UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA

RENATA DOS SANTOS ALMEIDA

*Similaridades entre o transcriptoma humano e murino focando
genes situados em regiões de susceptibilidade ao diabetes mellitus
do tipo 1*

**RIBEIRÃO PRETO
2012**

RENATA DOS SANTOS ALMEIDA

*Similaridades entre o transcriptoma humano e murino focando
genes situados em regiões de susceptibilidade ao diabetes mellitus
do tipo 1*

Dissertação apresentada à 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 Aleixo
da Silva Passos Júnior

RIBEIRÃO PRETO
2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTES TRABALHOS, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Almeida, Renata dos Santos

Similaridades entre o transcriptoma humano e murino focando genes situados em regiões de susceptibilidade ao diabetes *mellitus* do tipo 1.

Ribeirão Preto, 2012.

95p.

Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Área de concentração: Genética.

Orientador: Passos, Geraldo Aleixo

1. Diabetes do tipo 1. 2. *Microarrays*. 3. Expressão gênica. 3. Bioinformática. 4. Genes de susceptibilidade. 5. Homólogos.

FOLHA DE APROVAÇÃO

RENATA DOS SANTOS ALMEIDA

Similaridades entre o transcriptoma humano e murino focando genes situados em regiões de susceptibilidade ao diabetes *mellitus* do tipo 1

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências.
Área de Concentração: Genética

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Aos meus queridos pais, José Almeida e Maria do Carmo, pelo exemplo de força e coragem, que sempre me inspirou a lutar pelos meus ideais; pela compreensão, confiança e estímulo a mim conferidos para a concretização de mais uma jornada educacional.

AGRADECIMENTOS

A Deus, por ter me dado força, coragem e capacidade para a realização deste trabalho.

Aos meus pais, José Almeida e Maria do Carmo, pelo estímulo permanente, valores morais e educacionais a mim conferidos, sem os quais eu não teria chegado até aqui. Pelo amor incondicional e conselhos proporcionados, permitindo que meus sonhos se tornassem realidade. Sem vocês, nada disso seria possível. Aos meus irmãos Felipe e Fernanda, amigos e companheiros, pelo incentivo e confiança em minhas realizações. Espero repor os preciosos momentos que foram subtraídos de todos vocês.

Ao Prof. Dr. Geraldo Passos pela orientação, paciência e oportunidade de trabalhar no Laboratório de Imunogenética Molecular, e por ter acreditado no meu crescimento profissional, contribuindo para minha formação científica.

Aos pesquisadores participantes do projeto temático FAPESP, ao qual este trabalho está inserido: Profa. Dra. Elza Tiemi Sakamoto-Hojo e Prof. Dr. Eduardo Antônio Donadi.

À Adriane Feijó pela amizade, apoio e ensinamentos constantes, por ter compartilhado muitos de seus conhecimentos sobre a análise de *microarrays*, importantes no meu aprendizado, e também pela significativa contribuição na realização deste trabalho.

Às amigas Nathalia Bispo César e Thaís Arns pela ajuda e incentivo contínuos, discussões de papers, trocas de idéias e contribuições neste trabalho. Foram ótimos os momentos vividos durante este mestrado.

Às colegas do Laboratório de Imunogenética Molecular Amanda Assis, Claudia Macedo, Cristhiana Collares, Erna Domingues, Flávia Porto, Janaína Dernowsek, Juliana Massaro, Nicole Pezzi, Paula Barbim e Thaís Fornari pelo apoio e convivência durante o mestrado. Agradeço especialmente à Juliana, Chris e Amanda pelos incentivos, sugestões e amizade.

Aos colegas do Laboratório de Mutagênese, Danilo Jordão Xavier, Paula Takahashi, Fernanda Caetano e Leonardo Franchi pela alegria e receptividade. Em especial, a Danilo e Paulinha que sempre compartilharam idéias e me auxiliaram quando necessário.

Ao Programa de Pós-graduação em Genética da Faculdade de Medicina de Ribeirão Preto – USP, especialmente ao coordenador, Prof. Dr. Ademilson Spencer Egea Soares e à vice-coordenadora, Profa. Dra. Silvana Giuliatti. Agradeço aos professores pela contribuição em minha formação científica e aos colegas de pós-graduação, e à Secretaria do Departamento, em especial às secretárias Susie Nalon, Silvia Consiglieri e Maria Aparecida Oliveira pelo apoio dispensado durante o mestrado.

À FAPESP pelo apoio financeiro necessário para a realização deste trabalho e a CAPES pela bolsa concedida.

Aos membros da banca examinadora pelas contribuições, críticas e sugestões.

Às amigas do Laboratório de Genética e Biologia Molecular de Fungos, Gabriela Persinoti, Elza Lang e Maíra Pompeu pela feliz convivência durante o mestrado.

Agradeço particularmente a Larissa Gomes e Nalu Peres que com apoio e força constantes, estiveram sempre presentes. Todos os momentos vividos durante esses anos estarão sempre em minha memória.

A todos meus familiares, pelo apoio e presença em minha vida, mesmo à distância. Agradeço especialmete à minha prima e amiga Juliana Santos pelo estímulo e suporte durante a realização deste trabalho.

A todos aqueles que, mesmo não tendo sido mencionados, contribuíram para a realização deste trabalho durante esta jornada educacional.

Muito Obrigada.
Renata Almeida.

“Quando a gente acha que tem todas as respostas, vem a vida e muda todas as perguntas...”

Luís Fernando Veríssimo

RESUMO

ALMEIDA, R. S. **Similaridades entre o transcriptoma humano e murino focando genes situados em regiões de susceptibilidade ao diabetes *mellitus* do tipo 1.** 2012. 95p. Dissertação de Mestrado – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil, 2012.

O diabetes *mellitus* do tipo 1 (DM1) é uma doença autoimune que se desenvolve a partir da ação combinada de múltiplos fatores genéticos e ambientais, sendo caracterizada pela perda seletiva das células β produtoras de insulina nas ilhotas pancreáticas em indivíduos geneticamente susceptíveis. O HLA de classe II, localizado no cromossomo humano 6p21.3, representa uma das regiões genômicas mais importantes associadas ao DM1, embora evidências apontem para a participação de diversos outros loci na susceptibilidade à doença. Essas regiões cromossômicas poderiam apresentar genes funcionalmente ativos com perfis transcricionais semelhantes ao camundongo *Mus musculus*, muito utilizado como modelo animal para o estudo de doenças humanas. Para testar esta hipótese, foi realizada análise dos perfis transcricionais de linfócitos periféricos provenientes de pacientes com DM1 e camundongos NOD (*Non-obese diabetic*) diabéticos, focando os genes situados em regiões de susceptibilidade. Foram utilizados dados de *microarrays* do genoma funcional completo da plataforma Agilent (*Whole genome one-color Agilent 4x44k*) de dezenove pacientes e oito controles, e de camundongos NOD pré-diabéticos e diabéticos. A linhagem NOD foi utilizada no estudo, pois representa um modelo experimental para o estudo do diabetes autoimune e desenvolve a doença espontaneamente. Para a análise dos dados de *microarrays* foi utilizado o *software GeneSpring GX* e os programas *Cluster* e *TreeView*. A modulação transcricional dos genes foi estabelecida comparando-se os pacientes com os controles, e os animais diabéticos com os pré-diabéticos. Os loci de susceptibilidade ao DM1 humano foram definidos utilizando-se uma tabela curada disponível no banco T1DBase (<http://t1dbase.org>) e seus correspondentes murinos, segundo o banco de dados *Homology Maps* (<http://www.ncbi.nlm.nih.gov/projects/homology/maps/>). Foram considerados para o estudo apenas os pares de homólogos com expressão em linfócitos humanos e murinos conforme os dados disponíveis no banco *BioGPS* (<http://biogps.org>). Avaliamos se os genes murinos estavam situados em regiões de susceptibilidade (*Idd*) ao DM1 do camundongo, utilizando-se o T1DBase. Todos os genes humanos selecionados com homologia com camundongo foram mapeados quanto à localização cromossômica, buscando-se por regiões de sintenia entre as duas espécies por meio da ferramenta *Synteny* disponível no banco de dados *Ensembl Genome Browser* (<http://www.ensembl.org/>). Os pares de homólogos situados em regiões sintênicas foram então verificados quanto à similaridade de sequência de DNA e identidade protéica entre humano e camundongo, utilizando-se dados do *HomoloGene* (<http://www.ncbi.nlm.nih.gov/homologene/>). Foram analisados 463 genes humanos, dos quais 73 apresentaram correspondência em camundongo e expressão em linfócitos T. Dos 73 genes identificados, 31 deles apresentaram mesma modulação de *fold change* entre as duas espécies, com 12 mapeados em regiões de susceptibilidade murinas (*Idd*). Dos doze genes em regiões *Idd*, 4 se apresentaram induzidos: APOM (Apom), COL11A2 (Col11a2), HLA-DOB (H2-Ob) e PRR3 (Prr3); e 8 reprimidos: CYP21A2 (Cyp21a1), STK19 (Stk19), PHTF1 (Phtf1), RSBN1 (Rsb1), CDSN (Cdsn), TRIM39 (Trim39), VARS2 (Vars2) e IL21 (Il21). Foram identificados 59 genes em regiões de sintenia humano-camundongo, com 58 apresentando similaridade na sequência de DNA acima de 70% entre as duas espécies, e identidade de sequência de aminoácidos das respectivas proteínas variando de 61,4 a 99,7%. Esses resultados demonstram que a maioria dos genes estudados apresenta conservação funcional, como indicado pelo alto grau de identidade das proteínas. Além disso, evidenciou-se um compartilhamento de perfis de expressão de genes em regiões de susceptibilidade ao DM1 humano e murino, contribuindo para um melhor conhecimento das semelhanças entre o modelo animal (linhagem NOD) e o diabetes autoimune humano.

Palavras-chave: Diabetes do tipo 1, *microarrays*, expressão gênica, bioinformática, genes de susceptibilidade, homólogos.

ABSTRACT

ALMEIDA, R. S. **Similarities between human and mouse transcriptomes focusing genes positioned in type 1 diabetes *mellitus* susceptibility regions.** 2012. 95p. Master's Dissertation – Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2012.

Type 1 diabetes (T1D) is a common autoimmune disease that arises from multiple genetic and environmental risk factors. It is characterized by selective loss of insulin-producing β -cells in the pancreatic islets in genetically susceptible individuals. The HLA class II locus on human chromosome 6p21.3 represents one of the most important genomic regions associated with T1D but there are evidences for the participation of several others along the chromosomes, which also contribute to disease susceptibility. These chromosomal regions may harbor functional genes with transcriptional profiles similar to those presented by the mouse *Mus musculus*, an animal model highly used in researches of human diseases. To test this hypothesis, it was performed a transcriptome profiling analysis of peripheral lymphocytes from T1D patients and diabetic NOD (*Non-obese diabetic*) mice focusing those genes positioned in chromosomal susceptibility regions. To perform this analysis, whole genome one-color Agilent 4x44k microarrays were used from 19 patients and 8 controls and from pre-diabetic and diabetic NOD mice. NOD strain was used since it is a well established mouse model for T1D. *GeneSpring GX* software and *Cluster* and *TreeView* programs were applied for microarray data analysis. The transcriptional modulation was established by comparisons of diabetic patients versus controls and diabetic versus pre-diabetic animals. The human T1D susceptibility loci were defined using a curated human dataset available in T1DBase (<http://t1dbase.org>) and mouse counterparts according to Homology Maps database (<http://www.ncbi.nlm.nih.gov/projects/homology/maps/>). For the present study we considered only homolog gene pairs with expression in human and mouse lymphocytes taking into account data available in *BioGPS* database (<http://biogps.org>). Following these procedures, mouse genes were checked for chromosome location in order to subserve the establishment of syntenic regions. The human-mouse syntenic regions were defined using Synteny tool from Ensembl Genome Browser (<http://www.ensembl.org/>). The homolog gene pairs located in human-mouse syntenic regions were checked for DNA sequence similarity and protein identity according to HomoloGene data (<http://www.ncbi.nlm.nih.gov/homologene/>). The 463 human genes were analyzed with 73 presenting murine counterparts and expression in lymphocytes. From these, 31 genes presented same fold change modulation in both species, with 12 of them mapped in murine diabetes susceptibility regions (*Idd*). The 12 genes in *Idd* regions showed different transcriptional modulations, 4 featured up-regulation: APOM (Apom), COL11A2 (Col11a2), HLA-DOB (H2-Ob) e PRR3 (Prr3); and 8 down-regulation: CYP21A2 (Cyp21a1), STK19 (Stk19), PHTF1 (Phtf1), RSBN1 (Rsb1), CDSN (Cdsn), TRIM39 (Trim39), VARS2 (Vars2) e IL21 (Il21). 51 genes were identified in human-mouse syntenic regions with 58 presenting DNA sequence similarity above 70% and protein identity ranging from 61,4 to 99,7%. These results show that most genes studied present functional conservation, as indicated by the high degree of identity of the proteins. Additionally, it was observed shared expression profiles between human and murine T1D susceptibility regions. These results contribute to a better understanding of similarities between the animal model (NOD mouse strain) and human autoimmune diabetes.

Keywords: Type 1 diabetes, *microarrays*, gene expression, bioinformatics, susceptibility genes, homologs.

LISTA DE ILUSTRAÇÕES

Figura 1 – Etiologia do diabetes do tipo 1: quebra de tolerância imunológica	17
Figura 2 – Modelo da patogênese e história natural do diabetes <i>mellitus</i> do tipo 1	18
Figura 3 – Baixa frequência de variantes e susceptibilidade à doença.....	21
Figura 4 – Loci de susceptibilidade ao DM1 mapeados até o ano 2009 em função da OR para os alelos de risco em cada genes candidato	25
Figura 5 – Visão geral das etapas realizadas em experimentos de <i>microarrays</i> para análise de transcriptomas.....	29
Figura 6 – Representação gráfica das relações de sentença conservada entre os genomas humano e murino.....	31
Figura 7 – Análise da pureza dos linfócitos T CD3 ⁺ periféricos	41
Figura 8 – Densitometria das amostras de RNA isoladas de linfócitos T CD3 ⁺ periféricos murinos	42
Figura 9 – Densitometria da amostra de RNA de um paciente com DM1 isolada de PBMCs	42
Figura 10 – <i>Pipeline</i> das análises dos dados de <i>microarrays</i> humano e murino realizadas para busca de similaridades entre as duas espécies (item 4.9.1)	47
Figura 11 – <i>Pipeline</i> das análises realizadas para a comparação de expressão dos genes humanos e murinos (item 4.9.2)	48
Figura 12 – Diagrama de Venn representando o <i>screening</i> realizado na seleção de genes expressos em linfócitos T localizados em regiões de susceptibilidade ao DM1 comuns entre humanos e camundongos.....	51
Figura 13 – Gráfico ilustrativo gerado pelo banco de dados BioGPS (http://biogps.org) mostrando a representação tecidual do gene TRIM39, com os nomes de cada tecido, órgão e/ou tipos celulares onde esse gene é expresso.....	52
Figura 14 – Comparação dos perfis de expressão gênica entre indivíduos doentes e controles após aplicação de algoritmo de agrupamento hierárquico	57
Figura 15 – Genes identificados em regiões sintênicas nos cromossomos humanos 3 e 10, e 2 e 9 murinos	61
Figura 16 – Genes identificados em regiões sintênicas nos cromossomos humanos 6, 15, 17 e 19, e 9, 11 e 17 murinos.....	62
Figura 17 – Genes identificados em regiões sintênicas nos cromossomos humanos 1 e 19, e 3 e 7 murinos	63
Figura 18 – Genes identificados em regiões sintênicas nos cromossomos humanos 4, 11 e 12, e 3, 5, 7 e 10 murinos	64

LISTA DE TABELAS

Tabela 1 – Dados demográficos dos controles utilizados no estudo	37
Tabela 2 – Principais características clínicas dos pacientes com DM1	38
Tabela 3 – Genes comuns entre humano e camundongo com expressão em linfócitos T.....	53
Tabela 4 – Genes identificados em regiões de sintenia humano/camundongo	58
Tabela 5 – Similaridade de sequência de DNA e identidade de proteína entre os genes humanos e murinos situados em regiões de sintenia	65

SUMÁRIO

1 INTRODUÇÃO	15
1.1 O Diabetes <i>Mellitus</i> do Tipo 1.....	15
1.2 Patogênese do diabetes <i>mellitus</i> do tipo 1	16
1.3 O Papel do Timo no diabetes <i>mellitus</i> do tipo 1	18
1.4 Metodologias Utilizadas no Mapeamento Genético do DM1	20
1.5 Genes Envolvidos com o DM1.....	22
1.5.1 O antígeno leucocitário humano (HLA).....	22
1.5.2 Genes não-MHC	23
1.6 O Modelo Animal <i>Non-Obese Diabetic</i> (NOD).....	25
1.7 Análise de transcriptomas pela tecnologia de <i>microarrays</i>	27
1.8 Comparação de Transcriptomas entre Espécies Distintas	29
 2 HIPÓTESE.....	 33
 3 OBJETIVOS	 35
3.1 Objetivo Geral	35
3.2 Objetivos Específicos	35
 4 MATERIAL E MÉTODOS	 37
4.1 Pacientes com Diabetes <i>mellitus</i> do Tipo 1 (DM-1) e Controles	37
4.2 Linhagem de Camundongos <i>Non-Obese Diabetic</i> (NOD).....	38
4.3 Avaliação Laboratorial do Diabetes Autoimune nos Camundongos NOD.....	39
4.4 Coleta de Sangue e Isolamento de Células Mononucleares do Sangue Periférico (PBMCs) de Pacientes e Controles.....	39
4.5 Isolamento de Linfócitos T CD3 ⁺ Periféricos Murinos e Verificação de Pureza por Citometria de Fluxo.....	40
4.6 Extração de RNA Total e Avaliação da Integridade	41
4.7 <i>Oligo-microarrays</i>	42
4.8 Aquisição das Imagens de Hibridação.....	43
4.9 Análise dos Dados de <i>Microarrays</i>	43
4.9.1 Normalização, agrupamento das amostras, filtragens e <i>fold change</i>	43
4.9.2 Busca por genes humanos situados em regiões de susceptibilidade e comparação com camundongos NOD	45

5 RESULTADOS	50
5.1 Análise dos Dados de <i>Microarrays</i> Humano e Murino.....	50
5.2 Análise de Genes Humanos Presentes em Regiões de Susceptibilidade ao DM1 e Comparação com a Linhagem de Camundongos NOD.....	50
5.3 Análise de Regiões Sintênicas Homem/Camundongo	58
5.4 Análise das sequências de DNA e de proteínas referentes a genes humanos e murinos situados em regiões de sintenia entre as duas espécies	65
 6 DISCUSSÃO	 68
 7 CONCLUSÕES.....	 81
 8 REFERÊNCIAS BIBLIOGRÁFICAS	 83

Introdução

1 INTRODUÇÃO

1.1 O Diabetes *Mellitus* do Tipo 1

O diabetes *mellitus* compreende um grupo de doenças metabólicas que apresenta alterações no metabolismo de carboidratos, gorduras e proteínas, sendo caracterizado por hiperglicemia crônica decorrente de falhas na produção de insulina, na ação desta ou em ambos. Diversos processos patogênicos têm sido relatados como responsáveis por tais alterações, entre eles a destruição das células β pancreáticas e anormalidades que levam à resistência à insulina (ADA, 2011; ALBERTI & ZIMMET, 1998).

O distúrbio de etiologia autoimune decorrente da destruição das células β produtoras de insulina localizadas no pâncreas denomina-se diabetes *mellitus* do tipo 1 (DM1) (ADA, 2011; HALLER *et al.*, 2005). Tais células funcionam como “termostatos” distintos que agem como sensores de glicose para liberar insulina, a fim de manter a homeostase no sangue (BLUESTONE *et al.*, 2010).

O DM1 também pode ocorrer devido a causas desconhecidas, sem histórico de processo autoimune, sendo representado por uma minoria dos pacientes diagnosticados com tal doença. Os indivíduos acometidos por etiologia autoimune ou pela forma idiopática apresentam diferentes graus de deficiência insulínica (ADA, 2011).

Tal carência na produção de insulina acarreta perda de controle dos níveis de glicose no sangue, exigindo a administração exógena do hormônio para sobrevivência do indivíduo (POLYCHRONAKOS & LI, 2011). Mesmo com os tratamentos para reposição da insulina, complicações vasculares, neurológicas e visuais podem se desenvolver a longo prazo, levando a perda na qualidade de vida e incapacidade significativa (BLUESTONE *et al.*, 2010; PLANAS *et al.*, 2010).

O diabetes autoimune acomete principalmente crianças e adolescentes, ainda que se manifeste em indivíduos com idade senil (ADA, 2011).

Segundo estimativas realizadas pela Organização Mundial de Saúde em 2000 o número de indivíduos com diabetes no mundo representa 177 milhões, tendo aumentado significativamente comparado aos anos de 1985 (30 milhões) e 1995 (135 milhões). No geral, o diabetes tem sido considerado uma doença em expansão com projeção para se tornar uma das principais causas de morte e incapacidade no mundo nos próximos vinte e cinco anos (WHO, Fact Sheet N°236).

Com relação ao diabetes do tipo 1 foi observada uma mudança significativa em sua incidência ao longo da segunda metade do século 20, com destaque para o aumento no número de casos diagnosticados em crianças com idade inferior a 5 anos (PATTERSON *et al.*, 2009; GALE, 2002).

Além disso, estudos epidemiológicos mostram que um número cada vez maior de indivíduos têm sido diagnosticados com a doença ao redor do mundo, tendo a Finlândia e a Sardenha na Itália as maiores taxas de incidência: >40 e $37,8/10^5$, respectivamente (BORCHERS *et al.*, 2010). Na Europa, países da região Central e Oriental que apresentavam baixas taxas de incidência revelaram um rápido aumento anual em tais índices entre 1989 e 2003, com projeção para aumento até 2015 (PATTERSON *et al.*, 2009).

Em um estudo realizado por Patterson e colaboradores (2009), estima-se que em 2020 o número de novos casos de DM1 seja 24.400 na Europa, sendo esperado que a incidência em crianças com idade entre 0 e 4 anos duplique entre 2005 e 2020.

No Brasil, em uma pesquisa conduzida por Negrato e colaboradores (2010), foi observada uma incidência anual de $10,4/10^5$ pessoas entre 1986 e 2006 em crianças com idade inferior ou igual a 14 anos na cidade de Bauru, estado de São Paulo. A média de casos registrados aumentou de $2,82/10^5$ em 1987 para $18,49/10^5$ em 2002, o que representa um aumento de 6,56 vezes na população estudada.

O fato da incidência do diabetes tipo 1 ter aumentado significativamente no mundo torna necessário mais esforços das agências de saúde para realização de estudos epidemiológicos, a fim de mapear a extensão da doença e entender sua patogênese para prevenir e melhorar a qualidade de vida, fornecendo tratamento médico mais adequado.

1.2 Patogênese do diabetes *mellitus* do tipo 1

O DM1 é reconhecidamente uma doença autoimune com manifestações clínicas como hiperglicemia, poliúria, polidipsia, cetoacidose e perda de peso, sendo caracterizada histologicamente por insulite e lesões nas ilhotas de Langerhans (HALLER *et al.*, 2005).

Diversos mecanismos imunológicos e genéticos e fatores ambientais são considerados relevantes para a manifestação da doença, entretanto as causas precisas desencadeantes de autoimunidade ainda não foram elucidadas.

Inúmeros processos celulares e moleculares têm sido propostos como importantes para o desenvolvimento do DM1, entre eles: quebra de tolerância central (tolerância a antígenos próprios e eliminação de linfócitos T autorreativos pela seleção negativa no timo) (ZUCHELLI *et al.*,

2005); expressão alterada do complexo principal de histocompatibilidade (MHC); mimetismo molecular (ROEP *et al.*, 2002); proliferação homeostática exagerada de células T autorreativas (KING *et al.*, 2004); migração deletéria de células dendríticas para os linfonodos pancreáticos promovendo a ativação de células T diabetogênicas (TURLEY *et al.*, 2003), entre outros processos. A Figura 1 ilustra alguns mecanismos envolvidos na etiologia do DM1.

Figura 1 – Etiologia do diabetes do tipo 1: quebra de tolerância imunológica. Em um indivíduo com susceptibilidade genética, autoantígenos pancreáticos são processados por células apresentadoras de antígenos (APC) residentes que migram para os linfonodos drenantes do pâncreas e apresentam tais antígenos a linfócitos T CD4⁺ autorreativos. Estes se tornam ativados, sofrem proliferação por expansão clonal e se dirigem para as células β pancreáticas nas ilhotas. Neste local, as células efectoras recrutam os linfócitos CD8⁺ citotóxicos e outras células inflamatórias, o que resulta em insulite destrutiva. As APCs e os linfócitos T produzem citocinas pró-inflamatórias, levando a um maior recrutamento e ativação de células T. Os linfócitos B estão envolvidos na fase inicial da doença, assim como as APCs, e produzem autoanticorpos contra antígenos. MHC – complexo principal de histocompatibilidade. (Adaptado de WALDRON-LYNCH & HEROLD, 2011).

Em relação ao diagnóstico do DM1, além das características clínicas, marcadores da destruição autoimune das células β (como autoanticorpos) têm sido muito utilizados para distinguir este tipo de diabetes das outras formas (PATEL & MACEROLLO, 2010).

Autoanticorpos contra a insulina (IAA), a descarboxilase do ácido glutâmico 65 (GAD65A), células da ilhota (ICA) e a proteína 2 associada a insulinoma (IA-2A) são os mais conhecidos e detectados (HALLER *et al.*, 2005), sendo identificado recentemente anticorpos contra o transportador de zinco 8 (ZnT8A) (WENZLAU *et al.*, 2007).

Embora a presença de anticorpos seja um indicativo de autoimunidade e de progressão para a doença (GRIJSE *et al.*, 2010; BORG *et al.*, 1997), geralmente quando o DM1 é clinicamente detectado grande parte da massa de células β pancreáticas está destruída como constatado pelos baixos níveis plasmáticos de peptídeo C, um indicador da produção endógena de insulina (PATEL & MACEROLLO, 2010; MATHIS & GAGLIA, 2009).

Isto sugere que os processos iniciais de autoimunidade já ocorreram, o que implica em um número limitado de tratamentos disponíveis para reversão da doença, tornando necessária a realização de mais pesquisas nessa área, a fim de identificar marcadores que revelem as etapas principais do DM1 (MATHIS & GAGLIA, 2009). A Figura 2 mostra as principais fases relacionadas ao desenvolvimento do DM1.

Figura 2 – Modelo da patogênese e história natural do diabetes mellitus do tipo 1. Modelo moderno que atualiza o tradicional pela inclusão de informações adquiridas com o avanço no entendimento dos papéis exercidos pela genética, imunologia e ambiente na história natural do DM1. IAA – autoanticorpos contra insulina; GADA – autoanticorpos contra GAD; IA-2A – autoanticorpos contra IA-2; ICA – autoanticorpos contra células da ilhota; FPIR – primeira fase de resposta à insulina; IVGTT – teste intravenoso de tolerância à glicose. (Modificado de HALLER *et al.*, 2005).

1.3 O Papel do Timo no diabetes mellitus do tipo 1

A discriminação do próprio/não-próprio é uma propriedade essencial do sistema imune que dirige uma variedade de mecanismos efetores contra agentes patogênicos, enquanto ignora os constituintes próprios do organismo. Somente quando a tolerância ao próprio está finamente balanceada é que a integridade do organismo é garantida (KYEWSKI & DERBINSKI, 2004).

O timo é o principal sítio de desenvolvimento das células T e considerado de fundamental importância para a tolerância do sistema imune ao próprio – tolerância central. Perturbações na função tímica podem levar à perda de tolerância e autoimunidade agressiva (NOWELL *et al.*, 2007).

O timo maduro é um ambiente celular altamente dinâmico, apresentando linfócitos T em desenvolvimento (timócitos) e uma variedade de células do estroma, incluindo fibroblastos, células endoteliais, células dendríticas e um epitélio especializado, entre outras (NOWELL *et al.*, 2007; BLACKBURN & MANLEY, 2004).

O estroma tímico apresenta duas regiões principais definidas histologicamente: o córtex e a medula, que exibem diferentes tipos de células epiteliais com características morfológicas e fenotípicas distintas (NOWELL *et al.*, 2007; BLACKBURN & MANLEY, 2004).

No córtex, ocorre a chamada seleção positiva, processo caracterizado pelo enriquecimento de timócitos potencialmente reativos a antígenos estranhos apresentados pelas moléculas do complexo principal de histocompatibilidade próprio. As células epiteliais corticais (*Cortical thymic epithelial cells – cTECs*) são as principais responsáveis por esse processo. Os linfócitos T em desenvolvimento bem-sucedidos nesta etapa são direcionados à medula tímica (TAKAHAMA, 2006).

No compartimento medular, os timócitos interagem com as células medulares tímicas (*Medullary thymic epithelial cells – mTECs*) que exibem expressão gênica promíscua (*Promiscuous Gene Expression – PGE*), fenômeno caracterizado pela expressão de antígenos tecido-específicos (*Tissue-specific antigens – TSAs*) com consequente exposição dos peptídeos próprios pelo MHC, o que contribui para a tolerância das células T ao próprio (MAGALHÃES *et al.*, 2006; KYEWSKI & DERBINSKI, 2004).

A PGE apresentada pelas mTECs a seleção negativa dos linfócitos autorreativos, onde os timócitos que reagem fortemente ao MHC/peptídeo próprio são eliminados por apoptose e aqueles que reagem fracamente, são direcionados para a corrente sanguínea ou vasos linfáticos para desempenharem suas funções na periferia. Tal processo contribui para a deleção de células T autorreativas, evitando que reações de autoimunidade sejam desencadeadas no organismo (TAKAHAMA, 2006).

A PGE está em parte sob o controle do fator de transcrição Aire (do inglês *autoimmune regulator*). Mutações no gene Aire causam um distúrbio autoimune sistêmico conhecido como síndrome APECED (do inglês *Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy*) (OMIM: 240300) em que os indivíduos afetados podem apresentar além de outras anomalias o diabetes do tipo 1 (KLUGER *et al.*, 2012).

Segundo Kishimoto e Sprent (2001), os camundongos NOD (*Non-obese diabetic*, linhagem geneticamente susceptível ao DM1 apresentam tolerância central defectiva, o que contribuiria para o surgimento de diabetes autoimune nesses animais.

Em um estudo realizado por Fornari e colaboradores (2010) com timo de camundongos NOD, foi observado que o gene Aire estava reprimido nos animais jovens pré-diabéticos e que genes codificantes de autoantígenos específicos (TSAs) relacionados ao DM1 também estavam reprimidos, sugerindo que a desregulação cronológica de genes importantes na seleção negativa podem estar associados ao desenvolvimento de DM1 em camundongos.

Isso demonstra que além dos estudos direcionados a genes de susceptibilidade ao DM1, é importante considerar a investigação dos mecanismos envolvidos na indução de tolerância central.

1.4 Metodologias Utilizadas no Mapeamento Genético do DM1

Por ser uma doença complexa, o DM1 envolve a participação combinada de múltiplos fatores genéticos e ambientais, com o número e ação dos genes atuantes na patologia ainda não estão completamente determinados (CONCANNON *et al.*, 2009). Entretanto, é reconhecido que a susceptibilidade genética desempenha um papel importante na doença como demonstrado pela concordância excedente a 50% entre gêmeos monozigóticos com até 60 anos de idade (REDONDO *et al.*, 2008), comparado aos dizigóticos com uma taxa de 3,8% em indivíduos finlandeses (HYTTINEN *et al.*, 2003).

Os primeiros estudos genéticos em DM1 ocorreram na primeira metade da década de 1970 (CUDWORTH & WOODROW, 1975; NERUP *et al.*, 1974; SINGAL & BLAJCHMAN, 1973), indicando a associação da região do antígeno leucocitário humano (*Human leukocyte antigen – HLA*) à doença.

Cerca de dez anos após tais achados, cientistas evidenciaram a participação de outro locus na susceptibilidade genética ao DM1: a região do gene da insulina (INS) no braço curto do cromossomo 11 (BELL *et al.*, 1984), atualmente considerado o segundo maior contribuinte para o risco da doença de acordo com dados de GWAS (POCIOT *et al.*, 2010), sendo o maior fator de risco associado ao HLA (POLYCHRONAKOS & LI, 2011).

Outros estudos realizados nas últimas décadas buscaram determinar quais alelos HLA são responsáveis pela associação com o DM1, além de vários outros loci genéticos, fora desta região e daquela do gene da insulina, que devem contribuir para o risco da doença (NOBLE & ERLICH, 2012).

A maioria deles foi conduzida baseando-se em três metodologias: os estudos de ligação (famílias ou apenas pares de irmãos afetados); os de genes candidatos e aqueles que abrangem o genoma de maneira global (*Genome-wide association studies – GWAS*) (CONCANNON *et al.*, 2009; INVERNAZZI & GERSHWIN, 2009).

A análise de ligação é baseada na transmissão de dois ou mais genes ligados em um cromossomo para a mesma célula filha (loci genéticos herdados juntos) com associação de um traço fenotípico observado em famílias (INVERNAZZI & GERSHWIN, 2009).

Os estudos de ligação têm obtido sucesso no mapeamento de genes e variantes que apresentam alta penetrância na doença como em distúrbios monogênicos, entretanto como o DM1 não apresenta padrão de herança mendeliana e as variantes causais contribuem individualmente pouco para o risco da doença, essa abordagem se torna limitada na identificação de loci de susceptibilidade (VISSCHER *et al.*, 2012; CONCANNON *et al.*, 2009). A Figura 3 ilustra o risco relativo em função da frequência de alelos em diversas doenças, desde as monogênicas até as multifatoriais.

Figura 3 – Baixa frequência de variantes e susceptibilidade à doença. (Modificado de MCCARTHY *et al.*, 2008).

Há vários anos os estudos de genes candidatos têm sido largamente realizados por permitirem a identificação de genes e variantes comuns envolvidos em doenças e características complexas, sendo considerada uma estratégia complementar à análise de ligação (HIRSCHHORN & DALY, 2005).

Tal abordagem se baseia no estudo de genes selecionados a partir de sua localização em regiões de ligação descritas ou de evidências que demonstram sua contribuição para o risco da doença, buscando-se determinar se eles estão pouco ou muito representados nos casos da doença em comparação a controles e se apresentam polimorfismos dentro ou próximos a eles que provoquem alterações funcionais (HIRSCHHORN & DALY, 2005; TABOR *et al.*, 2002).

Entretanto, quando os mecanismos patofisiológicos subjacentes às doenças não são claramente conhecidos ou o objetivo é expandir a busca por variantes causais sem conhecimento prévio de um gene específico, a presente metodologia se torna inadequada no *screening* de outros componentes genéticos envolvidos na doença (HIRSCHHORN & DALY, 2005).

Assim, os estudos de associação global do genoma (GWAS) preenchem essa lacuna, pois não requerem informações prévias de associações genéticas e permitem a investigação de genomas em níveis de resolução anteriormente indisponíveis (VISSCHER *et al.*, 2012; PEARSON & MANOLIO, 2008).

Neste método de análise, centenas de milhares de variantes genéticas são genotipadas simultaneamente utilizando *chips* de SNPs (*Single nucleotide polymorphism* – *SNP*) que cobrem suficientemente o genoma, buscando relacionar alelos de genes a doenças ou características quantitativas (PEARSON & MANOLIO, 2008).

Os GWASs se apoiam no princípio do desequilíbrio de ligação (*Linkage disequilibrium* – *LD*) em nível populacional (VISSCHER *et al.*, 2012), definido como a associação não aleatória entre alelos de loci adjacentes (ARDLIE *et al.*, 2002).

Todos esses métodos juntamente com outros tipos de análise, como a análise de expressão gênica por PCR em tempo real e *microarrays*, contribuíram significativamente para um melhor entendimento dos mecanismos genéticos envolvidos no DM1. A seguir serão detalhados alguns fatores genéticos identificados ao longo das últimas décadas com um papel relevante na etiologia da doença.

1.5 Genes Envolvidos com o DM1

1.5.1 O antígeno leucocitário humano (HLA)

O HLA compreende um grupo de moléculas codificadas pelos genes do complexo principal de histocompatibilidade (*Major histocompatibility complex* – *MHC*) localizados no cromossomo 6p21.3. Tais moléculas são responsáveis pela apresentação de antígenos às

células do sistema imune, como os linfócitos T CD4⁺ e CD8⁺, sendo classificados em três grupos: MHC de classe I, MHC de classe II e MHC de classe III (TROWSDALE, 2002).

Os genes HLA são altamente polimórficos, sendo sua diversidade genética responsável pela capacidade do sistema imunológico em responder a quase todos os patógenos que entram em contato com o organismo humano e murino.

A região codificante desses genes contribui com cerca de 50% do risco genético associado ao diabetes autoimune, com a maior parte da susceptibilidade genética atribuída aos loci DR e DQ que pertencem à classe II (MEHERS & GILLESPIE, 2008), sendo os haplótipos DRB1*04-DQB1*0302 e DRB1*03 encontrados em cerca de 90% dos casos da doença (TODD, 2010).

Em uma pesquisa conduzida por Erlich e colaboradores para o Consórcio Internacional de Diabetes do Tipo 1 (2008) em 644 famílias com indivíduos afetados pelo DM1, foi observada uma razão de chances (*Odds ratio* – OR) de 8,39 para o haplótipo DRB1*0401-DQA1*03(01, 02 ou 03)-DQB1*0302 e 0,3 para DRB1*0401-DQA1*03(01, 02 ou 03)-DQB1*0301, demonstrando que o alelo DQB1*0302 é um relevante alelo de susceptibilidade.

A OR é uma medida de risco relativo utilizada em estudos de caso-controle que pode ser definida como a probabilidade de exposição a uma variante genética de susceptibilidade em casos comparada aquela em controles (HIRSCHHORN & DALY, 2005; WANG *et al.*, 2008).

O estudo citado também constatou que os alelos DRB1 conferem risco à doença, uma vez que haplótipo o DRB1*0403-DQA1*0301-DQB1*0302 apresenta uma OR de 0,27, enquanto o DRB1*0401-DQA1*03(01, 02 ou 03)-DQB1*0302, uma OR de 8,39.

Tais achados demonstram ainda uma possível interação genética entre os alelos DRB1 e DQB1, revelando a importância de alelos específicos associados ao haplótipo na determinação do risco relativo para a doença (ERLICH *et al.*, 2008).

Outros genes do HLA, como HLA-A e -B (NEJENTSEV *et al.*, 2007), também foram associados ao DM1. Além da região do HLA, outros loci contribuem de maneira importante para o desenvolvimento do DM1, embora em menor grau, correspondendo aos 50% restantes da susceptibilidade genética associada à doença (MEHERS & GILLESPIE, 2008).

1.5.2 Genes não-MHC

A insulina é considerada o autoantígeno primário no DM1 humano e em camundongos NOD, com seu gene contribuindo significativamente para o risco da doença (WALDRON-LYNCH & HEROLD, 2011).

VNTRs (número variável de repetições *em tandem*) localizados próximo ao promotor do gene da insulina (INS) têm sido associados à doença (DURINOVIC-BELLÓ *et al.*, 2010; WALTER *et al.*, 2003).

Foi observado que o VNTR de classe I composto por 26 a 63 unidades de repetição está envolvido na expressão alterada da insulina no timo humano, indicando que baixos níveis de expressão do gene podem estar envolvidos na falha de indução de tolerância ao autoantígeno, o que culminaria na liberação de linfócitos T autorreativos do timo para a periferia (PUGLIESE *et al.*, 1997; VAFIADIS *et al.*, 1997).

Depois da insulina, o gene PTPN22 (situado no cromossomo 1p13.2) é o que confere maior risco relativo à doença, apresentando uma OR muito próxima a apresentada pelo gene da insulina (POCIOT *et al.*, 2010). O PTPN22 codifica uma tirosina fosfatase linfóide (LYP) que participa da regulação negativa de sinais derivados do receptor de célula T (TCR), atuando como supressor da ativação de linfócitos T (CONCANNON *et al.*, 2009; BOTTINI *et al.*, 2004).

Um SNP no nucleotídeo 1858(C>T) do gene PTPN22 que gera uma mudança no aminoácido codificado (Arg620Trp) foi associado ao DM1 (BOTTINI *et al.*, 2004) e replicado (SMYTH *et al.*, 2004), e certamente leva a uma alteração no mecanismo responsável pela ativação de linfócitos T.

Além de PTPN22 e INS, os loci CTLA4 no cromossomo 2q33 e IL2RA no 10p15 também foram associados em vários estudos ao diabetes do tipo 1 (QU *et al.*, 2009a; QU *et al.*, 2009b; LOWE *et al.*, 2007; NISTICÒ *et al.*, 1996).

Com o surgimento dos GWAS, tornou-se possível a identificação de inúmeros outros loci (BARRET *et al.*, 2009; HAKONARSON *et al.*, 2008; HAKONARSON *et al.*, 2007; SMYTH *et al.*, 2006), antes desconhecidos, com uma maior acurácia, sendo atualmente aceito que cerca de 50 loci genéticos fora do MHC contribuem para o risco do diabetes autoimune (POCIOT *et al.*, 2010) (<http://www.t1dbase.org>). A Figura 4 mostra alguns genes não-HLA mapeados como loci de susceptibilidade ao DM1.

Grande parte das regiões associadas por GWAS atingem megabases de tamanho e apresentam inúmeros genes cujo papel na doença permanece desconhecido, uma vez que, geralmente, nos intervalos associados apenas um ou mais genes têm sido inferidos como candidatos para o DM1. Embora tais genes não sejam considerados candidatos, eles poderiam apresentar alterações funcionais devido a polimorfismos, podendo contribuir para o risco ou desenvolvimento da doença (TODD, 2010).

Outro aspecto que deve ser mencionado é o fato de tais estratégias não revelarem como as regiões genômicas associadas participam da etiopatogenia do DM1, sendo importante considerar se elas são ou não transcritas em RNAs mensageiros (mRNAs).

Assim, a combinação de estudos de expressão gênica com a localização genômica dos loci previamente identificados abre perspectivas para a descoberta de genes em loci de características quantitativas (*Quantitative trait loci – QTL*) (MELANITOU, 2005).

Figura 4 – Loci de susceptibilidade ao DM1 mapeados até o ano 2009 em função da OR para os alelos de risco em cada gene candidato. As cores representam o ano de descoberta dos genes candidatos. * Genes não-MHC expressos nas ilhotas pancreáticas humanas. (Adaptado de POCIOT *et al.*, 2010).

1.6 O Modelo Animal *Non-Obese Diabetic* (NOD)

A linhagem de camundongos NOD representa o modelo animal cujo desenvolvimento do diabetes autoimune é o mais próximo ao DM1 humano. Tal linhagem é isogênica e foi desenvolvida no Laboratório Shionogi (Japão), durante um programa de cruzamento endogâmico para estabelecer uma sublinhagem susceptível à catarata (*Cataract and small eyes – CTS*) a partir de camundongos *outbred* JCl:ICR, uma colônia derivada de camundongos *Swiss* (TANIGUCHI *et al.*, 2007; MAKINO *et al.*, 1980).

Durante essa seleção da linhagem CTS, foi observado o primeiro caso de diabetes dependente de insulina (*Insulin-dependent diabetes – Idd*) em uma fêmea da vigésima geração de cruzamentos, sem susceptibilidade à catarata. Tal sublinhagem foi submetida a uma série de endocruzamentos até que camundongos desenvolvendo a doença espontaneamente foram estabelecidos – linhagem *non-obese diabetic* (NOD), com uma incidência cumulativa de diabetes em torno de 60-80% em fêmeas e menos de 20% em machos, ambos com 30 semanas de idade (TANIGUCHI *et al.*, 2007; LEITER *et al.*, 1987).

Assim como no DM1 humano, os camundongos NOD diabéticos apresentam polidipsia, poliúria, perda de peso, glicosúria, hiperglicemia e deficiência de insulina (KIKUTANI & MAKINO, 1992).

A incidência da doença é maior em animais mantidos em condições livres de patógenos específicos (*Specific pathogen-free – SPF*), fenômeno cuja explicação precisa permanece desconhecida, embora seja sugerido que a exposição a antígenos exógenos exerça um ajuste fino no sistema imune e, assim, proteja contra alergias, distúrbios autoimunes e outras doenças relacionadas ao desequilíbrio do sistema imunológico (OKADA *et al.*, 2010; ANDERSON & BLUESTONE, 2005).

Muito dos mecanismos patogênicos envolvidos no DM1 foram melhor compreendidos devido a utilização dessa linhagem e suas variações, como por exemplo, a importância das células dendríticas e dos macrófagos no início da doença, e a contribuição dos linfócitos T CD4⁺ e CD8⁺ na destruição das células β produtoras de insulina, sendo observado uma infiltração gradual de tais células no pâncreas, com surgimento de peri-insulite e progressão para insulite severa, culminando no diabetes clínico (LEHUEN *et al.*, 2010; DELOVITCH & SINGH, 1997).

Além disso, assim como em humanos, os camundongos NOD, apresentam linfócitos T autorreativos à insulina, GAD65, IA-2, entre outros, exibindo também anticorpos contra tais proteínas (ANDERSON & BLUESTONE, 2005).

O diabetes autoimune murino também é uma doença poligênica, com a participação de mais de 30 loci genéticos, identificados por diversas metodologias, entre elas a análise de ligação. Algumas dessas regiões de susceptibilidade (designadas *Idd*) foram associadas à doença em linhagens congênicas, conferindo resistência ao DM1 (DRIVER *et al.*, 2011).

Apesar do enorme número de regiões associadas, apenas alguns genes tiveram seus papéis descritos na patogenia do DM1, com a maior contribuição genética atribuída a alelos do MHC. Além desses genes, o gene da insulina (*Ins2*), o *Ptpn22* e o *Ctla4* são

alguns loci de susceptibilidade que também foram identificados no DM1 murino (THAYER *et al.*, 2010).

Várias regiões de susceptibilidade murinas apresentam sintenia com as humanas, como por exemplo, a *Idd1* no cromossomo 17 com a região 6p21 humana; a *Idd3* no cromossomo 3 com a região 4q27 humana; a *Idd5.1* no cromossomo 1 com a região 2q33.2 humana, entre outras (<http://www.tlbase.org.br>) (DRIVER *et al.*, 2011).

Com o conhecimento de que regiões de susceptibilidade são compartilhadas entre o DM1 humano e murino, é importante considerar se elas são ou não transcricionalmente ativas nas duas espécies e de que forma contribuem funcionalmente para o risco da doença, e assim conhecer melhor as semelhanças entre o modelo animal e o DM1 humano, e prever se estratégias terapêuticas baseadas na linhagem NOD são aplicáveis na espécie humana.

Assim, nossa visão é que ao compararmos perfis de expressão gênica transcricional de pacientes DM1 com camundongos NOD, poderemos descrever regiões sintênicas homem-camundongo implicadas no DM1. Para este estudo, optamos pela tecnologia dos *microarrays* com a qual podemos ganhar novos dados neste sentido.

1.7 Análise de transcriptomas pela tecnologia de *microarrays*

O transcriptoma compreende o conjunto completo de RNAs expressos em uma célula ou tecido em uma dada situação, incluindo transcritos alternativos e não-codificantes, cujas alterações refletem as condições fisiológicas ou patológicas em que o organismo se encontra (RUAN *et al.*, 2004).

A análise global dos níveis de expressão gênica transcricional representa uma poderosa ferramenta na investigação do comportamento dos genes em condições específicas, como um estágio do desenvolvimento, a exposição a drogas ou condições patológicas, incluindo o câncer e as doenças autoimunes. A tecnologia de *microarrays* é uma das ferramentas mais utilizadas para esse fim.

Esta metodologia permite a análise simultânea de milhares de transcritos em um único experimento, viabilizando a comparação de diferentes dados de expressão gênica entre duas amostras de RNA, como por exemplo, de um controle e de um paciente. Os *microarrays* se baseiam na ligação por complementaridade de bases de RNAs fita simples aos fragmentos de DNA, RNA ou oligonucleotídeos depositados em uma lâmina, (BRAZMA & VILO, 2001). A Figura 5 ilustra o princípio geral da técnica.

É importante notar que mais do que a descoberta de novos transcritos e/ou novos genes, o que é possibilitado pelo sequenciamento de DNA/RNA, a técnica de *microarrays* possibilita a descoberta de novas funções de genes já conhecidos. E é nesse sentido que trabalhamos.

A análise de transcriptomas complementa as metodologias genéticas baseadas em associação por DNA, utilizadas no mapeamento de genes envolvidos em doenças autoimunes, incluindo o DM1, uma vez que pode revelar alterações na regulação de genes e processos biológicos envolvidos na doença e apresentados em diversos órgãos e tecidos, como o sangue, baço, pâncreas e linfonodos pancreáticos (PLANAS *et al.*, 2010).

De fato, pesquisas conduzidas em células mononucleares do sangue periférico (PBMCs) e em sangue total, empregando a tecnologia de *microarrays*, têm demonstrado padrões distintos de expressão gênica entre indivíduos pré-diabéticos e diabéticos, com a identificação de genes regulados por interferon (IFN), vias biológicas relacionadas ao metabolismo celular e outros processos, associadas ao diabetes autoimune (REYNIER *et al.*, 2010; WANG *et al.*, 2008; KAIZER *et al.*, 2007).

Em relação ao modelo animal da doença, um estudo realizado por Irie e colaboradores (2008) em linfócitos T CD4⁺ do baço de camundongos NOD e NOD.*Idd3/5* identificou dois genes como novos candidatos para o DM1: Cd55 e Acadl. Tais loci de susceptibilidade foram revelados através da análise global da expressão gênica por *microarrays*.

Nosso grupo vem contribuindo nesta área há anos com trabalhos mostrando que certos loci de susceptibilidade ao lúpus eritematoso sistêmico (TREVISAN *et al.*, 2006; PEREIRA *et al.*, 2004), artrite reumatóide ou diabetes *mellitus* do tipo 1 (RASSI *et al.*, 2008) são regiões cromossômicas que transcrevem RNAs mensageiros.

Isso demonstra que a análise de regiões associadas por DNA em combinação com experimentos de *microarrays* é uma estratégia relevante que pode atribuir funções a genes que ainda não haviam sido associados a doenças autoimunes, incluindo o DM1.

Figura 5 – Visão geral das etapas realizadas em experimentos de *microarrays* de cDNA para análise de transcriptomas. Montagem realizada a partir de:
http://www.microarray.lu/en/MICROARRAY_Overview.shtml
<http://en.wikipedia.org/wiki/File:Heatmap.png>

1.8 Comparação de Transcriptomas entre Espécies Distintas

Os estudos comparativos do perfil transcricional de diferentes espécies têm sido propostos para a investigação dos mecanismos moleculares envolvidos na evolução de transcriptomas e também sua influência em caracteres morfológicos (ROMERO *et al.*, 2012).

Durante as últimas três décadas, as pesquisas na área de evolução molecular focaram essencialmente nos padrões de sequência de DNA e proteínas para a inferência de relações filogenéticas entre espécies. Só recentemente pôde-se estudar a evolução dos padrões gerais de expressão gênica devido ao surgimento das tecnologias para análise de transcriptomas (LIAO & ZHANG, 2006).

A comparação dos níveis globais de expressão gênica entre espécies fornece informações sobre a evolução dos processos de regulação gênica, incluindo as pressões

evolutivas e a extensão das mudanças proporcionadas pelos diferentes mecanismos associados à variação regulatória (ROMERO *et al.*, 2012).

Em um estudo realizado por Zheng-Bradley e colaboradores (2010), foi observado que dados de expressão gênica humana e murina provenientes de amostras de diferentes tecidos apresentaram mesmo padrão de agrupamento em uma análise de componente principal (PCA). Em ambas as espécies, as amostras analisadas agruparam predominantemente em quatro categorias: amostras de linhagens celulares, músculo/coração, sistema nervoso e amostras de fígado, revelando que os padrões globais de expressão gênica são conservados evolutivamente entre humanos e camundongos.

Outra pesquisa realizada com estas duas espécies demonstrou que os níveis de expressão gênica de pares de ortólogos estão correlacionados com os níveis de divergência na sequência de DNA, enquanto que os perfis transcricionais tecido-específicos não apresentam tal correlação, demonstrando que genes altamente expressos ou expressos em um grande número de tecidos evoluem mais lentamente (JORDAN *et al.*, 2005).

Ortólogos são genes presentes em diferentes espécies, que evoluíram de um gene ancestral comum por especiação e geralmente mantêm uma função similar durante o curso da evolução (GRIGORYEV *et al.*, 2004).

Outro aspecto que precisa ser considerado neste tipo de abordagem é a observação do comportamento transcricional de regiões sintênicas entre duas espécies. O conhecimento das divergências de expressão entre humanos e camundongos fornece informações sobre a importância dos genes em processos biológicos essenciais para os organismos e que foram mantidos ao longo da evolução, além de ser interessante para a comunidade científica que utiliza modelos animais no estudo de doenças humanas (LIAO & ZHANG, 2006).

O termo sintenia se refere a loci gênicos situados em um mesmo cromossomo sem levar em consideração se eles foram ou não estabelecidos em ligação genética pela análise clássica de ligação (PASSARGE *et al.*, 1999).

A sintenia conservada corresponde a uma região em que grupos de genes encontram-se localizados em um cromossomo de maneira conservada, sendo observada tal organização em diferentes espécies sem a necessidade dos genes estarem em ordem precisa (GREGORY *et al.*, 2002).

A visualização de sintenia conservada demonstra que as relações entre diferentes genomas são complexas, com a região de um cromossomo em uma espécie podendo corresponder a vários segmentos em diferentes cromossomos de outra espécie e vice-versa (HUNT *et al.*, 2004).

Diversos mecanismos genéticos têm sido implicados na evolução de cromossomos, como translocações, inversões, fissões, fusões, duplicações e deleções de segmentos, todos contribuindo com o processo de especiação (RIESEBERG, 2001).

Segundo Church e colaboradores (2009), mais de 80% dos genes humanos apresentam ortólogos murinos e segmentos genômicos atingindo mais de 300 Kb (kibobases) de tamanho estão em sintenia conservada com o genoma do camundongo. A Figura 6 é uma representação gráfica das relações de sintenia conservada entre as espécies humana e murina.

Figura 6 – Representação gráfica das relações de sintenia conservada entre os genomas humano e murino.

As cores indicam os cromossomos murinos, com as linhas representando a orientação dos segmentos (parte superior esquerda para parte inferior à direita é direito, enquanto direito de cima para baixo é invertido). As regiões em branco não foram cobertas pelos alinhamentos, formando um segmento ≥ 300 kb. Os triângulos vermelhos correspondem aos centrômeros humanos. (Adaptado de CHURCH *et al.*, 2009).

Todos esses avanços no mapeamento de genes e na avaliação de transcriptomas foram permitidos devido ao desenvolvimento de programas de bioinformática capazes de analisar com sucesso o crescente número de dados gerados pelo sequenciamento de genomas e experimentos de *microarrays*. Além disso, as ferramentas de bioinformática acopladas à análise de regiões sintênicas e sua integração com dados de expressão gênica em larga escala fornecerá um maior conhecimento das relações genéticas entre humano e camundongo.

Nosso grupo já tem experiência nessa área, descrevendo regiões sintênicas transcricionalmente ativas durante a diferenciação osteoblástica de células-tronco mesenquimais adultas humanas e murinas (DERNOWSEK, 2010).

Diante disso, no presente trabalho propomos avaliar a modulação transcricional em regiões de susceptibilidade ao DM1, comparando-se as espécies humana e murina. Para isto, utilizamos a tecnologia de *microarrays* e ferramentas de bioinformática para análise de dados do transcriptoma.

Hipótese

2 HIPÓTESE

Devido à colinearidade de genomas que ocorre entre humanos e camundongos, esperamos encontrar um perfil de expressão transcricional próximo entre as duas espécies focando em genes presentes em regiões sintênicas de susceptibilidade ao diabetes *mellitus* do tipo 1.

Objetivos

3 OBJETIVOS

3.1 Objetivo Geral

O objetivo geral deste trabalho foi descrever o perfil de expressão transcricional de genes localizados em regiões cromossômicas de susceptibilidade ao diabetes *mellitus* do tipo 1 (DM1) humano e em regiões correspondentes no camundongo (*Mus musculus*).

3.2 Objetivos Específicos

- Avaliar o perfil de expressão gênica transcricional em linfócitos periféricos de pacientes diabéticos e de camundongos NOD, focando os genes situados nos *loci* de susceptibilidade humanos através do uso da tecnologia de *microarrays*;
- Fazer uso de programas de bioinformática dedicados à análise dos dados de *microarrays*: o pacote de programas *GeneSpring GX* para estabelecer os perfis de expressão gênica; o ambiente R para realizar ajustes nas matrizes de expressão e os programas *Cluster* e *TreeView* para a análise de agrupamento hierárquico dos dados humanos e murinos;
- Comparar os dados de expressão de linfócitos de camundongos NOD com os de pacientes DM1, focando as regiões de susceptibilidade humana e suas correspondentes murinas, pelo uso do T1DBase na identificação de tais regiões;
- Identificar regiões cromossômicas sintênicas entre as duas espécies estudadas que apresentam genes com mesmo perfil de expressão, utilizando a ferramenta *Synteny* disponível na seção *Comparative Genomics* do banco de dados *Ensembl Genome Browser*.
- Realizar busca de similaridade de sequência de DNA e identidade de proteína nos genes identificados em regiões sintênicas entre as duas espécies, utilizando o banco de dados *HomoloGene*.

Material e Métodos

4 MATERIAL E MÉTODOS

4.1 Pacientes com Diabetes *mellitus* do Tipo 1 (DM-1) e Controles

Foram selecionados pacientes com idade entre 18 e 37 anos de ambos os sexos pelos Profs. Drs. Milton César Foss e Maria Cristina Foss-Freitas da Divisão de Endocrinologia do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP). Como controles foram estudados oito indivíduos normais na faixa etária de 47 a 65 anos, de ambos os sexos, sem histórico familiar de DM-1 e/ou DM-2, apresentando níveis normais de glicemia no momento da coleta de sangue. Todos os pacientes e controles foram informados sobre o objetivo do estudo e assinaram termo de consentimento livre e esclarecido. Esse projeto foi aprovado pelo Comitê de Ética do HCFMRP-USP (Processo no. 9153/2008). As tabelas 1 e 2 mostram os dados demográficos dos controles e pacientes utilizados no estudo. O presente trabalho é parte integrante do projeto temático FAPESP nº 08/56594-8 sob coordenação do Prof. Dr. Geraldo A. S. Passos Junior.

Tabela 1 – Dados demográficos dos controles utilizados no estudo.

Controle	Idade (anos)	Sexo	Glicemia (mg/dL)	HbA1c (%) ^a
Co1	51	F ^b	NA	NA
Co4	65	M	NA	NA
Co5	52	F	100	5.4
Co6	58	M	98	5.5
Co11	47	F	89	5.6
Co12	53	M	88	5.7
Co15	55	M	95	5.3
Co16	48	F	84	5.4

^a HbA1c (%) – Hemoglobina glicada em porcentagem. ^b M – sexo masculino; F – sexo feminino.

Tabela 2 – Principais características clínicas dos pacientes com DM1.

Paciente	Idade (anos)	Sexo	Tempo da doença (anos)	Glicemia (mg/dL)	HbA1c (%) ^a
DM1-1	36	M ^b	11	213	10,8
DM1-2	23	M	13	197	8,3
DM1-3	24	M	6	260	10
DM1-4	18	M	8	23	7,2
DM1-5 ^c	23	M	20	178	10,1
DM1-6	21	F	8	223	7,8
DM1-7	29	M	2	59	11,1
DM1-8	30	M	14	47	8,9
DM1-9	21	M	16	66	9
DM1-10	28	F	5	193	12,5
DM1-11	29	M	3	225	9,8
DM1-12	27	F	10	257	10,4
DM1-13	37	F	7	82	8,4
DM1-14	24	F	14	293	8,5
DM1-15	22	F	13	143	8,3
DM1-16	18	M	5	60	9,5
DM1-17 ^c	25	F	6	85	10,5
DM1-18	23	M	11	123	10,3
DM1-19	25	M	8	162	7,7

^a HbA1c (%) – Hemoglobina glicada em porcentagem. ^b M – sexo masculino; F – sexo feminino. ^c Pacientes usam metformina.

4.2 Linhagem de Camundongos *Non-Obese Diabetic* (NOD)

Os camundongos da linhagem NOD foram adquiridos no biotério do CEMIB da Universidade Estadual de Campinas (UNICAMP) onde são mantidos em condições *Specific pathogen free* (SPF). Os animais foram transferidos para o Laboratório de Imunogenética Molecular no Departamento de Genética da Faculdade de Medicina de Ribeirão Preto (FMRP-USP) em mini-isoladores Alesco com entrada de ar filtrado (filtro de 0,45 μ). Os mini-isoladores foram previamente autoclavados assim como a ração e a água, garantindo o fornecimento *ad libitum* sob condições SPF. Além disso, os mini-isoladores foram mantidos em câmara isoladora Alesco também com entrada de ar filtrado (filtro de 0,45 μ) em um

ambiente com temperatura constante de $22^{\circ}\text{C} \pm 2^{\circ}\text{C}$ e ciclo de iluminação de 12 horas. Este procedimento garantiu a manutenção dos animais sob condições de assepsia.

O suplemento polivitamínico Glicopan Pet (Vetnil, São Paulo, Brasil) previamente esterilizado por filtração (filtro de $0,45\ \mu$) foi fornecido juntamente com a água para compensar as perdas ocorridas durante a autoclavagem da ração. Neste estudo, foram utilizadas fêmeas, uma vez que a incidência do diabetes é maior (60-80%) do que em machos (20-30%) como observa Anderson e Bluestone (2005). O presente projeto foi realizado mediante aprovação da Comissão de Ética em Experimentação Animal da FMRP-USP (Processo no. 120/2008).

4.3 Avaliação Laboratorial do Diabetes Autoimune nos Camundongos NOD

Os animais foram avaliados por determinações de glicemia utilizando-se o aparelho *Accu Check - Active Kit* (Roche Diagnóstica Brasil Ltda, São Paulo, Brasil). A leitura do valor foi realizada aplicando-se uma gota de sangue periférico em papel reativo que era inserido no leitor óptico do aparelho. O resultado era expresso em concentração de glicose em mg/dL.

Os animais cujos valores de glicemia excederam 250 mg/dL foram considerados diabéticos. Fêmeas com quatro semanas de idade foram utilizadas como controles sendo consideradas pré-diabéticas, uma vez que seu índice glicêmico era inferior a 250 mg/dL.

4.4 Coleta de Sangue e Isolamento de Células Mononucleares do Sangue Periférico (PBMCs) de Pacientes e Controles

Foram coletados 20-30 mL de sangue venoso periférico de cada paciente e cada controle em tubos contendo EDTA (1,8 mg/mL) (Sigma, St. Louis, MO, USA), sendo as amostras mantidas sob refrigeração até serem utilizadas para a separação de PBMCs. O processamento do sangue foi realizado logo após a coleta. Adotamos o protocolo descrito por Bøyum (1968), sendo a fração mononuclear isolada por centrifugação em gradiente de densidade utilizando Ficoll-Paque (GE Healthcare, Piscataway, NJ, USA).

Os procedimentos realizados para a separação celular incluíram adição de mesmo volume de salina 0,9% estéril à cada amostra de sangue, sendo a solução resultante depositada sobre Ficoll-Paque na proporção de 2:1. Em seguida, as amostras foram centrifugadas por 15 minutos a $1042 \times g$ em temperatura ambiente, obtendo-se ao final um anel de células mononucleares localizado na interface do Ficoll e plasma sanguíneo. As células desta fase

foram colhidas com auxílio de pipeta Pasteur e submetidas a duas lavagens com salina 0,9% estéril, sendo o *pellet* obtido ressuspensão em salina 0,9% tratado com dietilpirocarbonato (DEPC) (Sigma-Aldrich Brasil Ltda, São Paulo, Brasil). As células separadas de pacientes e controles foram imediatamente submetidas à extração de RNA total.

4.5 Isolamento de Linfócitos T CD3⁺ Periféricos Murinos e Verificação de Pureza por Citometria de Fluxo

Os camundongos NOD pré-diabéticos e diabéticos foram sacrificados em câmara de CO₂ e submetidos a cirurgia para remoção do baço. A maceração dos baços foi realizada em meio de cultura DMEM F10, sendo a suspensão celular resultante submetida à separação de linfócitos T CD3⁺ periféricos por esferas imunomagnéticas por seleção negativa utilizando o Pan T-cell Isolation Kit (Mouse) (Miltenyi Biotec Inc., Auburn, CA, USA), seguindo as instruções do fabricante. As células obtidas por esse processo foram utilizadas tanto para análise por citometria de fluxo quanto para extração de RNA total.

Os procedimentos realizados para a análise por *Fluorescence-activated cell sorter* (FACS) incluíram lavagem dos linfócitos T CD3⁺ murinos (1x10⁶ células/tubo) em 3 mL de PBS 1x e centrifugação a 400 x g por 10 minutos para obtenção dos *pellets*. Em seguida, o sobrenadante foi descartado e os botões de células foram ressuspensos em 100 µL de tampão de FACS com 40 µL de *FcBlock* (soro de coelho), sendo incubados por 30 minutos a 4°C e após esse período adicionado 1 µL de cada anticorpo monoclonal marcado: anti-CD3 (PE), anti-CD4 (PerCP) e anti-CD8 (FITC) (eBioscience, Inc., San Diego, CA, USA).

As amostras foram então incubadas por 45 minutos a 4°C, lavadas em 2 mL de tampão de FACS e centrifugadas para obtenção dos *pellets*. A lavagem foi repetida, descartou-se o tampão e foram adicionados 150 µL de PBS com 1% de formol para fixação dos anticorpos. A leitura das marcações foi realizada no citômetro de fluxo FACSCalibur™ (BD, Franklin Lakes, NJ, USA).

Os dados obtidos foram analisados utilizando-se o programa estatístico *GraphPad Prism 5.0* (<http://www.graphpad.com/prism/Prism.htm>), sendo aplicado o teste *t-student* com correção de *Welch's* e $p < 0,05$.

A figura 7 mostra os resultados da análise de FACS, verificando-se grau de pureza de aproximadamente 88,3 % nos linfócitos T CD3⁺ periféricos isolados.

Figura 7 – Análise da pureza dos linfócitos T CD3⁺ periféricos por FACS. As preparações apresentaram pureza superior a 88% como mostra o perfil (resultado típico).

4.6 Extração de RNA Total e Avaliação da Integridade

As amostras de RNA total de PBMCs humanas foram preparadas utilizando-se o reagente Trizol (Life Technologies, Carlsbad, CA, USA) seguindo as instruções do fabricante. Em todas as amostras (pacientes e controles), foi utilizado 1 mL do reagente para cada 5 mL de sangue. O RNA obtido das amostras foi suspenso em água tratada com DEPC (Sigma-Aldrich Brasil Ltda, São Paulo, Brasil) e armazenado em freezer a -80°C para manutenção de integridade.

As amostras de RNA total de linfócitos T CD3⁺ murinos foram preparadas utilizando-se o *mirVana*TM *PARIS*TM Kit (Life Technologies, Carlsbad, CA, USA) seguindo as instruções do fabricante. O RNA obtido das amostras foi suspenso em água tratada com DEPC (Sigma-Aldrich Brasil Ltda, São Paulo, Brasil) e armazenado em freezer à -80°C para manutenção de integridade.

A quantificação e pureza das extrações foram avaliadas por meio de absorbâncias e cálculo das razões, utilizando-se um espectrofotômetro *NanoDrop ND-1000* (NanoDrop Products, Wilmington, DE, USA). Foram utilizadas apenas preparações livres de fenol ($A_{260}/A_{230} \sim 1,8$) e de proteínas ($A_{260}/A_{280} = 1,8-2,0$).

A integridade das amostras foi avaliada por meio de eletroforese microfluídica utilizando-se o chip *RNA 6000 Nano* do aparelho *Agilent 2100 Bioanalyzer* (Agilent Technologies, Santa Clara, CA, USA) seguindo as instruções do fabricante. Este sistema calcula o índice (*RNA Integrity Number – RIN*) com escala de 0 a 10 baseado na qualidade da preparação de RNA. Só foram incluídas amostras com índice RIN superior ou igual a 8,0 (Figura 8 e 9).

Figura 8 – Densitometria das amostras de RNA isoladas de linfócitos T CD3⁺ periféricos murinos. A figura mostra as frações de rRNA 28S, 18S e 4-5S. *RNA integrity number* (RIN) = 10,0.

Figura 9 – Densitometria da amostra de RNA de um paciente com DM1 isolada de PBMCs. A figura mostra as frações de rRNA 28S, 18S e 4-5S. *RNA integrity number* (RIN) = 9,5.

4.7 Oligo-microarrays

Para avaliação do transcriptoma dos pacientes com DM1, controles e camundongos NOD pré-diabéticos e diabéticos foram utilizadas lâminas de *microarrays* (Agilent Technologies, Santa Clara, CA, USA) contendo cada uma quatro regiões com 41.000 oligos (60 mer) representando transcritos e 3.000 controles internos. O formato dos *microarrays* foi então denominado de *4x44 K oligo-arrays*. Neste trabalho foi utilizada a metodologia monocolor Cy3 (cianina 3). Todos os procedimentos realizados para amplificação, marcação e hibridação dos mRNAs das amostras

com os *microarrays* seguiram as recomendações do fabricante, sendo utilizado o *Kit Agilent Quick Amp Labeling* (Agilent Technologies, Santa Clara, CA, USA).

As amostras de RNA controle (*spike-in*) foram misturadas a 500 ng de RNA total para transcrição reversa com utilização da enzima *Moloney Murine Leukemia Virus – Reverse Transcriptase (MMLV-RT)*. Além disso, foram utilizados *primers* adaptados com timinas consecutivas acopladas a um promotor T7 para síntese de cDNAs dupla fita. Após essa reação, a enzima *T7 RNA Polimerase* foi adicionada com nucleotídeos marcados com o fluorocromo Cy3 para síntese de cRNAs (RNA complementar) e posterior amplificação.

As amostras de cRNAs foram então purificadas e, em seguida, submetidas à hibridação às lâminas de *microarray* utilizando-se câmaras de hibridação *Agilent Microarray Hybridization Chambers* (Agilent Technologies, Santa Clara, CA, USA) que foram incubadas em forno rotativo (Agilent Technologies) a 65°C por 17 horas. Após o período de hibridação, foram realizadas lavagens das lâminas com os tampões *GE Wash Buffer 1* e *GE Wash Buffer 2* fornecidos no kit para posterior aquisição das imagens. Os mesmos procedimentos utilizados para aquisição dos *microarrays* foram realizados para as amostras humanas e murinas.

As lâminas de *microarray* utilizadas para RNA humano foram do produto *Whole Human Gene Expression Microarray Kit (4x44 K)* com AMADID 014850 e as utilizadas para RNA de camundongos, do produto *Whole Mouse Gene Expression Microarray Kit (4x44 K)* com AMADID 014868.

4.8 Aquisição das Imagens de Hibridação

As imagens de hibridação dos *arrays Agilent* foram obtidas utilizando-se o *scanner* para *microarray* modelo *Agilent Bundle®* por meio do *software Feature Extraction* (Agilent Technologies, Santa Clara, CA, USA). O mesmo procedimento foi realizado para as amostras humanas e murinas.

4.9 Análise dos Dados de *Microarrays*

4.9.1 Normalização, agrupamento das amostras, filtragens e *fold change*

Para a análise dos dados de *microarrays* foi utilizado o *software GeneSpring GX* (Agilent Technologies, Santa Clara, CA, USA), que fornece ferramentas estatísticas acessíveis e eficazes para a rápida visualização e análise de dados de transcriptomas,

genomas, proteomas e metabolomas. A licença adquirida neste projeto junto a Agilent foi apenas para análise de transcriptomas.

Inicialmente, os dados obtidos por meio do *software Feature Extraction* foram submetidos a um *threshold* cuja função é remover valores de expressão muito pequenos ou negativos antes da transformação dos dados em escala logarítmica, evitando assim valores muito negativos ou perdidos nos dados transformados. O limiar utilizado foi 1, que é o *default* do programa *GeneSpring*. Posteriormente, foi realizada a normalização dos dados, onde o método quantil foi utilizado. Este método torna igual a distribuição dos valores de expressão de todas as amostras em um experimento. Estes procedimentos compreendem a etapa de pré-processamento dos dados.

Em seguida a essa etapa, um experimento é criado e os dados de expressão gênica podem ser analisados segundo outros parâmetros como *Interpretations* e *Quality Control*. Assim, os dados gerados no pré-processamento foram inicialmente submetidos a um agrupamento de acordo com a doença (em humanos, todos os controles foram agrupados como *Control* e todos os pacientes como *T1D [Type 1 Diabetes]*; Em camundongos, todos os animais pré-diabéticos foram agrupados como *Pre-diabetic NOD* e todos os diabéticos como *Diabetic NOD* e uma interpretação foi criada. A partir deste agrupamento, foram aplicados dois tipos de filtros aos dados: por expressão (*filter probsets by expression*) e por *flags* (*filter probsets by flags*), que compreendem a filtragem das entidades (genes) de acordo com seus valores de intensidade de sinal e a filtragem das *flags* – atributos que indicam a qualidade dos genes, respectivamente. O Percentile (máximo de 75,0 e mínimo de 20,0) nos dados normalizados e *flags* apenas presentes foram os parâmetros utilizados na filtragem por expressão e por *flags*, respectivamente.

Após esses ajustes, as matrizes de expressão foram retiradas do *software GeneSpring* e as etapas seguintes foram realizadas no ambiente R (R DEVELOPMENT CORE TEAM, 2011). Inicialmente, os genes sem símbolo foram removidos das listas de *microarray* e foi realizada mediana dos dados, a fim de eliminar eventuais problemas de múltiplas sondas para o mesmo gene, mantendo-se apenas um valor de expressão para cada símbolo gênico.

Os dados obtidos foram submetidos à análise de *fold change* cuja função é identificar genes com diferenças de expressão entre duas condições levando-se em consideração um dado ponto de corte ou limiar. Como neste projeto avaliou-se a modulação transcricional de genes associados com o diabetes *mellitus* do tipo 1 no homem, foi estabelecido um *fold change* igual a 1,0 para todas as análises, buscando-se observar qualquer diferença de expressão entre os indivíduos doentes e saudáveis. Para essa análise, calculamos a mediana a

fim de manter apenas uma amostra para cada condição e a razão absoluta das intensidades de sinal normalizadas (sem escala logarítmica) entre as condições doença e controle (em humanos, *T1D vs control*; em camundongos, *diabetic NOD vs pre-diabetic NOD*) foi calculada.

4.9.2 Busca por genes humanos situados em regiões de susceptibilidade e comparação com camundongos NOD

Os genes em regiões de susceptibilidade ao DM1 humano foram selecionados utilizando o banco de dados *T1DBase* (<http://t1dbase.org>) que fornece tanto uma lista das regiões associadas com base no desequilíbrio de ligação como os genes ali situados. Nesta base de dados, também são disponibilizadas as regiões cromossômicas de susceptibilidade identificadas no camundongo. O *T1DBase* é um banco de dados curado resultante da colaboração entre grupos de pesquisa, que integra diferentes bases de dados, ferramentas e dados da literatura confiáveis, facilitando a pesquisa nesta área (BURREN *et al.*, 2011). Neste trabalho, foi utilizada uma tabela com os *loci* associados ao DM1 humano com atualização em 20 de janeiro de 2011 (Versão 4.2), compreendendo 53 regiões distribuídas em 22 cromossomos.

Para a busca dos genes homólogos no camundongo foi utilizado o banco de dados *Homology Maps* (<http://www.ncbi.nlm.nih.gov/projects/homology/maps/>) (acesso em 3 de abril de 2012) que fornece mapas comparativos entre humanos, camundongos (*M. musculus*) e ratos (*Rattus norvegicus*), utilizando as predições de homologia realizadas pela ferramenta do NCBI *HomoloGene* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene>). Neste trabalho, foi utilizada uma tabela com os genes humanos como referência na comparação com camundongos. Os 463 genes humanos codificantes de proteínas localizados nas regiões de susceptibilidade disponíveis no *T1DBase* bem como seus homólogos murinos foram procurados nas listas de genes dos *microarrays* humano e murino obtidas após a análise de *fold change* (item 4.11.1).

Como neste projeto foram utilizadas populações celulares mistas (PBMCs humanas) e linfócitos T CD3⁺ murinos, foi realizada uma busca no banco de dados *BioGPS* (<http://biogps.org>) (WU *et al.*, 2009) que indica a expressão gênica em tecidos e tipos celulares específicos, verificando-se os pares de homólogos selecionados na etapa anterior e mantendo-se apenas aqueles com expressão acima da mediana em linfócitos T humanos e murinos. Dessa maneira, procurou-se focar somente nos genes expressos em linfócitos T de

humanos e camundongos. As bases de dados utilizadas na análise foram *GeneAtlas U133A* e *GeneAtlas MOE430* para humano e camundongo, respectivamente.

Após esse *screening*, foram identificados os genes que apresentaram mesma regulação de *fold change* (*up-* ou *down-regulation*) entre as duas espécies e selecionados aqueles que também foram mapeados em regiões de susceptibilidade murinas (*Idd – Insulin-dependent diabetes*), utilizando uma tabela do *TIDBase* que fornece as regiões associadas ao DM1 por ligação em camundongo com atualização em 5 de dezembro de 2011 (Versão 4.8).

Os genes selecionados por triagem no *BioGPS* foram submetidos à análise de agrupamento hierárquico (*Hierarchical Clustering – HCL*), que consiste em um método aglomerativo, onde os perfis de expressão mais similares são unidos para formar um grupo. O agrupamento foi realizado tanto por entidades (genes) quanto por amostras (sem mediana) e a distância métrica utilizada foi a *Pearson uncentered*, sendo selecionado *average linkage* como método de agrupamento. Estes procedimentos foram realizados utilizando-se os programas *Cluster 3.0* e *TreeView 1.60* (EISEN *et al.*, 1998, <http://rana.lbl.gov/EisenSoftware.htm>).

Outro procedimento realizado foi a análise dos genes obtidos do *screening* pelo *BioGPS* quanto à localização cromossômica, buscando-se por regiões de sintenia entre homem e camundongo. A ferramenta *MapView* (SAYERS *et al.*, 2012, <http://www.ncbi.nlm.nih.gov/mapview/>) disponível no banco de dados *NCBI* e a ferramenta *Synteny* disponível na seção *Comparative Genomics* do *Ensembl Genome Browser* (FLICEK *et al.*, 2012, <http://www.ensembl.org/>) foram utilizados na identificação dessas regiões.

Finalmente, os genes identificados em regiões de sintenia entre as duas espécies foram submetidos à busca de similaridade de sequência de DNA e identidade de proteína, utilizando-se o banco de dados *HomoloGene* com atualização em 16 de abril de 2012. As figuras 10 e 11 ilustram o *pipeline* das análises computacionais dos dados deste trabalho.

Figura 10 – Pipeline das análises dos dados de *microarray* humano e murino realizadas para posterior busca de similaridades entre as duas espécies (item 4.9.1).

Figura 11 – *Pipeline* das análises realizadas para a comparação de expressão dos genes humanos e murinos (item 4.9.2).

Resultados

5 RESULTADOS

5.1 Análise dos Dados de *Microarrays* Humano e Murino

Após a obtenção dos dados de *microarrays* humanos e murinos pelo *software Feature Extraction*, estes foram analisados separadamente utilizando o *software GeneSpring GX*, sendo normalizados e aplicados os filtros e parâmetros. Ao final, foram identificados 15.089 genes humanos e 14.856 genes de camundongo. Após esse procedimento, foi calculada a mediana das amostras e realizada transformação dos dados de expressão em função exponencial e cálculo de *fold change*, sendo os dados resultantes utilizados na análise comparativa humano/camundongo.

5.2 Análise de Genes Humanos Presentes em Regiões de Susceptibilidade ao DM1 e Comparação com a Linhagem de Camundongos NOD

A análise de genes humanos situados em regiões de susceptibilidade ao DM1 foi realizada a partir de uma lista disponível no *TIDBase*. Dos 463 genes codificantes de proteínas localizados nas 53 regiões cromossômicas de susceptibilidade, 306 foram identificados nos dados de *microarray* humano deste trabalho, sendo que 267 deles apresentaram homologia com camundongo de acordo com o *Homology Maps*. Finalmente, dos 267 homólogos murinos, apenas 165 foram identificados nos dados de *microarrays* murinos deste trabalho.

Além disso, foram selecionados apenas os pares de homólogos com expressão em linfócitos T humanos e murinos, segundo o banco de dados *BioGPS* (<http://biogps.org>). Foi observado que 73 genes satisfizeram esse parâmetro (exemplo do gene TRIM39, Figura 13) e 31 deles apresentaram mesma modulação de *fold change* entre as duas espécies. Destes 31 genes, 12 foram encontrados em regiões de susceptibilidade (*Idd – insulin dependet-diabetes*) também no camundongo (Tabela 3). A Figura 12 mostra um diagrama de Venn com o número de genes em cada etapa realizada durante a seleção dos genes de linfócitos T comuns entre as duas espécies.

A Tabela 3 mostra todos os genes que foram selecionados no trabalho com as regiões de susceptibilidade em humano e camundongo, as modulações de *fold change* e a localização cromossômica deles. Dos 73 genes com expressão em linfócitos T, os genes AIRE, APOM,

ASCL2, ATXN2, BRCC3, CCR2, CCR3, CDSN, COL11A2, CYP21A2, GLI1, GRB7, HLA-DOB, HOXA7, ICAM4, IFIH1, IKZF3, IL21, KIF5A, PHTF1, PRR3, RAB5B, RNF41, RSNB1, SH2B3, SIRPB1, STK19, TCAP, TRIM39, VARS2 e XCR1 apresentaram mesma modulação de *fold change* homem/camundongo, sendo 16 genes reprimidos e 15, induzidos, embora alguns tenham apresentado uma modulação muito sutil não atingindo *fold change* de 1,1. A maior parte desses genes estão localizados na região do MHC humano (cromossomo 6p21) (29,03% ou 9 genes), bem como na região de MHC murina no cromossomo 17 B1. Dos 31 genes com mesma modulação entre as duas espécies, 12 deles também estão em regiões *Idd* murinas: Apom, Coll1a2, Cyp21a1, H2-Ob e Stk19 na *Idd1*; Il21 na *Idd3*; Phtf1 e Rsnb1 na *Idd18.2*; e Cdsn, Prr3, Trim39 e Vars2 na *Idd24*.

Além disso, alguns genes com modulação diferente entre humanos e camundongos também foram mapeados em regiões *Idd* murinas: Egfl8, Hspa11, Ng23, Tnf e Tnxb na *Idd1*; Adamts7 e Rasgrf1 na *Idd2*; 4932438A13Rik na *Idd3*; Ctla4 na *Idd5.1*; Rasgrp1 na *Idd13*; Pou5f1, Trim15 e Trim26 na *Idd24*; e Aff3 na *Idd26*.

No agrupamento hierárquico (Figura 14), os 73 genes demonstraram ser bons marcadores na separação dos indivíduos diabéticos (camundongos diabéticos e pacientes) dos não diabéticos, uma vez que em humanos os controles agruparam juntos e em camundongo, os pré-diabéticos também.

Figura 12 – Diagrama de Venn representando o *screening* realizado na seleção de genes expressos em linfócitos T e localizados em regiões de susceptibilidade ao DM1 comuns entre humanos e camundongos.

Figura 13 - Gráfico ilustrativo gerado pelo banco de dados BioGPS (<http://biogps.org>) mostrando a representação tecidual do gene TRIM39, com os nomes de cada tecido, órgão e/ou tipos celulares onde esse gene é expresso. (A) Representação tecidual em humanos. (B) Representação tecidual em camundongos. Em destaque (retângulo vermelho), as células T.

Tabela 3 – Genes comuns entre humano e camundongo com expressão em linfócitos T.

<i>Homo sapiens</i>				<i>Mus musculus</i>			
Gene	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a	Homólogo	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a
ACAD10	12q24.12	12q24.12	1,222944557	Acad10	-	5 F	-1,047032337
ADAMTS7	15q25.1	15q25.1	1,009657305	Adamts7	<i>Idd2</i>	9 E3.1	-1,01401395
AFF3	2q11.2	2q11.2	1,713297636	Aff3	<i>Idd26</i>	1B	-1,154804793
AIRE ^b	21q22.3	21q22.3	-1,035920292	Aire	-	10 C1	-2,59312904
APOM	MHC	MHC	1,145144059	Apom	<i>Idd1</i>	17 B1	1,166412673
ASCC2	22q12.2	22q12.2	1,254152687	Ascc2	-	11 A1	-1,537147917
ASCL2	11p15.5	11p15.5	1,400290058	Ascl2	<i>ND^c</i>	7 F5	1,996124907
ATXN2	12q24.12	12q24.12	1,10407716	Atxn2	-	5 F	1,341293537
BRCC3	Xq28	Xq28	-1,443411846	Brcc3	-	X A7.3	-1,279830054
C6orf26	MHC	MHC	-1,45759357	Ng23	<i>Idd1</i>	17 B1	1,185803862
CCR2	3p21.31	3p21.31	-2,045996958	Ccr2	-	9 F4	-1,893961494
CCR3	3p21.31	3p21.31	1,05257461	Ccr3	-	9 F4	1,197252958
CDK2	12q13.2	12q13.2	1,121312033	Cdk2	-	10 D3	-1,826878262
CDSN	MHC	MHC	-1,142526003	Cdsn	<i>Idd24</i>	17 B1	-1,234905718
COL11A2	MHC	MHC	1,289015668	Col11a2	<i>Idd1</i>	17 B1	1,134834173
CTLA4	2q33.2	2q33.2	1,52720899	Ctla4	<i>Idd5.1</i>	1 C2	-2,768829082
CYP21A2	MHC	MHC	-1,015398476	Cyp21a1	<i>Idd1</i>	17 B1	-1,955538582
CYP27B1	12q13.3	12q14.1	-1,259976128	Cyp27b1	-	10 D3	3,534097642
DDC	7p12.2	7p12.1	-1,356847518	Ddc	-	11 A1	1,027572545
EGFL8	MHC	MHC	-1,037699878	Egfl8	<i>Idd1</i>	17 B1	1,223729761
FUT1	19q13.4	19q13.33	-1,181160323	Fut1	-	7 B4	1,562167336

<i>Homo sapiens</i>				<i>Mus musculus</i>			
Gene	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a	Homólogo	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a
FUT2	19q13.4	19q13.33	-1,041289808	Fut2	-	7 B4	1,734840826
FYCO1	3p21.31	3p21.31	-1,250610903	Fyco1	-	9 F4	1,307200101
GLI1	12q13.3	12q13.3	1,088093777	Gli1	-	10 D3	1,264372708
GRB7	17q12	17q12	-1,100648647	Grb7	-	11 D	-1,508670344
HLA-DOB	MHC	MHC	1,87085254	H2-Ob	<i>Idd1</i>	17 B1	5,357959217
HOXA7	7p15.2	7p15.2	-1,873880821	Hoxa7	-	6 B3	-2,237033715
HSPA1L	MHC	MHC	-1,069052719	Hspa1l	<i>Idd1</i>	17 B1	1,151051433
ICAM4	19p13.2	19p13.2	-1,048731964	Icam4	-	9 A3	-5,276961654
ICAM5	19p13.2	19p13.2	-1,17715129	Icam5	-	9 A3	1,399405435
IFIH1	2q24.2	2q24.2	-1,886822879	Ifih1	-	2 C1.3	-3,009625964
IGF2	11p15.5	11p15.5	1,010038685	Igf2	<i>ND</i>	7 F5	-2,544308836
IKZF3	17q12	17q12	1,27935505	Ikzf3	-	11 D	1,597157956
IKZF4	12q13.2	12q13.2	-1,071372745	Ikzf4	-	10 D3	1,027294481
IL21	4q27	4q27	-1,476926468	Il21	<i>Idd3</i>	3 B	-6,105176483
IL2RA	10p15.1	10p15.1	1,552936512	Il2ra	-	2 A1	-2,579258641
INHBC	12q13.3	12q13.3	1,218965251	Inhbc	-	10 D3	-2,978117344
INHBE	12q13.3	12q13.3	-1,107916919	Inhbe	-	10 D3	1,372101017
INS	11p15.5	11p15.5	-1,062634236	Ins2	<i>ND</i>	7 F5	1,488176353
KIAA1109	4q27	4q27	1,224354214	4932438A13Rik	<i>Idd3</i>	3 B	-1,017805929
KIF5A	12q13.3	12q13.3	1,107369737	Kif5a	-	10 D3	1,111067795
KIFC1	MHC	MHC	1,259870981	Kifc1	-	17 B1	-2,30164048
LTF	3p21.31	3p21.31	-1,464147602	Ltf	-	9 F3	5,398320004
MYL2	12q24.12	12q24.11	-1,270502576	Myl2	-	5 F	1,83575229

<i>Homo sapiens</i>				<i>Mus musculus</i>			
Gene	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a	Homólogo	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a
NEUROD2	17q12	17q12	-1,332011049	Neurod2	-	11 D	1,055526866
NFATC2IP	16p11.2	16p11.2	1,360879662	Nfatc2ip	ND	7 F3	-1,91375255
PA2G4	12q13.2	12q13.2	1,02788474	Pa2g4	-	10 D3	-2,147548182
PHTF1	1p13.2	1p13.2	-1,418289702	Phtf1	Idd18.2	3 F2.2	-1,164075393
POU5F1	MHC	MHC	1,314445659	Pou5f1	Idd24	17 B1	-1,009914739
PRKCQ	10p15.1	10p15.1	1,180504876	Prkcq	-	2 A1	-1,108633417
PRR3	MHC	MHC	1,520425893	Prr3	Idd24	17 B1	1,123732579
PTPN11	12q24.12	12q24.13	1,704406549	Ptpn11	-	5 F	-1,799940978
RAB5B	12q13.2	12q13.2	1,13656016	Rab5b	-	10 D3	1,713893424
RASGRF1	15q25.1	15q25.1	-1,495583179	Rasgrf1	Idd2	9 E3.1	1,003998212
RASGRP1	15q14	15q14	1,018136857	Rasgrp1	Idd13	2 E5	-1,049812005
RNF41	12q13.2	12q13.2-13.3	1,033819271	Rnf41	-	10 D3	1,24857285
RSBN1	1p13.2	1p13.2	-1,899914506	Rsb1	Idd18.2	3 F2.2	-3,414318158
SH2B3	12q24.12	12q24.12	1,148577098	Sh2b3	-	5 F	1,548942941
SIRPB1	20p13	20p13	1,178389924	LOC100038947	-	3 A1	2,483277004
STK19	MHC	MHC	-1,120074349	Stk19	Idd1	17 B1	-1,424219489
TCAP	17q12	17q12	-1,182858909	Tcap	-	11 D	-1,267565589
TH	11p15.5	11p15.5	1,425579702	Th	ND	7 F5	-1,013924339
TMRSS3	21q22.3	21q22.3	1,627647908	Tmrss3	-	17 A3.3	-5,82143167
TNF	MHC	MHC	11,05588416	Tnf	Idd1	17 B1	-1,668097943
TNXB	MHC	MHC	1,57860946	Tnxb	Idd1	17 B1	-1,235768924
TRAFD1	12q24.12	12q24.13	1,325622045	Trafd1	-	5F	-1,20743707
TRIM15	MHC	MHC	1,028885757	Trim15	Idd24	17 B1	-1,001149413

<i>Homo sapiens</i>				<i>Mus musculus</i>			
Gene	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a	Homólogo	Região de Susceptibilidade	Citobanda	Modulação (Fold change) ^a
TRIM26	MHC	MHC	1,544019743	Trim26	<i>Idd24</i>	17 B1	-1,829188305
TRIM39	MHC	MHC	-1,091011267	Trim39	<i>Idd24</i>	17 B1	-1,276839643
VAR2	MHC	MHC	-1,258990154	Vars2	<i>Idd24</i>	17 B1	-1,205758516
VPS52	MHC	MHC	1,168126891	Vps52	-	17 B1	-1,600976354
XCR1	3p21.31	3p21.31	1,431161021	Xcr1	-	9 F4	1,897404666
ZBTB22	MHC	MHC	1,352788218	Zbtb22	-	17 B1	-1,07577659

^aA modulação gênica é referente à comparação dos indivíduos diabéticos em relação aos controles. ^bAs faixas indicam genes que apresentam mesma modulação entre as duas espécies. ^cND – Não determinada, região de susceptibilidade murina sem designação.

5.3 Análise de Regiões Sintênicas Homem/Camundongo

Os genes obtidos do *screening* pelo *BioGPS* foram submetidos à busca por localização cromossômica e verificados se estavam situados em regiões de sintenia homem/camundongo. Dos 73 genes, 59 foram identificados em regiões de sintenia humano-camundongo.

A Tabela 4 mostra os 59 genes mapeados em regiões sintênicas humano-camundongo com o intervalo genômico do bloco em megabases (Mb).

Tabela 4 – Genes identificados em regiões de sintenia homem/camundongo.

<i>Homo sapiens</i>			<i>Mus musculus</i>		
Citobanda	Gene de interesse	Região sintênica em Mb	Citobanda	Gene de interesse	Região sintênica em Mb
1p13.2	PHTF1	Cr 1: 93,9 – 120,7	3 F2.2	Phtf1	Cr 3: 97,7 – 122,3
1p13.2	RSBN1		3 F2.2	Rsbn1	
3p21.31	CCR2	Cr 3: 37,3 – 46,4	9 F4	Ccr2	Cr 9: 118,4 – 124,0
3p21.31	CCR3		9 F4	Ccr3	
3p21.31	FYCO1		9 F4	Fyco1	
3p21.31	XCR1	Cr 4: 122,2 – 141,2	9 F4	Xcr1	Cr 3: 36,1 – 52,0
4q27	KIAA1109		3 B	4932438A13Rik	
4q27	IL21		3 B	Il21	
6p22.1	TRIM15	Cr 6: 29,3 – 33,3	17B1	Trim15	Cr 17: 34,0 – 37,6
6p22.1	TRIM26		17B1	Trim26	
6p22.1	TRIM39		17B1	Trim39	
6p21.33	PRR3		17B1	Prr3	
6p21.33	VAR2		17B1	Vars2	
6p21.33	CDSN		17B1	Cdsn	
6p21.33	POU5F1		17B1	Pou5f1	
6p21.33	TNF		17B1	Tnf	
6p21.33	APOM		17B1	Apom	
6p21.33	C6orf26		17B1	Ng23	
6p21.33	HSPA1L		17B1	Hspa1l	
6p21.33	STK19		17B1	Stk19	
6p21.33	CYP21A2		17B1	Cyp21a1	
6p21.33	TNXB		17B1	Tnxb	
6p21.32	EGFL8		17B1	Egfl8	
6p21.32	HLA-DOB		17B1	H2-Ob	
6p21.32	COL11A2		17B1	Col11a2	
6p21.32	VPS52		17B1	Vps52	
6p21.32	ZBTB22		17B1	Zbtb22	
10p15.1	IL2RA	Cr 10: 5,9 – 27,2	2A1	Il2ra	Cr 2: 3,0 – 22,0
10p15.1	PRKCQ		2A1	Prkcq	
11p15.5	IGF2	Cr 11: 0,2 – 3,1	7 F5	Igf2	Cr 7: 148,0 – 150,8
11p15.5	INS		7 F5	Ins2	
11p15.5	TH		7 F5	Th	
11p15.5	ASCL2	Cr 12: 55,4 – 104,4	7 F5	Ascl2	Cr 10: 86,1 – 129,8
12q13.2	CDK2		10 D3	Cdk2	
12q13.2	RAB5B		10 D3	Rab5b	
12q13.2	IKZF4		10 D3	Ikzf4	
12q13.2	PA2G4		10 D3	Pa2g4	
12q13.2	RNF41		10 D3	Rnf41	
12q13.3	INHBC		10 D3	Inhbc	
12q13.3	INHBE		10 D3	Inhbe	

<i>Homo sapiens</i>			<i>Mus musculus</i>		
Citobanda	Gene de interesse	Região sintênica em Mb	Citobanda	Gene de interesse	Região sintênica em Mb
12q13.3	GLI1		10 D3	Gli1	
12q13.3	KIF5A		10 D3	Kif5a	
12q14.1	CYP27B1		10 D3	Cyp27b1	
12q24.11	MYL2		5 F	Myl2	
12q24.12	SH2B3		5 F	Sh2b3	
12q24.12	ATXN2	Cr 12: 110,5 –	5 F	Atxn2	Cr 5: 115,3 –
12q24.12	ACAD10	121,5	5 F	Acad10	123,1
12q24.13	TRAFD1		5 F	Trafd1	
12q24.13	PTPN11		5 F	Ptpn11	
15q25.1	ADAMTS7	Cr 15: 79,0 – 80,2	9 E3.1	Adamts7	Cr 9: 89,1 – 90,1
15q25.1	RASGRF1		9 E3.1	Rasgrf1	
17q12	NEUROD2		11 D	Neurod2	
17q12	TCAP	Cr 17: 36,4 – 43,6	11 D	Tcap	Cr 11: 97,1 – 103,5
17q12	GRB7		11 D	Grb7	
17q12	IKZF3		11 D	Ikzf3	
19p13.2	ICAM4	Cr 19: 8,9 – 11,7	9 A3	Icam4	Cr 9: 18,3 – 21,9
19p13.2	ICAM5		9 A3	Icam5	
19q13.33	FUT2	Cr 19: 48,8 – 51,9	7 B4	Fut2	Cr 7: 50,7 – 53,2
19q13.33	FUT1		7 B4	Fut1	

* Cr – cromossomo.

A figura 15A mostra que os genes humanos FYCO1, XCR1, CCR3 e CCR2 situados no cromossomo 3p21.31 foram mapeados em uma região sintênica ao camundongo, que compreende o intervalo genômico 37,3 Mb – 46,4 Mb. Tal região é sintênica ao bloco murino 118,4 Mb – 124,0 Mb presente no cromossomo 9 e que apresenta os genes homólogos Fyco1, Xcr1, Ccr3 e Ccr2 na mesma ordem na citobanda F4.

A figura 15B mostra os genes IL2RA e PRKCQ presentes no cromossomo 10p15.1 como integrantes do bloco humano 5,9 Mb – 27,2 Mb que é sintênico à região murina que abrange o intervalo genômico 3,0 Mb – 22,0 Mb no cromossomo 2. Tal região murina apresenta os genes homólogos Il2ra e Prkcq na citobanda A1 em ordem inversa à observada no cromossomo humano 10.

A figura 16A mostra que os genes humanos ADAMTS7 e RASGRF1 situados no cromossomo 15q25.1 foram mapeados em uma região sintênica ao camundongo, que compreende o intervalo genômico 79,0 Mb – 80,2 Mb. Esta região é sintênica ao bloco murino 89,1 Mb – 90,1 Mb presente no cromossomo 9 e apresenta os genes homólogos Adamts7 e Rasgrf1 na citobanda E3.1 em ordem inversa à observada no cromossomo humano 15.

A Figura 16B mostra os genes TRIM15, TRIM26 e TRIM39 presentes no cromossomo 6p22.1, PRR3, VARS2, CDSN, POU5F1, TNF, APOM, C6orf26, HSPA1L, STK19, CYP21A2 e TNXB no 6p21.33, e EGFL8, HLA-DOB, COL11A2, VPS52 e ZBTB22

no 6p21.32 como integrantes do bloco humano 29,3 Mb – 33,3 Mb que é sintênico à região murina com intervalo genômico 34,0 Mb – 37,6 Mb no cromossomo 17. Tal região murina apresenta os genes homólogos Trim15, Trim26, Trim39, Prr3, Vars2, Cdsn, Pou5f1, Tnf, Apom, Ng23, Hspa11, Stk19, Cyp21a1, Tnxb, Egfl8, H2-Ob, Col11a2, Vps52 e Zbtb22 na citobanda B1 em ordem inversa à observada no cromossomo humano 6.

A Figura 16C mostra que os genes humanos NEUROD2, TCAP, GRB7 e IKZF3 situados no cromossomo 17q12 foram mapeados em um bloco sintênico ao genoma murino, compreendendo o intervalo genômico 36,4 Mb – 43,6 Mb. A região sintênica murina abrange o intervalo de 97,1 Mb – 103,5 Mb presente no cromossomo 11 e apresenta os homólogos Neurod2, Tcap, Grb7 e Ikzf3 na citobanda D na mesma ordem observada no cromossomo humano 17.

A figura 16D mostra os genes ICAM4 e ICAM5 presentes no cromossomo 19p13.2 como integrantes do bloco humano 8,9 Mb – 11,7 Mb que é sintênico à região murina com intervalo genômico 18,3 Mb – 21,9 Mb no cromossomo 9. Esta região murina apresenta os homólogos Icam4 e Icam5 na citobanda A3 com a mesma ordem observada no cromossomo humano 19.

A figura 17A mostra que os genes humanos FUT2 e FUT1 situados no cromossomo 19q13.33 foram mapeados em um bloco sintênico ao genoma murino, compreendendo o intervalo genômico 48,8 Mb – 51,9 Mb. A região sintênica murina abrange o intervalo de 50,7 Mb – 53,2 Mb presente no cromossomo 7 e apresenta os genes homólogos Fut2 e Fut1 na citobanda B4 em ordem inversa à observada no cromossomo humano 19.

A figura 17B mostra os genes PHTF1 e RSBN1 localizados no cromossomo 1p13.2 e mapeados como integrantes do bloco humano 93,9 Mb – 120,7 Mb que é sintênico à região murina com o intervalo 97,7 Mb – 122,3 Mb no cromossomo 3. A região murina conservada apresenta os genes homólogos Phtf1 e Rsbn1 situados na citobanda F2.2 em ordem inversa à observada no cromossomo humano 1.

A figura 18A mostra que os genes humanos KIAA1109 e IL21 situados no cromossomo 4q27 foram mapeados em um bloco sintênico ao camundongo, compreendendo o intervalo genômico 122,2 Mb – 141,2 Mb. A região sintênica murina abrange o intervalo de 36,1 Mb – 52,0 Mb presente no cromossomo 3 e apresenta os genes homólogos Il21 e 4932438A13Rik na citobanda B em ordem inversa à observada no cromossomo humano 4.

A figura 18B mostra os genes CDK2, RAB5B, IKZF4, PA2G4 e RNF41 no cromossomo 12q13.2, INHBC, INHBE, GLI1 e KIF5A no 12q13.3, e CYP27B1 no 12q14.1 como integrantes do bloco humano 55,4 Mb – 104,4 Mb que é sintênico à região murina com

intervalo genômico 86,1 Mb – 129,8 Mb no cromossomo 10. Tal região murina apresenta os genes homólogos *Cdk2*, *Rab5b*, *Ikzf4*, *Pa2g4*, *Rnf41*, *Inhbc*, *Inhbe*, *Gli1*, *Kif5a* e *Cyp27b1* na citobanda D3 em ordem inversa à observada no cromossomo humano 12.

A figura 18C mostra que os genes humanos *IGF2*, *INS*, *TH* e *ASCL2* situados no cromossomo 11p15.5 foram mapeados em uma região sintênica ao camundongo, abrangendo o intervalo genômico 0,2 Mb – 3,1 Mb. A região sintênica murina compreende o intervalo de 148,0 Mb – 150,8 Mb presente no cromossomo 7 e apresenta os genes homólogos *Igf2*, *Ins2*, *Th* e *Ascl2* na citobanda F5 com a mesma ordem observada no cromossomo humano 11.

Por último, a figura 18D mostra os genes *MYL2* no cromossomo 12q24.11, *SH2B3*, *ATXN2* e *ACAD10* na citobanda 12q24.12, e *TRAFD1* e *PTPN11* na 12q24.13 como integrantes do bloco humano 110,5 Mb – 121,5 Mb que é sintênico à região murina com intervalo genômico 115,3 Mb – 123,1 Mb no cromossomo 5. Tal região murina apresenta os genes homólogos *Myl2*, *Sh2b3*, *Atxn2*, *Acad10*, *Trafd1* e *Ptpn11* na citobanda F em ordem inversa à observada no cromossomo humano 12.

Figura 15 – Genes identificados em regiões sintênicas nos cromossomos humanos 3 e 10, e 2 e 9 murinos. Caixa cinza, bloco compartilhado entre as duas espécies; verde, genes reprimidos; vermelho, genes induzidos. Sublinhados apresentam mesma modulação humano/camundongo; * presença de um ou mais genes após o gene em questão; Hsa, *Homo sapiens*; Mmu, *Mus musculus*.

Figura 16 – Genes identificados em regiões sintênicas nos cromossomos humanos 6, 15, 17 e 19, e 9, 11 e 17 murinos. Caixa cinza, bloco compartilhado entre as duas espécies; verde, genes reprimidos; vermelho, genes induzidos. Sublinhados apresentam mesma modulação humano/camundongo; * presença de um ou mais genes após o gene em questão; Hsa, *Homo sapiens*; Mmu, *Mus musculus*.

Figura 17 – Genes identificados em regiões sintênicas nos cromossomos humanos 1 e 19, e 3 e 7 murinos. Caixa cinza, bloco compartilhado entre as duas espécies; verde, genes reprimidos; vermelho, genes induzidos. Sublinhados apresentam mesma modulação humano/camundongo; * presença de um ou mais genes após o gene em questão; Hsa, *Homo sapiens*; Mmu, *Mus musculus*.

Figura 18 – Genes identificados em regiões sintênicas nos cromossomos humanos 4, 11 e 12, e 3, 5, 7 e 10 murinos. Caixa cinza, bloco compartilhado entre as duas espécies; verde, genes reprimidos; vermelho, genes induzidos. Sublinhados apresentam mesma modulação humano/camundongo; * presença de um ou mais genes após o gene em questão; Hsa, *Homo sapiens*; Mmu, *Mus musculus*.

5.4 Análise das sequências de DNA e de proteínas referentes a genes humanos e murinos situados em regiões de sintenia entre as duas espécies

Os genes identificados em regiões de sintenia homem/camundongo foram avaliados quanto a similaridade de sequência de DNA e identidade de proteína, utilizando-se o banco de dados *HomoloGene*. Dos 59 genes sintênicos, 58 apresentaram similaridade de sequência de DNA variando de 71,3% a 93,6% e identidade protéica variando de 61,4% a 99,7%. De acordo com o banco de dados atual, o gene humano RAB5B não se encontra mais no mesmo grupo de homologia do camundongo, sendo o gene humano pertencente ao *HomoloGene* 104027, e o gene murino, ao *HomoloGene* 131219.

A Tabela 5 mostra os 58 genes mapeados em regiões sintênicas humano-camundongo com o nível de similaridade de sequência de DNA e a identidade da proteína entre as duas espécies.

Tabela 5 – Similaridade de sequência de DNA e identidade de proteína entre os genes humanos e murinos situados em regiões de sintenia entre as duas espécies.

Gene (Homólogo murino)	Identidade (%) <i>H. sapiens</i> vs <i>M. musculus</i>	
	DNA	Proteína
PHTF1 (Phtf1)	90,1	93,7
RSBN1 (Rsb1)	93,6	95,3
CCR2 (Ccr2)	80,7	79,7
CCR3 (Ccr3)	77,0	70,1
FYCO1 (Fyco1)	79,8	78,2
XCR1 (Xcr1)	76,3	72,3
KIAA1109 (4932438A13Rik)	90,7	96,7
IL21 (Il21)	76,9	63,0
TRIM15 (Trim15)	78,8	76,2
TRIM26 (Trim26)	85,3	88,9
TRIM39 (Trim39)	90,4	97,5
PRR3 (Prr3)	85,5	83,0
VARS2 (Vars2)	84,4	84,2
CDSN (Cdsn)	71,3	70,7
POU5F1 (Pou5f1)	83,7	86,1
TNF (Tnf)	82,9	80,2
APOM (Apom)	84,5	81,8
C6orf26 (Ng23)	78,8	71,9
HSPA1L (Hspa1l)	85,3	94,7
STK19 (Stk19)	82,4	85,8
CYP21A2 (Cyp21a1)	77,7	72,7
TNXB (Tnxb)	73,2	68,6
EGFL8 (Egfl8)	80,1	81,9
HLA-DOB (H2-Ob)	81,1	75,6
COL11A2 (Col11a2)	87,6	93,6
VPS52 (Vps52)	89,0	98,2
ZBTB22 (Zbtb22)	83,9	87,4
IL2RA (Il2ra)	72,2	61,4
PRKCQ (Prkcq)	87,3	94,9

Gene (Homólogo murino)	Identidade (%) <i>H. sapiens</i> vs <i>M. musculus</i>	
	DNA	Proteína
IGF2 (Igf2)	86,0	82,3
INS (Ins2)	82,4	81,8
TH (Th)	83,0	88,3
ASCL2 (Ascl2)	77,7	74,1
CDK2 (Cdk2)	92,5	99,0
RAB5B*	-	-
IKZF4 (Ikzf4)	92,4	96,9
PA2G4 (Pa2g4)	92,1	99,0
RNF41 (Rnf41)	92,2	99,7
INHBC (Inhbc)	81,1	76,4
INHBE (Inhbe)	83,7	82,8
GLI1 (Gli1)	84,0	86,1
KIF5A (Kif5a)	90,9	98,1
CYP27B1 (Cyp27b1)	81,3	81,7
MYL2 (Myl2)	85,3	96,4
SH2B3 (Sh2b3)	79,8	78,4
ATXN2 (Atxn2)	89,0	90,8
ACAD10 (Acad10)	80,7	80,2
TRAJD1 (Trajd1)	81,3	76,6
PTPN11 (Ptpn11)	87,2	99,5
ADAMTS7 (Adamts7)	78,4	74,2
RASGRF1 (Rasgrf1)	84,4	85,0
NEUROD2 (Neurod2)	92,2	98,2
TCAP (Tcap)	85,4	90,4
GRB7 (Grb7)	86,2	91,0
IKZF3 (Ikzf3)	85,6	86,1
ICAM4 (Icam4)	74,3	70,0
ICAM5 (Icam5)	82,6	85,5
FUT2 (Fut2)	78,9	80,5
FUT1 (Fut1)	77,3	78,8

* Não é mais considerado como homólogo *H. sapiens*/ *Mus musculus*.

Discussão

6 DISCUSSÃO

O presente trabalho se refere ao estudo comparativo da expressão transcricional de RNAs mensageiros (perfil do transcriptoma) de células do sistema imune de pacientes humanos com diabetes *mellitus* do tipo 1 (DM1) e de camundongos *M. musculus* da linhagem *non-obese diabetic* (NOD).

O objetivo deste tipo de comparação é em primeiro lugar, estabelecer um paralelo entre as duas espécies, focando os genes e/ou as regiões cromossômicas sintênicas de susceptibilidade. Em segundo lugar, tentaremos dar um sentido evolutivo aos padrões de expressão transcricional.

Neste estudo, utilizamos linfócitos periféricos humanos e murinos, uma vez que essas células apresentam um potencial de responder a mudanças no corpo e em seu ambiente, atuando como “sentinelas” e sendo informativas em relação a eventos fisiológicos e patológicos no organismo (REYNER *et al.*, 2010).

Além disso, já foi observado que as células do sangue periférico expressam grande proporção dos genes humanos, tornando possível o estudo de mudanças clinicamente relevantes na expressão gênica de maneira reprodutível (REYNER *et al.*, 2010; LIEW *et al.*, 2006).

Das 44.000 sequências analisadas neste trabalho pela tecnologia de *microarrays*, apontamos 463 que representam genes (mRNAs) situados em regiões cromossômicas humanas consideradas como regiões de susceptibilidade ao DM1. Neste conjunto de sequências humanas, identificamos 73 que são compartilhadas com camundongos. Isto já nos indicou que seria possível analisar e/ou comparar a expressão de um grupo de genes compartilhados entre *H. sapiens* e *M. musculus* no contexto do DM1.

Destes 73 genes³¹ apresentaram mesma modulação de *fold change* entre as duas espécies, sendo 12 deles (APOM, COL11A2, CYP21A2, HLA-DOB, STK19, PHTF1, RSNB1, CDSN, PRR3, TRIM39, VARS2, IL21) encontrados em regiões de susceptibilidade também em *M. musculus*.

Os genes APOM, COL11A2, CYP21A2, HLA-DOB, STK19, CDSN, PRR3, TRIM39 e VARS2 estão todos localizados no cromossomo humano 6p21, região do complexo principal de histocompatibilidade (MHC). Nota-se que a região do MHC é uma das mais associadas com o DM1 e principal contribuinte para a doença tanto em seres humanos como em camundongos NOD (BLUESTONE *et al.*, 2010). Esses genes encontram-se mapeados no

cromossomo murino 17 B1, região ortóloga ao MHC humano, que apresenta regiões *Idds* distintas.

Tais genes são integrantes de diferentes classes do MHC humano e murino: TRIM39 (Trim39), PRR3 (Prr3), VARS2 (Vars2) e CDSN (Cdsn) são genes de classe I; APOM (Apom), STK19 (Stk19) e CYP21A2 (Cyp21a1) pertencem à classe III e, HLA-DOB (H2-Ob) e COL11A2 (Col11a2) são constituintes do MHC de classe II (HORTON *et al.*, 2004).

A região do MHC de classe III consiste em um denso arranjo de genes localizado entre o MHC de classe I e o de classe II em primatas e roedores. Alguns de seus genes, como os da cascata do sistema complemento C4, C2, e o fator B, desempenham um papel na imunidade inata, enquanto outros, como o gene da valina tRNA sintetase (VARS), parecem não exercer função especializada na resposta imune, embora sejam responsáveis por outras funções importantes na célula (XIE *et al.*, 2003).

O MHC de classe III também tem sido referido como MHC central devido à sua localização e é considerado a região mais densa em genes e mais conservada do MHC em mamíferos, apresentando também genes envolvidos na imunidade adaptativa, genes de resposta inflamatória como o fator de necrose tumoral (TNF) e as linfotoxinas A e B, e HSP70 (FLAJNIK & KASAHARA, 2001; ALLCOCK *et al.*, 2000). Além disso, também é considerado a região mais densa em genes do genoma humano, abrangendo aproximadamente 700 kb e apresentando cerca de 60 genes em humano e 61 genes em camundongo (DEAKIN *et al.*, 2006; XIE *et al.*, 2003).

O gene APOM codifica a apolipoproteína apoM pertencente o grupo de proteínas associadas a lipoproteínas, principalmente encontrada nas lipoproteínas de alta densidade (HDL – *High-density lipoproteins*), embora também esteja presente em outras subclasses como as lipoproteínas ricas em triglicerídeos (TGRLP – *Triglycerides-rich lipoproteins*) e aquelas de baixa densidade (LDL – *Low density lipoproteins*) (DAHLBÄCK & NIELSEN, 2009; CHRISTOFFERSEN *et al.*, 2006; XU & DAHLBÄCK, 1999).

Em seres humano e em camundongos, APOM está situado entre os genes BAG6 (BAT3) e GPANK1 (BAT4), genes localizados em um cluster de transcritos associados ao HLA-B (H2-Q1 em camundongo) (DAHLBÄCK & NIELSEN, 2009; XU & DAHLBÄCK, 1999).

Diversos estudos em modelos animais têm sugerido o envolvimento de apoM no metabolismo de HDL e demonstrado que ela apresenta propriedades anti-aterogênicas (CHRISTOFFERSEN *et al.*, 2008; WOLFRUM *et al.*, 2005). Wolfrum e colaboradores (2005) demonstraram que camundongos deficientes em apoM acumularam colesterol em

grandes partículas de HDL além de apresentarem conversão prejudicada do HDL em pre β -HDL, uma subclasse de apolipoproteínas pobres em lipídeos, que atuam como aceptores chave do colesterol periférico celular. Além disso, os pesquisadores superexpressaram a proteína em camundongos deficientes de receptor de LDL (*Ldl*^{-/-}) e mostraram que os animais apresentaram menos lesões ateroscleróticas quando submetidos a uma dieta rica em colesterol.

Neste trabalho, as análises realizadas revelaram uma sutil indução de APOM tanto em seres humanos como em camundongos quando indivíduos diabéticos (pacientes ou camundongos) foram comparados com controles (pacientes saudáveis ou animais pré-diabéticos).

Com relação ao envolvimento no DM1, Wu e colaboradores (2009) conduziram um estudo em indivíduos chineses e caucasianos suecos diabéticos buscando polimorfismos na região promotora do gene APOM. Demonstraram que polimorfismos nesta região altera a atividade do promotor do gene e conferem susceptibilidade ao DM1. Em um estudo realizado com uma doença também autoimune – a artrite reumatóide – em pacientes coreanos, foi observado que polimorfismos no promotor do gene APOM também conferem risco à doença (HU *et al.*, 2011).

Os genes CYP21A2 e STK19 fazem parte do módulo gênico RCCX que pode estar presente na população humana em uma ou mais cópias em cada cromossomo 6. O módulo representa os quatro genes consecutivos STK19 (também conhecido como RP), C4, CYP21 e TNX. O gene C4 pode estar representado por uma, duas ou três cópias na região do MHC de classe III, entretanto, frequentemente está presente em duas cópias – C4A e C4B, que codificam dois isotipos do componente do sistema complemento C4. TNX é representado por TNXA e TNXB, o primeiro considerado um fragmento gênico de TNXB, abrangendo o íntron 32 até o éxon 45 do gene; e o último, o gene funcional, sem deleções que codifica uma proteína de matriz extracelular – a tenascina XB (WU *et al.*, 2009; YU *et al.*, 2000; YANG *et al.*, 1999).

O gene CYP21 é representado por CYP21A2 e CYP21A1P, o primeiro considerado o gene funcional codificando a enzima 21-hidroxilase, e o último, um pseudogene, codificando uma proteína defeituosa. A enzima 21-hidroxilase faz parte da família de oxidases citocromo P450 e participa do processo de biossíntese do cortisol, um hormônio esteróide muito importante fisiologicamente, cuja deficiência tem sido considerada um dos fatores mais relevantes na manifestação da hiperplasia adrenal congênita (HUYNH *et al.*, 2009).

O gene STK19 também se apresenta duplicado no genoma humano, correspondendo ao gene funcional STK19 (ou RP1) e ao pseudogene STK19P (ou RP2) que consiste de um fragmento de RP1, mais especificamente os dois últimos éxons do gene, e codifica uma proteína truncada (CHANG & LEE, 2011; WU *et al.*, 2009). STK19 codifica uma quinase que fosforila resíduos de serina e treonina, e se localiza no núcleo, apresentando assim um possível papel no controle transcricional nuclear (GOMEZ-ESCOBAR *et al.*, 1998).

O módulo gênico RCCX que constitui os quatro genes citados anteriormente variam em número de cópias (mono, bi ou trimodular, mais raramente em quatro cópias) e pode estar associado a distúrbios autoimunes ou a doenças como a hiperplasia adrenal congênita e a síndrome Ehlers Danlos, uma desordem do tecido conectivo (YU *et al.*, 2000; BURCH *et al.*, 1997). Além desta variação no número de cópias do módulo, há também uma diferença na configuração dos genes presentes, podendo ocorrer ausência de algum gene funcional ou mais de uma cópia de um dos genes funcionais, porém ausência de outro e presença apenas do pseudogene, ou ainda genes quiméricos como TNXA/TNXB e CYP21A2/CYP21A1P, todos esses arranjos (incluindo o número de módulos) sendo possíveis devido à ocorrência de *crossing-overs* desiguais (LEE, 2005; YU *et al.*, 2000; RUPERT *et al.*, 1999; YANG *et al.*, 1999).

Os genes TNX, CYP21, C4 e STK19 também estão em bloco no camundongo, entretanto o padrão de duplicação não é idêntico ao de humano e no caso do gene Cyp21, o gene Cyp21a1 é o gene funcional e o Cyp21a2-ps, o pseudogene (YU *et al.*, 2000).

No desenvolvimento deste estudo, observamos repressão de CYP21A2 (Cyp21a1 em camundongo) e STK19 tanto em seres humanos como em camundongos quando indivíduos diabéticos (pacientes ou camundongos) foram comparados com controles (pacientes saudáveis ou animais pré-diabéticos).

Os genes CYP21A2 e STK19 e seus homólogos funcionais murinos apresentam proteínas evolutivamente conservadas como observada pela identidade de sequência de 76,9% e 83,8%, respectivamente (XIE *et al.*, 2003; YU *et al.*, 2000). Além disso, como eles apresentaram mesma modulação entre seres humanos e camundongos, e estão próximos no genoma, podem estar sob a regulação de sequências similares nas duas espécies.

Tais genes apesar de estarem situados em regiões de susceptibilidade ao DMI humano e murino, não exercem claramente um papel no desenvolvimento da doença. Entretanto, o módulo gênico do qual são integrantes já foi associado à artrite reumatóide juvenil (RUPERT *et al.*, 1999) e lúpus eritematoso sistêmico (WU *et al.*, 2009), doenças de caráter autoimune como o diabetes *mellitus* do tipo I.

A região do MHC de classe II é a mais centromérica do MHC humano e apresenta genes com função relacionada à resposta imune adaptativa, como por exemplo os envolvidos na apresentação de antígenos exógenos às células T CD4⁺. Compreende os genes polimórficos α e β , os genes codificantes dos componentes do proteasoma induzido por interferon γ (PSMB8 e PSMB9), transportadores associados à via de apresentação de antígenos de classe I TAP1 e TAP2, entre outros. Além disso, esta região apresenta uma sub-região chamada de classe II estendida que inclui genes relacionados e não relacionados ao sistema imune (HORTON *et al.*, 2004; FLAJNIK & KASAHARA, 2001; ALLCOCK *et al.*, 2000).

A região do MHC de classe II é similar em seres humano e em camundongos em relação à disposição dos genes, considerada sintênica e apresenta grupos de genes ortólogos. Entretanto, alguns genes estão presentes em múltiplas cópias como resultado de duplicações espécie-específicas geradas desde o processo de especiação (KUMÁNOVICS *et al.*, 2003; ALLCOCK *et al.*, 2000).

Os genes COL11A2 e HLA-DOB identificados no presente trabalho fazem parte do MHC de classe II e apresentam os genes ortólogos H2-Ob e Col11a2 no genoma murino. COL11A2 é um gene que integra o MHC de classe II estendido e codifica uma das cadeias do colágeno tipo XI, a cadeia $\alpha 2$ (HORTON *et al.*, 2004; SPRANGER, 1998).

O colágeno tipo XI é uma molécula heterotrimérica formada pelas cadeias $\alpha 1$, $\alpha 2$ e $\alpha 3$, que constitui uma pequena parte do colágeno presente na cartilagem de diversos órgãos e tecidos e tem sido relacionado a uma família de displasias espondilo-epifisárias, como a displasia oto-espondilo-megaepifisária – OSMED, a síndrome de Stickler e outras, devido a mutações nos genes responsáveis pela síntese de qualquer uma das cadeias polipeptídicas formadoras deste tipo de colágeno (BOWEN *et al.*, 2008; SPRANGER, 1998).

Neste trabalho, as análises realizadas revelaram uma sutil indução de COL11A2 tanto em seres humano como em camundongos. Com relação ao envolvimento no DM1, em um estudo com 2.321 famílias de diversos grupos étnicos (sendo a maioria caucasiana) apresentando indivíduos com a doença, Eike e colaboradores (2009) identificaram três regiões com associação ao DMI, sendo uma delas delimitada pelos genes HLA-DPB1/COL11A2/RING1. Entretanto, destacam que mais pesquisas devem ser realizadas a fim de replicar tais achados em outras populações e de conduzir um mapeamento fino nas regiões descritas ou próximo a elas.

Em outro estudo, Bowen e colaboradores (2008) demonstraram a presença das cadeias $\alpha 1$ e $\alpha 2$ do colágeno tipo XI nos tecidos normal e maligno de cólon, e afirmaram que apesar desse tipo de colágeno ser funcionalmente caracterizado em cartilagem, ele pode ser

importante no desenvolvimento inicial de doenças inflamatórias intestinais, como a colite ulcerativa e a doença de Chron, doenças que também são autoimunes.

O HLA-DOB é um gene integrante do MHC de classe II não-clássico, que codifica a cadeia β do heterodímero HLA-DO. Seu homólogo murino H2-Ob também forma um heterodímero que faz par com a cadeia α chamada H2-Oa. O polipeptídeo codificado por esse gene apresenta um alto grau de conservação entre as duas espécies (mais de 70% de identidade na sequência de aminoácidos), sugerindo que ele desempenhe um importante papel funcional na célula (DENZIN *et al.*, 2005; HomoloGene, <http://www.ncbi.nlm.nih.gov/homologene/>, acesso em 3 de julho de 2012).

O HLA-DOB participa da via de processamento e apresentação de antígeno, exercendo um papel modulador no HLA-DM (H2-M) que está envolvido no carregamento de peptídeo nos compartimentos endossomais das células apresentadoras de antígenos profissionais (DENZIN *et al.*, 2005; BROCKE *et al.*, 2003).

Apesar desse gene ser preferencialmente expresso em células B humanas e murinas, antígenos do MHC de classe II podem ser expressos em linfócitos T ativados por estimulação antigênica ou mitogênica (HENG *et al.*, 2008; TONELLE *et al.*, 1985).

Neste estudo, foi observada uma indução considerável de HLA-DOB tanto em seres humanos como em camundongos.

Yi e colaboradores (2011) demonstraram que a superexpressão de H2-Ob em células dendríticas de camundongos NOD apresentou um papel protetor no desenvolvimento da doença. Os camundongos que superexpressaram o gene apresentaram insulite reduzida e não desenvolveram o diabetes autoimune devido à apresentação alterada de um repertório do MHCII-peptídeo próprio, provavelmente evitando a ativação de células T diabetogênicas.

Embora nossas análises tenham revelado indução do gene na doença em seres humanos e camundongos, vale ressaltar que foram observados os níveis de mRNA e não os protéicos. Assim, os mRNAs de HLA-DOB podem estar sob regulação de miRNAs ou outros elementos que participam do controle pós-transcricional.

A região do MHC de classe I apresenta genes de classe I clássico, de classe I não-clássico e genes não-classe I, que desempenham funções tanto no sistema imune como funções não relacionadas a ele. Aqueles de classe I clássica são altamente polimórficos, ubiquamente expressos e estão envolvidos na apresentação de antígenos às células T CD8⁺ e também na resposta imune mediada pelas células *natural killer* (NK). (HORTON *et al.*, 2004; ALLCOCK *et al.*, 2000).

Os genes de MHC classe I não-clássico são expressos em um número limitado de tecidos, apresentam pouco polimorfismo e desempenham funções imunológicas, algumas especializadas, mas também funções não relacionadas ao sistema imune. Em humano, vários genes não-classe I se encontram localizados entre o HLA-C e o HLA-E, como fatores de transcrição, genes reguladores de crescimento celular, genes relacionados à patogênese da psoríase, entre outros, sendo considerados ortólogos entre espécies de diferentes ordens de mamíferos (KUMÁNOVICS *et al.*, 2003).

De acordo com a hipótese da estrutura, os diversos genes de classe I são considerados derivados de um gene ancestral comum que sofreu duplicações independentes na espécie humana e na murina, enquanto que os não-classe I já existiam antes da especiação (ALLCOCK *et al.*, 2000).

Os genes CDSN, PRR3, TRIM39 e VARS2 identificados nas análises fazem parte da região do MHC de classe I na espécie humana e na murina (*M. musculus*). TRIM39 é um gene pertencente à família de motivo tripartite, onde o motivo que dá nome à família consiste em três domínios ligadores de zinco: um domínio RING (*Really interesting new gene*) e dois B-box; apresentando uma região *coiled-coil* em seguida aos domínios citados (LEE *et al.*, 2009; REYMOND *et al.*, 2001).

A proteína TRIM39 foi recentemente identificada como uma proteína ligadora de MOAP-1, um fator apoptótico que sofre degradação constitutivamente pelo sistema ubiquitina-proteasoma e é induzido sob estímulo de apoptose pela ocorrência de inibição do processo de poli-ubiquitinação (LEE *et al.*, 2009). No estudo realizado por Lee e colaboradores (2009), foi observado que TRIM39 aumenta a estabilidade de MOAP-1 e modula a sensibilidade de células induzidas à apoptose por etoposídeo.

No presente trabalho, o gene TRIM39 se apresentou levemente reprimido em ambas as espécies na doença em relação à condição controle. Tal gene não apresenta um envolvimento claro no diabetes autoimune, apesar de estar situado em regiões de susceptibilidade previamente descritas em seres humanos e em camundongos.

Entretanto, recentemente um estudo conduzido em pacientes e indivíduos saudáveis japoneses identificou um SNP no gene TRIM39 e o associou à doença de Behçet, uma patologia crônica de caráter inflamatório e sistêmico, caracterizada por úlceras orais recorrentes, úlceras genitais, lesões cutâneas e uveíte (KURATA *et al.*, 2010).

O gene CDSN codifica uma proteína secretada por queratinócitos chamada corneodesmosina que é incorporada nos desmossomos e sua degradação tem sido sugerida como essencial para ocorrência do processo normal de descamação (WU *et al.*, 2011). Já o

gene PRR3 codifica o polipeptídeo 3 rico em prolina, também conhecido como CAT56, e o VARS2, também conhecido como VARS2L, codifica uma possível valil-tRNA sintetase mitocondrial (<http://www.ncbi.nlm.nih.gov/>).

Nas análises realizadas, os genes CDSN e VARS2 estão reprimidos em seres humanos e em camundongos, enquanto o gene PRR3 se apresentou induzido.

Em relação ao CDSN, um polimorfismo no gene foi associado à psoríase (AHNINI *et al.*, 1999), uma doença autoimune que acomete a pele, e um SNP no gene PRR3 próximo ao HLA-E foi associado à espondilite anquilosante (PALADINI *et al.*, 2009), também autoimune.

Os outros três genes identificados neste estudo com mesma modulação entre as duas espécies e em regiões de susceptibilidade ao DMI também no modelo animal foram RSBN1, PHTF1 e IL21, todos eles situados fora da região do MHC, estando os dois primeiros localizados na citobanda 13.2 do braço curto do cromossomo 1 humano e no cromossomo murino 3 F2, e o último, na citobanda 27 do braço longo do cromossomo humano 4 e na murina 3 B.

O gene RSBN1 codifica uma proteína preferencialmente expressa em espermátides, que apresenta dois domínios: um *homeobox* e um rico em prolina que provavelmente atua na ativação da transcrição. Além da presença de regiões com tais características, a proteína se localiza no núcleo, o que sugere seu papel na regulação da expressão gênica celular (TAKAHASHI *et al.*, 2004). Apesar desse gene ter expressão caracterizada em testículo, segundo os dados de expressão gênica do *HaemAtlas* (WATKINS *et al.*, 2009) disponíveis no T1DBase, linfócitos T CD8⁺ e CD4⁺ humanos apresentam alta expressão de mRNAs desse gene, embora seu papel nessas células ainda seja desconhecido.

O gene PHTF1 é um gene conservado evolutivamente desde *Drosophila* até mamíferos e codifica um polipeptídeo que apresenta um homeodomínio, sendo considerado pertencente à família de genes *homeobox* (MANUEL *et al.*, 2000). PHTF1 é principalmente expresso em testículo, embora apresente menores níveis de mRNAs em linfócitos T humanos e murinos segundo o banco de dados *BioGPS* (WU *et al.*, 2009). Além disso, foi observado que PHTF1 codifica uma proteína integral de membrana localizada no retículo endoplasmático de células germinativas masculinas em maturação (OYHENART *et al.*, 2003).

Neste estudo, observamos que RSBN1, PHTF1 e IL21 estão reprimidos em diabéticos em relação aos controles tanto nos dados de expressão humanos como nos murinos.

Em relação ao IL21, este codifica uma citocina produzida preferencialmente por células T CD4⁺ ativadas, embora também seja expressa em células T *natural killer* (NKT) ativadas, células T CD8⁺ sob condições específicas, entre outras, e que está envolvida na modulação da proliferação de células que apresentam o receptor IL-21R, como por exemplo, as células T (MONTELEONE *et al.*, 2009). Seu correspondente em camundongo integra uma das regiões de susceptibilidade ao DM1 mais fortes, localizada fora da região genômica do MHC, a *Idd3* (MCGUIRE *et al.*, 2009).

Il21 tem sido considerado, juntamente com Il2, gene candidato envolvido na patogênese do diabetes autoimune experimental responsável pela associação da região *Idd3* à doença (DRIVER *et al.*, 2011) devido a evidências reportadas em diversos estudos, demonstrando que: Il21 é significativamente mais expresso (níveis de mRNAs) em células T CD4⁺ isoladas do baço de camundongos NOD, bem como em linfócitos infiltrados do pâncreas, em relação a camundongos NOD^{B6.Idd3}, uma linhagem congênica que apresenta o intervalo genômico da *Idd3* correspondente ao encontrado na linhagem C57BL/6 (MCGUIRE *et al.*, 2009); tal citocina é necessária para o desenvolvimento do diabetes, uma vez que animais (*background* NOD) *knockout* para o receptor de Il-21 (IL-21R) apresentaram proteção completa contra a doença, não desenvolvendo-a (SUTHERLAND *et al.*, 2009); Il-21 é importante para a infiltração de células no pâncreas, uma vez que camundongos NOD submetidos à neutralização de Il-21 apresentaram redução na insulite e reversão do diabetes quando sujeitos a um transplante de ilhotas pancreáticas em conjunto com terapia para neutralizar a interleucina em questão (MCGUIRE *et al.*, 2011).

Apesar dos níveis de expressão gênica de IL-21 estarem reduzidos na comparação realizada neste trabalho, está claro que tal gene não é o único componente envolvido na patogênese do DM1, uma vez que a doença sofre influência de diversos fatores genéticos e ambientais. Além disso, foi observado que as *Idd3* e *Idd1* sozinhas não são suficientes para o desenvolvimento do DM1 em camundongos congênicos (F2 de B6.NOD-*H-2*^{g7} x C3H/He) para essas regiões (IKEGAMI *et al.*, 2004).

Além dos doze genes comentados anteriormente, dois genes de particular importância para o desenvolvimento da doença foram identificados neste trabalho com modulação transcricional diferente entre as duas espécies: CTLA4 e IL2RA. CTLA4 codifica uma proteína que desempenha importantes funções no sistema imune, como a regulação da resposta de linfócitos T efetores e a manutenção da função de células T reguladoras (Tregs) (DRIVER *et al.*, 2011). Já o gene IL2RA codifica a cadeia α do receptor de IL2, sendo muito

conhecido por ser expresso nas células Tregs, embora outros tipos celulares também apresentem expressão do gene (DRIVER *et al.*, 2011).

CTLA4 está localizado na região de susceptibilidade humana 2q33.2 e na murina, *Idd5.1*, sendo considerado gene candidato para as regiões humana e murina. Já o gene IL2RA está situado na região de susceptibilidade humana 10p15.1, não sendo mapeado em região de susceptibilidade murina, embora também exerça um papel importante no diabetes desenvolvido em camundongos NOD (DENDROU & WICKER, 2008).

Neste trabalho, observamos que CTLA4 está induzido nos dados de *microarrays* humanos de acordo com a comparação realizada, e que em camundongos NOD o gene se apresentou muito reprimido na doença em relação à condição controle. Esses resultados são interessantes já que CTLA4 é um gene muito expresso em células Treg (HENG *et al.*, 2008) e participa da regulação negativa da ativação de células T (NISTICÒ *et al.*, 1996).

Em relação a IL2RA, nossas análises revelaram indução do gene nos dados de expressão humanos, enquanto nos dados murinos o gene se apresentou reprimido. Em relação aos camundongos NOD diabéticos, uma repressão do gene em questão poderia indicar uma menor proporção na população de células Tregs, fundamentais na regulação de células T autorreativas (LOWE *et al.*, 2007). Em um estudo realizado por Lowe e colaboradores (2007), dois SNPs encontrados no gene IL2RA em pacientes diabéticos com DM1 foram associados a menores níveis séricos de IL2RA solúvel.

Outro achado interessante em nosso trabalho foram os 59 genes em regiões de sintenia entre seres humanos e camundongos. Observamos que a maioria deles são integrantes da região de MHC humana e da murina, sendo a segunda região mais representada a humana 12q13 e a murina 10 D3. Uma grande região do braço longo do cromossomo 12 humano, abrangendo a 13, tem sido caracterizada com uma das regiões mais densas em genes do genoma humano (XIE *et al.*, 2003).

Além disso, cinquenta e oito deles apresentaram identidade de sequência de aminoácidos acima de 70%, o que demonstra que os produtos desses genes devem exercer funções similares em ambas as espécies, apresentando conservação funcional. Como esses resultados são observados no presente estudo conduzido sobre o diabetes autoimune, podemos sugerir que tais genes, principalmente aqueles com mesma modulação entre as duas espécies, devem ser importantes na doença já que foram mantidos na espécie humana com similar organização genômica após aproximadamente 100 milhões de anos de separação da espécie murina (JORDAN *et al.*, 2004) e vários deles estão situados em regiões de susceptibilidade ao DM1 em ambas as espécies.

Em relação à comparação dos perfis de expressão gênica entre seres humanos e camundongos, observamos que a maioria dos genes identificados neste trabalho (42) não apresentou modulação semelhante entre as duas espécies. Vale salientar que a observação de diferenças nos níveis de expressão gênica é particularmente difícil de interpretar, uma vez que as explicações podem ser confundidas por influências ambientais e genéticas. Dessa maneira, poderíamos assumir que divergências nas experiências vivenciadas por espécies distintas podem levar a perturbações na regulação gênica, bem como a um aumento na variação dos níveis de expressão gênica (ROMERO *et al.*, 2012).

Diferentemente, genes com baixa variação nos níveis de expressão gênica entre espécies estão provavelmente sob a ação de seleção estabilizadora de regulação gênica, sendo mais resistentes a mudanças ambientais (ROMERO *et al.*, 2012).

Além disso, vale lembrar que as células humanas e murinas utilizadas neste trabalho foram linfócitos periféricos, mostrando que os resultados obtidos podem refletir diferenças no sistema imune das duas espécies. Tais divergências seriam atribuídas à distância evolutiva entre humano e camundongo, mas também às diferenças no tamanho, expectativa de vida e nichos ecológicos entre as duas espécies, fazendo com que seus sistemas imunológicos enfrentem patógenos diferentes (MESTA & HUGHES, 2004).

Diversos estudos vêm sendo realizados sobre a evolução do transcriptoma, onde alguns pesquisadores afirmam que as taxas de diferenças nos níveis de expressão gênica entre espécies são similares em pseudogenes e genes codificantes de proteínas de primatas, sugerindo um modelo de evolução puramente neutra, ou seja, sem forças seletivas (KHAITOVICH *et al.*, 2004); enquanto outros afirmam que as mudanças no nível de expressão gênica observadas em pares de ortólogos humanos e murinos são significativamente menores do que aquelas apresentadas por genes aleatórios das duas espécies em questão, sugerindo um modelo de evolução neutra acompanhado de pressão seletiva (JORDAN *et al.*, 2004; LIAO & ZHANG, 2006).

Embora seja possível realizar este tipo de estudo devido ao desenvolvimento de tecnologias de alta resolução que permitem a obtenção de dados de expressão gênica em larga escala, como os *microarrays* e, mais recentemente, a metodologia de RNAseq (WANG *et al.*, 2009), tal abordagem ainda representa um desafio para os pesquisadores (ROMERO *et al.*, 2012).

Apesar das diferenças na modulação de *fold change* observadas nos dados de expressão gênica humano e murino deste trabalho, foi possível estabelecer um paralelo da doença entre as duas espécies. Os doze genes da lista final foram encontrados em regiões de

susceptibilidade tanto em humano como em camundongo, e alguns deles já foram melhor estudados em camundongos NOD e em pacientes diabéticos como comentado anteriormente.

Além disso, vale salientar que esse tipo de abordagem baseado em dados de *oligo microarrays* de genes humanos e murinos é especialmente importante para a pesquisa biomédica, pois possibilita um maior entendimento das bases moleculares das doenças, bem como um maior conhecimento das semelhanças entre humano e camundongo, fornecendo subsídios para o desenvolvimento de terapias em modelos animais aplicáveis na es humana (MESTA & HUGHES, 2004).

Conclusões

7 CONCLUSÕES

Neste trabalho conseguimos estabelecer um paralelo entre o DM1 humano e o murino em relação à expressão de RNAs mensageiros de genes de susceptibilidade à doença em regiões sintênicas entre as duas espécies. Observamos que a maioria dos genes estudados pertence à região do complexo principal de histocompatibilidade, demonstrando que tais genes são os que mais compartilham semelhanças no perfil transcricional.

Observamos que os genes selecionados para comparação do perfil transcricional entre as duas espécies apresentam um alto grau de identidade em seus produtos proteicos, evidenciando uma conservação funcional e o possível papel de tais genes no desenvolvimento do diabetes autoimune, com destaque para aqueles com mesma modulação transcricional entre humano e camundongo, e situados em regiões de susceptibilidade nas duas espécies.

Os resultados obtidos contribuem para uma visão mais abrangente dos mecanismos moleculares envolvidos na etiologia da autoimunidade, demonstrando que além do compartilhamento de regiões de susceptibilidade associadas por DNA, há um compartilhamento dos perfis de expressão de genes entre as duas espécies, em que papel na doença ainda é pouco conhecido.

Isso evidencia a importância das análises comparativas de dados do transcriptoma entre espécies na descoberta de novos mecanismos envolvidos no controle genético molecular dos processos biológicos relacionados ao DM1, facilitando a seleção de genes de interesse para futuras pesquisas sobre a contribuição funcional desses na doença.

Referências Bibliográficas

8 REFERÊNCIAS BIBLIOGRÁFICAS

ADA – AMERICAN DIABETES ASSOCIATION. Diagnosis and classification of diabetes *mellitus*. **Diabetes Care**. v. 34, supplement 1, p. 62-69, 2011.

AHNINI, R. T. *et al.* Novel genetic association between the corneodesmosin (MHC S) gene and susceptibility to psoriasis. **Human Molecular Genetics**. V. 8, n. 6, p. 1135-1140, 1999.

ALBERTI, K. G.; ZIMMET, P. Z. Definition, diagnosis and classification of diabetes *mellitus* and its complications. Part 1: diagnosis and classification of diabetes *mellitus* provisional report of a WHO consultation. **Diabetic Medicine**. V. 15, n. 7, p. 539-553, 1998.

ALLCOCK, R. J. N.; MARTIN, A. M.; PRICE, P. The mouse as a model for the effects of MHC genes on human disease. **Immunology Today**. V.21, n. 7, p. 328-332, 2000.

ANDERSON, M. S.; BLUESTONE, J. A. The NOD mouse: a model of immune dysregulation. **Annual Review of Immunology**. V. 23, p. 447-485, 2005.

ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Nature Reviews Genetics**. V. 3, n. 4, p. 299-309, 2002.

BARRETT, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. **Nature Genetics**. V. 41, n. 6, p. 703-707, 2009.

BELL, G. I.; HORITA, S.; KARAM, J. H. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes *mellitus*. **Diabetes**. V. 33, n. 2, p. 176-183, 1984.

BLACKBURN, C. C.; MANLEY, N. R. Developing a new paradigm for thymus organogenesis. **Nature Reviews Immunology**. V. 4, n. 4, p. 278-289, 2004.

BLUESTONE, J. A.; HEROLD, K.; EISENBARTH, G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. **Nature**. V. 464, n. 7293, p. 1293-1300, 2010.

BORCHERS, A. T.; UIBO, R.; GERSHWIN, M. E. The geoepidemiology of type 1 diabetes. **Autoimmunity Reviews**. V. 9, n. 5, p. A355-365, 2010.

BORG, H.; FERNLUND, P.; SUNDKVIST, G. Protein tyrosine phosphatase-like protein IA2-antibodies plus glutamic acid decarboxylase 65 antibodies (GADA) indicates autoimmunity as frequently as islet cell antibodies assay in children with recently diagnosed diabetes *mellitus*. **Clinical Chemistry**. V. 43, n. 12, p. 2358-2363, 1997.

BOTTINI, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. **Nature Genetics**. V. 36, n. 4, p. 337-338, 2004.

BOWEN, K. B. *et al.* Immunohistochemical localization of collagen type XI $\alpha 1$ and $\alpha 2$ chains in human colon tissue. **Journal of Histochemistry & Cytochemistry**. V. 56, n. 3, p. 275-283, 2008.

BØYUM, A. Isolation of mononuclear cells and granulocytes from blood. **Scandinavian Journal of Clinical and Laboratory Investigation**. V. 21, Supplementum 97, p. S77-S89, 1968.

BRAZMA, A.; VILO, J. Gene expression data analysis. **FEBS Letters**. V. 480, n. 1, p. 17-24, 2000.

BROCKE, P. *et al.* Downmodulation of antigen presentation by H2-O in B cell lines and primary B lymphocytes. **European Journal of Immunology**. V. 33, n. 2, p. 411-421, 2003.

BURCH, G. H. *et al.* Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. **Nature Genetics**. V. 17, n., p. 104-108, 1997.

BURREN, O. S. *et al.* Development of an integrated genome informatics, data management and workflow infrastructure: A toolbox for the study of complex disease genetics. **Human Genomics**. V. 1, n. 2, p. 98-109, 2004.

BURREN, O. S. *et al.* T1Dbase: update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. **Nucleic Acids Research**. V. 39, Database issue, p. D997-D1001, 2011.

CHANG, S.-F.; LEE, H.-H. Analysis of the CYP21A2 Gene with Intergenic Recombination and Multiple Gene Deletions in the RCCX Module. **Genetic Testing and Molecular Biomarkers**. V. 15, n. 1-2, p. 1-8, 2011.

CHRISTOFFERSEN, C.; DAHLBÄCK, B.; NIELSEN, L. B. Apolipoprotein M: progress in understanding its regulation and metabolic functions. **Scandinavian Journal of Clinical and Laboratory Investigation**. V. 66, n. 7, p. 631-637, 2006.

CHRISTOFFERSEN, C. *et al.* Effect of apolipoprotein M on high density lipoprotein metabolism and atherosclerosis in low density lipoprotein receptor knock-out mice. **The Journal of Biological Chemistry**. V. 283, n. 4, p. 1839-1847, 2008.

CHURCH, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. **PLOS Biology**. V. 7, n. 5, e1000112, 2009.

CONCANNON, P.; RICH, S. S.; NEPOM, G. T. Genetics of type 1A diabetes. **The New England Journal of Medicine**. V. 360, n. 16, p. 1646-1654, 2009.

CURDWORTH, A. G.; WOODROW, J. C. Evidence for HL-A-linked genes in “juvenile” diabetes *mellitus*. **British Medical Journal**. V. 3, n. 5976, p. 133-135, 1975.

DAHLBÄCK, B.; NIELSEN, L. B. Apolipoprotein M affecting lipid metabolism or just catching a ride with lipoproteins in the circulation? **Cellular and Molecular Life Sciences: CMLS**. V. 66, n. 4, p. 559-564, 2009.

DEAKIN, J. E. *et al.* Evolution and comparative analysis of the MHC Class III inflammatory region. **BMC Genomics**. V. 7, 281, 2006.

DE GRIJSE, J. *et al.* Predictive power of screening for antibodies against insulinoma-associated protein 2 beta (IA-2beta) and zinc transporter-8 to select first-degree relatives of type 1 diabetic patients with risk of rapid progression to clinical onset of the disease: implications for prevention trials. **Diabetologia**. V. 53, n. 3, p. 517-524, 2010.

DELOVITCH, T. L.; SINGH, B. The nonobese diabetic mouse as a model of autoimmune diabetes: immune dysregulation gets the NOD. **Immunity**. V. 7, n. 6, p. 727-738, 1997.

DENDROU, C. A.; WICKER, L. S. The IL-2/CD25 pathway determines susceptibility to T1D in humans and NOD mice. **Journal of Clinical Immunology**. V. 28, n. 6, p. 685-696, 2008.

DENZIN, L. K. *et al.* Right place, right time, right peptide: DO keeps DM focused. **Immunological Reviews**. V. 207, p. 279-292, 2005.

DERNOWSEK, J. A. **Colinearidade de transcriptomas durante a diferenciação osteoblástica de células tronco mesenquimais adultas humanas e murinas**. Dissertação (Mestrado em Genética) – Universidade de São Paulo, Ribeirão Preto, 2010.

DRIVER, J. P.; SERREZE, D. V.; CHEN, Y. G. Mouse models for the study of autoimmune type 1 diabetes: a NOD to similarities and differences to human disease. **Seminars in Immunopathology**. V. 33, n. 1, p. 67-87, 2011.

DURINOVIC-BELLÓ, I. *et al.* Insulin gene VNTR genotype associates with frequency and phenotype of the autoimmune response to proinsulin. **Genes and Immunity**. V. 11, n. 2, p. 188-193, 2010.

EIKE, M. C. *et al.* Conditional analyses on the T1DGC MHC dataset: novel associations with type 1 diabetes around HLA-G and confirmation of HLA-B. **Genes and Immunity**. V. 10, n. 1, p. 56-67, 2009.

EISEN, M. B. *et al.* Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences of the United States of America (PNAS)**. V. 95, n. 25, p. 14863-14868, 1998.

ERLICH, H. *et al.* HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. **Diabetes**. V. 57, n. 4, p. 1084-1092, 2008.

FLAJNIK, M. F.; KASAHARA, M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. **Immunity**. V. 15, n. 3, p. 351-362, 2001.

FLICEK, P. *et al.* Ensembl 2012. **Nucleic Acids Research**. V. 40, Database issue, p. D84-D90, 2012.

FORNARI, T. A. *et al.* Age-related deregulation of Aire and peripheral tissue antigen genes in the thymic stroma of non-obese diabetic (NOD) mice is associated with autoimmune type 1 diabetes mellitus (DM-1). **Molecular and Cellular Biochemistry**. V. 342, n. 1-2, p. 21-28, 2010.

GALE, E. A. M. The rise of childhood type 1 diabetes in the 20th century. **Diabetes**. V. 51, n. 12, p. 3353-3361, 2002.

GOMEZ-ESCOBAR, N. *et al.* The G11 gene located in the major histocompatibility complex encodes a novel nuclear serine/threonine protein kinase. **The Journal of Biological Chemistry**. V. 273, n. 47, p. 30954-30960, 1998.

GREGORY, S. G. *et al.* A physical map of the mouse genome. **Nature**. V. 418, n. 6899, p. 743-750, 2002.

GRIGORYEV, D. N. *et al.* Orthologous gene-expression profiling in multi-species models: search for candidate genes. **Genome Biology**. V. 5, n. 5, R34, 2004.

HAKONARSON, H. *et al.* A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. **Nature**. V. 448, n. 7153, p. 591-594, 2007.

HAKONARSON, H. *et al.* A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. **Diabetes**. V. 57, n. 4, p. 1143-1146, 2008.

HALLER, M. J.; ATKINSON, M. A.; SCHATZ, D. Type 1 diabetes *mellitus*: etiology, presentation, and management. **Pediatric Clinics of North America**. V. 52, n. 6, p. 1553-1578, 2005.

HENG, T. S. *et al.* The Immunological Genome Project: networks of gene expression in immune cells. **Nature Immunology**. V. 9, n. 10, p. 1091-1094, 2008.

HIRSCHHORN, J. N.; DALY, M. J. Genome-wide association studies for common diseases and complex traits. **Nature Reviews Genetics**. V. 6, n. 2, p. 95-108, 2005.

HORTON, R. *et al.* Gene map of the extended human MHC. **Nature Reviews Genetics**. V. 5, n. 12, p. 889-899, 2004.

HU, H.-J. *et al.* Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis. **Experimental and Molecular Medicine**. V. 43, n. 11, p. 613-621, 2011.

HUNT, E. *et al.* The visual language of synteny. **OMICS: A Journal of Integrative Biology**. V. 8, n. 4, p. 289-305, 2004.

HUYNH, T. *et al.* The clinical and biochemical spectrum of congenital adrenal hyperplasia secondary to 21-hydroxylase deficiency. **Clinical Biochemist Reviews**. V. 30, n. 2, p. 75-86, 2009.

HYTTINEN, V. *et al.* Genetic liability of type 1 diabetes and the onset age among 22,650 young finnish twin pairs: a nationwide follow-up study. **Diabetes**. V. 52, n. 4, p. 1052-1055, 2003.

IKEGAMI, H. *et al.* Idd1 and Idd3 are necessary but not sufficient for development of type 1 diabetes in NOD mouse. **Diabetes Research and Clinical Practice**. V. 66, Supplement 1, p. S85-90, 2004.

INVERNIZZI, P.; GERSHWIN, M. E. The genetics of human autoimmune disease. **Journal of Autoimmunity**. V. 33, n. 3-4, p. 290-299, 2009.

IRIE, J. *et al.* Genome-wide microarray expression analysis of CD4⁺ T Cells from nonobese diabetic congenic mice identifies Cd55 (Daf1) and Acadl as candidate genes for type 1 diabetes. **The Journal of Immunology**. V. 180, n. 2, p. 1071-1079, 2008.

JORDAN, I. K.; MARÍÑO-RAMÍREZ, L.; KOONIN, E. V. Evolutionary significance of gene expression divergence. **Gene**. V. 345, n. 1, p. 119-126, 2004.

KAIZER, E. C. *et al.* Gene expression in peripheral blood mononuclear cells from children with diabetes. **The Journal of Clinical Endocrinology and Metabolism**. V. 92, n. 9, p. 3705-3711, 2007.

KHAITOVICH, P. *et al.* A neutral model of transcriptome evolution. **PLOS Biology**. V. 2, n. 5, E132, 2004.

KIKUTANI, H.; MAKINO, S. The murine autoimmune diabetes model: NOD and related strains. **Advances in Immunology**. V. 51, p. 285-322, 1992.

KING, C. *et al.* Homeostatic expansion of T cells during immune insufficiency generates autoimmunity. **Cell**. V. 117, n. 2, p. 265-277, 2004.

KISHIMOTO, H.; SPRENT, J. A defect in central tolerance in NOD mice. **Nature Immunology**. V. 2, n. 11, p. 1025-1031, 2001.

KLUGER, N.; RANKI, A.; KROHN, K. APECED: is this a model for failure of T cell and B cell tolerance? **Frontiers in Immunology**. V. 3, 232, 2012.

KUMÁNOVICS, A.; TAKADA, T.; LINDAHL, K. F. Genomic organization of the mammalian MHC. **Annual Review of Immunology**. V. 21, p. 629-657, 2001.

KURATA, R. *et al.* TRIM39 and RNF39 are associated with Behçet's disease independently of HLA-B*51 and -A*26. **Biochemical and Biophysical Research Communications**. V. 401, n. 4, p. 533-537, 2010.

KYIEWSKI, B.; DERBINSKI, J. Self-representation in the thymus: an extended view. **Nature Reviews Immunology**. V. 4, n. 9, p. 688-698, 2004.

LEE, H.-H. Chimeric CYP21P/CYP21 and TNXA/TNXB genes in the RCCX module. **Molecular Genetics and Metabolism**. V. 84, n. 1, p. 4-8, 2005.

LEE, S. S. *et al.* TRIM39 is a MOAP-1-binding protein that stabilizes MOAP-1 through inhibition of its poly-ubiquitination process. **Experimental Cell Research**. V. 315, n. 7, p. 1313-1325, 2009.

LEHUEN, A. *et al.* Immune cell crosstalk in type 1 diabetes. **Nature Reviews Immunology**. V. 10, n. 7, p. 501-513, 2010.

LEITER, E. H.; PROCHAZKA, M.; COLEMAN, D. L. The non-obese diabetic (NOD) mouse. **American Journal of Pathology**. V. 128, n. 2, p. 380-383, 1987.

LIAO, B.-Y.; ZHANG, J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. **Molecular Biology and Evolution**. V. 23, n. 3, p. 530-540, 2006.

LIEW, C.-C. *et al.* The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. **Journal of Laboratory and Clinical Medicine**. V. 147, n. 3, p. 126-132, 2006.

LOWE, C. E. *et al.* Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. **Nature Genetics**. V. 39, n. 9, p. 1074-1082, 2007.

MAGALHÃES, D. A. *et al.* Promiscuous gene expression in the thymus: the root of central tolerance. **Clinical and Developmental Immunology**. V. 13, n. 2-4, p. 81-99, 2006.

MAKINO, S. *et al.* Breeding of nonobese, diabetic strain of mice. **Jikken Dobutsu**. V. 29, n. 1, p. 1-13, 1980.

MANUEL, A. *et al.* Molecular characterization of a novel gene family (PHTF) conserved from *Drosophila* to mammals. **Genomics**. V. 64, n. 2, p. 216-220, 2000.

MATHIS, D.; GAGLIA, J. Autoimmune diabetes. In: WEISSLEDER, R. *et al.* **Molecular Imaging**. 1st ed., PMPH-USA, 2009.

MCCARTHY, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. **Nature Reviews Genetics**. V. 9, n. 5, p. 356-369, 2008.

MCGUIRE, H. M. *et al.* Loss of parity between IL-2 and IL-21 in the NOD Idd3 locus. **PNAS**. V. 106, n. 46, p. 19438-19443, 2009.

MCGUIRE, H. M. *et al.* Interleukin-21 is critically required in autoimmune and allogeneic responses to islet tissue in murine models. **Diabetes**. V. 60, n. 3, p. 867-875, 2011.

MEHERS, K. L.; GILLESPIE, K. M. The genetic basis for type 1 diabetes. **British Medical Bulletin**. V. 88, n. 1, p. 115-129, 2008.

MELANITOU, E. Functional genomics in early autoimmunity. **Annals of The New York Academy of Sciences**. V. 1050, p. 64-72, 2005.

MESTAS, J.; HUGHES, C. C. W. Of mice and not men: differences between mouse and human immunology. **The Journal of Immunology**. V. 172, n. 5, p. 2731-2738, 2004.

MONTELEONE, G.; PALLONE, F.; MACDONALD, T. T. Interleukin-21 as a new therapeutic target for immune-mediated diseases. **Trends in Pharmacological Sciences**. V. 30, n. 8, p. 441-447, 2009.

NEGRATO, C. A. *et al.* Temporal trends in incidence of Type 1 diabetes between 1986 and 2006 in Brazil. **Journal of Endocrinology Investigation**. V. 33, n. 6, p. 373-377, 2010.

NEJENTSEV, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. **Nature**. V. 450, n. 7171, p. 887-892, 2007.

NERUP, J. *et al.* HL-A antigens and diabetes *mellitus*. **The Lancet**. V. 2, n. 7885, p. 864-866, 1974.

NISTICÒ, L. *et al.* The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. **Human Molecular Genetics**. V. 5, n. 7, p. 1075-1080, 1996.

NOBLE, J. A.; ERLICH, H. A. Genetics of type 1 diabetes. **Cold Spring Harbor Perspectives in Medicine**. V. 2, n. 1, a007732, 2012.

NOWELL, C. S. *et al.* Thymus and parathyroid organogenesis. In: LANZA, R.; LANGER, R.; VACANTI, J. **Principles of Tissue Engineering**. 3rd ed., Elsevier, 2007.

OKADA, H. *et al.* The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. **Clinical and Experimental Immunology**. V. 160, n. 1, p. 1-9, 2010.

OYHENART, J. *et al.* Phtf1 is an integral membrane protein localized in an endoplasmic reticulum domain in maturing male germ cells. **Biology of Reproduction**. V. 68, n. 3, p. 1044-1053, 2003.

PALADINI, F. *et al.* HLA-E gene polymorphism associates with ankylosing spondylitis in Sardinia. **Arthritis Research and Therapy**. V. 11, n. 6, p. R171, 2009.

PASSARGE, E.; HORSTHEMKE, B.; FARBER, R. A. Incorrect use of the term synteny. **Nature Genetics**. V. 23, n. 4, p. 387, 1999.

PATEL, P.; MACEROLLO, A. Diabetes *mellitus*: diagnosis and screening. **American Family Physician**. V. 81, n. 7, p. 863-870, 2010.

PATTERSON, C. C. *et al.* Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. **Lancet**. V. 373, n. 9680, p. 2027-2033, 2009.

PEARSON, T. A.; MANOLIO, T. A. How to interpret a genome-wide association study. **JAMMA**. V. 299, n. 11, p. 1335-1344, 2008.

PEREIRA, E. *et al.* Immunosuppressive therapy modulates T lymphocyte gene expression in patients with systemic lupus erythematosus. **Immunology**. V. 113, n. 1, p. 99-105, 2004.

PLANAS, R.; PUJOL-BORREL, R.; VIVES-PI, M. Global gene expression changes in type 1 diabetes: insights into autoimmune response in the target organ and in the periphery. **Immunology Letters**. V. 133, n. 2, p. 55-61, 2010.

POCIOT, F. *et al.* Genetics of type 1 diabetes: what's next? **Diabetes**. V. 59, n. 7, p. 1561-1571, 2010.

POLYCHRONAKOS, C.; LI, Q. Understanding type 1 diabetes through genetics: advances and prospects. **Nature Reviews Genetics**. V. 12, n. 11, p. 781-792, 2011.

PUGLIESE, A. *et al.* The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. **Nature Genetics**. V. 15, n. 3, p. 293-297, 1997.

QU, H. Q. *et al.* Remapping the type I diabetes association of the CTLA4 locus. **Genes and Immunity**. V. 10, supplement 1, p. S27-S32, 2009a.

QU, H. Q. *et al.* The type I diabetes association of the IL2RA locus. **Genes and Immunity**. V. 10, supplement 1, p. S42-S48, 2009b.

RASSI, D. M. *et al.* Gene expression profiles stratified according to type 1 diabetes *mellitus* susceptibility regions. **Annals of the New York Academy of Sciences**. V. 1150, p. 282-289, 2008.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. **R Foundation for Statistical Computing**. Vienna, Austria, 2011. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

REDONDO, M. J. *et al.* Concordance for islet autoimmunity among monozygotic twins. **The New England Journal of Medicine**. V. 359, n. 26, p. 2849-2850, 2008.

REYMOND, A. *et al.* The tripartite motif family identifies cell compartments. **The European Molecular Biology Organization (EMBO) Journal**. V. 20, n. 9, p. 2140-2151, 2001.

REYNIER, F. *et al.* Specific gene expression signature associated with development of autoimmune type-I diabetes using whole-blood microarray analysis. **Genes and Immunity**. V. 11, n. 3, p. 269-278, 2010.

RIESEBERG, L. H. Chromosomal rearrangements and speciation. **Trends in Ecology and Evolution**. V. 16, n. 7, p. 351-358, 2001.

ROEP, B. O. *et al.* Molecular mimicry in type 1 diabetes: immune cross-reactivity between islet autoantigen and human cytomegalovirus but not Cocksackie virus. **Annals of the New York Academy of Sciences**. V. 958, p. 162-165, 2002.

ROMERO, I. G.; RUVINSKY, I.; GILAD, Y. Comparative studies of gene expression and the evolution of gene regulation. **Nature Reviews Genetics**. V.13, n. 7, p. 505-516, 2012.

RUAN, Y. *et al.* Interrogating the transcriptome. **Trends in Biotechnology**. V. 22, n. 1, p. 23-30, 2004.

RUPERT, K. L.; RENNEBOHM, R. M.; YU, C. Y. An unequal crossover between the RCCX modules of the human MHC leading to the presence of a CYP21B gene and a tenascin TNXB/TNXA-RP2 recombinant between C4A and C4B genes in a patient with juvenile rheumatoid arthritis. **Experimental and Clinical Immunogenetics**. V. 16, n. 2, p. 81-97, 1999.

SAYERS, E. W. *et al.* Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**. V. 40, Database issue, p. D13-D25, 2012.

SINGAL, D. P.; BLAJCHMAN, M. A. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes *mellitus*. **Diabetes**. V. 22, n. 6, p. 429-432, 1973.

SMYTH, D. *et al.* Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. **Diabetes**. V. 53, n. 11, p. 3020-3023, 2004.

SMYTH, D. J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. **Nature Genetics**. V. 38, n. 6, p. 617-619, 2006.

SPRANGER, J. The type XI collagenopathies. **Pediatric Radiology**. V. 28, n. 10, p. 745-750, 1998.

SU, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. **PNAS**. V. 101, n. 16, p. 6062-6067, 2004.

SUTHERLAND, A. P. R. *et al.* Interleukin-21 is required for the development of type 1 diabetes in NOD mice. **Diabetes**. V. 58, n. 5, p. 1144-1155, 2009.

TABOR, H. K.; RISCH, N. J.; MYERS, R. M. Candidate-gene approaches for studying complex genetic traits: practical considerations. **Nature Reviews Genetics**. V. 3, n. 5, p. 391-397, 2002.

TAKAHAMA, Y. Journey through the thymus: stromal guides for T-cell development and selection. **Nature Reviews Immunology**. V. 6, n. 2, p. 127-135, 2006.

TAKAHASHI, T. *et al.* Rosbin: a novel homeobox-like protein gene expressed exclusively in round spermatids. **Biology of Reproduction**. V. 70, n. 5, p. 1485-1492, 2004.

TANIGUCHI, H.; MAKINO, S.; IKEGAMI, H. The NOD mouse and its related strains. In: SHAFRIR, E. **Animal Models of Diabetes. Frontiers in Research**. 2nd ed., CRC Press, 2007.

THAYER, T. C.; WILSON, S. B.; MATHEWS, C. E. Use of nonobese diabetic mice to understand type 1 diabetes. **Endocrinology and Metabolism Clinics of North America**. V. 39, n. 3, p. 541-561, 2010.

TODD, J. A. Etiology of type 1 diabetes. **Immunity**. V. 32, n. 4, p. 457-467, 2010.

TONNELLE, C.; DEMARS, R.; LONG, E. O. DO β : a new β chain gene in HLA-D with a distinct regulation of expression. **The EMBO Journal**. V. 4, n. 11, p. 2839-2847, 1985.

TREVISAN, G. L. *et al.* Using cDNA microarrays to identify human CD19+ B cell gene products (ESTs) originated from systemic lupus erythematosus susceptibility loci. **Autoimmunity**. V. 5, n. 5, p. 319-323, 2006.

TROWSDALE, J. The gentle art of gene arrangement: the meaning of gene clusters. **Genome Biology**. V. 3, n. 3, comment2002.1-2002.5, 2002.

TURLEY, S. *et al.* Physiological beta cell death triggers priming of self-reactive T cells by dendritic cells in a type-1 diabetes model. **The Journal of Experimental Medicine**. V. 198, n. 10, p. 1527-1537, 2003.

VAFIADIS, P. *et al.* Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. **Nature Genetics**. V. 15, n. 3, p. 289-292, 1997.

VISSCHER, P. M. *et al.* Five years of GWAS discovery. **The American Journal of Human Genetics**. V. 90, n. 1, p. 7-24, 2012.

WALDRON-LYNCH, F.; HEROLD, K. C. Immunomodulatory therapy to preserve pancreatic β -cell function in type 1 diabetes. **Nature Reviews Drug Discovery**. V. 10, n. 6, p. 439-452, 2011.

WALTER, M. *et al.* IDDM2/insulin VNTR modifies risk conferred by IDDM1/HLA for development of type 1 diabetes and associated autoimmunity. **Diabetologia**. V. 46, n. 5, p. 712-720, 2003.

WANG, J.; WICKER, L. S.; SANTAMARIA, P. IL-2 and its high-affinity receptor: genetic control of immunoregulation and autoimmunity. **Seminars in Immunology**. V. 21, n. 6, p. 363-371, 2009.

WANG, X. *et al.* Identification of a molecular signature in human type 1 diabetes *mellitus* using serum and functional genomics. **The Journal of Immunology**. V. 180, n. 3, p. 1929-1937, 2008.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**. V. 10, n. 1, p. 57-63, 2009.

WATKINS, N. A. *et al.* A HaemAtlas: characterizing gene expression in differentiated human blood cells. **Blood**. V. 113, n. 19, p. e1-9, 2009.

WENZLAU, J. M. *et al.* The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. **PNAS**. V. 104, n. 43, p. 17040-17045, 2007.

WHO – World Health Organization. Fact sheet N°236. **Diabetes: the cost of diabetes**. Disponível em: < <http://www.who.int/mediacentre/factsheets/fs236/en/> > Acesso em: 23 de julho de 2012.

WOLFRUM, C.; POY, M. N.; STOFFEL, M. Apolipoprotein M is required for pre β -HDL formation and cholesterol efflux to HDL and protects against atherosclerosis. **Nature Medicine**. V. 11, n. 4, p. 418-422, 2005.

WU, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. **Genome Biology**. V. 10, n. 11, p. R130, 2009.

WU, Y. *et al.* Association of -619C/T polymorphism in CDSN gene and psoriasis risk: a meta-analysis. **Genetics and Molecular Research**. V. 10, n.4, p. 3632-3640, 2011.

WU, Y. L. *et al.* Molecular basis of complete complement C4 deficiency in two north-african families with systemic lupus erythematosus (SLE). **Genes and Immunity**. V. 10, n. 5, p. 433-445, 2009.

WU, X. *et al.* Apolipoprotein M promoter polymorphisms alter promoter activity and confer the susceptibility to the development of type 1 diabetes. **Clinical Biochemistry**. V. 42, n. 1-2, p. 17-21, 2009.

XIE, T. *et al.* Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. **Genome Research**. V. 13, n. 12, p. 2621-2636, 2003.

XU, N.; DAHLBÄCK, B. A novel human apolipoprotein (apoM). **The Journal of Biological Chemistry**. V. 274, n. 44, p. 31286-31290, 1999.

YANG, Z. *et al.* Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module). **The Journal of Biological Chemistry**. V. 274, n. 17, p. 12147-12156, 1999.

YI, W. *et al.* Targeted regulation of self-peptide presentation prevents type I diabetes in mice without disrupting general immunocompetence. **The Journal of Clinical Investigations**. V. 120, n. 4, p. 1324-1336, 2010.

YU, C. Y. *et al.* The human and mouse MHC class III region: a parade of 21 genes at the centromeric segment. **Immunology Today**. V. 21, n. 7, p. 320-328, 2000.

ZHENG-BRADLEY, X. *et al.* Large scale comparison of global gene expression patterns in human and mouse. **Genome Biology**. V. 11, n. 12, R124, 2010.

ZUCCHELLI, S. *et al.* Defective central tolerance induction in NOD mice: genomics and genetics. **Immunity**. V. 22, n. 3, p. 385-396, 2005.