RBM23

C14orf112

O gene LIN28B modula os genes:

HBA1

RAPGEF1

RBM23

C14orf112

CCDC22

SCOC

O gene EFHC1 modula os genes:

HPS5

IMAGE:137208

Inicialmente, o Mandala considera a primeira sentença do arquivo nº 1 (escrita em azul): “o gene ZNF236 modula os genes” e confere todo o arquivo nº 2 em

busca da mesma sentença. Quando a encontra, o programa armazena numa lista todos os genes localizados logo abaixo (neste caso, marcados em vermelho, PLCE1, RPL24, C5orf42, RBM23 e GPR172) e começa um novo laço: considera o primeiro gene abaixo daquela sentença no arquivo nº 1 (HMGCLL1, em verde claro) e verifica se ele encontra-se na lista armazenada. Como não obtém sucesso, parte para o segundo gene abaixo daquela sentença no arquivo nº 1 (GPR172, alaranjado) e, ao perceber sua ocorrência na lista guardada, registra esta interação. Ele procede da mesma forma para os outros genes, obtendo êxito apenas com RBM23.

Terminada esta rodada de verificação, o programa segue para a segunda sentença do arquivo nº 1 (escrita em roxo): “o gene LIN28B modula os genes” e, mais uma vez, varre a totalidade do arquivo nº 2 até que encontre a mesma sentença. Depois disso, armazena numa lista todos os genes dispostos logo abaixo (no caso, em verde escuro, HBA1, RAPGEF1, RBM23, C14orf112, CCDC22 e SCOC) e inicia mais uma rotina: considera o primeiro gene abaixo daquela sentença no arquivo nº 1 (CCDC22, em marrom) e busca o mesmo na lista de genes guardados. Por ter sucesso, registra a interação e depois procede da mesma forma para todos os genes abaixo daquela sentença no arquivo nº 1, alcançando sucesso com SCOC e RAPGEF1.

Por último, o Mandala considera a sentença “O gene HLA-DRB1 modula os genes” do primeiro arquivo (escrita em azul claro) e não consegue encontrá-la no segundo. Termina assim o procedimento de comparação dos arquivos, sendo evidenciadas as interações regulatórias comuns às duas redes gênicas. O resultado final é apresentado da seguinte forma:

O gene ZNF236 modula os genes:
GPR172A
RBM23

O gene LIN28B modula os genes:
CCDC22
SCOC
RAPGEF1

4.5 Determinação das Interações Regulatórias Características de Cada Grupo de Pacientes

Após o programa Mandala ter fornecido a descrição textual de todas as redes gênicas obtidas a partir do programa GeneNetwork, pôde ser realizada então a busca das interações regulatórias comumente encontradas nas redes dos pacientes que, conforme determinado previamente, integravam um mesmo grupo.

Porém, como o programa Mandala é capaz de comparar apenas duas redes gênicas de cada vez, e um único grupo chegava a reunir até dez componentes (ou seja, dez redes), como é o caso daquele formado pelos pacientes com epítipo compartilhado em dose única ou dupla (SE HLA-DRB1), foi necessário recorrer a scripts especialmente desenvolvidos para comparar mais de duas redes numa única rodada.

Assim, foram escritos códigos em linguagem de programação Perl com instruções para manipular os arquivos texto resultantes do programa Mandala, de forma a identificar todas as interações gênicas que fossem comuns aos vários participantes de um mesmo grupo.

Portanto, considerando o conjunto dos dez pacientes apresentando o epítipo compartilhado em dose única ou dupla (grupo HLA-DRB1 SE), foi possível constatar que 11 genes localizados em *loci* de suscetibilidade à AR realizaram as mesmas interações com outros 39 genes em pelo menos nove daqueles pacientes. A rede de regulação transcricional envolvendo todos estes genes é representada na figura 13 a seguir.

Para interpretar uma rede de interações como essa, é necessário saber que a seta representa a dependência condicional, inferida estatisticamente, entre os níveis de expressão dos dois genes interconectados. Tenha em mente que a partir de dados experimentais obtidos por meio da tecnologia de cDNA microarrays, um modelo estatístico Bayesiano foi utilizado para afirmar que os níveis de expressão do gene

para o qual a seta indica depende, direta ou indiretamente, dos níveis de expressão do gene pelo qual a seta se origina.

Agora, considerando os nove pacientes nos quais puderam ser detectados anticorpos contra peptídeos citrulinados cíclicos (grupo anti-CCP), foi possível identificar que 10 genes encontrados em LS à AR estabeleciam as mesmas interações com outros 12 genes em pelo menos oito daqueles pacientes. A rede de regulação transcricional correspondente pode ser conferida na figura 14 adiante.

Por outro lado, ao considerar os nove pacientes com doença muito ativa (grupo DAS28 > 5.1), observou-se que 10 genes pertencentes a LS para AR obedeceram a um mesmo padrão de interações regulatórias com outros 29 genes em, no mínimo, oito daqueles pacientes (veja a figura 15).

Tomando o grupo DMARDs + anti-TNF, que é composto pelos cinco pacientes tratados com um bloqueador de TNF- α , Infliximabe, combinado ao Metotrexato, pôde-se constatar o seguinte: na rede de todos os pacientes, 8 genes localizados em LS à AR eram responsáveis por regular os níveis de expressão de outros 11. Observe a figura 16 na sequência.

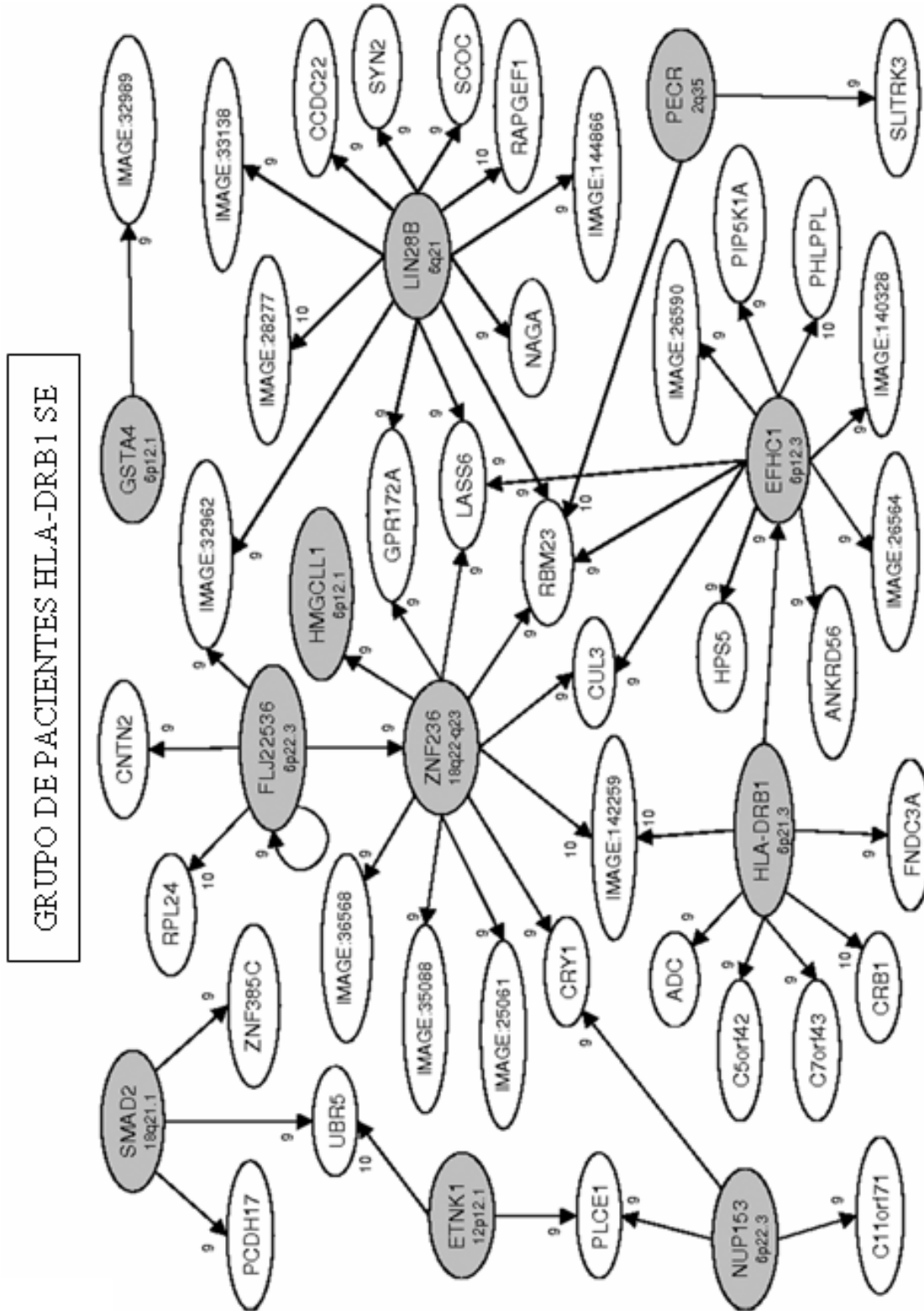


Figura 13 – Rede de regulação transcricional inferida a partir dos dados de Expressão gênica dos pacientes reunidos no grupo HLA-SE. Os genes localizados em *loci* de suscetibilidade à AR aparecem na cor cinza. As setas indicam a direção da interação gênica. Os números correspondem à quantidade de pacientes na qual a interação ocorre.

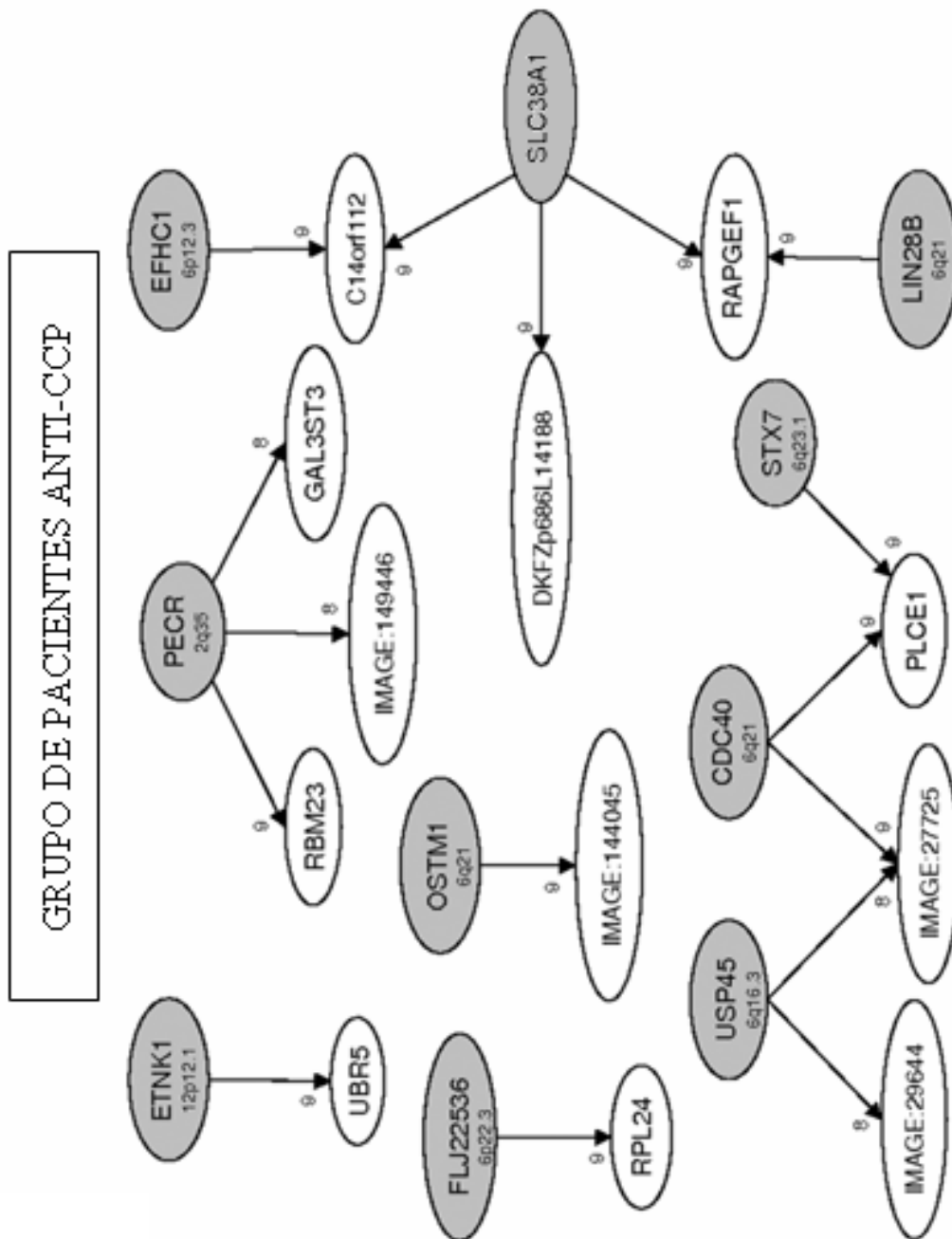


Figura 14 – Rede de regulação transcricional inferida a partir dos dados de expressão gênica dos pacientes reunidos no grupo anti-CCP. Os genes localizados em *loci* de suscetibilidade à AR aparecem na cor cinza. As setas indicam a direção da interação gênica. Os números correspondem à quantidade de pacientes na qual a interação ocorre.

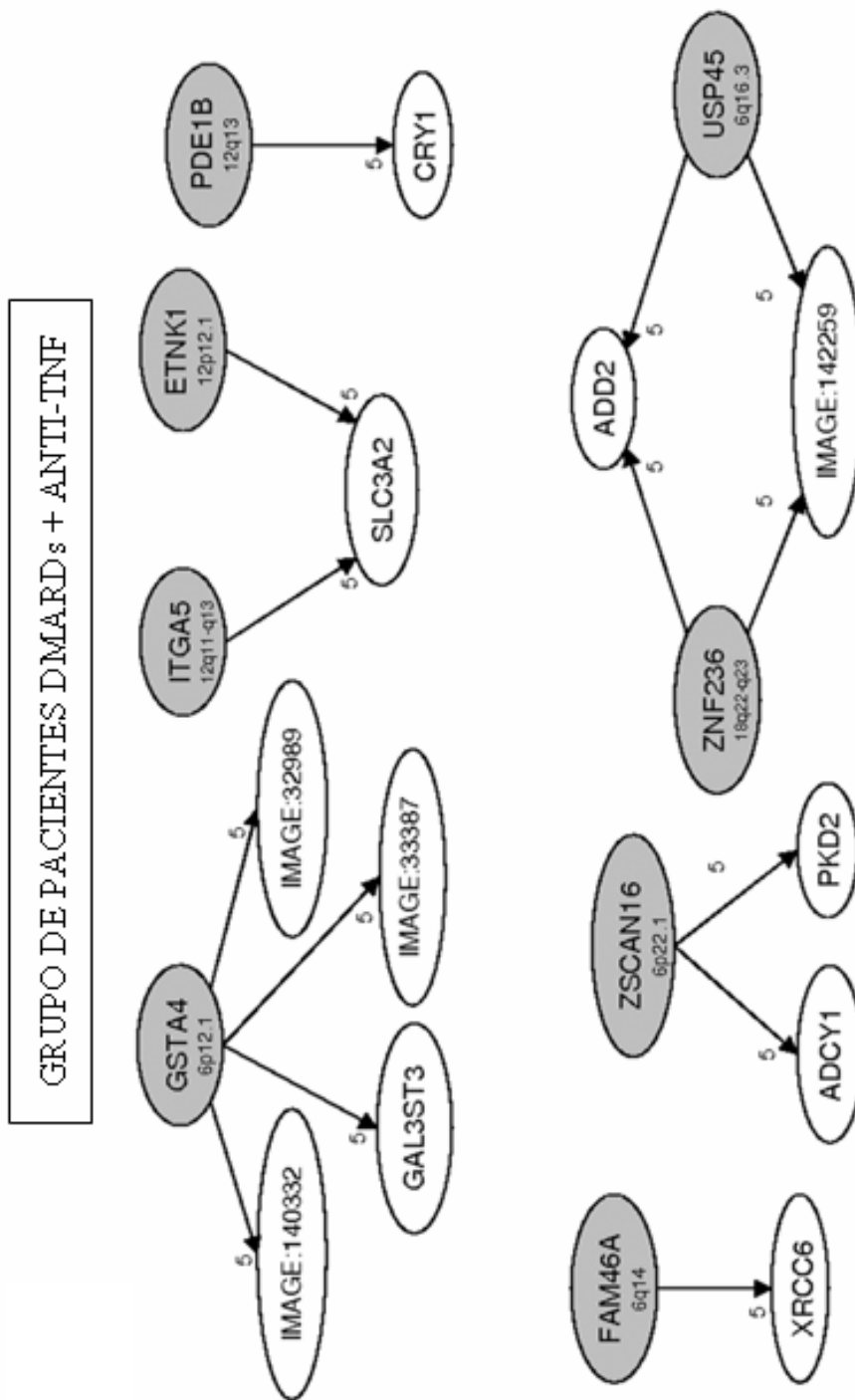


Figura 16 – Rede de regulação transcricional inferida a partir dos dados de expressão gênica dos pacientes reunidos no grupo DMARDs + anti-TNF. Os genes localizados em *loci* de suscetibilidade à AR aparecem na cor cinza. As setas indicam a direção da interação gênica. Os números correspondem à quantidade de pacientes na qual a interação ocorre.

Todos os genes acima representados foram submetidos à pesquisa em bancos de dados públicos, tais como Gene Ontology (THE GENE ONTOLOGY CONSORTIUM, 2000), GeneCards (REBHAN *et al.*, 1998), KEGG (KANEHISA & GOTO, 2000), OMIM (McKUSICK, 1998), dentre outros que são compilados na ferramenta SOURCE, desenvolvida na Stanford University, Palo Alto – CA, e cujo endereço eletrônico é <http://smd.stanford.edu/cgi-bin/source/sourceSearch>.

Do total de 113 genes participantes das redes apresentadas nas figuras 13-16, 78 puderam ser caracterizados com relação aos processos biológicos em que estão envolvidos, às funções moleculares pelas quais são responsáveis e também quanto à localização celular.

De maneira geral, a maioria dos genes está envolvida em processos metabólicos de macromoléculas, da transcrição do DNA até a modificação das proteínas sintetizadas. Destaca-se também o metabolismo de lipídeos. Outro processo biológico no qual participa um considerável número de genes é a sinalização celular, representada tanto por receptores de membrana responsáveis pela transdução de sinais quanto pelas cascatas bioquímicas de sinalização intracelular. Desenvolvimento e homeostase celulares também mobilizam boa parte destes genes, conforme pode ser observado na figura 17 adiante.

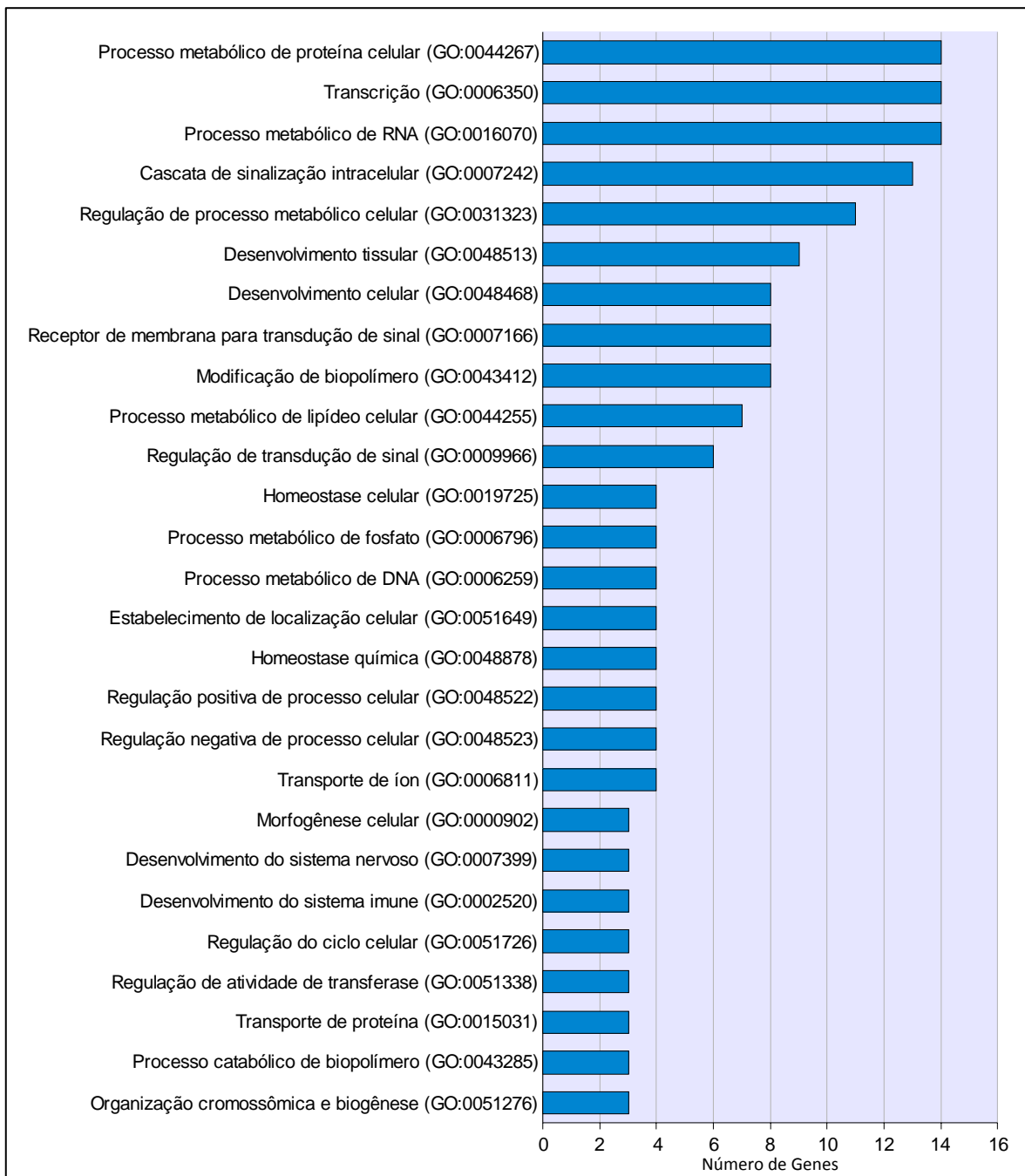


Figura 17 – Gráfico demonstrativo do número de genes representados nas redes de regulação transcricional que participam de cada processo biológico celular (identificado no banco de dados Gene Ontology por meio de uma série numérica).

Já com relação à função molecular, é possível afirmar que a maior parcela dos genes sintetizam ligantes de proteínas, exercendo papéis como o de transferase ou de hidrolase. Também tem destaque a ligação de íons metálicos, principalmente os de zinco e cálcio. Moléculas que interagem com DNA ou RNA possuem alguma representatividade, como pode ser averiguado na figura a seguir:

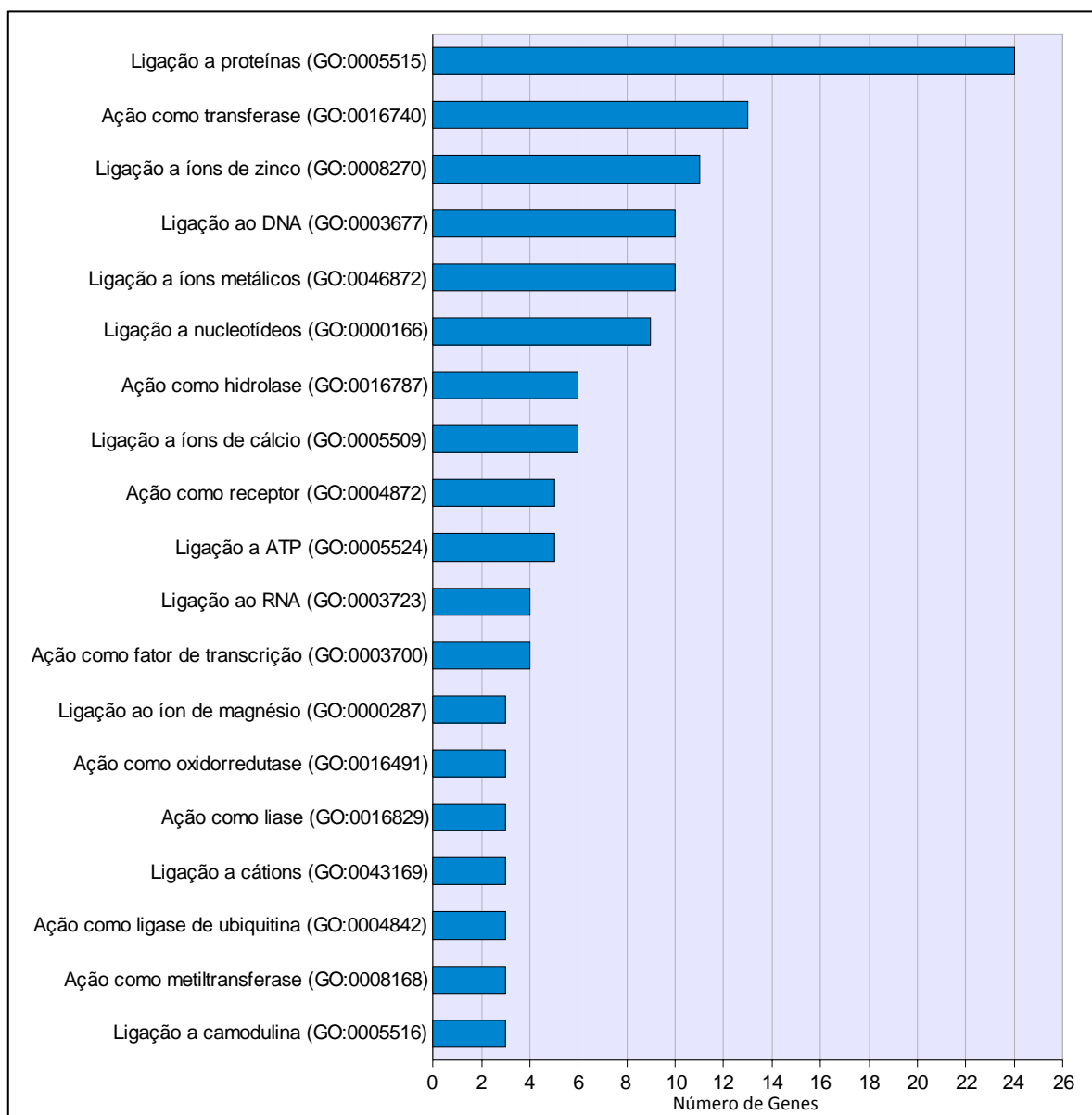


Figura 18 – Gráfico demonstrativo do número de genes representados nas redes de regulação transcricional que são responsáveis por cada função molecular (identificada no banco de dados Gene Ontology por meio de uma série numérica).

Quando se analisa a localização celular, a quantidade majoritária dos produtos gênicos pode ser encontrada no núcleo. Outras sequências codificam proteínas que integram a membrana da célula ou a de organelas, bem como o complexo de Golgi e o retículo endoplasmático. Veja a próxima figura:

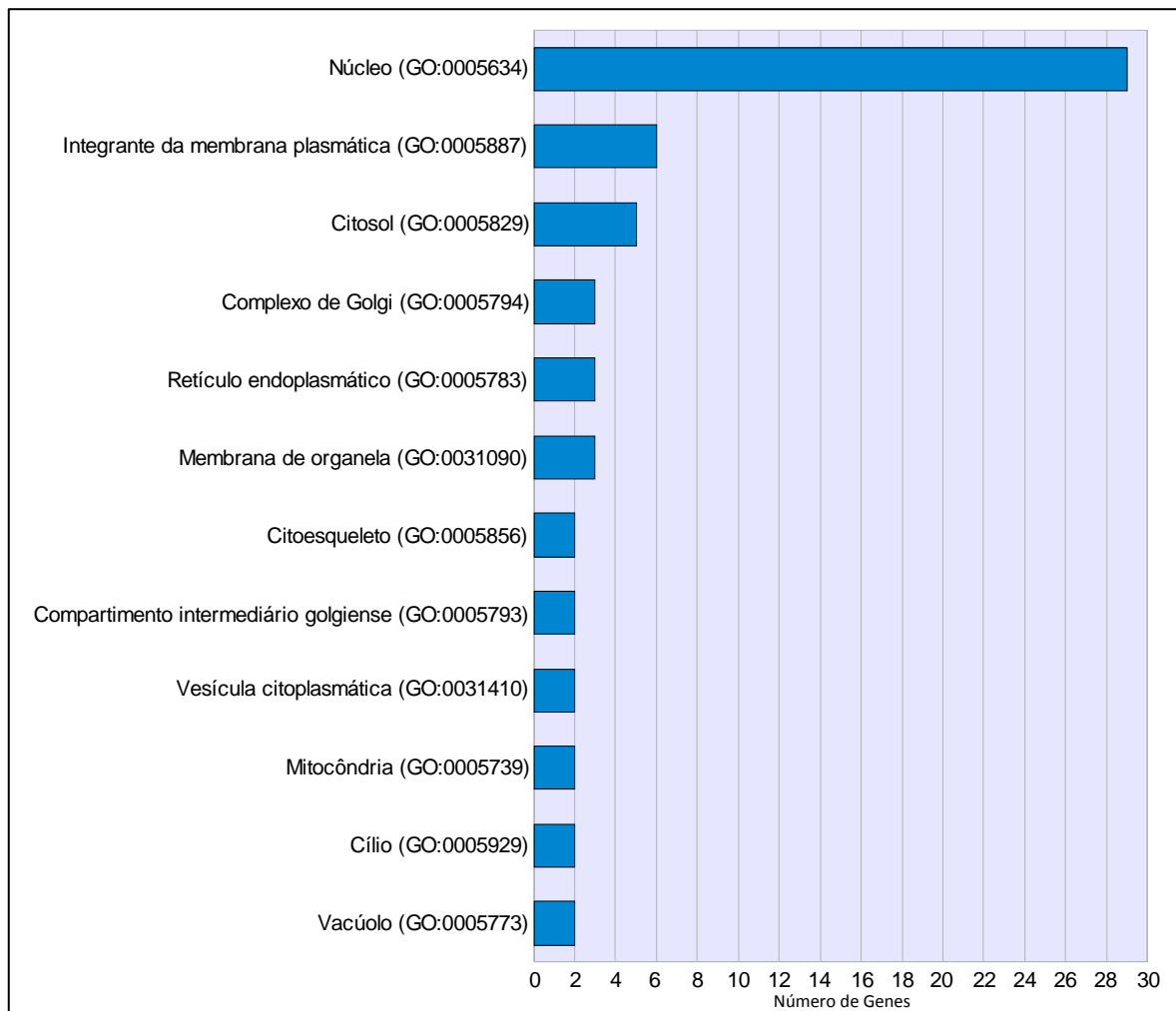


Figura 19 – Gráfico demonstrativo do número de genes representados nas redes de regulação transcricional que terão seu produto localizado numa região celular característica (identificada no banco de dados Gene Ontology por meio de uma série numérica).

A seguir, é apresentada uma tabela onde é possível conferir o símbolo, nome, número de acesso no GeneBank, localização cromossômica e a descrição sumária de cada um dos 113 genes participantes daquelas redes de regulação transcricional. As sequências localizadas em *loci* de suscetibilidade à AR aparecem com o símbolo em negrito.

Tabela 4: Genes cujos mRNAs participam das redes de regulação transcricional apresentadas nas figuras 13 - 16.

Símbolo	Nome do Gene	Número de Acesso	Citobanda	Descrição Sumária
ADCY1	Adenylate cyclase 1 (brain)	R14114.1	7p13-p12	cAMP biosynthetic process
ADD2	Adducin 2 (β)	R19254.1	2p14-p13	immune system development
ADIPOR1	Adiponectin receptor 1	R62584.1	1p36.13-q41	serve as receptors for adiponectin, fatty acid oxidation and glucose uptake by adiponectin
ANAPC2	Anaphase-promoting complex subunit 2	R33244.1	9q34.3	cellular macromolecule biosynthetic process, termed the anaphase promoting complex (APC)
ANKRD56	Ankyrin repeat domain 56	R24589.1	4q21.1	unknown
C11orf71	Chromosome 11 open reading frame 71	R19933.1	11q14.2-q14.3	unknown
C12orf51	Chromosome 12 open reading frame 51	R71499.1	12q24.13	unknown
C14orf112	Chromosome 14 open reading frame 112	R18677.1	14q24.2	unknown
C5orf42	Chromosome 5 open reading frame 42	R37164.1	5p13.2	unknown
C6orf166	Chromosome 6 open reading frame 166	R69874.1	6q15	unknown
C6orf170	Chromosome 6 open reading frame 170	R73756.1	6q22.31	regulation of signal transduction and hydrolase activity
C7orf43	Chromosome 7 open reading frame 43	R21690.1	7q22.1	unknown
CCDC22	Coiled-coil domain containing 22	R46105.1	Xp11.23	unknown
CCL28	Chemokine (C-C motif) ligand 28	R36539.1	5p12	immunoregulatory and inflammatory processes
CDC14A	CDC14 cell division cycle 14 homologue A (<i>S. cerevisiae</i>)	T74426.1	1p21	dephosphorylation, regulation of progression through cell cycle
CDC40	Cell division cycle 40 homologue (<i>S. cerevisiae</i>)	R70791.1	6q21	mRNA splicing
CEP350	Centrosomal protein 350 kDa	R25352.1	1p36.13-q41	required for anchoring microtubules to centrosomes and for integrity of microtubule network
CEP72	Centrosomal protein 72 kDa	R37706.1	5p15.33	member of the leucine-rich-repeat superfamily of proteins, localized to the centrosome
CHST10	Carbohydrate sulfotransferase 10	R16195.1	2q11.2	carbohydrate metabolic process, learning and/or memory
CNTN2	Contactin 2 (axonal)	R14151.1	1q32.1	cell adhesion, neurogenesis
COMMD2	COMM domain containing 2	R23418.1	3q25.1	unknown
CORO1B	Coronin, actin-binding protein, 1B	R36510.1	11q13.1	regulation of cell motility, signal transduction
CPEB3	Cytoplasmic polyadenylation element-binding protein 3	R36999.1	10q23.32	unknown
CPSF6	Cleavage and polyadenylation specific factor 6, 68 kDa	R18985.1	12q15	required for 3' RNA cleavage and polyadenylation processing
CRB1	Crumbs homologue 1 (<i>Drosophila</i>)	R35113.1	1q31-q32.1	eye photoreceptor cell differentiation
CRY1	Cryptochrome 1 (photolyase-like)	R19031.1	12q23-q24.1	DNA repair, regulation of transcription, visual perception

Tabela 4: Continuação

Símbolo	Nome do Gene	Número de Acesso	Citobanda	Descrição Sumária
CTBS	Chitobiase, di-N-acetyl-	R56854.1	1p22	cellular polysaccharide metabolic process
DDX49	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49	R26193.1	19p12	unknown
DHRS7B	Dehydrogenase/reductase (SDR family) member 7B	R36668.1	17p12	metabolic process
DKFZp686L-14188	Hypothetical gene supported by BX538329	R76726.1	8q21.3	unknown
EFHC1	EF-hand domain (C-terminal) containing 1	R68083.1	6p12.3	positive regulation of programmed cell death, cation homeostasis
EHMT1	Euchromatic histone-lysine <i>N</i> -methyltransferase 1	R38436.1	9q34.3	chromosome organization and biogenesis, protein methylation
EMP2	Epithelial membrane protein 2	R71514.1	16p13.2	cell development
ETNK1	Ethanolamine kinase 1	R75813.1	12p12.1	phospholipid and ethanolamine biosynthetic process
EXPH5	Exophilin 5	R70812.1	11q22.3	intracellular protein transport and cellular localization
FAM46A	Family with sequence similarity 46, member A	R71124.1	6q14	unknown
FAM70A	Family with sequence similarity 70, member A	R39187.1	Xq24	unknown
FBXL3	F-box and leucine-rich repeat protein 3	R37265.1	13q22	protein ubiquitination
FLJ22536	Hypothetical <i>locus</i> LOC401237	R15860.1	6p22.3	unknown
FNDC3A	Fibronectin type III domain containing 3A	R36431.1	13q14.2	mediates spermatid-sertoli adhesion during spermatogenesis (by similarity)
FRMPD2L2	FERM and PDZ domain containing 2 like 2	R20356.1	10q11.21	unknown
GAL3ST3	Galactose-3-Osulfotransferase 3	R14066.1	11q13.1	proteoglycan and carbohydrate biosynthetic process
GART	Phosphoribosylglycinamide formyltransferase	R14198.1	21q22.1 21q22.11	purine ribonucleoside monophosphate biosynthetic process
GPR172A	G protein-coupled receptor 172A	R18827.1	8q24.3	cell surface receptor for porcine endogenous retrovirus subgroup A
GPSM2	G protein signaling modulator 2 (AGS3 like, <i>C. elegans</i>)	R25539.1	1p13.3	cell surface receptor linked signal transduction
GSTA4	Glutathione S-transferase A4	T66320.1	6p12.1	response to stress
GTF3C4	General transcription factor IIIC, polypeptide 4, 90 kDa	R17802.1	9q34.13	transcription
HBB	Hemoglobin, β	R69941.1	11p15.5	positive regulation of nitric oxide biosynthetic process, determines the structure of the β -chain in adult hemoglobin
HLA-DRB1	Major histocompatibility complex (MHC), class II, DR β -3	H50622.1	6p21.3	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II
HMGCLL1	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase-like 1	R34667.1	6p12.1	metabolic process

Tabela 4: Continuação

Símbolo	Nome do Gene	Número de Acesso	Citobanda	Descrição Sumária
HPS5	Hermansky–Pudlak syndrome 5	R71594.1	11p14	organelle organization and biogenesis, pigmentation
HSD11B2	Hydroxysteroid (11- β) dehydrogenase 2	R35081.1	16q22	glucocorticoid biosynthetic process
INPP5E	Inositol polyphosphate-5 phosphatase, 72 kDa	R36586.1	9q34.3	lipid metabolic process
ITGA5	Integrin, α -5 (fibronectin receptor, α -polypeptide)	R32403.1	12q11-q13	integrin-mediated signaling pathway, cell adhesion
KCNT2	Potassium channel, subfamily T, member 2	R51505.1	1q31.3	potassium ion transport
LASS6	LAG1 homologue, ceramide synthase 6	R25481.1	2q24.3	regulation of transcription DNA-dependent
LIN28B	Lin-28 homologue B (<i>C. elegans</i>)	R64358.1	6q21	regulation of transcription DNA-dependent
LOC100129345	Hypothetical protein LOC100129345	R24238.1	14q32.2	unknown
LOC285954	Hypothetical protein LOC285954	R33696.1	7p13	unknown
LOC56251	Hypothetical protein YH95C04	R38312.1		unknown
LOC654350	Similar to rat ribosomal protein L9 homologue	R24333.1	15q22.31	unknown
MALT1	Mucosa-associated lymphoid tissue lymphoma translocation gene 1	T80158.1	18q21	positive regulation of cytokine production during adaptative immune response, activation of nuclear factor (NF)- κ B cascade
MOBP	Myelin-associated oligodendrocyte basic protein	R13982.1	3p22.1	nervous system development
N4BP1	NEDD4-binding protein 1	R28614.1	16q12.1	unknown
NAGA	N-acetylgalactosaminidase, α -	R25825.1	22q13-qter 22q11	carbohydrate metabolic process
NCAPG	Non-SMC condensin I complex, subunit G	—	4p15.33	mitotic sister chromatid segregation, M phase
NECAB2	N-terminal EF-hand calcium binding protein 2	T80509.1	16q23.3	unknown
NLK	Nemo-like kinase	R70769.1	17q11.2	negative regulation of Wnt receptor signaling pathway, regulation of transcription
NNAT	Neuronatin	R63918.1	20q11.2-q12	brain development
NOL7	Nucleolar protein 7, 27 kDa	R69895.1	6p23	unknown
NPY1R	Neuropeptide Y receptor Y1	R19478.1	4q31.3-q32	G protein signaling, coupled to cAMP nucleotide second messenger
NUP153	Nucleoporin 153 kDa	R40059.1	6p22.3	establishment of localization, transport
OSTM1	Osteopetrosis-associated transmembrane protein 1	R69603.1	6q21	myeloid leukocyte differentiation
PCDH17	Protocadherin 17	R49279.1	13q21.1	homophilic cell adhesion
PDE1B	Phosphodiesterase 1B, calmodulin-dependent	H51117.1	12q13	programmed cell death, visual learning
PECR	Peroxisomal trans-2-enoyl-CoA reductase	H84466.1	2q35	regulation of apoptosis, fatty acid biosynthetic process
PHF1	PHD finger protein 1	R37685.1	6p21.3	transcriptional repressor

Tabela 4: Continuação

Símbolo	Nome do Gene	Número de Acesso	Citobanda	Descrição Sumária
PHLPL	PH domain and leucine rich repeat protein phosphatase-like	R25700.1	16q22.3	dephosphorylation
PIP5K1A	Phosphatidylinositol-4-phosphate 5-kinase, type I, α	—	1q22-q24	glycerophospholipid metabolic process
PKD2	Polycystic kidney disease 2 (autosomal dominant)	R38199.1	4q21-q23	negative regulation of progression through cell cycle, JAK-STAT cascade, calcium
PLCE1	Phospholipase C, epsilon 1	R40791.1	10q23	positive regulation of MAPK activity, G protein signaling coupled to IP3 second messenger, diacylglycerol biosynthetic process, regulation of cell growth
POLR1E	Polymerase (RNA) I polypeptide E, 53 kDa	R54238.1	9p13.2	rRNA transcription
POLR3E	Polymerase (RNA) III (DNA directed) polypeptide E, 80 kD	R18902.1	16p12.1	transcription
POMP	Proteasome maturation protein	R38655.1	13q12.3	immune response, protein folding
PPCDC	Phosphopantothencysteine decarboxylase	R33457.1	15q24.2	coenzyme A biosynthetic process
PREPL	Prolyl endopeptidase-like	R15785.1	2p21	proteolysis
RAB37	RAB37, member RAS oncogene family	R25709.1	17q25.1	small GTPase-mediated signal transduction
RAPGEF1	Rap guanine nucleotide exchange factor (GEF) 1	R76487.1	9q34.3	transmembrane receptor protein tyrosine kinase signaling pathway
RBJ	Rab and DnaJ domain Containing	H06266.1	2p23.3	small GTPase mediated signal transduction
RBM23	RNA-binding motif protein 23	R42182.1	14q11.2	increases transcription of steroid-responsive transcriptional reporters in a hormone-dependent manner
S1PR1	Sphingosine-1-phosphate receptor 1	R20666.1	1p21	receptor for the lysosphingolipid sphingosine 1-phosphate (S1P)
SCOC	Short coiled-coil protein	R34436.1	4q31.1	unknown
SERPINH1	Serpin peptidase inhibitor, clade H (heat shock protein 47), member 1 (collagen-binding protein 1)	R71093.1	11q13.5	molecular chaperone involved in the maturation of collagen molecules (autoantibodies to this protein have been found in patients with rheumatoid arthritis)
SETBP1	SET-binding protein 1	R17655.1	18q21.1	unknown
SETDB1	SET domain, bifurcated 1	R37131.1	1q21	establishment and/or maintenance of chromatin architecture, blastocyst development
SLC38A1	Solute carrier family 38, member 1	R20471.1/ R66555.1	12q13.11	glutamine transport
SLC3A2	Solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	T80267.1	11q13	tryptophan transport, calcium ion transport, cell growth
SMAD2	SMAD family member 2	R33715.1	18q21.1	regulation of transcription, mesoderm formation

Tabela 4: Continuação

Símbolo	Nome do Gene	Número de Acesso	Citobanda	Descrição Sumária
SMAP2	Stromal membrane-associated GTPase-activating protein 2	R20203.1	1p35.3-p34.1	unknown
SPEG	SPEG complex <i>locus</i>	R25020.1	2q35	protein amino acid phosphorylation, muscle development
STX7	Syntaxin 7	R73740.1	6q23.1	post-Golgi vesicle mediated transport
SYN2	Synapsin II	R38909.1/ R49277.1	3p25	neurotransmitter secretion
TBC1D22B	TBC1 domain family, member 22B	T66272.1	6p21.2	regulation of Ras GTPase activity, regulation of Rab protein signal transduction
TCN1	Transcobalamin I (vitamin B12-binding protein, R binder family)	R11710.1	11q11-q12	cobalamin transport
TRIM27	Tripartite motif-containing 27	H08440.1	6p22	repression of gene transcription, spermatogenesis
USP45	Ubiquitin specific peptidase 45	R15391.1	6q16.3	proteolysis involved in cellular protein catabolic process
UBR5	Ubiquitin protein ligase E3 component n-recognin 5	R18607.1	8q22	ubiquitin-dependent protein catabolic process
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6 (Ku autoantigen, 70 kDa)	T79976.1	22q13.2-q13.31	double-strand break repair via nonhomologous end joining, positive regulation of transcription
ZNF236	Zinc finger protein 236	R25834.1	18q22-q23	regulation of transcription DNA-dependent
ZNF385C	Zinc finger protein 385C	R20240.1	17q21.2	unknown
ZNF511	Zinc finger protein 511	R68665.1	10q26.3	regulation of transcription DNA-dependent
ZNF718	Zinc finger protein 718	R37117.1	4p16.3	unknown
ZSCAN16	Zinc finger and SCAN domain containing 16	R35096.1	6p22.1	regulation of transcription DNA-dependent

Discussão

5 DISCUSSÃO

5.1 Considerações Iniciais

As doenças autoimunes, incluindo a artrite reumatoide, impõem grandes desafios à investigação médica e biológica devido principalmente à heterogeneidade de características e sua etiologia multifatorial. O desequilíbrio no controle central da tolerância, o qual envolve a expressão tímica de genes codificando autoantígenos tissulares (fenômeno conhecido como expressão gênica promíscua), tem sido apontado como um dos principais mecanismos responsáveis por tal disfunção imunológica (KYEWSKI & DERBINSKI, 2004; MAGALHÃES *et al.*, 2006). Além disso, fatores genéticos têm sido fortemente associados à suscetibilidade a doenças autoimunes, como os alelos HLA-DRB1, localizados na região cromossômica 6p21.3, que conferem suscetibilidade à AR dentre outras doenças. Vários fatores ambientais, como tabagismo e infecções, também poderiam desencadear uma condição autoimunológica nos indivíduos geneticamente suscetíveis.

O estudo de genes candidatos e análises por *genome-wide linkage*, os quais representam abordagens primordiais para a descoberta de genes relevantes na AR (OROZCO *et al.*, 2006; BARTON *et al.*, 2008; BOWES & BARTON, 2008; LIU *et al.*, 2008), têm possibilitado a identificação de diferentes regiões cromossômicas que, de forma semelhante à região HLA, apresentam genes desempenhando um importante papel de suscetibilidade à AR. Como exemplo, estudos de *genome-wide scan* revelaram um novo gene responsável pela transdução de sinal linfocitária, chamado de PTPN22 (*protein tyrosine phosphatase nonreceptor type 2*). Situado na região 1p13.3-p13.1, ele parece estar envolvido com várias doenças autoimunes, incluindo AR, LES e diabetes mellitus do tipo 1 (DM1) (CANTOR *et al.*, 2004; JOHN *et al.*, 2004; CHOI *et al.*, 2006).

Apesar de sua reconhecida contribuição para a investigação dos *loci* de suscetibilidade à AR, as técnicas de análise considerando genes candidatos ou por *genome-wide linkage* são incapazes de elucidar todos os mecanismos envolvidos em sua patologia. Além dos chamados estudos de associação (YAMAMOTO & YAMADA, 2005; WANG *et al.*, 2005), a análise da expressão gênica em larga escala tem se destacado nesta tarefa, pois genes localizados em regiões de suscetibilidade provavelmente encontram-se diferencialmente expressos na AR.

A partir da tecnologia de cDNA microarrays, o Grupo de Imunogenética Molecular tem demonstrado importantes características do perfil transcricional de doenças autoimunes, incluindo a descrição de como a expressão gênica de linfócitos T é afetada pelo tratamento imunossupressor de pacientes com LES (PEREIRA *et al.*, 2004); a determinação da localização genômica de novas sequências (ESTs) originárias de *loci* de suscetibilidade para LES (TREVISAN *et al.*, 2006); a associação entre diferentes genótipos HLA de classe II e a expressão gênica diferencial de pacientes com DM1 (RASSI *et al.*, 2006); a meta-análise de dados de expressão gênica provenientes de AR e LES, de modo a identificar suas assinaturas moleculares comuns e exclusivas (SILVA *et al.*, 2007); além da reconstrução de redes transcricionais durante a ontogenia do timo (MACEDO *et al.*, 2008).

Num dos trabalhos recentemente divulgados pelo grupo, a metodologia dos microarrays permitiu estabelecer a associação existente entre o perfil de expressão gênica diferencial de pacientes com AR e características tais como tipo de resposta imune, nível de atividade da doença e eficácia do tratamento. Muitos dos genes então identificados pertencem a regiões de suscetibilidade à doença; outros participam de diferentes processos biológicos, como transdução de sinais, fosforilação e apoptose (JUNTA *et al.*, 2008). Este estudo contribuiu para dissecar a complexidade molecular envolvida na AR, além de promover um melhor entendimento acerca da influência genética sobre o fenótipo da doença e o tipo de resposta terapêutica.

A técnica de cDNA microarrays é responsável por mensurar a concentração de mRNA nos tecidos inflamados ou em células mononucleares do sangue periférico (*peripheral blood mononuclear cells* – PBMCs), constituindo-se uma ferramenta útil para avaliar as diferenças nos níveis de expressão gênica da AR (TOONEN *et al.*, 2008). Enquanto as amostras oriundas da membrana sinovial podem fornecer indícios mais precisos sobre os fatores patogênicos envolvidos na doença crônica, as PBMCs têm sua utilização difundida pela facilidade de acesso às amostras, o que é essencial na prática experimental. Elas são consideradas elementos-chave da autodefesa, agindo inclusive como repórteres da condição imunológica vigente no organismo, propriedades que as tornam relevantes na descoberta de novos marcadores moleculares, bem como na investigação de variantes de doenças e avaliação da resposta a diferentes tratamentos (GU *et al.*, 2002; SHAFFER *et al.*, 2006; YAMAGATA *et al.*, 2006).

No presente estudo, foram avaliados os perfis transcricionais característicos de PBMCs de quinze pacientes com AR utilizando-se a tecnologia dos cDNA microarrays. Considerando que os níveis de expressão gênica podem ser influenciados pelo nível de atividade da doença (DAS28), pela presença de marcadores autoimunes (anticorpos anti-CCP) e de fatores genéticos conferindo suscetibilidade (HLA-DRB1 SE), e mesmo devido à administração medicamentosa (DMARDs + anti-TNF), os pacientes com AR foram estratificados de acordo com estas variáveis.

Após a análise estatística, foi construído um diagrama de Venn (figura 7) a fim de estabelecer quais são os genes significativamente modulados sob cada uma dessas condições. Os genes com expressão diferencial em cada condição foram então submetidos à modelagem das interações regulatórias entre si, de maneira a produzir uma rede transcricional característica de cada variável (figuras 13-16). Por “rede transcricional característica” entenda-se o conjunto de interações gênicas encontradas na maioria dos pacientes agrupados segundo determinada variável. Por exemplo, considerando o conjunto dos nove pacientes com DAS28 > 5.1, foram

selecionadas apenas as interações regulatórias presentes em, no mínimo, oito desses pacientes. No caso dos cinco pacientes tratados com DMARDs + anti-TNF, considerou-se exclusivamente as interações regulatórias observadas na totalidade dos indivíduos. Os principais aspectos das redes gênicas obtidas serão discutidos nos parágrafos a seguir.

5.2 Sobre as Redes de Regulação Transcricional

Em primeiro lugar, considere os genes modulados dentre os dez pacientes portadores do epítipo compartilhado (*shared epitope* – SE) em dose única ou dupla, cujos mRNAs estabelecem a rede transcricional representada na figura 13. Pode-se destacar a presença de vários nós gênicos (reconhecidos como genes com grande atividade regulatória), como, por exemplo, o gene LIN28B (*lin28 homologue B*), localizado na região 6q21, que corresponde a um dos *loci* de suscetibilidade (LS) à AR. Capaz de interagir com a molécula de DNA e regular sua transcrição, o produto gênico de LIN28B controla os níveis de expressão de outros doze genes, incluindo CCDC22, SYN2, SCOC, RAPGEF1, NAGA, RBM23, LASS6 e GPR172A, todos situados em diferentes cromossomos.

Ainda no cromossomo 6, diametralmente oposto ao gene LIN28B, mais especificamente no LS 6p22.3, está o gene FLJ22536 (*hypothetical locus LOC401237*), que, embora tenha sua função desconhecida, mostra-se regulando a funcionalidade do nó gênico ZNF236 (*zinc finger protein 236*). Localizado no LS 18q22-q23, este é responsável por sintetizar uma proteína reativa ao íon zinco, além de se ligar ao DNA de modo a controlar o processo de transcrição. Na rede em questão, o gene ZNF236 modula dez genes, que incluem RBM23, LASS6 e GPR172A (igualmente regulados pelo gene LIN28B, conforme mencionado acima), bem como CRY1, CUL3 e HMGCLL1, sendo que a este último, encontrado no LS 6p12.1, se atribui a participação em vários processos metabólicos.

Bem próximo ao ZNF236, no LS 18q21.1, localiza-se o gene SMAD2 (*SMAD family member 2*), codificador de uma proteína que, após ser processada, adentra o núcleo e liga-se a promotores alvo de forma a impedir o processo de transcrição juntamente com outros cofatores. Na rede em questão, observa-se que ele modula os níveis de expressão dos genes PCDH17, UBR5 e ZNF385C.

Agindo como mediador de várias cascatas de sinalização (KEGG pathways: hsa04110, hsa 04310, hsa04350, hsa05200; KEGG gene: hsa:4087), o gene SMAD2 destaca-se por participar da via de sinalização do TGF- β (*transforming growth factor beta*), o qual é responsável pelos processos biológicos de proliferação e diferenciação celulares, desempenhando papel crucial no sistema imune, diabetes, câncer e patologias cardíacas.

Com relação ao sistema imune, acredita-se que o TGF- β esteja envolvido no desenvolvimento de linfócitos T imunorreguladores CD4+CD25+ e linfócitos T helper 17 (Th17) (CHEN *et al.*, 2003; YANG *et al.*, 2008). Ele também está fortemente associado à síndrome de Marfan (OMIM # 154700) (BOILEAU *et al.*, 2005), doença caracterizada pela falta de fibrilina-1, um tipo de glicoproteína colágena que, juntamente com elastina e fibronectina, constitui a matriz extracelular de tecidos conjuntivos, como a membrana sinovial.

É notável o fato dos genes LIN28B, ZNF236 e SMAD2, cujos produtos gênicos são capazes de se ligar à molécula de DNA de maneira a controlar o processo de transcrição, aparecerem na rede de interações exatamente como nós gênicos, cada um regulando os níveis de expressão de vários genes concomitantemente. Neste caso, as interações gênicas inferidas unicamente de acordo com seus níveis de expressão poderiam ser então explicadas com base em interações físicas de moléculas.

Observe na figura 13 que, além de serem regulados por LIN28B e ZNF236, ambos os genes RBM23 e LASS6 apresentam seus níveis de expressão influenciados pelo nó gênico EFHC1 (*EF-hand domain containing 1*), pertencente ao LS 6p12.3. Incumbido de promover a morte celular programada e a homeostase de

cátions, o EFHC1 também estende seu controle sobre os genes HPS5, ANKRD56, PHLPP1, PIP5K1A, dentre outros. Perceba ainda que sua transcrição é modulada pelo nó gênico HLA-DRB1 (*major histocompatibility complex – MHC, class II, DR β -1*), situado no LS 6p21.3, o qual controla a expressão dos seguintes genes: ADC, C5orf42, C7orf43, CRB1, FNDC3A e *expressed sequence tag* ID: 142259.

O alelo HLA-DRB1 é responsável por codificar a cadeia β (DRB), que, juntamente com uma cadeia α (DRA), constitui a molécula HLA de classe II. Este heterodímero, ancorado através da membrana plasmática de células apresentadoras de antígenos, tem a função de expor peptídeos derivados de proteínas extracelulares (previamente processados no interior das APCs), auxiliando numa das principais tarefas do sistema imune: discernir entre antígenos próprios e não próprios. Assim, variantes DRB1 estão diretamente envolvidas no adequado funcionamento do sistema defensivo do organismo, desempenhando um importante papel de suscetibilidade a doenças autoimunes, motivo pelo qual são reconhecidas como epítomos compartilhados da artrite reumatoide.

É muito interessante o fato do HLA-DRB1 aparecer regulando os níveis de expressão do gene FNDC3A (*fibronectin type III domain containing 3A*), que pertence a uma família de proteínas contendo múltiplos domínios de fibronectina tipo III e uma região terminal rica em prolina, estando supostamente ancoradas na membrana celular (OBHOLZ *et al.*, 2006).

Mais especificamente, a proteína sintetizada por FNDC3A possui um total de nove domínios idênticos, cuja sequência é homóloga à fibronectina tipo III, que corresponde a um segmento de aminoácidos repetido de quinze a dezessete vezes na fibronectina constituinte da matriz extracelular (PANKOV & YAMADA, 2002; CRAIG *et al.*, 2004).

Também formada por um segmento do tipo I, repetido doze vezes, e um módulo do tipo II, que aparece duas vezes, a molécula de fibronectina desempenha funções de adesão, dispersão, crescimento e diferenciação celulares, participando

da formação estrutural de vários tecidos, como cartilagem articular e membrana sinovial (RUOSLAHTI, 1988).

No entanto, uma vez degradada por proteinases, a fibronectina também adquire atividade catalítica, podendo ser demonstrado que seus fragmentos com ligação a heparina COOH-terminal, heparina NH₂-terminal, e gelatina NH₂-terminal estimulam a condrólise de cartilagem, induzindo a sua destruição (HOMANDBERG *et al.*, 1992; YASUDA & POOLE, 2002). De fato, fragmentos de fibronectina com 30 a 200 kDA foram encontrados em titulações elevadas no líquido sinovial e na cartilagem de pacientes com osteoartrite e AR (XIE *et al.*, 1992; HOMANDBERG *et al.*, 1998). De acordo com YASUDA e colaboradores, na artrite reumatoide os fragmentos de fibronectina ligantes a heparina COOH-terminal induziriam, via moléculas de integrina $\alpha 4\beta 1$, a produção de metaloproteinases de matriz (MMP-1, MMP-3 e MMP-13) nos fibroblastos sinoviais, causando progressiva destruição da membrana (YASUDA *et al.*, 2003).

A dependência condicional apresentada pelo FNDC3A com relação aos níveis de expressão gênica do epítipo compartilhado HLA-DRB1 é considerada de especial interesse. Uma investigação mais aprofundada desta interação regulatória poderia promover um melhor entendimento acerca da função molecular dos *loci* de suscetibilidade à artrite reumatoide, como também a outras doenças autoimunes.

A seguir, considere a rede representada na figura 14, coposta por vinte e dois genes cujos mRNAs estabelecem as interações regulatórias características dos nove pacientes com anticorpos anti-CCP. Nela, é possível identificar quatro interações gênicas que também aparecem no grupo de pacientes com SE positivo: o gene PECR, pertencente ao LS 2q35, possui ação regulatória sobre RBM23; ETNK1, localizado no LS 12p12.1, modula o gene UBR5; o gene LIN28B controla os níveis de expressão gênica de RAPGEF1; e, finalmente, FLJ22536 apresenta-se regulando o gene RPL24.

Por outro lado, dentre as particularidades da rede, pode-se mencionar a presença do gene USP45 (*ubiquitin-specific peptidase 45*), pertencente ao LS 6q16.3,

que, envolvido no processo biológico de proteólise dependente de ubiquitina, é capaz de influenciar a transcrição das *expressed sequence tags* ID: 29644 e ID: 27725. Este último, inclusive, é igualmente regulado pelo gene CDC40 (*cell division cycle 40 homologue*), situado no LS 6q21, que participa da segunda reação catalítica envolvida no processo de *splicing* de pré-mRNAs, estando possivelmente relacionado com a progressão do ciclo celular (DAHAN & KUPIEC, 2002).

Outro nó gênico que aparece pela primeira vez é o chamado SLC38A1 (*solute carrier family 38, member 1*), encontrado no LS 12q13.11, cuja função está associada ao transporte do aminoácido glutamina, precursor do neurotransmissor glutamato. Na rede em questão, o SLC38A1 modula os níveis de expressão gênica de C14orf112 (*alias* do gene COX16, *cytochrome c oxidase assembly factor homologue*), DKFZp686L14188 e RAPGEF1.

Destaca-se também a presença do gene OSTM1 (*osteopetrosis-associated transmembrane protein 1*), localizado no LS 6q21, o qual codifica uma proteína com região N-terminal rica em leucina, cujas propriedades permitem sua ligação a membros da subfamília A participantes da família RGS (*regulator of the G-protein signaling*), além de funcionar como ligase da ubiquitina E3. O gene OSTM1, que na rede em questão aparece modulando a *expressed sequence tag* ID: 144045, desempenha papel fundamental na maturação e função de osteoclastos e melanócitos, de modo que sua depleção pode ocasionar uma doença autossômica recessiva conhecida como osteopetrose maligna infantil (OMIM # 259720) (DEL FATORE *et al.*, 2008).

Agora, considere o conjunto de genes modulados dentre os nove pacientes exibindo doença muito ativa ($DAS28 > 5.1$), cujos níveis de expressão permitiram a reconstrução da rede regulatória representada na figura 15. Destacam-se alguns nós gênicos, já conhecidos, realizando novas interações transcricionais em decorrência da reorganização dos pacientes segundo o nível de atividade da doença.

O gene LIN28B mantém a modulação de RBM23 e RAPGEF1, mas agora estabelece controle sobre os níveis de expressão gênica de C14orf112, HBA1 e as *expressed sequence tags* ID: 53110 e ID: 38915. Já o nó gênico ZNF236 conserva sua interação com RBM23, mas aparece interferindo na transcrição de PLCE1, RPL24 e C5orf42. O nó gênico HLA-DRB1 continua regulando FNDC3A, mas neste caso passa a se relacionar com quatro novos genes: C12orf51 e as *expressed sequence tags* ID: 27415, ID: 27941 e ID: 142878. O nó gênico USP45, além de continuar modulando as *expressed sequence tags* ID: 27725 e ID: 29644, agora estende sua influência sobre os níveis de expressão dos seguintes genes: UBR5, SERPINH1, LOC100129345, MOBP e FNDC3A. O gene SLC38A1, por sua vez, segue controlando C14orf112 e RAPGEF1, mas estabelece novas interações com RBM23, EXPH5, RBJ e com as *expressed sequence tags* ID: 39294 e ID: 137208. Perceba também que ETNK1 surge pela primeira vez como nó gênico, modulando EBR5 e mais sete genes distintos: CHD2, POLR1E, NOL7, EMP2, as *expressed sequence tags* ID: 27725 e ID: 137208, além de NUP153, que está localizado no LS 6p22.3.

Finalmente, leve em consideração a rede gênica representada na figura 16, a qual especifica as interações regulatórias comumente encontradas no conjunto dos pacientes tratados com Metotrexato combinado ao Infliximabe, que é um agente biológico com ação bloqueadora de TNF- α (DMARDs + anti-TNF). Nela podem ser notadas algumas dependências transcricionais envolvendo *loci* de suscetibilidade que ainda não apareceram em nenhuma das redes discutidas anteriormente.

É o caso do gene PDE1B (*phosphodiesterase 1B, calmodulin-dependent*), que, localizado no LS 12q13, faz parte de uma família de genes com ação catalítica dependente de calmodulina, os chamados *calmodulin-dependent cyclic nucleotide phosphodiesterases* (CaM-PDEs). A calmodulina representa o arquétipo de um conjunto de proteínas ativadas pelo cálcio, e as PDEs são enzimas que ajudam a retirar o grupamento fosfato de nucleotídeos purínicos – guanosina monofosfato cíclico (GMPc) e adenosina monofosfato cíclico (AMPc) –, transformando-os nos

respectivos nucleosídeos (KEGG pathway: hsa00230, KEGG gene: hsa:5153, KEGG enzyme #3.1.4.17). Na rede em questão, o PDE1B é capaz de influenciar a transcrição do gene CRY1, relacionado ao reparo de DNA e à percepção visual, participando do ritmo circadiano de mamíferos (KEGG pathway: hsa04710, KEGG gene: hsa:1407).

Outro exemplo é o gene ITGA5 (*integrin, α -5*), contíguo ao gene PDE1B, uma vez que está localizado no LS 12q11-q13. Ele é responsável por sintetizar a cadeia alfa do tipo 5, que será clivada no meio extracelular para, juntamente com a cadeia beta do tipo 1, formar uma molécula de integrina capaz de se ligar à fibronectina (KEGG pathway: hsa04512, KEGG gene: hsa:3678).

Sandya e colaboradores observaram um aumento progressivo e sincronizado dos níveis de integrina $\alpha 5\beta 1$ e fibronectina no líquido sinovial e na cartilagem articular de ratos induzidos a desenvolver artrite por meio de injeções de colágeno tipo II. Para os autores, as alterações paralelas nos níveis de integrina $\alpha 5\beta 1$ e fibronectina de condrócitos constituiriam um dos mecanismos moleculares durante a progressão da artrite (SANDYA *et al.*, 2007).

De acordo com Alsaleh e colegas, as vias de sinalização da integrina $\alpha 5\beta 1$ estimulariam a secreção de moléculas denominadas BAFF pelos sinoviócitos, as quais apresentam titulações elevadas em amostras do líquido sinovial de pacientes com artrite reumatoide (ALSALEH *et al.*, 2007). Além disso, foi demonstrado que a superprodução de BAFF por células dendríticas e macrófagos pode desempenhar um papel crucial na patogênese de artrite em modelos experimentais (PERS *et al.*, 2005; ZHANG *et al.*, 2005).

O gene ITGA5 mostra-se, portanto, associado a processos de catabolismo, i.e., de degradação da matriz extracelular constituinte da membrana sinovial e da cartilagem articular.

Na rede em questão, o gene ITGA5 aparece juntamente com ETNK1 modulando os níveis de expressão de SLC3A2 (*solute carrier family 3, member 2*). Curiosamente, este gene codifica uma proteína ancorada através da membrana celular que se

associa a integrinas para mediar vias de sinalização relacionadas ao crescimento celular normal e à tumorigênese.

A equipe de Chloé Féral demonstrou que o gene SLC3A2 (CD98hc) está relacionado à formação da matriz de fibronectina *in vivo* e que sua depleção resultou na diminuição da montagem da matriz extracelular *in vitro*, fenômeno este não dependente da disponibilidade de fibronectina ou de sua afinidade com integrinas. Os autores explicam que a proteína sintetizada pelo SLC3A2, ao se ligar com integrinas (principalmente $\alpha 5\beta 1$, $\alpha IIb\beta 3$ e $\alpha v\beta 3$), dispara cascatas de sinalização intracelular que ativam moléculas chamadas RhoA, responsáveis por promover a contratilidade celular necessária à montagem da matriz de fibronectina (FÉRAL *et al.*, 2007).

O gene SLC3A2 é evidenciado, portanto, como participante de processos anabólicos, i.e., de síntese ou construção da matriz extracelular formadora dos tecidos conjuntivos.

A rede de interações em questão mostra claramente a dependência condicional entre os níveis de expressão gênica de ITGA5 e SLC3A2, os quais 1) codificam proteínas que interajam fisicamente e 2) estão relacionadas à manutenção da matriz extracelular de fibronectina, fenômeno que depende do equilíbrio entre processos anabólicos e catabólicos (YASUDA, 2006).

É muito interessante o fato de estes genes aparecerem tão somente na rede comum aos pacientes tratados com DMARDs + anti-TNF, permitindo sugerir que os benefícios decorrentes do tipo de administração medicamentosa, tais como a redução na destruição da cartilagem articular, poderiam estar relacionados a alterações no perfil de expressão de genes importantes na artrite reumatoide, tais como ITGA5 e SLC3A2.

5.2 Sobre o Programa Mandala

A investigação de grafos complexos consiste numa tarefa muito difícil caso não esteja disponível uma ferramenta dedicada. Sem dúvida, é praticamente impossível analisar visualmente milhares de interações gênicas em busca de alguma informação biológica consistente. No contexto deste trabalho, a comparação de quinze redes transcricionais diferentes (cada uma composta por centenas de genes), com o objetivo de identificar as interações regulatórias comuns aos pacientes com artrite reumatoide, só foi viável graças ao desenvolvimento do programa Mandala.

De fato, a disponibilidade desta ferramenta causou um impacto bastante positivo às pesquisas então realizadas pelo Grupo de Imunogenética Molecular. Por exemplo, num trabalho recentemente desenvolvido pela Dra. Claudia Macedo, o programa Mandala foi utilizado para interpretar as redes transcricionais envolvidas na chamada expressão gênica promíscua (PGE) de autoantígenos tissulares por mTECs (*medullary thymic epithelial cells*), fenômeno essencial à formação de um repertório de células T tolerantes a antígenos próprios.

A autora percebeu que os genes expressando autoantígenos tissulares periféricos (PTAs) realmente estavam interconectados na rede. Dentre eles, destaca-se o gene *Gucy2d* (*guanylate cyclase 2d*), que aparecia modulando 74 genes positivamente e 8 negativamente. Quando fora silenciado o gene *Aire* (*autoimmune regulator*), o qual desempenha papel fundamental na PGE por funcionar como fator de transcrição de muitos PTAs, *Gucy2d* teve sua expressão diminuída, mas passou a induzir 78 e a reprimir 120 genes. A explicação está no fato do gene *Gucy2d* apresentar regiões *T-* e *G-boxes* à montante do códon ATG correspondentes aos sítios de ligação da proteína AIRE (MACEDO *et al.*, 2009).

O programa Mandala também foi útil nas pesquisas realizadas por Adriane Feijó Evangelista, que, ao longo de seu mestrado, investigou as redes de regulação transcricional envolvidas no processo de diferenciação carcinogênica de células tronco mesenquimais humanas (MSC). A autora observou que a ocorrência de algumas interações regulatórias dependia da origem das MSC: enquanto os nós gênicos ALCAM e TP53 apareciam apenas nas redes características de MSC obtidas a partir da medula óssea, os nós NRCAM e PI3K1 eram exclusivos das redes gênicas encontradas nas MSC provenientes do cordão umbilical.

Por outro lado, foi possível determinar que os genes PRKDC e GNG12 apresentam grande atividade regulatória em ambas as redes. Além disso, a identificação de sequências que interagem com fatores de transcrição reforçam a possibilidade de ocorrência das interações propostas, as quais, segundo a autora, dão “uma ideia panorâmica do mecanismo de controle observado nas MSC”, podendo decidir seu destino à diferenciação normal ou ao câncer (EVANGELISTA, 2009).

A receptividade da nova ferramenta pelos integrantes do laboratório acabou motivando a sua disponibilização a toda comunidade científica. Juntamente com o Dr. Pedro R. Rodrigues Prado, analista de sistemas do Departamento de Genética (FMRP), foi então desenvolvida uma página web com as mesmas funcionalidades do programa Mandala.

Portando arquivos referentes à lista de genes e à matriz regulatória, obtidas a partir do software GeneNetwork, o usuário é capaz de representar uma rede gênica no formato textual, bem como obter as interações regulatórias comuns a duas redes. As opções quanto ao tipo de orientação no qual o resultado será organizado (ou seja, a partir dos genes que modulam ou que são modulados) foram mantidas. Além disso, nessa versão é possível escolher se os genes serão representados por meio do símbolo ou da respectiva identificação numérica (ID),

que não é redundante. O site já está disponível no endereço: <http://www.rge.fmrp.usp.br/mandala>.

Com a possibilidade de utilização do programa Mandala por qualquer pessoa no mundo interessada em analisar as redes gênicas obtidas a partir de um software tão difundido quanto o GeneNetwork, espera-se ajudar, de alguma forma, a dar ainda mais visibilidade à produção científica da Universidade de São Paulo.

Conclusões

6 CONCLUSÕES

- ❖ Neste trabalho, dados de expressão gênica obtidos por meio da tecnologia de cDNA microarrays foram utilizados para inferir as redes de regulação transcricional envolvendo *loci* de suscetibilidade à artrite reumatoide que são características de pacientes organizados de acordo com o nível de atividade da doença, presença do epítipo compartilhado HLA-DRB1 e de anticorpos contra peptídeos citrulinados cíclicos, além do tipo de medicamento recebido.
- ❖ Dentre as interações gênicas observadas, destaca-se a regulação transcricional exercida pelo epítipo compartilhado HLA-DRB1, localizado no *loci* de suscetibilidade 6p21.3, sobre o gene FNDC3A, responsável por codificar uma proteína com domínios da molécula de fibronectina, a qual participa da montagem da matriz extracelular de tecidos conjuntivos, como membrana sinovial e cartilagem articular.
- ❖ Outra interação gênica de grande interesse é a regulação transcricional exercida pelo gene ITGA5, pertencente ao *loci* de suscetibilidade 12q11-q13, sobre SLC3A2. As proteínas sintetizadas por ambos os genes interagem fisicamente e estão associadas à manutenção da matriz extracelular de fibronectina.
- ❖ Também foi desenvolvido o programa Mandala, capaz de fornecer uma representação textual para as redes gênicas obtidas a partir do software GeneNetwork, bem como identificar as interações comuns a duas redes regulatórias. Ele está disponível no endereço <http://www.rge.fmrp.usp.br/mandala>.

Referências Bibliográficas

7 REFERÊNCIAS BIBLIOGRÁFICAS

- Agrawal S, Misra R, Aggarwal A. Autoantibodies in rheumatoid arthritis: association with severity of disease in established RA. *Clinical Rheumatology* 26(2): 201-204, 2007.
- Aho K, Koskenvuo M, Tuominen J, Kaprio J. Occurrence of rheumatoid arthritis in a nationwide series of twins. *Journal of Rheumatology* 13(5): 899-902, 1986.
- Alamanos Y and Drosos AA. Epidemiology of adult rheumatoid arthritis. *Autoimmunity Reviews* 4(3): 130-136, 2005.
- Alegre ML, Frauwirth KA, Thompson CB. T-cell regulation by CD28 and CTLA-4. *Nature Reviews Immunology* 1(3): 220-228, 2001.
- Aletaha D and Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clinical and Experimental Rheumatology* 23(5 Suppl 39): S100-S108, 2005.
- Alsaleh G, Messer L, Semaan N, Boulanger N, Gottenberg JE, Sibia J, Wachsmann D. BAFF synthesis by rheumatoid synoviocytes is positively controlled by $\alpha 5\beta 1$ integrin stimulation and is negatively regulated by tumor necrosis factor α and toll-like receptor ligands. *Arthritis & Rheumatism* 56(10): 3202-3214, 2007.
- Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. The American

- Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatism* 31(3): 315-324, 1988.
- Bachmaier K, Neu N, de la Maza LM, Pal S, Hessel A, Penninger JM. Chlamydia infections and heart disease linked through antigenic mimicry. *Science* 283(5406): 1335-1339, 1999.
- Bansal M, Della Gatta G, di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* 22(7): 815-822, 2006.
- Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Gibbons L, Plant D; Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P; YEAR consortium, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Worthington J. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Human Molecular Genetics* 17(15): 2274-2279, 2008.
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nature Genetics* 37(4): 382-390, 2005.
- Bengtsson H, Jönsson G, Vallon-Christersson J. Calibration and assessment of channel-specific biases in microarray data with extended dynamical range. *BMC Bioinformatics* 5: 177, 2004.
- Bernot A. In: *Genome Transcriptome and Proteome Analysis*, John Wiley and Sons, Somerset - NJ, 2004.

- Bertolo MB and Magalhães EP. Artrite reumatóide. In: *Medicina Ambulatorial*, by Lopes AC, Ward LS and Guariento ME (ed.), Atheneu, São Paulo - SP, 2006.
- Boileau C, Jondeau G, Mizuguchi T, Matsumoto N. Molecular genetics of Marfan syndrome. *Current Opinion in Cardiology* 20(3): 194-200, 2005.
- Bowes J and Barton A. Recent advances in the genetics of RA susceptibility. *Rheumatology* 47(4): 399-402, 2008.
- Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pavelka K, van Vollenhoven R, Sharp J, Perez JL, Spencer-Green GT. The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis & Rheumatism* 54(1): 26-37, 2006.
- Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25(2):169-193, 2000.
- Cantor RM, Yuan J, Napier S, Kono N, Grossman JM, Hahn BH, Tsao BP. Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13, and 17q21-23 and novel evidence at 3p24, 10q23-24, 13q32, and 18q22-23. *Arthritis & Rheumatism* 50(10): 3203-3210, 2004.
- Chen W, Jin W, Hardegen N, Lei KJ, Li L, Marinos N, McGrady G, Wahl SM. Conversion of peripheral CD4⁺CD25⁻ naive T cells to CD4⁺CD25⁺ regulatory T cells by TGF-beta induction of transcription factor Foxp3. *Journal of Experimental Medicine* 198(12): 1875-1886, 2003.

- Choi SJ, Rho YH, Ji JD, Song GG, Lee YH. Genome scan meta-analysis of rheumatoid arthritis. *Rheumatology* 45(2): 166-170, 2006.
- Choi SJ, Rho YH, Ji JD, Song GG, Lee YH. Genome scan meta-analysis of rheumatoid arthritis. *Rheumatology* 45(2): 166-170, 2006.
- Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, Keystone EC, Loveless JE, Burmester GR, Cravets MW, Hesse EW, Shaw T, Totoritis MC. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis & Rheumatism* 54(9): 2793-2806, 2006.
- Combe B, Codreanu C, Fiocco U, Gaubitz M, Geusens PP, Kvien TK, Pavelka K, Sambrook PN, Smolen JS, Wajdula J, Fatenejad S. Etanercept and sulfasalazine, alone and combined, in patients with active rheumatoid arthritis despite receiving sulfasalazine: a double-blind comparison. *Annals of the Rheumatic Diseases* 65(10): 1357-1362, 2006.
- Constantinides A and Mostoufi N. *Numerical Methods for Chemical Engineers With Matlab Applications*, Prentice-Hall Inc., Upper Saddle River - NJ, 1999.
- Craig D, Gao M, Schulten K, Vogel V. Tuning the mechanical stability of fibronectin type III modules through sequence variations. *Structure* 12(1): 21-30, 2004.
- Criswell LA, Merlino LA, Cerhan JR, Mikuls TR, Mudano AS, Burma M, Folsom AR, Saag KG. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. *American Journal of Medicine* 112(6): 465-471, 2002.

- Criswell LA, Saag KG, Mikuls TR, Cerhan JR, Merlino LA, Lum RF, Pfeiffer KA, Woehl B, Seldin MF. Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. *Annals of the Rheumatic Diseases* 65(9): 1163-1167, 2006.
- Dahan O and Kupiec M. Mutations in genes of *Saccharomyces cerevisiae* encoding pre-mRNA splicing factors cause cell cycle arrest through activation of the spindle checkpoint. *Nucleic Acids Reseach* 30(20): 4361-4370, 2002.
- Davis PM, Abraham R, Xu L, Nadler SG, Suchard SJ. Abatacept binds to the Fc receptor CD64 but does not mediate complement-dependent cytotoxicity or antibody-dependent cellular cytotoxicity. *Journal of Rheumatology* 34(11): 2204-2210, 2007.
- Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of HLA to rheumatoid arthritis. *Clinical Genetics* 36(3): 178-182, 1989.
- Del Fattore A, Cappariello A, Teti A. Genetics, pathogenesis and complications of osteopetrosis. *Bone* 42(1): 19-29, 2008.
- D'haeseleer P. *Reconstructing Gene Networks from Large Scale Gene Expression Data*, (Ph.D. dissertation), University of New Mexico, Albuquerque - NM, 2000.
- D'haeseleer P, Liang S, Somogyi R. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16(8): 707-726, 2000.
- di Bernardo D, Thompson M, Gardner T, Chobot S, Eastwood E, Wojtovich A, Elliott S, Schaus S, Collins J. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nature Biotechnology* 23(3): 377-383, 2005.

- Draghici S. *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC Press, Danvers - MA, 2003.
- Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nature Genetics* 21(1): 10-14, 1999.
- Edwards JC and Cambridge G. Sustained improvement in rheumatoid arthritis following a protocol designed to deplete B lymphocytes. *Rheumatology* 40(2): 205-211, 2001.
- Evangelista AF. *Redes de regulação transcricional evidenciam vias de carcinogênese durante a diferenciação osteoblástica de células-tronco mesenquimais*. Dissertação de Mestrado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto - SP, 2009.
- Féral CC, Zijlstra A, Tkachenko E, Prager G, Gardel ML, Slepak M, Ginsberg MH. CD98hc (SLC3A2) participates in fibronectin matrix assembly by mediating integrin signaling. *Journal of Cell Biology* 178(4): 701-711, 2007.
- Ferri C and Zignego AL. Relation between infection and autoimmunity in mixed cryoglobulinemia. *Current Opinion in Rheumatology* 12(1): 53-60, 2000.
- Firestein GS. Evolving concepts of rheumatoid arthritis. *Nature* 423(6937): 356-361, 2003.
- Firestein GS, Tsai V, Zvaifler NJ. Cellular immunity in the joints of patients with rheumatoid arthritis and other forms of chronic synovitis. *Rheumatic Diseases Clinics of North America* 13(2): 191-213, 1987.
- Fong KY, Boey ML, Koh WH, Feng PH. Cytokine concentrations in the synovial fluid and plasma of rheumatoid arthritis patients: correlation with bony erosions. *Clinical and Experimental Rheumatology* 12(1): 55-58, 1994.

- Fransen J and Van Riel PL. The Disease Activity Score and the EULAR response criteria. *Clinical and Experimental Rheumatology* 23 (5 Suppl 39): S93-S99, 2005.
- Fraser A and Burnell D. *Computer Models in Genetics*, McGraw-Hill, New York - NY, 1970.
- Fraser A. Simulation of genetic systems by automatic digital computers. I. Introduction. *Australian Journal of Biological Science* 10: 484-491, 1957.
- Gardner T, di Bernardo D, Lorenz D, Collins J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301(5629): 102-105, 2003.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5(10): R80, 2004.
- Glökler J and Angenendt P. Protein and antibody microarray technology. *Journal of Chromatography, B. Analytical Technologies in the Biomedical and Life Sciences* 797(1-2): 229-240, 2003.
- Goldbach-Mansky R, Lee J, McCoy A, Hoxworth J, Yarboro C, Smolen JS, Steiner G, Rosen A, Zhang C, Ménard HA, Zhou ZJ, Palosuo T, Van Venrooij WJ, Wilder RL, Klippel JH, Schumacher HR Jr, El-Gabalawy HS. Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset. *Arthritis Research* 2(3): 236-243, 2000.

- Gordon DA, Stein JL , Broder I. The extra-articular features of rheumatoid arthritis: a systemic analysis of 127 cases. *American Journal of Medicine* 54(4): 445-452, 1973.
- Gorman JD, David-Vaudey E, Pai M, Lum RF, Criswell LA. Particular HLA-DRB1 shared epitope genotypes are strongly associated with rheumatoid vasculitis. *Arthritis & Rheumatism* 50(11): 3476-3484, 2004.
- Goronzy JJ, Matteson EL, Fulbright JW, Warrington KJ, Chang-Miller A, Hunder GG, Mason TG, Nelson AM, Valente RM, Crowson CS, Erlich HA, Reynolds RL, Swee RG, O'Fallon WM, Weyand CM. Prognostic markers of radiographic progression in early rheumatoid arthritis. *Arthritis & Rheumatism* 50(1): 43-54, 2004.
- Grassi W, De Angelis R, Lamanna G, Cervini C. The clinical features of rheumatoid arthritis. *European Journal of Radiology* 27(Suppl 1): S18-S24, 1998.
- Grossman JM and Brahn E. Rheumatoid arthritis: current clinical and research directions. *Journal of Women's Health* 6(6): 627-638, 1997.
- Gu J, Märker-Hermann E, Baeten D, Tsai WC, Gladman D, Xiong M, Deister H, Kuipers JG, Huang F, Song YW, Maksymowych W, Kalsi J, Bannai M, Seta N, Rihl M, Crofford LJ, Veys E, De Keyser F, Yu DT. A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients. *Rheumatology* 41(7): 759-766, 2002.
- Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences* 97(17): 9390-9395, 2000.

- Hadler NM and Gillings DB. *Arthritis and Society: The Impact of Musculoskeletal Diseases*, Butterworth-Heinemann, Boston – MA, 1985.
- Holland JH. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor - MI, 1975.
- Homandberg GA, Meyers R, Xie DL. Fibronectin fragments cause chondrolysis of bovine articular cartilage slices in culture. *Journal of Biological Chemistry* 267(6): 3597-3604, 1992.
- Homandberg GA, Wen C, Hui F. Cartilage damaging activities of fibronectin fragments derived from cartilage and synovial fluid. *Osteoarthritis Cartilage* 6(4): 231-244, 1998.
- Hutchinson D, Shepstone L, Moots R, Lear JT, Lynch MP. Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA. *Annals of the Rheumatic Diseases* 60(3): 223-227, 2001.
- Ignatavicius DD. Rheumatoid arthritis and the older adult. *Geriatric Nursing* 22(3): 139-142, 2001.
- Ihaka R and Gentleman R. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3): 299-314, 1996.
- John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. *American Journal of Human Genetics* 75(1): 54-64, 2004.

- Junta CM, Sandrin-Garcia P, Fachin-Saltoratto AL, Mello SS, Oliveira RD, Rassi DM, Giuliatti S, Sakamoto-Hojo ET, Louzada-Junior P, Donadi EA, Passos GA. Differential gene expression of peripheral blood mononuclear cells from rheumatoid arthritis patients may discriminate immunogenetic, pathogenic and treatment features. *Immunology* 127(3): 365-372, 2009.
- Kanehisa M and Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28(1): 27-30, 2000.
- Karlson EW, Lee IM, Cook NR, Manson JE, Buring JE, Hennekens CH. A retrospective cohort study of cigarette smoking and risk of rheumatoid arthritis in female health professionals. *Arthritis & Rheumatism* 42(5): 910-917, 1999.
- Kauffman SA. *Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, 1993.
- Knight DM, Trinh H, Le J, Siegel S, Shealy D, McDonough M, Scallon B, Moore MA, Vilcek J, Daddona P, Ghrayeb J. Construction and initial characterization of a mouse-human chimeric anti-TNF antibody. *Molecular Immunology* 30(16): 1443-1453, 1993.
- Kremer JM, Genant HK, Moreland LW, Russell AS, Emery P, Abud-Mendoza C, Szechinski J, Li T, Ge Z, Becker JC, Westhovens R. Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a randomized trial. *Annals of Internal Medicine* 144(12): 865-876, 2006.
- Kyewski B and Derbinski J. Self-representation in the thymus: an extended view. *Nature Reviews Immunology* 4(9): 688-698, 2004.

- Lennon G, Auffray C, Polymeropoulos M, Soares MB. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 33(1): 151-152, 1996.
- Liang S, Fuhrman S, Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symposium on Biocomputing* 3: 18-29, 1998.
- Linsley PS, Brady W, Urnes M, Grosmaire LS, Damle NK, Ledbetter JA. CTLA-4 is a second receptor for the B cell activation antigen B7. *Journal of Experimental Medicine* 174(3): 561-569, 1991.
- Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, Smolen JS, Weisman M, Emery P, Feldmann M, Harriman GR, Maini RN. Infliximab and methotrexate in the treatment of rheumatoid arthritis. *New England Journal of Medicine* 343(22): 1594-1602, 2000.
- Liu C, Batliwalla F, Li W, Lee A, Roubenoff R, Beckman E, Khalili H, Damle A, Kern M, Furie R, Dupuis J, Plenge RM, Coenen MJ, Behrens TW, Carulli JP, Gregersen PK. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Molecular Medicine* 14(9-10): 575-581, 2008.
- Louzada-Júnior P, Freitas MV, Oliveira RD, Deghaide NH, Conde RA, Bertolo MB, Donadi EA. A majority of Brazilian patients with rheumatoid arthritis HLA-DRB1 alleles carry both the HLA-DRB1 shared epitope and anti-citrullinated peptide antibodies. *Brazilian Journal of Medical and Biological Research* 41(6): 493-499, 2008.
- Macedo C, Evangelista AF, Magalhães DA, Fornari TA, Linhares LL, Junta CM, Silva GL, Sakamoto-Hojo ET, Donadi EA, Savino W, Passos GA. Evidence

- for a network transcriptional control of promiscuous gene expression in medullary thymic epithelial cells. *Molecular Immunology* 46(16): 3240-3244, 2009.
- Macedo C, Magalhães DA, Tonani M, Marques MC, Junta CM, Passos GA. Genes that code for T cell signaling proteins establish transcriptional regulatory networks during thymus ontogeny. *Molecular and Cellular Biochemistry* 318(1-2): 63-71, 2008.
- Magalhães DA, Silveira EL, Junta CM, Sandrin-Garcia P, Fachin AL, Donadi EA, Sakamoto-Hojo ET, Passos GA. Promiscuous gene expression in the thymus: the root of central tolerance. *Clinical & Developmental Immunology* 13(2-4): 81-99, 2006.
- Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, Smolen J, Emery P, Harriman G, Feldmann M, Lipsky P. Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. *The Lancet* 354(9194): 1932-1939, 1999.
- Margolin A, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Della Favera R, Califano A. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl 1): S7, 2006.
- Masi AT. Articular patterns in the early course of rheumatoid arthritis. *American Journal of Medicine* 75(6A): 16-26, 1983.
- MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, Silman AJ. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis & Rheumatism* 43(1): 30-37, 2000.

- McIntyre KW, Stepan GJ, Kolinsky KD, Benjamin WR, Plocinski JM, Kaffka KL, Campen CA, Chizzonite RA, Kilian PL. Inhibition of interleukin 1 (IL-1) binding and bioactivity in vitro and modulation of acute inflammation in vivo by IL-1 receptor antagonist and anti-IL-1 receptor monoclonal antibody. *Journal of Experimental Medicine* 173(4): 931-939, 1991.
- McKusick VA. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders (12th edition)*, Johns Hopkins University Press, Baltimore – MD, 1998.
- Mello SS. *Análise dos Dados de Expressão Gênica em Larga Escala: Considerações Sobre a Análise e Processamento dos Dados e Métodos Estatísticos para a Determinação de Genes Diferencialmente Expressos em Experimentos de DNA Microarrays Fluorescentes (relatório técnico)*, Universidade de São Paulo, Ribeirão Preto – SP, 2004.
- Mladenovic V, Domljan Z, Rozman B, Jajic I, Mihajlovic D, Dordevic J, Popovic M, Dimitrijevic M, Zivkovic M, Champion G, Musikic P, Löw-Friedrich I, Oed C, Seifert H, Strand V. Safety and effectiveness of leflunomide in the treatment of patients with active rheumatoid arthritis. Results of a randomized, placebo-controlled, phase II study. *Arthritis & Rheumatism* 38(11): 1595-1603, 1995.
- Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85(1): 1-15, 2005.
- Mohler KM, Torrance DS, Smith CA, Goodwin RG, Stremmer KE, Fung VP, Madani H, Widmer MB. Soluble tumor necrosis factor (TNF) receptors are effective therapeutic agents in lethal endotoxemia and function simultaneously as both TNF carriers and TNF antagonists. *Journal of Immunology* 151(3): 1548-1561, 1993.

- Murali-Krishna K, Altman JD, Suresh M, Sourdive DJ, Zajac AJ, Miller JD, Slansky J, Ahmed R. Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* 8(2): 177-187, 1998.
- Obholz KL, Akopyan A, Waymire KG, MacGregor GR. FNDC3A is required for adhesion between spermatids and Sertoli cells. *Developmental Biology* 298(2): 498-513, 2006.
- Orozco G, Rueda B, Martin J. Genetic basis of rheumatoid arthritis. *Biomedicine & Pharmacotherapy* 60(10): 656-662, 2006.
- Panayi GS. B cells: a fundamental role in the pathogenesis of rheumatoid arthritis? *Rheumatology* 44(Suppl 2): ii3-ii7, 2005.
- Pankov R and Yamada KM. Fibronectin at a glance. *Journal of Cell Science* 115(Pt 20): 3861-3863, 2002.
- Pereira E, Tamia-Ferreira MC, Cardoso RS, Mello SS, Sakamoto-Hojo ET, Passos GA, Donadi EA. Immunosuppressive therapy modulates T lymphocyte gene expression in patients with systemic lupus erythematosus. *Immunology* 113(1): 99-105, 2004.
- Pers JO, Daridon C, Devauchelle V, Jousse S, Saraux A, Jamin C, Youinou P. BAFF overexpression is associated with autoantibody production in autoimmune diseases. *Annals of the New York Academy of Sciences* 1050: 34-39, 2005.
- Pinheiro GRC. Instrumentos de medida da atividade da artrite reumatoide - por que e como empregá-los. *Revista Brasileira de Reumatologia* 47(5): 362-365, 2007.
- Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint

- counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis & Rheumatism* 38(1): 44-48, 1995.
- Rassi DM, Junta CM, Fachin AL, Sandrin-Garcia P, Mello SS, Fernandes AP, Deghaide NN, Foss-Freitas MC, Foss MC, Sakamoto-Hojo ET, Passos GA, Donadi EA. Is HLA class II profile relevant for the study of large-scale differentially expressed genes in type 1 diabetes mellitus patients? *Annals of the New York Academy of Sciences* 1079: 305-309, 2006.
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14(8): 656-664, 1998.
- Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Research* 11(1): 3-11, 2001.
- Ruoslahti E. Fibronectin and its receptors. *Annual Review of Biochemistry* 57: 375-413, 1988.
- Saag KG, Cerhan JR, Kolluri S, Ohashi K, Hunninghake GW, Schwartz DA. Cigarette smoking and rheumatoid arthritis severity. *Annals of the Rheumatic Diseases* 56(8): 463-469, 1997.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34(2): 374-378, 2003.

- Sakamoto-Hojo ET, Mello SS, Cardoso RS, Passos GAS. Utilização de genômica funcional e proteômica em mutagênese. In: *Mutagênese Ambiental*, by Ribeiro LG, Salvadori DMF and Marques EK (ed.), Editora Ulbra, Canoas - RS, 2003.
- Salfeld J, Kaymakçalan Z, Tracey D, Roberts A, Kamen R. Generation of fully human anti-TNF antibody D2E7. *Arthritis & Rheumatism* 41(Suppl 9): S57, 1998.
- Sandya S, Achan MA, Sudhakaran PR. Parallel changes in fibronectin and alpha5beta1 integrin in articular cartilage in type II collagen-induced arthritis. *Indian Journal of Biochemistry & Biophysics* 44(1): 14-18, 2007.
- Savageau MA. Biochemical system analysis I: some mathematical properties of the rate law for the component enzymatic reactions. *Journal of Theoretical Biology* 25: 365-369, 1969a.
- Savageau MA. Biochemical system analysis II: the steady-state solutions for an n-pool system using a power-law approximation. *Journal of Theoretical Biology* 5: 370-379, 1969b.
- Savageau MA. *Biochemical System Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley, Reading - MA, 1976.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235): 467-470, 1995.
- Scherer MT, Ignatowicz L, Winslow GM, Kappler JW, Marrack P. Superantigens: bacterial and viral proteins that manipulate the immune system. *Annual Review of Cell Biology* 9: 101-128, 1993.

- Shaffer AL, Wright G, Yang L, Powell J, Ngo V, Lamy L, Lam LT, Davis RE, Staudt LM. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunological Reviews* 210: 67-85, 2006.
- Shmerling RH, Delbanco TL. How useful is the rheumatoid factor? An analysis of sensitivity, specificity, and predictive value. *Archives of Internal Medicine* 152(12): 2417-2420, 1992.
- Shovman O, Gilburd B, Zandman-Goddard G, Sherer Y, Orbach H, Gerli R, Shoenfeld Y. The diagnostic utility of anti-cyclic citrullinated peptide antibodies, matrix metalloproteinase-3, rheumatoid factor, erythrocyte sedimentation rate, and C-reactive protein in patients with erosive and non-erosive rheumatoid arthritis. *Clinical and Developmental Immunology* 12(3): 197-202, 2005.
- Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, Ollier WE. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *British Journal of Rheumatology* 32(10): 903-907, 1993.
- Silva GL, Junta CM, Mello SS, Garcia PS, Rassi DM, Sakamoto-Hojo ET, Donadi EA, Passos GA. Profiling meta-analysis reveals primarily gene coexpression concordance between systemic lupus erythematosus and rheumatoid arthritis. *Annals of the New York Academy of Sciences* 1110: 33-46, 2007.
- Smith C. Heritability of liability and concordance in monozygous twins. *Annals of Human Genetics* 34: 85-91, 1970.
- Smith WL. Prostanoid biosynthesis and mechanisms of action. *American Journal of Physiology* 263(2 Pt 2): F181-F191, 1992.

- Smyth GK and Speed T. Normalization of cDNA microarray data. *Methods* 31(4): 265-273, 2003.
- Somogyi R and Sniegowski C. Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulations. *Complexity* 1(6): 45-63, 1996.
- Spirtes P, Glymour C, Scheines R, Kauffman S, Aimale V, Wimberly F. Constructing Bayesian network models of gene expression networks from microarray data. In: *Proceedings of the Atlantic Symposium on Comparative Biology, Genome Information Systems and Technology*, 2000.
- Stekel D. *Microarray Bioinformatics*, Cambridge University Press, Cambridge, 2003.
- Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, Alfredsson L. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Annals of the Rheumatic Diseases* 62(9): 835-841, 2003.
- Storey JD and Tibshirani R. Statistical methods for detecting differential gene expression. In: *Functional Genomics: Methods and Protocols*, by Brownstein MJ and Khodursky AB (ed.), *Methods in Molecular Biology series*, Humana Press, Totowa - NJ, 2003.
- Strand V, Cohen S, Schiff M, Weaver A, Fleischmann R, Cannon G, Fox R, Moreland L, Olsen N, Furst D, Caldwell J, Kaine J, Sharp J, Hurley F, Loew-Friedrich I. Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. *Archives of Internal Medicine* 159(21): 2542-2550, 1999.

- The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genetics* 25(1): 25-29, 2000.
- Thomas JG, Olson JM, Tapscott SJ, Zhao LP. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research* 11(7): 1227-1236, 2001.
- Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, Suzuki M, Nagasaki M, Ohtsuki M, Ono M, Furukawa H, Nagashima M, Yoshino S, Mabuchi A, Sekine A, Saito S, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics* 35(4): 341-348, 2003.
- Toonen EJ, Barrera P, Radstake TR, van Riel PL, Scheffer H, Franke B, Coenen MJ. Gene expression profiling in rheumatoid arthritis: current concepts and future directions. *Annals of the Rheumatic Diseases* 67(12): 1663-1669, 2008.
- Topfer F, Gordon T, McCluskey J. Intra- and inter-molecular spreading of autoimmunity involving the nuclear self-antigens La (SS-B) and Ro (SS-A). *Proceedings of the National Academy of Sciences* 92(3): 875-879, 1995.
- Torres NV and Voit EO. *Pathway Analysis and Optimization in Metabolic Engineering*, Cambridge University Press, Cambridge, 2003.
- Toussirot E, Auge B, Tiberghien P, Chabod J, Cedoz JP, Wendling D. HLA-DRB1 alleles and shared amino acid sequences in disease susceptibility and severity in patients from eastern France with rheumatoid arthritis. *Journal of Rheumatology* 26(7): 1446-1451, 1999.

- Trevisan GL, Rassi DM, Baião AM, Sandrin-Garcia P, Mello SS, Tamia-Ferreira MC, Junta CM, Fachin AL, Marques MM, Sakamoto-Hojo ET, Donadi EA, Passos GA. Using cDNA microarrays to identify human CD19(+) B cell gene products (ESTs) originated from systemic lupus erythematosus susceptibility loci. *Autoimmunity Reviews* 5(5): 319-323, 2006.
- Turesson C, Jacobsson L, Bergstrom U. Extra-articular rheumatoid arthritis: prevalence and mortality. *Rheumatology* 38(7): 668-674, 1999.
- Turesson C, Schaid DJ, Weyand CM, Jacobsson LTH, Goronzy JJ, Petersson IF, Sturfelt G, Nyhäll-Wåhlin BM, Truedsson L, Dechant SA, Matteson EL. The impact of HLA-DRB1 genes on extra-articular disease manifestations in rheumatoid arthritis. *Arthritis Research & Therapy* 7(6): R1386-R1393, 2005.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* 98(9): 5116-5121, 2001.
- Van der Heijde DM, van 't Hof M, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *Journal of Rheumatology* 20(3): 579-581, 1993.
- Van Gestel AM, Haagsma CJ, van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis & Rheumatism* 41(10): 1845-1850, 1998.
- Venables PJW, Maini RN. *Clinical features of rheumatoid arthritis*. In: *UpToDate*, by Rose BD (ed.), *UpToDate*, Wellesley - MA, 2006.
- Verpoort KN, van Gaalen FA, van der Helm-van Mil AH, Schreuder GM, Breedveld FC, Huizinga TW, de Vries RR, Toes RE. Association of HLA-

- DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. *Arthritis & Rheumatism* 52(10): 3058-3062, 2005.
- Veuculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 270(2535): 484-487, 1995.
- Voit EO. *Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity*, Van Nostrand Reinhold, New York - NY, 1991.
- Voit EO. *Computational Analysis of Biochemical Systems*, Cambridge University Press, Cambridge, 2000.
- Von Bodungen U, Lechner F, Pfister H, Vogt HR, Cheevers WP, Bertoni G, Jungi TW, Peterhans E. Immunohistology of the early course of lentivirus-induced arthritis. *Clinical & Experimental Immunology* 111(2): 384-390, 1998.
- Wang WYS, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics* 6(2): 109-118, 2005.
- Weinblatt ME, Kremer JM, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Fox RI, Jackson CG, Lange M, Burge DJ. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *New England Journal of Medicine* 340(4): 253-259, 1999.
- Westhovens R, Cole JC, Li T, Martin M, Maclean R, Lin P, Blaisdell B, Wallenstein GV, Aranda R, Sherrer Y. Improved health-related quality of life for rheumatoid arthritis patients treated with abatacept who have inadequate response to anti-TNF therapy in a double-blind, placebo-controlled, multicentre randomized clinical trial. *Rheumatology* 45(10): 1238-1246, 2006.

- Wilder RL. Adrenal and gonadal steroid hormone deficiency in the pathogenesis of rheumatoid arthritis. *Journal of Rheumatology* 44((Suppl): 10-12, 1996.
- Wolbink GJ, Vis M, Lems W, Voskuyl AE, de Groot E, Nurmohamed MT, Stapel S, Tak PP, Aarden L, Dijkmans B. Development of antiinfleximab antibodies and relationship to clinical response in patients with rheumatoid arthritis. *Arthritis & Rheumatism* 54(3): 711-715, 2006.
- Wolf JR, Hu Y, Hirschhorn RR, Steiner SM. MyoD and regulation of prostaglandin H synthase. *Experimental Cell Research* 207(2): 439-441, 1993.
- Wolfe F, Cathey MA, Roberts FK. The latex test revisited. Rheumatoid factor testing in 8,287 rheumatic disease patients. *Arthritis & Rheumatism* 34(8): 951-960, 1991.
- Woolley DE and Tetlow LC. Observations on the microenvironmental nature of cartilage degradation in rheumatoid arthritis. *Annals of the Rheumatic Diseases* 56(3): 151-161, 1997.
- Wu CC, Huang HC, Juan HF, Chen ST. GeneNetwork: an interactive tool for reconstruction of genetic networks using microarray data. *Bioinformatics* 20(18): 3691-3693, 2004.
- Wucherpfennig KW. Mechanisms for the induction of autoimmunity by infectious agents. *Journal of Clinical Investigation* 108(8): 1097-1104, 2001.
- Xie DL, Meyers R, Homandberg GA. Fibronectin fragments in osteoarthritic synovial fluid. *Journal of Rheumatology* 19(9): 1448-1452, 1992.
- Yamagata T, Benoist C, Mathis D. A shared gene-expression signature in innate-like lymphocytes. *Immunological Reviews* 210(1): 52-66, 2006.

- Yamamoto K and Yamada R. Genome-wide single nucleotide polymorphism analyses of rheumatoid arthritis. *Journal of Autoimmunity* 25(Suppl 1): 12-15, 2005.
- Yang LI, Anderson DE, Baecher-Allan C, Hastings WD, Bettelli E, Oukka M, Kuchroo VK, Hafler DA. IL-21 and TGF- β are required for differentiation of human T(H)17 cells. *Nature* 454(7202): 350-352, 2008.
- Yang YH and Thorne N. Normalization for two-color cDNA microarray data. In: *Science and Statistics: A Festschrift for Terry Speed, Goldstein DR (ed.), IMS Lecture Notes, Monograph Series 40*: 403-418, 2003.
- Yasuda T and Poole AR. A fibronectin fragment induces type II collagen degradation by collagenase through an interleukin-1-mediated pathway. *Arthritis & Rheumatism* 46(1): 138-148, 2002.
- Yasuda T, Shimizu M, Nakagawa T, Julovi SM, Nakamura T. Matrix metalloproteinase production by COOH-terminal heparin-binding fibronectin fragment in rheumatoid synovial cells. *Laboratory Investigation, a Journal of Technical Methods and Pathology* 83(2): 153-162, 2003.
- Yasuda T. Cartilage destruction by matrix degradation products. *Modern Rheumatology/The Japan Rheumatism Association* 16(4): 197-205, 2006.
- Young A and Koduri G. Extra-articular manifestations and complications of rheumatoid arthritis. *Best Practice & Research - Clinical Rheumatology* 21(5): 907-927, 2007.
- Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20(18): 3594-3603, 2004.

Zhang X, Park CS, Yoon SO, Li L, Hsu YM, Ambrose C, Choi YS. BAFF supports human B cell differentiation in the lymphoid follicles through distinct receptors. *International Immunology* 17(6): 779-788, 2005.

Zhao ZS, Granucci F, Yeh L, Schaffer PA, Cantor H. Molecular mimicry by Herpes Simplex virus-type1: autoimmune disease after viral infection. *Science* 279(5355): 1344-1347, 1998.

Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using proteome chips. *Science* 293(5537): 2101-2105, 2001.

Apêndice A

/*CÓDIGO-FONTE DO PROGRAMA MANDALA*/

```

#include<stdio.h>           /* Bibliotecas padrão */
#include<stdlib.h>
#include<string.h>

struct Nome{               /* Estrutura global: lista de genes */
    char gene[25];
};

struct Posicao{            /* Estrutura global: matriz regulatória */
    char matriz[2000];
};

struct Frase{            /* Estrutura global: frases dos arquivos a serem comparados */
    char palavra[25];
};

void grafo_UMA_COR()     /* Função para interpretação textual de grafos cujas interações não assumem */
{                         /* valores positivo ou negativo. Ou seja, neste caso, os genes apenas modulam ou não */
    FILE *fp;
    FILE *fs;             /* Variáveis locais: apontadores de arquivo */
    FILE *ft;

    int a=0;
    int b;
    int c;                /* Variáveis locais: inteiros */
    int x;
    int y;
    int opcao_pesquisa;

    char nome_matriz_regulatoria[50];
    char nome_lista_genes[50];
    char nome_resultado[50];    /* Variáveis locais: strings [tamanho máximo] */
    char genes[25];

    struct Nome *nome;      /* Variáveis locais: apontadores para as estruturas globais */
    struct Posicao *posicao;

    if((nome=(struct Nome*)malloc(2000*sizeof(struct Nome)))==NULL) /* Procedimento para garantir */
    {                         /* alocação suficiente de memória */
        printf("Não foi possível alocar memória suficiente!\n"); /* da estrutura global Nome */
        exit(1);             /* 2000 é o núm. máximo de genes */
    }                         /* manipulados pelo GeneNetwork */
}

```

```

if((posicao=(struct Posicao*)malloc(2000*sizeof(struct Posicao)))==NULL) /* Procedimento para garantir */
{
    printf("Não foi possível alocar memória suficiente!\n"); /* alocação suficiente de memória */
    exit(1); /* da estrutura global Posicao */
} /* 2000 é o núm. máximo de genes */
/* manipulados pelo GeneNetwork */

printf("\nDigite o nome do arquivo correspondente à lista de genes:\n"); /* Pede ao usuário entrar com o */
scanf("%s", &nome_lista_genes); /* arquivo com a lista de genes */
if (( fp = fopen(nome_lista_genes, "r")) == NULL) /* Verifica se o arquivo existe */
{
    printf("\nNão foi possível abrir o arquivo! Verifique se o nome fornecido está correto.\n");
    exit(1); /* Se não, emite mensagem de erro e aborta */
}

printf("\n");
while(!feof(fp)) /* Lê o arquivo com a lista de genes até o final */
{
    a++;
    fgets(genes,25,fp);
    strcpy(nome[a].gene,genes);
    printf("%s", nome[a].gene);
}
printf("O número de genes nesta lista é: %d\n", a); /* Imprime na tela o número total de genes na lista */

printf("\nDigite o nome do arquivo correspondente à matriz regulatória:\n"); /* Pede ao usuário entrar com o */
scanf("%s", &nome_matriz_regulatoria); /* o arquivo da matriz regulatória */
if(( fs = fopen(nome_matriz_regulatoria, "r")) == NULL) /* Verifica se o arquivo existe */
{
    printf("\nNão foi possível abrir o arquivo! Verifique se o nome fornecido está correto.\n");
    exit(1); /* Se não, emite mensagem de erro e aborta */
}

printf("\nDigite o nome do arquivo que receberá o resultado da análise.\n"); /* Pede ao usuário digitar o */
printf("O arquivo deve ter extensão '.txt'.\n"); /* nome do arquivo de resultados */
scanf("%s", &nome_resultado);
if (( ft = fopen(nome_resultado, "w")) == NULL) /* Procedimento de verificação */
{
    printf("O arquivo que receberá o resultado não pode ser aberto!\n");
    exit(1);
}

for(b=1; b<=a; b++) /* Lê o arquivo com a matriz regulatória até o final */
{
    for(c=1; c<=(a+2); c++)
    {
        posicao[b].matriz[c]=fgetc(fs);
    }
}

```

```

printf("\n");
printf("*****\n");
printf("* Deseja realizar a pesquisa a partir dos: *\n");
printf("* *\n");
printf("* 1 - Genes que modulam *\n");
printf("* 2 - Genes que são modulados *\n");
printf("*****\n\n");
scanf("%d", &opcao_pesquisa);
while(opcao_pesquisa<1 || opcao_pesquisa>2) /* Verifica se o usuário digitou uma opção válida (1 ou 2) */
{
    printf("\nDigite 1 ou 2!\n"); /* Se não, confirma junto ao usuário */
    scanf("%d", &opcao_pesquisa);
}

if(opcao_pesquisa==1) /* Opção 1: o resultado será apresentado da seguinte maneira: */
{ /* O gene X modula os genes Y, Z, etc */
    printf("A pesquisa foi realizada a partir dos genes que modulam.\n");
    for(x=1; x<=a; x++) /* Varre toda a matriz regulatória, de cima para baixo... */
    {
        if(x!=a)
        {
            fprintf(ft, "O gene %smodula os genes:\n", nome[x].gene);
            for(y=1; y<=a; y++) /* ... e da esquerda para direita, respectivamente */
            {
                if(posicao[x].matriz[y]=='1') /* Se a matriz indica a existência de interação */
                {
                    fprintf(ft, "%s", nome[y].gene); /* Correlaciona os genes envolvidos */
                }
            }
            fprintf(ft, "\n-----\n\n"); /* Imprime separador */
        }
        else /* Laço especialmente desenvolvido para o último registro da lista */
        { /* de genes, faz o mesmo que o laço imediatamente anterior */
            fprintf(ft, "O gene %s\nmodula os genes:\n", nome[x].gene);
            for(y=1; y<=a; y++)
            {
                if(posicao[x].matriz[y]=='1')
                {
                    fprintf(ft, "%s", nome[y].gene);
                }
            }
            fprintf(ft, "\n-----\n\n");
        }
    }
}
}

```

```

else
{
    if(opcao_pesquisa==2)          /* Opção 2: o resultado será apresentado da seguinte maneira: */
    {                               /* Os genes Y, Z e etc. são modulados pelo gene X */
        printf("A pesquisa foi realizada a partir dos genes que são modulados.\n");
        for(y=1; y<=a; y++)        /* Varre toda a matriz regulatória, da esquerda para direita... */
        {
            if(y!=a)
            {
                fprintf(ft, "O gene %s é modulado pelos genes:\n", nome[y].gene);
                for(x=1; x<=a; x++) /* ... e de cima para baixo, respectivamente */
                {
                    if(posicao[x].matriz[y]=='1') /* Se a matriz indica que há interação */
                    {
                        fprintf(ft, "%s", nome[x].gene); /* Correlaciona os genes */
                    } /* envolvidos */
                }
                fprintf(ft, "\n-----\n\n"); /* Imprime separador */
            }
            else /* Laço especialmente desenvolvido para o último registro da lista */
            { /* de genes, faz o mesmo que o laço imediatamente anterior */
                fprintf(ft, "O gene %s\n é modulado pelos genes:\n", nome[y].gene);
                for(x=1; x<=a; x++)
                {
                    if(posicao[x].matriz[y]=='1')
                    {
                        fprintf(ft, "%s", nome[x].gene);
                    }
                }
                fprintf(ft, "\n-----\n\n"); /* Imprime separador */
            }
        }
    }
}

printf("\nPronto!!\n"); /* Imprime na tela */
printf("\nAbra o arquivo resultante da análise com o programa\n");
printf("Microsoft Office Word ou com o aplicativo WordPad MFC.\n");

fclose(fs);
fclose(fp); /* Fecha os apontadores para arquivos */
fclose(ft);

} /* Termina a função */

```

```

void grafo_DUAS_CORES()      /* Função para interpretação textual de grafos cujas interações assumem */
{                             /* valores positivo ou negativo. Procede da mesma maneira que a função anterior, */
    FILE *fp;                 /* mas, neste caso, considera que os genes podem induzir ou reprimir */
    FILE *fs;
    FILE *ft;

    int a=0;
    int b;
    int c;
    int x;
    int y;
    int opcao_pesquisa;

    char nome_matriz_regulatoria[50];
    char nome_lista_genes[50];
    char nome_resultado[50];
    char genes[25];

    struct Nome *nome;
    struct Posicao *posicao;

    if((nome=(struct Nome*)malloc(2000*sizeof(struct Nome)))==NULL)
    {
        printf("Não foi possível alocar memória suficiente!\n");
        exit(1);
    }

    if((posicao=(struct Posicao*)malloc(2000*sizeof(struct Posicao)))==NULL)
    {
        printf("Não foi possível alocar memória suficiente!\n");
        exit(1);
    }

    printf("\nDigite o nome do arquivo correspondente à lista de genes:\n");
    scanf("%s", &nome_lista_genes);
    if (( fp = fopen(nome_lista_genes, "r") ) == NULL)
    {
        printf("\nNão foi possível abrir o arquivo! Verifique se o nome fornecido está correto.\n");
        exit(1);
    }

    while(!feof(fp))
    {
        a++;
        fgets(genes,25,fp);
        strcpy(nome[a].gene,genes);
        printf("%s", nome[a].gene);
    }
    printf("\n\nO número de genes nesta lista é: %d\n", a);

```

```

printf("\nDigite o nome do arquivo correspondente à matriz regulatória:\n");
scanf("%s", &nome_matriz_regulatoria);
if(( fs = fopen(nome_matriz_regulatoria, "r")) == NULL)
{
    printf("\nNão foi possível abrir o arquivo! Verifique se o nome fornecido está correto.\n");
    exit(1);
}

printf("\nDigite o nome do arquivo que receberá o resultado da análise.\n");
printf("O arquivo deve ter extensão '.txt'.\n");
scanf("%s", &nome_resultado);
if (( ft = fopen(nome_resultado, "w")) == NULL)
{
    printf("O arquivo que receberá o resultado não pode ser aberto!\n");
    exit(1);
}

for(b=1; b<=a; b++)
{
    for(c=1; c<=(a+2); c++)
    {
        posicao[b].matriz[c]=fgetc(fs);
    }
}

printf("\n");
printf("*****\n");
printf("* Deseja realizar a pesquisa a partir dos:          *\n");
printf("*                                                    *\n");
printf("* 1 - Genes que modulam                               *\n");
printf("* 2 - Genes que são modulados                          *\n");
printf("*****\n\n");
scanf("%d", &opcao_pesquisa);
while(opcao_pesquisa<1 || opcao_pesquisa>2)
{
    printf("\nDigite 1 ou 2!\n");
    scanf("%d", &opcao_pesquisa);
}

if(opcao_pesquisa==1)
{
    printf("A pesquisa foi realizada a partir dos genes que modulam.\n");

    fprintf(ft, "-----\n");
    fprintf(ft, "|                RESULTADO DA PESQUISA                |\n");
    fprintf(ft, "-----\n\n");
    fprintf(ft, "-----\n");
    fprintf(ft, "| GENES QUE INDUZEM | \n");
    fprintf(ft, "-----\n\n");
}

```

```

for(x=1; x<=a; x++)          /* Varre toda a matriz regulatória, de cima para baixo... */
{
    if(x!=a)
    {
        fprintf(ft, "O gene %sinduz os genes:\n", nome[x].gene);
        for(y=1; y<=a; y++)    /* ... e da esquerda para direita, respectivamente */
        {
            if(posicao[x].matriz[y]=='1')    /* Se a matriz indica que ocorre indução */
            {
                fprintf(ft, "%s", nome[y].gene);    /* Correlaciona os genes envolvidos */
            }
        }
        fprintf(ft, "\n\n");
    }
    else    /* Laço especialmente desenvolvido para o último registro da lista */
    {    /* de genes, faz o mesmo que o laço imediatamente anterior */
        fprintf(ft, "O gene %s\ninduz os genes:\n", nome[x].gene);
        for(y=1; y<=a; y++)
        {
            if(posicao[x].matriz[y]=='1')
            {
                fprintf(ft, "%s", nome[y].gene);
            }
        }
        fprintf(ft, "\n\n");
    }
}

fprintf(ft, "\n-----\n");
fprintf(ft, "| GENES QUE REPRIMEM | \n");
fprintf(ft, "-----\n\n");

for(x=1; x<=a; x++)          /* Reinicia a leitura da matriz regulatória na mesma ordem */
{
    if(x!=a)
    {
        fprintf(ft, "O gene %sreprime os genes:\n", nome[x].gene);
        for(y=1; y<=a; y++)
        {
            if(posicao[x].matriz[y]=='2')    /* Se a matriz indica que ocorre repressão */
            {
                fprintf(ft, "%s", nome[y].gene);    /* Correlaciona os genes envolvidos */
            }
        }
        fprintf(ft, "\n\n");
    }
}

```



```

else
{
    fprintf(ft, "O gene %s\nreprime os genes:\n", nome[x].gene);
    for(y=1; y<=a; y++)
    {
        if(posicao[x].matriz[y]=='2')
        {
            fprintf(ft, "%s", nome[y].gene);
        }
    }
    fprintf(ft, "\n\n");
}
}
else
{
    if(opcao_pesquisa==2)
    {
        printf("A pesquisa foi realizada a partir dos genes que são modulados.\n");
        fprintf(ft, "-----\n");
        fprintf(ft, "|                RESULTADO DA PESQUISA                |\n");
        fprintf(ft, "-----\n\n");

        fprintf(ft, "-----\n");
        fprintf(ft, "| GENES QUE SÃO INDUZIDOS |\n");
        fprintf(ft, "-----\n\n");

        for(y=1; y<=a; y++) /* Varre toda a matriz regulatória, da esquerda para direita... */
        {
            if(y!=a)
            {
                fprintf(ft, "O gene %s é induzido pelos genes:\n", nome[y].gene);
                for(x=1; x<=a; x++) /* ... e de cima para baixo, respectivamente */
                {
                    if(posicao[x].matriz[y]=='1') /* Se a matriz indica que há indução */
                    {
                        fprintf(ft, "%s", nome[x].gene); /* Correlaciona os genes */
                        /* envolvidos */
                    }
                }
                fprintf(ft, "\n\n");
            }
            else
            {
                fprintf(ft, "O gene %s\né induzido pelos genes:\n", nome[y].gene);
                for(x=1; x<=a; x++)
                {
                    if(posicao[x].matriz[y]=='1')
                    {
                        fprintf(ft, "%s", nome[x].gene);
                    }
                }
            }
        }
    }
}

```

```

        }
    }
    fprintf(ft, "\n\n");
}

fprintf(ft, "\n-----\n");
fprintf(ft, "| GENES QUE SÃO REPRIMIDOS |\n");
fprintf(ft, "-----\n\n");

for(y=1; y<=a; y++) /* Reinicia a leitura da matriz regulatória na mesma ordem */
{
    if(y!=a)
    {
        fprintf(ft, "O gene %s é reprimido pelos genes:\n", nome[y].gene);
        for(x=1; x<=a; x++)
        {
            if(posicao[x].matriz[y]=='2') /* Se a matriz indica que ocorre repressão */
            {
                fprintf(ft, "%s", nome[x].gene); /* Correlaciona os genes */
                /* envolvidos */
            }
        }
        fprintf(ft, "\n\n");
    }
    else
    {
        fprintf(ft, "O gene %s\n é reprimido pelos genes:\n", nome[y].gene);
        for(x=1; x<=a; x++)
        {
            if(posicao[x].matriz[y]=='2')
            {
                fprintf(ft, "%s", nome[x].gene);
            }
        }
        fprintf(ft, "\n\n");
    }
}

}

printf("\nPronto!!\n");
printf("\nAbra o arquivo resultante da análise com o programa\n");
printf("Microsoft Office Word ou com o aplicativo WordPad MFC.\n");

fclose(fs);
fclose(fp);
fclose(ft);

}

```

```

void compara_REDES()          /* Função para comparação de duas redes regulatórias */
{
    FILE *fp;
    FILE *fs;                 /* Variáveis locais: apontadores de arquivo */
    FILE *ft;

    char primeira_rede[50];
    char segunda_rede[50];
    char resultado_comparacao[50];          /* Variáveis locais: strings [tamanho máximo] */
    char frase_primeiro_arquivo[50];
    char frase_segundo_arquivo[50];

    int a=0;
    int b=0;
    int x;                     /* Variáveis locais: inteiros */
    int y;
    int g;
    int h;

    struct Frase *frase1;     /* Variáveis locais: apontadores para estrutura global */
    struct Frase *frase2;

    if((frase1=(struct Frase*)malloc(1000000*sizeof(struct Frase)))==NULL) /* Procedimento para garantir */
    {                                                                    /* alocação suficiente de memória */
        printf("Não foi possível alocar memória suficiente!\n");          /* da estrutura global Frase */
        exit(1);                                                         /* 1000000 é o núm. máximo de */
    }                                                                    /* interações na primeira rede */

    if((frase2=(struct Frase*)malloc(1000000*sizeof(struct Frase)))==NULL) /* Idem ao anterior para a 2ª rede*/
    {
        printf("Não foi possível alocar memória suficiente!\n");
        exit(1);
    }

    printf("\nDigite o nome do arquivo com o resultado da primeira rede:\n"); /* Pedir para que o usuário entre */
    scanf("%s", &primeira_rede);                                             /* com o arquivo correspondente */
    if (( fp = fopen(primeira_rede, "r")) == NULL)                             /* Verifica se o arquivo existe */
    {
        printf("\nNão foi possível abrir o arquivo! Verifique se o nome fornecido está correto.\n");
        exit(1);                                                            /* Se não, emite mensagem de erro e aborta */
    }

    while(fgets(frase_primeiro_arquivo,50,fp)!=NULL) /* Lê todo o arquivo contendo a interpretação textual */
    {                                                                    /* para a primeira rede, armazenando suas frases */
        a++;
        strcpy(frase1[a].palavra, frase_primeiro_arquivo);
    }
}

```

```

printf("\nDigite o nome do arquivo com o resultado da segunda rede:\n"); /* Pede para que o usuário entre */
scanf("%s", &segunda_rede); /* com o arquivo correspondente */
if (( fs = fopen(segunda_rede, "r") ) == NULL) /* Verifica se o arquivo existe */
{
    printf("\nNão foi possível abrir o arquivo! Verifique se o nome fornecido está correto.\n");
    exit(1); /* Se não, emite mensagem de erro e aborta */
}

while(fgets(frase_segundo_arquivo,50,fs)!=NULL) /* Lê todo o arquivo contendo a interpretação textual */
{ /* para a primeira rede, armazenando suas frases */
    b++;
    strcpy(frase2[b].palavra, frase_segundo_arquivo);
}

if(strcmp(frase1[2].palavra, frase2[2].palavra)!=0) /* Procedimento para verificar se os */
{ /* arquivos são comparáveis entre si */
    printf("\nCertifique-se de que estes arquivos são comparáveis!\n");
    printf("Enquanto num aparece 'O gene (...) modula', no outro aparece 'O gene (...) é modulado'.\n");
    exit(1);
}

printf("\nDigite o nome do arquivo que receberá o resultado da comparação.\n"); /*Pede ao usuário digitar o */
printf("O arquivo deve ter extensão '.txt'.\n"); /* nome do arquivo de resultados */
scanf("%s", &resultado_comparacao);
if (( ft = fopen(resultado_comparacao, "w") ) == NULL) /* Procedimento de verificação */
{
    printf("O arquivo que receberá o resultado não pode ser aberto!\n");
    exit(1);
}

for(x=1;x<=a;x++) /* Rotina de comparação entre os arquivos */
{
    if(strstr(frase1[x].palavra, "O gene"))
    {
        for(y=1;y<=b;y++)
        {
            if(strcmp(frase1[x].palavra, frase2[y].palavra)==0)
            {
                g=x;
                do
                {
                    h=y;
                    do
                    {
                        if(strcmp(frase1[g].palavra, frase2[h].palavra)==0)
                        {
                            fprintf(ft, "%s", frase2[h].palavra);
                        }
                    }
                }
            }
        }
    }
}

```

```

        h++;
    }while(!strstr(frase2[h].palavra, "---"));

    g++;

    }while(!strstr(frase1[g].palavra, "---"));

    fprintf(ft, "\n-----\n\n"); /* Imprime separador */
    }
}

printf("\nPronto!!\n"); /* Imprime na tela */
printf("\nAbra o arquivo resultante da comparação com o programa\n");
printf("Microsoft Office Word ou com o aplicativo WordPad MFC.\n");

fclose(fs);
fclose(fp); /* Fecha os apontadores para arquivos */
fclose(ft);

} /* Termina a função */

int main(void) /* Função principal, responsável pelo direcionamento às três funções anteriores */
{
    int modulo_GeneNetwork; /* Variáveis locais: inteiros */
    int opcao;

    /* Recepção ao usuário com informações de autoria e instituição */
    printf(".....\n");
    printf("..... BEM VINDO AO PROGRAMA MANDALA!.....\n");
    printf(".....\n");
    printf("...Algoritmo para interpretação da matriz regulatória proveniente do GeneNetwork...\n");
    printf("...Autor: Guilherme Liberato da Silva.....\n");
    printf(".....Mestrando em Ciências área Genética.....\n");
    printf(".....Grupo de Imunogenética Molecular.....\n");
    printf(".....\n");
    printf(".....\n");
    printf(".....\n\n");

    printf("*****\n"); /* Imprime na tela o menu principal */
    printf("* Digite uma das opções a seguir: *\n");
    printf("*\n");
    printf("* 1 - Interpretar rede *\n");
    printf("* 2 - Comparar redes *\n");
    printf("*****\n\n");
    scanf("%d", &opcao);
}

```

```

if(opcao==1)      /* Se o usuário deseja interpretar uma rede gênica */
{
    printf("*****\n");
    printf("  * Entre com o número correspondente ao módulo de inferência      *\n");
    printf("  * utilizado no programa GeneNetwork:                               *\n");
    printf("  *                                                                     *\n");
    printf("  * 1 - Boolean Network                                                *\n");
    printf("  * 2 - Linear Model                                                    *\n");
    printf("  * 3 - S-Systems                                                       *\n");
    printf("  * 4 - Bayesian Network                                                *\n");
    printf("*****\n");
    scanf("%d", &modulo_GeneNetwork);

    while(modulo_GeneNetwork<1 || modulo_GeneNetwork>4) /* Verifica se o usuário digitou */
    {
        printf("\nDigite um número entre 1 e 4!\n");      /* uma opção válida (1 ou 2 ou 3 ou 4) */
        scanf("%d", &modulo_GeneNetwork);                /* Se não, confirma junto ao usuário */
    }

    switch(modulo_GeneNetwork) /* Direciona para a função de interpretação adequada, de acordo com */
    {
        /* o tipo de interação característico de cada módulo do GeneNetwork */
        case 1: printf("Módulo utilizado: Boolean Network\n"); grafo_UMA_COR(); break;

        case 2: printf("Módulo utilizado: Linear Model\n"); grafo_DUAS_CORES(); break;

        case 3: printf("Módulo utilizado: S-Systems\n"); grafo_DUAS_CORES(); break;

        case 4: printf("Módulo utilizado: Bayesian Network\n"); grafo_UMA_COR(); break;
    }
}
else
{
    if(opcao==2)      /* Caso o usuário queira comparar duas redes regulatórias... */
    {
        compara_REDES();          /* ... o programa invoca a função responsável */
    }
}
/* Fim da função principal */
}

```

Apêndice B

Genetic Susceptibility Loci in Rheumatoid Arthritis Establish Transcriptional Regulatory Networks with Other Genes

Guilherme Liberato Silva,^{a,*} Cristina Moraes Junta,^{a,*}
Elza Tiemi Sakamoto-Hojo,^b Eduardo Antonio Donadi,^c
Paulo Louzada-Junior,^c and Geraldo A. S. Passos^{a,d}

^a*Molecular Immunogenetics Group, Department of Genetics and* ^c*Clinical Immunology Division, Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil*

^b*Department of Biology, University of São Paulo, Ribeirão Preto, Brazil*

^d*Discipline of Genetics, Department of Morphology, Faculty of Dentistry, University of São Paulo, Ribeirão Preto, Brazil*

Linkage studies have identified the human leukocyte antigen (*HLA*)-*DRB1* as a putative rheumatoid arthritis (RA) susceptibility locus (SL). Nevertheless, it was estimated that its contribution was partial, suggesting that other non-*HLA* genes may play a role in RA susceptibility. To test this hypothesis, we conducted microarray transcription profiling of peripheral blood mononuclear cells in 15 RA patients and analyzed the data, using bioinformatics programs (significance analysis of microarrays method and GeneNetwork), which allowed us to determine the differentially expressed genes and to reconstruct transcriptional networks. The patients were grouped according to disease features or treatment with tumor necrosis factor blocker. Transcriptional networks that were reconstructed allowed us to identify the interactions occurring between RA SL and other genes, for example, *HLA-DRB1* interacting with *FNDC3A* (fibronectin type III domain containing 3A). Given that fibronectin fragments can stimulate mediators of matrix and cartilage destruction in RA, this interaction is of special interest and may contribute to a clearer understanding of the functional role of *HLA-DRB1* in RA pathogenesis.

Key words: rheumatoid arthritis; susceptibility loci; gene expression; gene networks

Introduction

Although environmental and genetic factors are believed to contribute to the development of autoimmune diseases, including rheumatoid arthritis (RA) [online Mendelian Inheritance in Man (OMIM) number: 180300],^{1,2}

they are strongly affected by susceptibility loci (SL) that are distributed on chromosomes 1q31-q32, 1p13, 1p36.13, 2q32.2-q32.3, 5q31, 6q23, 6p21.3, and 16p13 (for review, see OMIM at <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=180300>).

Patients with RA show considerable variation in disease progression and clinical outcomes. Although genetically complex, the genes in the human leukocyte antigen (*HLA*) class II region on chromosome 6p21.3 have been strongly associated with susceptibility to the disease. Several *HLA-DRB1* alleles (*0401, *0404, *0405, *0408, *0101, *0102, *1001,

Address for correspondence: Geraldo A. S. Passos, Molecular Immunogenetics Group, Department of Genetics, Faculty of Medicine of Ribeirão Preto, University of São Paulo, 14040-900, Ribeirão Preto, SP, Brazil. Voice: +55 16 36 02 3030; fax: + 55 16 36 33 00 69. passos@rge.fmrp.usp.br

*These authors contributed equally to this study.

and *1402) have been associated with RA, depending on the population studied.³ All of these DRB1 alleles share a highly conserved amino acid sequence between positions 70–74 (QKRAA, QRRAA, or RRRAA) in the third hypervariable region (HVR3), which forms part of the peptide-binding pocket in the DR heterodimer molecule. This conserved sequence has been commonly referred to as the RA-shared epitope (SE).

Although the presence of a double dose of HLA-DR SE alleles has been associated with the outcome of RA,^{4–6} their contribution to disease susceptibility is estimated to be approximately 30%, strongly suggesting the participation of non-HLA genes. Moreover, RA patients are complex and feature different laboratory, clinical, and treatment characteristics that can be used to predict disease severity.

For example, autoantibodies against cyclic citrullinated peptides (anti-CCP antibodies) are putative and good predictors of the development of RA; they are highly specific for the disease and can be detected years before the first clinical manifestations.^{7–9}

The presence of joint damage and signs of disease activity are predictors of RA progression. Measurement of the disease activity score 28 (DAS28) enables clinicians to monitor the disease course and establish high- and low-disease activity, helping to identify rapidly progressing patients. High-disease activity is strongly correlated with lower functional capacities.¹⁰

The introduction of monoclonal antibodies that block the inflammatory cytokine tumor necrosis factor (TNF) as biologic therapeutic agents has remarkably advanced the treatment of RA. TNF blockers (anti-TNF) are frequently prescribed following the failure of one or several disease-modifying antirheumatic drugs (DMARDs), such as methotrexate (MTX). While anti-TNF therapy holds great promise for many patients, a substantial percentage of patients (40–60%) do not respond to either DMARD or biologic therapy. Moreover, anti-TNF therapy is expensive and can be associ-

ated with serious side effects, such as increased risk of infection and malignancy.¹¹ Efforts have been concentrated on discriminating responders from nonresponding patients, particularly those who are prone to develop serious toxicity from TNF blockers.^{11,12}

For this study, we hypothesized that the presence of *HLA-DRB1* alleles, which strongly influence disease susceptibility, plays a role by interacting with other non-HLA genes, including those also considered to be susceptibility loci for RA.

Considering that RA is a complex disease with many different features that could influence large-scale gene expression (transcriptome profiling), the patients were grouped according to: (1) the presence of HLA-SE in a single or double dose, which may influence the outcome of RA; (2) anti-CCP antibodies in the serum, which is pathognomonic of RA; (3) disease activity as evaluated by the DAS28, which may be interpreted as a clinical phenotype; and (4) treatment with a TNF blocker, which may influence the transcriptome profiling.

We used peripheral blood mononuclear cells (PBMC) because they may act as reporters of ongoing chronic tissue inflammation via their gene expression. The microarray method was used and the data were analyzed by significance analysis of microarrays method (SAM) and GeneNetwork programs, which allowed us to establish the transcriptional profiles of RA patients stratified according to the features mentioned above. We also highlighted posttranscriptional interactions between some of the expressed genes in each of these comparisons.

Materials and Methods

Rheumatoid Arthritis Patients

All 15 patients fulfilled the 1987 revised criteria of the American College of Rheumatology for the diagnosis of RA.¹³ Of these, six exhibited SE in a double dose (SE/SE), four

in a single dose (SE/X), and five exhibited no SE (X/X). According to the DAS28,¹⁴ nine patients exhibited DAS28 > 5.0 and six had DAS28 < 5.0. Nine patients were positive for anti-CCP antibodies and six were not. All patients were using DMARDs and low doses of prednisone (<10 mg/day), and five were using a TNF blocker (infliximab 3–5 mg/Kg/dose). The patients were from urban southeastern Brazil, whose population is ethnically distinct and also an admixture of individuals between European Caucasians, Middle Easterners, and Black Africans. The study was approved by the Ethics Committee of the University Hospital of Ribeirão Preto, University of São Paulo, Brazil (protocol #8090/2001), and all patients provided an informed consent to participate in the study.

Anti-CCP Antibody Detection

The detection of anti-CCP2 IgG antibodies was performed using commercially available ELISA kits containing synthetic peptides (Quanta Lite anti-CCP 2, Inova, San Diego, CA). The ELISA was performed according to the manufacturer's instructions. Serum samples with results >20 U/mL were considered to be positive.

HLA-DRB1 Typing, RNA Preparation, and cDNA Microarray Hybridization

PBMCs were separated by Ficoll-Hypaque (Sigma, St. Louis, MO) gradient density. DNA was extracted using a classical salting-out procedure. HLA-DRB1 typing at a low- or high-resolution level was performed using PCR-amplified DNA hybridized with sequence-specific primers, using commercially available kits (One-Lambda, Canoga Park, CA).

Total RNA was extracted from PBMCs, using Trizol[®] (Invitrogen, Carlsbad, CA) reagent following the manufacturer's instructions. The integrity of the total RNA samples was evaluated by denaturing agarose gel electrophoresis under standard conditions and by Northern-

blot analysis, using a ³²P-labeled oligonucleotide probe recognizing the 28S rRNA fraction (data not shown).

The gene expression of PBMCs was assessed using glass slide cDNA microarrays prepared on silane-coated Ultra GAPS slides (#40015, Corning[®], New York, NY). The 4500 cDNA sequences were retrieved from the human-expressed sequence tag (EST) cDNA library (IMAGE Consortium, <http://image.hudsonalpha.org>). Microarrays were prepared based on published protocols, using PCR products from the cDNA clones¹⁵ and a Generation III Array Spotter (Amersham Molecular Dynamics, Sunnyvale, CA).

The cDNA complex probes derived from RA patients and reference samples were prepared by reverse transcription using 10 µg of total RNA labeled with the Cy3 fluorochrome, using the CyScribe post labeling kit (GE Healthcare Life Sciences, Sunnyvale, CA). A 15-h period was required for hybridization, followed by washing using an automatic slide processor system (ASP, Amersham Biosciences, Sunnyvale, CA). Microarrays were scanned using a Generation III laser scanner (Amersham Biosciences). As a reference for the hybridization procedure, we used equimolar quantities of cDNAs obtained from the total RNA of different human cell strains (Jurkat, HeLa, HEp-2, and U343). This approach allowed for the estimation of the relative amount of cDNA target sequence in each microarray spot.

cDNA Microarray Data Analysis

The microarray image quantification was performed using Spottfinder software (<http://www.tm4.org/>). The normalization process was carried out using the R platform (<http://www.r-project.org>), and statistical data were analyzed using the Multiexperiment Viewer (MeV) software (version 3.1; available online at <http://www.tm4.org/mev.html>).¹⁶

The SAM method¹⁷ was used for each independent sample, assigning a score to each gene on the basis of change in gene expression

relative to the standard deviation of repeated measurements. SAM uses permutations of the repeated measurements to estimate the percentage of genes identified by chance or the false discovery rate ($FDR = 0.005$). Its objective is to construct a scatter plot displaying the induced and repressed genes. The SAM program is available online at (<http://www-stat.stanford.edu/~tibs/SAM/>).

Data obtained using SAM were used to construct a Venn diagram identifying the common and exclusively expressed genes of each stratum (SE, anti-CCP, DAS28, treatment with DMARDs plus TNF blocker). Data mining of the genes was performed using SOURCE from the Stanford Microarray Database (<http://source.stanford.edu>).

Finally, the transcriptional regulatory networks were reconstructed using the GeneNetwork program,¹⁸ which can be obtained from the authors at (<http://idv.sinica.edu.tw/hchuang/GeneNetwork1.2Setup.exe>). This algorithm was used to compare the means of different gene expression values whose standard deviation did not overlap. Its objective was to compute a network that established relationships between all genes. Bayesian statistics for the network organization were used. In order to organize networks with only genes whose expression values were significant, values for the induced and repressed genes selected by the SAM statistics were used. In this case, of the 4500 sequences present on the microarray, only those differentially expressed in each of the four patient groups (*HLA-DRB1* SE positive, anti-CCP antibodies, $DAS28 > 5.0$, and treatment with DMARDs plus TNF blocker) were included in the calculations of the GeneNetwork software because of their statistical significance.

Results

Differentially Expressed Genes

From the set of 4500 cDNA sequences present on the microarray and used to trace

the differential gene expression profiling, we selected only those sequences that were positive in at least 86% (3876) of the microarray hybridizations for statistical analysis.

To identify the variation in gene expression of the full data set and to search for significant genes on the basis of differential expression, SAM analyses were performed for the following comparisons: SE haplotype groups (presence of SE in double or single dose versus no dose), anti-CCP antibodies (positive vs. negative), disease activity scores ($DAS28 > 5.0$ versus $DAS28 < 5.0$), and treatment features (DMARDs plus TNF blocker vs. only).

A Venn diagram was constructed to identify the common and exclusively modulated genes according to the above mentioned features (Fig. 1).

The presence of SE in a single or double dose identified 3675 differentially expressed genes when compared to patients with no SE. The presence of anti-CCP antibodies led to the differential expression of 3645 genes observed in patients exhibiting the autoantibody. Disease activity modulated the expression of 3493 genes that were observed in RA patients exhibiting active disease ($DAS28 > 5.0$). Treatment modulated 3134 genes identified in patients treated with DMARDs plus TNF blocker.

Gene Networks

As previously mentioned, the data of the significantly expressed genes according to the patient groups as revealed by the SAM statistics were entered into the GeneNetwork 1.2 program in order to visualize their possible relationships. All four groups of patients were dynamic in terms of interacting genes (summary information provided in Table 1). The Bayesian statistics used as a parameter to compare patient samples allowing reconstruction of the networks enabled us to visualize modulations of gene nodes over other genes, that is, a gene node g modulating a gene h ($g \rightarrow h$).

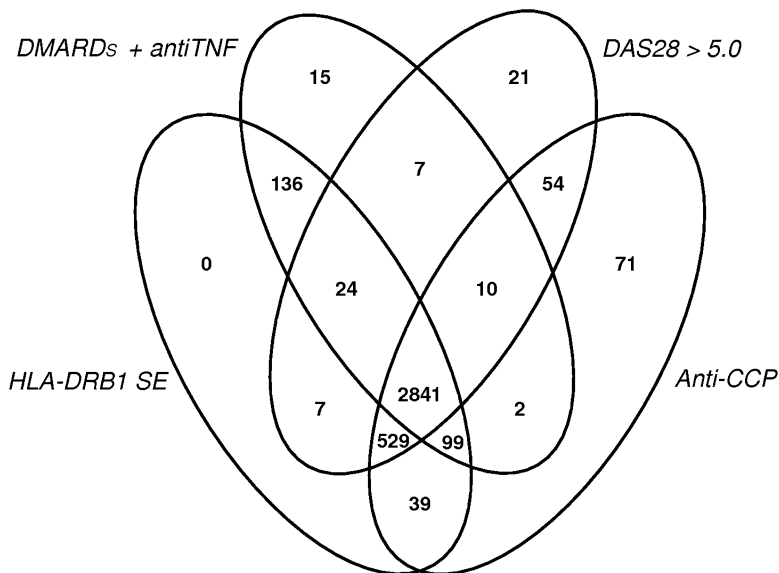


Figure 1. Venn diagram identifying the common and exclusively expressed genes of each stratum: shared epitope (SE), anticyclic citrullinated peptides (CCP) antibodies, disease activity score 28 (DAS28), and treatment with disease-modifying antirheumatic drugs (DMARDs) plus tumor necrosis factor (TNF) blocker.

Among the differentially expressed genes in the SE-positive patient group, a set of 52 was found to participate in their respective network, and of these, 11 (*SMA*, *FLJ22536*, *GST*, *ETNK1*, *ζNF236*, *LIN28B*, *NUP153*, *PECR*, *EFHC1*, and *HLADRBI*) represented gene nodes positioned on RA SL chromosomal regions (Fig. 2A).

The patient group with positive anti-CCP antibodies featured a set of 22 genes that participated in their respective network, and of these, nine (*PECR*, *FLJ22536*, *EFHC1*, *USP45*, *CDC40*, *STX7*, *ETNK1*, *OSTM1*, and *LIN28B*) were gene nodes positioned in RA SL (Fig. 2B).

The DAS28 > 5.0 patient group featured a set of 39 genes in their respective network, and of these, 10 (*USP45*, *ETNK1*, *PECR*, *NUP153*, *GSTA4*, *ζNF236*, *LIN28B*, *SLC38A1*, *EFHC1*, and *HLA-DRBI*) were gene nodes positioned in RA SL (Fig. 2C).

Finally, the group of patients treated with DMARDs plus TNF blocker featured 19 genes in their network, and of these, eight (*GSTA4*, *ζNF236*, *USP45*, *ζSCAN16*, *FAM46A*,

PDE1B, *ITGA5*, and *ETNK1*) were gene nodes positioned in RA susceptibility loci (Fig. 2D).

Discussion

Autoimmune diseases, including RA, represent one of greatest medical and biological challenges because of their multifactor origin and heterogeneous features. Imbalance in the control of central tolerance mechanism by means of thymic expression of genes that code for peripheral tissue antigens is gaining support as a mechanism for such immune dysfunction.^{19,20} Moreover, several environmental factors, such as lifestyle and food, may play a role as triggers in genetically susceptible individuals. In the case of RA—but not restricted to it—genetic factors strongly affect disease susceptibility, including *HLA-DRBI* alleles situated on the chromosome 6p21.3 region.

Candidate gene studies and genome-wide linkage analysis are the two primary approaches used to identify relevant genes in

TABLE 1. Genes Whose mRNAs Participate in the Networks Depicted in Figure 2A–D.

Gene Symbol	Gene Name	Accession Number	Cytoband	Summary Function
ADCY1	Adenylate cyclase 1 (brain)	R14114.1	7p13-p12	cAMP biosynthetic process
ADD2	Adducin 2 (β)	R19254.1	2p14-p13	immune system development
ADIPOR1	Adiponectin receptor 1	R62584.1	1p36.13-q41	serve as receptors for adiponectin, fatty acid oxidation and glucose uptake by adiponectin
ANAPC2	Anaphase-promoting complex subunit 2	R33244.1	9q34.3	cellular macromolecule biosynthetic process, termed the anaphase-promoting complex (APC)
ANKRD56	Ankyrin repeat domain 56	R24589.1	4q21.1	unknown
C11orf71	Chromosome 11 open reading frame 71	R19933.1	11q14.2-q14.3	unknown
C12orf51	Chromosome 12 open reading frame 51	R71499.1	12q24.13	unknown
C14orf112	Chromosome 14 open reading frame 112	R18677.1	14q24.2	unknown
C5orf42	Chromosome 5 open reading frame 42	R37164.1	5p13.2	unknown
C6orf166	Chromosome 6 open reading frame 166	R69874.1	6q15	unknown
C6orf170	Chromosome 6 open reading frame 170	R73756.1	6q22.31	regulation of signal transduction and hydrolase activity
C7orf43	Chromosome 7 open reading frame 43	R21690.1	7q22.1	unknown
CCDC22	Coiled-coil domain containing 22	R46105.1	Xp11.23	unknown
CCL28	Chemokine (C-C motif) ligand 28	R36539.1	5p12	immunoregulatory and inflammatory processes
CDC14A	CDC14 cell division cycle 14 homologue A (<i>Saccharomyces cerevisiae</i>)	T74426.1	1p21	dephosphorylation, regulation of progression through cell cycle
CDC40	Cell division cycle 40 homologue (<i>S. cerevisiae</i>)	R70791.1	6q21	mRNA splicing
CEP350	Centrosomal protein 350kDa	R25352.1	1p36.13-q41	required for anchoring microtubules to centrosomes and for integrity of microtubule network
CEP72	Centrosomal protein 72kDa	R37706.1	5p15.33	member of the leucine-rich-repeat superfamily of proteins, localized to the centrosome
CHST10	Carbohydrate sulfotransferase 10	R16195.1	2q11.2	Carbohydrate metabolic process, learning and/or memory
CNTN2	Contactin 2 (axonal)	R14151.1	1q32.1	cell adhesion, neurogenesis
COMMD2	COMM domain containing 2	R23418.1	3q25.1	unknown
CORO1B	Coronin, actin-binding protein, 1B	R36510.1	11q13.1	regulation of cell motility, signal transduction

Continued

TABLE 1. Continued

Gene Symbol	Gene Name	Accession Number	Cytoband	Summary Function
CPEB3	Cytoplasmic polyadenylation element-binding protein 3	R36999.1	10q23.32	unknown
CPSF6	Cleavage and polyadenylation specific factor 6, 68 kDa	R18985.1	12q15	required for 3' RNA cleavage and polyadenylation processing
CRB1	Crumbs homologue 1 (<i>Drosophila</i>)	R35113.1	1q31-q32.1	eye photoreceptor cell differentiation
CRY1	Cryptochrome 1 (photolyase-like)	R19031.1	12q23-q24.1	DNA repair, regulation of transcription, visual perception
CTBS	Chitobiase, di-N-acetyl-	R56854.1	1p22	cellular polysaccharide metabolic process
DDX49	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49	R26193.1	19p12	unknown
DHRS7B	Dehydrogenase/reductase (SDR family) member 7B	R36668.1	17p12	metabolic process
DKFZp686L-14188	Hypothetical gene supported by BX538329	R76726.1	8q21.3	unknown
EFHC1	EF-hand domain (C-terminal) containing 1	R68083.1	6p12.3	positive regulation of programmed cell death, cation homeostasis
EHMT1	Euchromatic histone-lysine N-methyltransferase 1	R38436.1	9q34.3	Chromosome organization and biogenesis, protein methylation
EMP2	Epithelial membrane protein 2	R71514.1	16p13.2	cell development
ETNK1	Ethanolamine kinase 1	R75813.1	12p12.1	Phospholipid and ethanolamine biosynthetic process
EXPH5	Exophilin 5	R70812.1	11q22.3	intracellular protein transport and cellular localization
FAM46A	Family with sequence similarity 46, member A	R71124.1	6q14	unknown
FAM70A	Family with sequence similarity 70, member A	R39187.1	Xq24	unknown
FBXL3	F-box and leucine-rich repeat protein 3	R37265.1	13q22	protein ubiquitination
FLJ22536	Hypothetical locus LOC401237	R15860.1	6p22.3	unknown
FNDC3A	Fibronectin type III domain containing 3A	R36431.1	13q14.2	unknown
FRMPD2L2	FERM and PDZ domain containing 2 like 2	R20356.1	10q11.21	unknown
GAL3ST3	Galactose-3-O-sulfotransferase 3	R14066.1	11q13.1	Proteoglycan and carbohydrate biosynthetic process

Continued

TABLE 1. Continued

Gene Symbol	Gene Name	Accession Number	Cytoband	Summary Function
GART	Phosphoribosylglycinamide formyltransferase	R14198.1	21q22.1 21q22.11	purine ribonucleoside monophosphate biosynthetic process
GPR172A	G protein-coupled receptor 172A	R18827.1	8q24.3	cell surface receptor for porcine endogenous retrovirus subgroup A
GPSM2	G protein signaling modulator 2 (AGS3-like, <i>Caenorhabditis elegans</i>)	R25539.1	1p13.3	cell surface receptor linked signal transduction
GSTA4	Glutathione S-transferase A4	T66320.1	6p12.1	response to stress
GTF3C4	General transcription factor IIIC, polypeptide 4, 90 kDa	R17802.1	9q34.13	transcription
HBB	Hemoglobin, β	R69941.1	11p15.5	positive regulation of nitric oxide biosynthetic process, determines the structure of the β -chain in adult hemoglobin
HLA-DRB1	Major histocompatibility complex (MHC), class II, DR β -3	H50622.1	6p21.3	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II
HMGCLL1	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase-like 1	R34667.1	6p12.1	metabolic process
HPS5	Hermansky-Pudlak syndrome 5	R71594.1	11p14	organelle organization and biogenesis, pigmentation
HSD11B2	Hydroxysteroid (11- β) dehydrogenase 2	R35081.1	16q22	glucocorticoid biosynthetic process
INPP5E	Inositol polyphosphate-5-phosphatase, 72 kDa	R36586.1	9q34.3	lipid metabolic process
ITGA5	Integrin, α -5 (fibronectin receptor, α -polypeptide)	R32403.1	12q11-q13	integrin-mediated signaling pathway, cell adhesion
KCNT2	Potassium channel, subfamily T, member 2	R51505.1	1q31.3	potassium ion transport
LASS6	LAG1 homologue, ceramide synthase 6	R25481.1	2q24.3	regulation of transcription DNA-dependent
LIN28B	Lin-28 homologue B (<i>C. elegans</i>)	R64358.1	6q21	regulation of transcription DNA-dependent
LOC100129345	Hypothetical protein LOC100129345	R24238.1	14q32.2	unknown
LOC285954	Hypothetical protein LOC285954	R33696.1	7p13	unknown
LOC56251	Hypothetical protein YH95C04	R38312.1		unknown
LOC654350	Similar to rat ribosomal protein L9 homologue	R24333.1	15q22.31	unknown

Continued

TABLE 1. Continued

Gene Symbol	Gene Name	Accession Number	Cytoband	Summary Function
MALT1	Mucosa-associated lymphoid tissue lymphoma translocation gene 1	T80158.1	18q21	positive regulation of cytokine production during adaptative immune response, activation of nuclear factor (NF)- κ B cascade
MOBP	Myelin-associated oligodendrocyte basic protein	R13982.1	3p22.1	nervous system development
N4BP1	NEDD4-binding protein 1	R28614.1	16q12.1	unknown
NAGA	N-acetylgalactosaminidase, α -	R25825.1	22q13-qter 22q11	carbohydrate metabolic process
NCAPG	Non-SMC condensin I complex, subunit G	—	4p15.33	mitotic sister chromatid segregation, M phase
NECAB2	N-terminal EF-hand calcium-binding protein 2	T80509.1	16q23.3	unknown
NLK	Nemo-like kinase	R70769.1	17q11.2	negative regulation of Wnt receptor signaling pathway, regulation of transcription
NNAT	Neuronatin	R63918.1	20q11.2-q12	brain development
NOL7	Nucleolar protein 7, 27 kDa	R69895.1	6p23	unknown
NPY1R	Neuropeptide Y receptor Y1	R19478.1	4q31.3-q32	G protein signaling, coupled to cAMP nucleotide second messenger
NUP153	Nucleoporin 153kDa	R40059.1	6p22.3	establishment of localization, transport
OSTM1	Osteopetrosis-associated transmembrane protein 1	R69603.1	6q21	myeloid leukocyte differentiation
PCDH17	Protocadherin 17	R49279.1	13q21.1	homophilic cell adhesion
PDE1B	Phosphodiesterase 1B, calmodulin-dependent	H51117.1	12q13	programmed cell death, visual learning
PECR	Peroxisomal trans-2-enoyl-CoA reductase	H84466.1	2q35	regulation of apoptosis, fatty acid biosynthetic process
PHF1	PHD finger protein 1	R37685.1	6p21.3	transcriptional repressor
PHLPPL	PH domain and leucine rich repeat protein phosphatase-like	R25700.1	16q22.3	dephosphorylation
PIP5K1A	Phosphatidylinositol-4-phosphate 5-kinase, type I, α	—	1q22-q24	glycerophospholipid metabolic process
PKD2	Polycystic kidney disease 2 (autosomal dominant)	R38199.1	4q21-q23	negative regulation of progression through cell cycle, JAK-STAT cascade, calcium ion transport
PLCE1	Phospholipase C, epsilon 1	R40791.1	10q23	positive regulation of MAPK activity, G protein signaling coupled to IP3 second messenger, diacylglycerol biosynthetic process, regulation of cell growth

Continued

TABLE 1. Continued

Gene Symbol	Gene Name	Accession Number	Cytoband	Summary Function
POLR1E	Polymerase (RNA) I polypeptide E, 53kDa	R54238.1	9p13.2	rRNA transcription
POLR3E	Polymerase (RNA) III (DNA directed) polypeptide E (80kD)	R18902.1	16p12.1	transcription
POMP	Proteasome maturation protein	R38655.1	13q12.3	immune response, protein folding
PPCDC	Phosphopantothenoylcysteine decarboxylase	R33457.1	15q24.2	coenzyme A biosynthetic process
PREPL	Prolyl endopeptidase-like	R15785.1	2p21	proteolysis
RAB37	RAB37, member RAS oncogene family	R25709.1	17q25.1	small GTPase-mediated signal transduction
RAPGEF1	Rap guanine nucleotide exchange factor (GEF) 1	R76487.1	9q34.3	transmembrane receptor protein tyrosine kinase signaling pathway
RBJ	Rab and DnaJ domain containing	H06266.1	2p23.3	small GTPase mediated signal transduction
RBM23	RNA-binding motif protein 23	R42182.1	14q11.2	increases transcription of steroid-responsive transcriptional reporters in a hormone-dependent manner
S1PR1	Sphingosine-1-phosphate receptor 1	R20666.1	1p21	receptor for the lysosphingolipid sphingosine 1-phosphate (S1P)
SCOC	Short coiled-coil protein	R34436.1	4q31.1	unknown
SERPINH1	Serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen-binding protein 1)	R71093.1	11q13.5	molecular chaperone involved in the maturation of collagen molecules (autoantibodies to this protein have been found in patients with rheumatoid arthritis)
SETBP1	SET-binding protein 1	R17655.1	18q21.1	unknown
SETDB1	SET domain, bifurcated 1	R37131.1	1q21	establishment and/or maintenance of chromatin architecture, blastocyst development
SLC38A1	Solute carrier family 38, member 1	R20471.1	12q13.11	glutamine transport
SLC38A1	Solute carrier family 38, member 1	R66555.1	12q13.11	glutamine transport
SLC3A2	Solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	T80267.1	11q13	tryptophan transport, calcium ion transport, cell growth
SMAD2	SMAD family member 2	R33715.1	18q21.1	regulation of transcription, mesoderm formation
SMAP2	Stromal membrane-associated GTPase-activating protein 2	R20203.1	1p35.3-p34.1	unknown

Continued

TABLE 1. Continued

Gene Symbol	Gene Name	Accession Number	Cytoband	Summary Function
SPEG	SPEG complex locus	R25020.1	2q35	protein amino acid phosphorylation, muscle development
STX7	Syntaxin 7	R73740.1	6q23.1	post-Golgi vesicle-mediated transport
SYN2	Synapsin II	R38909.1	3p25	neurotransmitter secretion
SYN2	Synapsin II	R49277.1	3p25	neurotransmitter secretion
TBC1D22B	TBC1 domain family, member 22B	T66272.1	6p21.2	regulation of Ras GTPase activity, regulation of Rab protein signal transduction
TCN1	Transcobalamin I (vitamin B12-binding protein, R binder family)	R11710.1	11q11-q12	cobalamin transport
TRIM27	Tripartite motif-containing 27	H08440.1	6p22	repression of gene transcription, spermatogenesis
USP45	Ubiquitin specific peptidase 45	R15391.1	6q16.3	proteolysis involved in cellular protein catabolic process
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6 (Ku autoantigen, 70 kDa)	T79976.1	22q13.2-q13.31	double-strand break repair via nonhomologous end joining, positive regulation of transcription
ZNF236	Zinc finger protein 236	R25834.1	18q22-q23	regulation of transcription DNA-dependent
ZNF385C	Zinc finger protein 385C	R20240.1	17q21.2	unknown
ZNF511	Zinc finger protein 511	R68665.1	10q26.3	regulation of transcription DNA-dependent
ZNF718	Zinc finger protein 718	R37117.1	4p16.3	unknown
ZSCAN16	Zinc finger and SCAN domain containing 16	R35096.1	6p22.1	regulation of transcription DNA-dependent

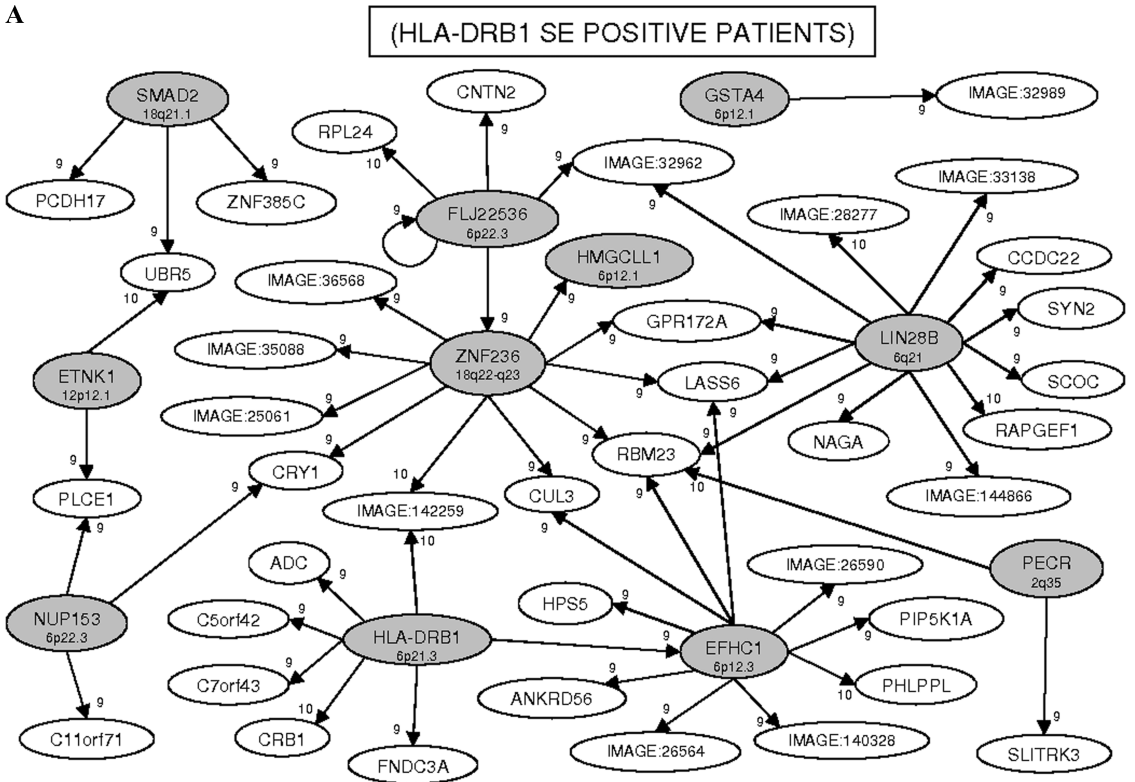
In bold, sequences localized on susceptibility loci. JAK-STAT, Janus kinases - signal transducer and activator of transcription; MAPK, mitogen-activated protein kinase; IP3, inositol phosphate second messenger.

RA.²¹⁻²⁵ These studies have identified large regions situated on different chromosomes that harbor genes other than HLA that play their roles in concert with *HLA-DRB1*, significantly contributing to the genetic susceptibility to RA. Recently, a novel and relevant lymphocyte signal transduction gene, *PTPN22* (protein tyrosine phosphatase nonreceptor type 2), situated on chromosome 1p13.3-p13.1, was identified by a genome-wide scan and shown to be shared by many autoimmune disorders, including RA, systemic lupus erythematosus (SLE), and diabetes mellitus type 1.²⁶⁻²⁸

We agree with Yamamoto and Yamada²¹ that candidate gene studies and linkage analysis are useful for locating susceptibility loci but alone cannot provide a complete understanding of the mechanisms of RA. Association studies using common allelic variants have been proposed as a powerful method for identifying common variants that underlie disease variants.²⁹

In our point of view, although their function is largely unknown, the non-HLA genes located in susceptibility regions are likely to be differentially expressed in RA (and other autoimmune

A



B

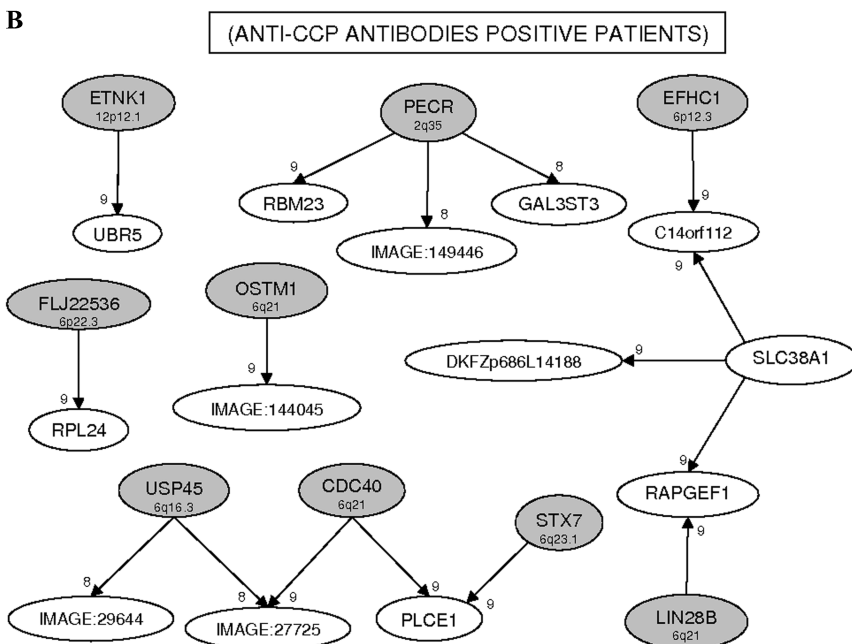


Figure 2. Transcriptional gene networks reconstructed from cDNA microarray gene expression data of the following patient groups. **(A)** SE haplotype, **(B)** presence of anti-CCP antibodies, **(C)** high DAS (DAS28 > 5.0), and **(D)** treatment with DMARDs plus TNF blocker. Arrows indicate the direction of gene interactions. Numerals correspond to the number of patients in which the interaction occurs; HLA, human leukocyte antigen.

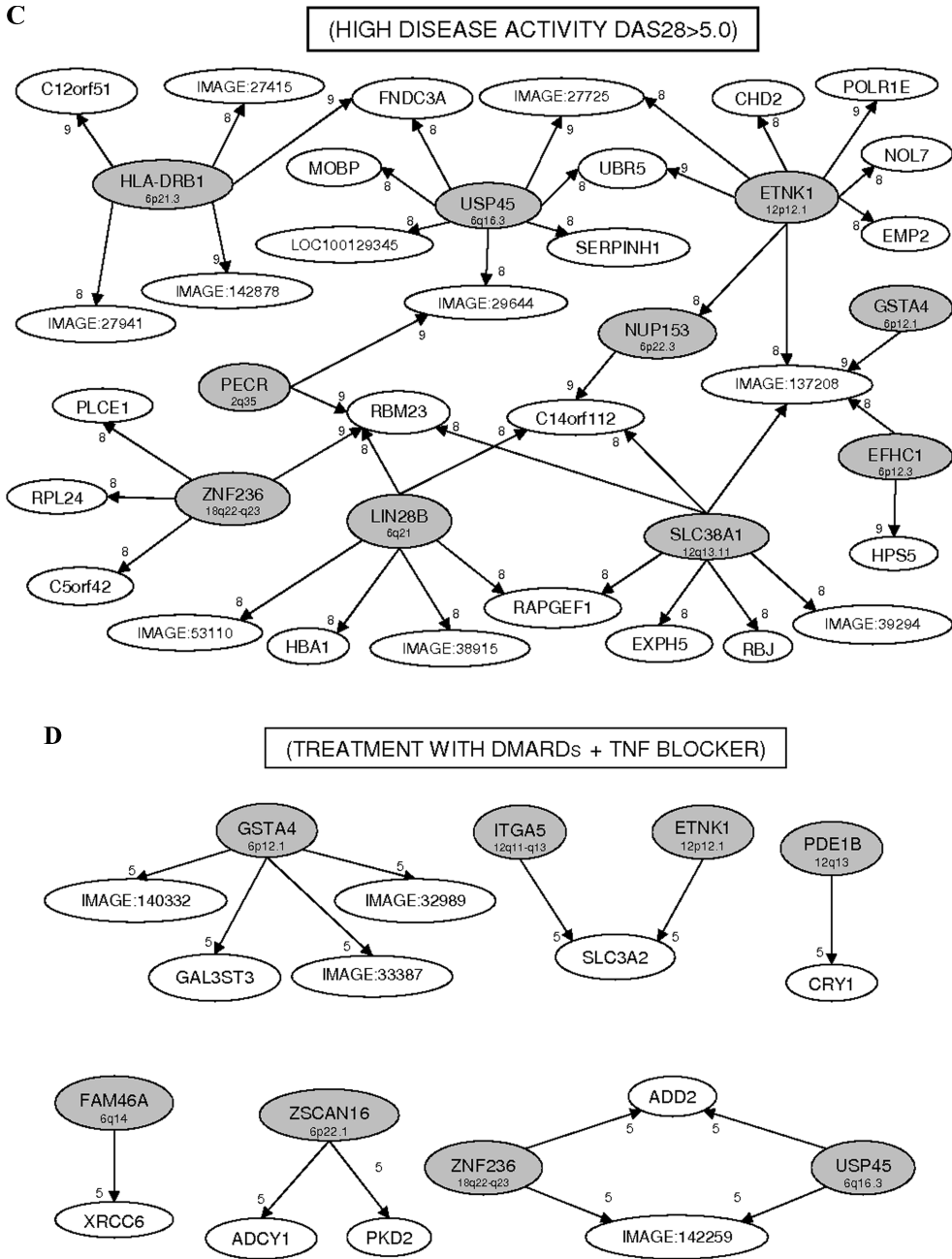


Figure 2. (Continued)

diseases), and this can be revealed by means of gene expression studies using microarray technology.

Our group has pursued this idea and demonstrated important features of the transcriptome profile in autoimmune diseases, such as how T lymphocyte gene expression is affected by im-

munosuppressive treatment in SLE patients,³⁰ the localization of new expressed sequence tags (ESTs) originating from susceptibility loci in SLE,³¹ the association of class II HLA genotyping with differential gene expression in patients with type I diabetes mellitus,³² profiling meta-analysis of large-scale gene expression

between RA and SLE,³³ and, recently, reconstructing transcriptional networks during thymus ontogeny.³⁴

In a recent study using the microarray method, we demonstrated an association of the differential gene expression profile of RA patients with their immunogenetic autoimmune response, disease activity, and treatment features. Many of the genes found were located in RA susceptibility regions, and others had roles in different biological processes, such as signal transduction, phosphorylation, and apoptosis.³⁵ The study contributed to dissecting the molecular complexity of RA and a better understanding of the influence of genes on the disease phenotype and response to treatment.

Microarray technology has been used to discriminate differences in gene expression profiles in tissue or PBMCs. Although tissue and cells obtained from the site of inflammation may more accurately reflect the ongoing pathogenetic features of chronic diseases, both synovial tissue and PBMCs have been used to evaluate differences in gene expression profiles in RA.^{32,36} If expression-based profiling is to be of practical importance, the ease of sample accessibility is crucial. In this context, peripheral immune cells are the key sentinels of host defense, being used to identify novel disease mediators, disease variants, and treatment responses.^{37–39} Because the gene expression profile may be influenced by disease activity, the presence or absence of relevant genetic (SE) and autoimmune biomarkers (anti-CCP antibodies), and even by the treatment used, in this study we evaluated the gene expression profile in PBMCs of RA patients stratified according to these variables. In order to identify differentially expressed genes influenced by each variable, a Venn diagram was constructed (Fig. 1). The modulated genes for each variable were then analyzed in terms of their interactions by means of their respective mRNAs transcripts (transcriptional networks) (Fig. 2A–D).

Since deregulation of the immune system has been implicated in the pathogenesis of autoimmune disorders, these genes represent

a fruitful path toward the understanding of their pathogenic features. In the following paragraphs, we will discuss the features of some of the important differentially expressed genes whose mRNAs establish network interactions for each variable evaluated in the present study.

Considering the modulated genes observed in RA patients with SE in a single or double dose (Fig. 2A), we pinpointed the following gene nodes that established a transcriptional network: the *ZNF236* (zinc finger protein 236) gene located at chromosome 18q22-q23, which is an RA SL region. Its encoded protein is a zinc ion-binding protein that interacts with DNA and regulates transcription. In the network, this gene interacted with 10 other genes, including *HGMCLL1*, an RA SL located at chromosome 6p12.1, and the *PECR* (peroxisomal trans-2-enoyl-CoA reductase) gene, located at the chromosome 2q35 RA SL region and encoding a mitochondrial/peroxisomal protein with oxidoreductase activity that plays a role in apoptosis regulation. Their respective mRNAs interact with two others (*SLITRK3* and *RBM23*). This network also included the *HLADRB1* allele, located at the chromosome 6p21.3 RA SL region. The *HLA-DRB1* allele belongs to the HLA class II β -chain paralogues. The class II molecule is a heterodimer consisting of an α - (DRA) and a β -chain (DRB), both anchored in the membrane. It plays a central role in the immune system by presenting peptides derived from extracellular proteins. Class II molecules are expressed in antigen-presenting cells, and DRB1 variants are considered to be RA SEs (shared epitopes), which play an important role in disease susceptibility.

Some of these differentially expressed genes have not yet been studied in association with RA.

The presence of anti-CCP antibodies (Fig. 2B), similar to the observations in SE-positive patients, influenced the *PECR* gene node, which interacts with *RBM23* and two other genes, including ID:149446 transcribed sequence and *GAL3ST3*. The *USP45* (ubiquitin-specific peptidase 45) gene, located

at the chromosome 6q16.3 RA SL region, encodes a protein participating in the ubiquitin-dependent protein catabolic process. The *OSTM1* (osteopetrosis-associated transmembrane protein 1) gene is located at chromosome 6q21, an RA SL, and encodes a protein that may be involved in the degradation of G proteins via the ubiquitin-dependent proteasome pathway. The encoded protein binds to members of subfamily A of the regulator of the G protein signaling (RGS) family through an N-terminal leucine-rich region. This protein also has a central *RING* finger-like domain and E3 ubiquitin ligase activity. This protein is highly conserved from flies to humans, and defects in the gene may cause the autosomal recessive infantile malignant form of osteopetrosis (OMIM 259720). In this network, *OSTM1* mRNA interacts with transcribed sequence ID 144045.

In the patient group with a high score for disease activity (DAS28 > 5.0) (Fig. 2C), we also observed the participation of the susceptibility loci *USP45* (6q16.3), *PECR* (2q35), and *ζNF236* (18q22-q23), although they established interactions with other genes. *HLA-DRB1* (6p21.3) and *USP45* (6q16.3) simultaneously interacted with the *FNDC3A* (fibronectin type III domain containing 3A) gene located at chromosome 13q14.2.

In the patient group treated with TNF blocker (Fig. 2D), we pinpointed the following gene nodes: The *GSTA4* (glutathione S-transferase A4) gene is located at chromosome 6p12.1, an RA susceptibility locus region, and showed interactions with four other genes, including the *GAL3ST3* (galactose-3-O-sulfotransferase 3) gene, which encodes a protein integral to the membrane and plays a role in carbohydrate binding. This network also featured *ζNF236* and *USP45* gene nodes, which simultaneously interacted with the transcribed sequence ID 142259 and the *ADD2* (adducing 2-β) gene, located at chromosome 2p14-p13.

Although some of these genes had been previously associated with RA pathogenesis, others were newly unveiled for future research.

Taken together, these results indicate that HLA SE profile, disease activity, treatment, and anti-CCP antibodies all influence the pattern of differential gene expression in RA. In addition, the RA susceptibility loci previously identified by linkage studies are in fact functional chromosomal regions whose mRNA transcripts establish transcriptional regulatory networks with other genes. In this regard, we highlight the interaction between *HLA-DRB1* and the *FNDC3A* gene located at chromosome 13q14.2, which was observed in SE-positive and DAS28 > 5.0 patients. Given that fibronectin fragments can stimulate mediators of matrix and cartilage destruction in RA patients,^{40,41} this kind of interaction is of special interest for better understanding the molecular function of susceptibility loci in autoimmune diseases.

Concluding Remarks

We realize that the cutting edge in autoimmune disease research is multifaceted. Immunobiology and medicine have a great challenge for the next decades to elucidate the decision fate of T cell responses toward autoaggressive or protective routes.⁴² Moreover, researchers are now exploring the possibility of microbial infections⁴³ and vaccination⁴⁴ as triggers of autoimmune diseases. In our point of view, all these environmental factors might elicit autoimmunity depending on the individual genetic background. The present work resumes a new avenue in autoimmunity research establishing a link between linkage studies and functional genomics of susceptibility loci.

Acknowledgments

We thank the patients who kindly collaborated with this research. This study received financial support from the Brazilian agencies FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), CNPq (Conselho Nacional de Desenvolvimento Científico e

Tecnológico), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). The cDNA clones used to prepare the microarrays were given by Dr. Catherine Nguyen from the INSERM U928 (Institut National de la Santé et de la Recherche Médicale, Marseille, France).

Conflicts of Interest

The authors declare no conflicts of interest.

References

- Ollier, W.E.R. 1995. Macgregor A. genetic epidemiology of rheumatoid arthritis. *Br. Med. Bull.* **51**: 267–285.
- Oliveira, R.D. *et al.* 2008. Share epitope, citrullinated cyclic peptide antibodies and smoking in Brazilian rheumatoid arthritis patients. *Clin. Rev. Allergy Immunol.* **34**: 325–332.
- Zanelli, E. *et al.* 2000. HLA class II association with rheumatoid arthritis: facts and interpretations. *Hum. Immunol.* **61**: 1254–1261.
- Gregersen, P.K. *et al.* 1987. The shared epitope hypothesis: an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis (current comment). *Arthritis Rheum.* **30**: 1205–1213.
- Montcel, S.T. *et al.* 2005. New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum.* **52**: 1063–1068.
- Thomson, G. 1995. Mapping disease genes: family-based association studies. *Am. J. Hum. Genet.* **57**: 487–498.
- Huizinga, T.W. *et al.* 2005. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum.* **52**: 3433–3438.
- Rantapaa-Dahlqvist, S. *et al.* 2003. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum.* **48**: 2741–2749.
- Louzada-Júnior, P. *et al.* 2008. A majority of Brazilian patients with rheumatoid arthritis HLA-DRB1 alleles carry both the HLA-DRB1 shared epitope and anti-citrullinated peptide antibodies. *Br. J. Med. Biol. Res.* **41**: 493–499.
- Jansen, L.M. *et al.* 2000. Predictors of functional status in patients with early rheumatoid arthritis. *Ann. Rheum. Dis.* **59**: 223–226.
- Hyrich, K.L. *et al.* 2006. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. *Rheumatology (Oxford)* **45**: 1558–1565.
- Plenge, R.M. & L.A. Criswel. 2008. Genetic variants that predict response to anti-tumor necrosis factor therapy in rheumatoid arthritis: current challenges and future directions. *Curr. Opin. Rheumatol.* **20**: 145–152.
- Arnett, F.C. *et al.* 1988. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**: 315–324.
- Prevoo, M.L. *et al.* 1995. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum.* **38**: 44–48.
- Hegde, P. *et al.* 2000. Concise guide to cDNA microarray analysis. *Biotechniques* **29**: 548–562.
- Saeed, A.I. *et al.* 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**: 374–378.
- Tusher, V.G. *et al.* 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**: 5116–5121.
- Wu, C.C. *et al.* 2004. GeneNetwork: an interactive tool for reconstruction of genetic network using microarray data. *Bioinformatics* **20**: 3691–3693.
- Kyewski, B. & J. Derbinski. 2004. Self-representation in the thymus: An extended view. *Nat. Rev. Immunol.* **4**: 688–698.
- Magalhães, D.A. *et al.* 2006. Promiscuous gene expression in the thymus: The root of central tolerance. *Clin. Dev. Immunol.* **13**: 81–99.
- Yamamoto, K. & R. Yamada. 2005. Genome-wide single nucleotide polymorphism analyses of rheumatoid arthritis. *J. Autoimmun.* **25**: 12–15.
- Orozco, G. *et al.* 2006. Genetic basis of rheumatoid arthritis. *Biomed. Pharmacother.* **60**: 656–662.
- Liu, C. *et al.* 2008. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol. Med.* **14**: 575–581.
- Bowes, J. & A. Barton. 2008. Recent advances in the genetics of RA susceptibility. *Rheumatology* **47**: 399–402.
- Barton, A. *et al.* 2008. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum. Mol. Genet.* **17**: 2274–2279.
- Cantor, R.M. *et al.* 2004. Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13 and 17q21-23 and novel evidence at 3p24,

- 10q23-24, 13q32 and 18q22-23. *Arthritis Rheum.* **50**: 3203–3210.
27. John, S. *et al.* 2004. Whole-genome scan in complex disease using 11,245 single-nucleotide polymorphisms: comparisons with microsatellites. *Am. J. Hum. Genet.* **75**: 54–64.
28. Choi, S.J. *et al.* 2006. Genome scan meta-analysis of rheumatoid arthritis. *Rheumatology* **45**: 166–170.
29. Wang, W.Y. *et al.* 2005. Genome-wide association studies. *Nat. Rev. Genet.* **6**: 109–118.
30. Pereira, E. *et al.* 2004. Immunosuppressive therapy modulates T lymphocyte gene expression in patients with systemic lupus erythematosus. *Immunology* **113**: 99–105.
31. Trevisan, G.L. *et al.* 2006. Using cDNA microarrays to identify human CD19+ B-cell gene products (ESTs) originated from systemic lupus erythematosus susceptibility loci. *Autoimmun. Rev.* **5**: 319–323.
32. Rassi, D.M. *et al.* 2006. Is HLA class II profile relevant for the study of large-scale differentially expressed genes in type I diabetes mellitus patients? *Ann. N. Y. Acad. Sci.* **1079**: 305–309.
33. Silva, G.L. *et al.* 2007. Profiling meta-analysis reveals primarily gene coexpression concordance between systemic lupus erythematosus and rheumatoid arthritis. *Ann. N. Y. Acad. Sci.* **1110**: 33–46.
34. Macedo, C. *et al.* 2008. genes that code for T cell signaling proteins establish transcriptional regulatory networks during thymus ontogeny. *Mol. Cell Biochem.* **318**: 63–71.
35. Junta, C.M. *et al.* 2008. Differential gene expression of peripheral blood mononuclear cells from rheumatoid arthritis patients may discriminate immunogenetic, pathogenic and treatment features. *Immunology* (in press). DOI: 10.1111/j.1365-2567.2008.03005x.
36. Toonen, E.J.M. *et al.* 2008. Gene expression profiling in rheumatoid arthritis; current concepts and future directions. *Ann. Rheum. Dis.* **67**: 1174–1177.
37. Shaffer, A.L. *et al.* 2006. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol. Rev.* **210**: 67–85.
38. Yamagata, T. *et al.* 2006. Shared gene-expression signature in innate-like lymphocytes. *Immunol. Rev.* **210**: 52–66.
39. Gu, J. *et al.* 2002. A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthritis patients. *Rheumatology* **41**: 759–766.
40. Yasuda, T. & A.R. Poole. 2002. A fibronectin fragment induces type II collagen degradation by collagenase through an interleukin 1 mediated pathway. *Arthritis Rheum.* **46**: 138–148.
41. Yasuda, T. 2006. Cartilage destruction by matrix degradation products. *Mod. Rheumatol.* **16**: 197–205.
42. Abbas, A.K. *et al.* 2007. Balancing autoaggressive and protective T cell responses. *J. Autoimmun.* **28**: 59–61.
43. Doria, A. *et al.* 2008. Exploring the complex relationships between infections and autoimmunity. *Autoimmun. Rev.* **8**: 89–91.
44. Conti, F. *et al.* 2008. Vaccination and autoimmune rheumatic diseases. *Autoimmun. Rev.* **8**: 124–128.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)