UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

O uso de banco de dados de ESTs para predizer e validar polimorfismos de base única em genes do sistema HLA classe I

Aluna: Thalita Cristina Figueiredo Cunha

Orientador: Prof. Dr. João Ricardo Mendes de Oliveira

RECIFE

2012

UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

O uso de banco de dados de ESTs para predizer e validar polimorfismos de base única em genes do sistema HLA classe I

Aluna: Thalita Cristina Figueiredo Cunha

Orientador: Prof. Dr. João Ricardo Mendes de Oliveira

Dissertação apresentada ao programa de Pós-graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do grau de Mestre em Ciências Biológicas. Sob a orientação do Professor Dr. João Ricardo Mendes de Oliveira.

RECIFE

2012

Cunha, Thalita Cristina Figueiredo

O uso de banco de dados de ESTs para predizer e validar polimorfismos de base única em genes do sistema HLA classe I/ Thalita Cristina Figueiredo Cunha. – Recife: O Autor, 2012.

73 folhas : il., fig., tab.
Orientador: João Ricardo Mendes de Oliveira
Dissertação (mestrado) – Universidade Federal de Pernambuco, Centro de Ciências Biológicas. Ciências Biológicas, 2012.

Inclui bibliografia

1. Genes 2. Bioinformática 3. Polimorfismo (genética) I. Título.

570.285 CDD (22.ed.) UFPE/CCB-2012-053

O uso de banco de dados de ESTs para predizer e validar polimorfismos de base única em genes do sistema HLA classe I

Thalita Cristina Figueiredo Cunha

Banca Examinadora

Prof. Dr. João Ricardo Mendes de Oliveira
Orientador – UFPE / Dept. Neuropsiquiatria/LIKA

Profa. Drª. Danyelly Bruneska Gondim Martins
UFPE / Dept. Bioquímica

Prof. Dr. Paulo Roberto Eleutério de Souza
UFRPE / Dept. de Biologia

RECIFE

2012

"*Feliz aquele que transfere o que sabe e aprende o que ensina*". Cora Coralina.

Agradecimentos

À minha mãe, Fátima, minha grande companheira e incentivadora. Obrigada pela dedicação a mim conferida e por estar sempre ao meu lado mesmo nos momentos estressantes.

Ao meu orientador, Prof. Dr. João Ricardo Mendes de Oliveira, por me aceitar no seu grupo de pesquisa e por todas as conversas acadêmicas e informais que contribuíram de forma significativa no meu crescimento tanto profissional quanto pessoal.

À minha família pelo apoio e por enaltecer minhas qualidades.

Ao meu namorado, Diénert Vieira, pelo carinho, companhia, incentivo e apoio que foram fundamentais para o desenvolvimento desta dissertação.

Aos meus queridos amigos da graduação de Ciências Biológicas - UFPB que compreendem a luta diária na busca de realizar nossos sonhos e desejos na profissão. Em especial, agradeço ao meu eterno "sócio" Bruno Chaussê por sempre me apoiar e incentivar.

Aos meus amigos, em especial à Íria Wiese, Bruno Barufaldi, Thaís Gaudêncio e Patrícia Keytth pelo apoio, incentivo, caronas, descontração, abrigo, entre outros fatores que foram fundamentais na execução do meu mestrado.

Aos companheiros de laboratório José Eriton, Danyllo Felipe, Lylyan Pimentel, Manuela Souza e Roberta Lemos pelos momentos de aprendizagem e descontração. Em especial à Manuela Souza e Roberta Lemos pelos ensinamentos sobre as ferramentas utilizadas na minha dissertação.

Aos membros da banca examinadora, por suas contribuições para o melhoramento do trabalho realizado.

À CAPES e PROPESQ-UFPE pelo apoio financeiro.

Resumo

Os genes HLA (Antígenos Leucocitários Humanos) estão localizados no MHC (Complexo Principal de Histocompatibilidade Humano). O MHC localiza-se no braço curto do cromossomo 6 (6p21.3), uma das regiões com os maiores níveis de polimorfismo do genoma e alta concentração gênica. Muitos dos polimorfismos encontrados nos genes HLA estão associados à rejeição de transplantes clínicos, suscetibilidade à doenças infecciosas, doenças auto-imunes e predisposição a um amplo espectro de doenças crônicas não infecciosas. O objetivo deste estudo foi predizer e validar SNPs presentes nas regiões codificantes dos genes HLA classe I, utilizando o banco de dados de ESTs humano. O CLCbio Workbench Combined® versão 5.7.1 foi inicialmente utilizado para construção do banco de dados de ESTs e recuperação dos arquivos de RNAm respectivamente a partir do Golden path of University of California Santa Cruz (UCSC) e National Center for Biotechnology Information (NCBI). Na etapa seguinte foram realizados múltiplos alinhamentos com auxílio do CUBO® versão 1.0.6, hardware com função de acelerar o processamento. O algoritmo utilizado foi o Smith-Waterman. Um número inicial de 18.029 ESTs foram alinhados com os RNAm dos genes HLA, sendo 3.287 ESTs selecionados após aplicação de parâmetros de estringência utilizados para minimizar erros de alinhamentos. A anotação dos SNPs encontrados nos ESTs selecionados revelou várias classes de variações: 94 transversões, 100 transições, 10 deleções e 17 inserções, totalizando 221 SNPs. Destes SNPs, 101 já estavam descritos no banco de dados do MHC (NCBI) e 120 são potenciais novos polimorfismos que podem indicar novos alelos. Alguns destes SNPs foram encontrados em mais de 200 diferentes ESTs. A maioria das variações foram encontradas nos exons 2 e 3 que codificam as estruturas moleculares alfa1 e alfa2 do receptor membranar, sítio de ligação dos peptídeos antigênicos. Estas estruturas são fundamentais para numerosas funções imunes. A validação virtual confirmou que algumas das variações identificadas foram previamente anotadas e confirmadas em amostras de DNA, demonstrando que este método utilizando ferramentas de Bioinformática é uma maneira viável para detectar variações genéticas em dados gerados em larga escala.

Palavras chave: Genes HLA classe I; Bioinformática; ESTs, Variações.

Abstract

Genes of the Human Leukocyte Antigen system (HLA) are located in the MHC Major Histocompatibility Complex (MHC), at the short arm of chromosome 6 (6p21.3). This region is one of the most polymorphic at the whole Human Genome and many of the polymorphisms found in the HLA genes are associated with clinical organs transplantation rejection, susceptibility to infectious diseases, autoimmune diseases and susceptibility to a broad spectrum of chronic non-infectious. The aim of this study was to predict and validate novel and previously known SNPs in the coding regions of HLA class I genes, using a database of human ESTs. The CLCbio Combined Workbench ® Version 5.7.1 was first used to build the database of ESTs and mRNA recovery files respectively from the Golden path of University of California Santa Cruz (UCSC) and National Center for Biotechnology Information (NCBI) in the next step multiple alignments were performed with the aid of CUBE® version 1.0.6, hardware to accelerate the processing function and the algorithm used was the Smith-Waterman. An initial number of 18,029 ESTs were aligned with the mRNA of HLA genes, but only 3,287 ESTs were selected after application of appropriate parameters of stringency used to minimize alignment errors. The annotation of the SNPs found in the selected ESTs revealed several classes of variations: 94 transversions, 100 transitions, 10 deletions and 17 insertions, a total of 221 SNPs. Of these SNPs, 101 were already described in the MHC database (NCBI) and 120 are potential new polymorphisms that may indicate new alleles. Some of these SNPs were found in more than 200 different ESTs. Most variations were found in exons 2 and 3 which encode the alpha1 and alpha2 molecular structures of membrane receptor, binding site of antigenic peptides. These structures are crucial for numerous immune functions. The virtual validation confirmed that some of the variations identified have been previously noted and confirmed in DNA samples, demonstrating that this methodology based on bioinformatics tools is a viable way to detect genetic variations in data generated on a large scale.

Keywords: HLA class I genes; Bioinformatics; ESTs; Variations.

Lista de Abreviaturas

BLAST – Basic Local Alignment Search Tool
cDNA – Complementar Desoxyribonucleic acid
CLC bio – CLC Combined Workbench
cSNPs – Coding Single Nucleotide Polymorphisms
dbMHC – Database Major Histocompatibility Complex
DNA – Desoxyribonucleic acid
DIPs – Deletion-Insertion Polymorphims
ESTs – Express Sequence Tags
HLA – Human Leucocyte Antigens
MHC – Major Histocompatibility Complex
NCBI – National Center for Biotechnology Information
nSNPs – Non-coding Single Nucleotide Polymorphisms
RNAm – Rybonucleic acid messeger
SNPs – Single Nucleotide Polymorphisms
STRs – Short Tandem Repeats
UCSC – University of California Santa Cruz
UTR – Untranslated region

Lista das Figuras

SUMÁRIO

INTRODUÇÃO

Os genes HLA (Antígenos Leucocitários Humanos) estão localizados no MHC (Complexo Principal de Histocompatibilidade Humano) que localiza-se no braço curto do cromossomo 6 (6p21.3), uma das regiões com os maiores níveis de polimorfismo do genoma e alta concentração gênica (Shiina, 2009). Os genes HLA classe I apresentam centenas de alelos em cada *locus* e as combinações ao acaso permitem uma acentuada variabilidade genotípica (polialelismo). Uma das possíveis explicações é de que o acentuado polimorfismo existe para garantir a sobrevivência das espécies (Turner, 2004).

A análise *in silico* tem sido proposta como um método de descoberta alternativo que utiliza grandes conjuntos de dados com informações potenciais de SNPs (polimorfismos de base única) que foram gerados com outros fins e que não tenham sido utilizados como fonte de informação de SNPs (Useche et al. 2001). Os SNPs localizados em regiões codificantes de genes podem modificar a sequência de aminoácidos através de substituições não sinônimas que influenciam na estrutura e na função protéica (Aouacheria et al. 2006). O uso de ESTs (etiquetas de sequências expressas) em estudos de identificação de SNPs tem a vantagem de facilitar o achado de polimorfismos diretamente ligados a estas regiões codificantes do gene. Além disto, a maioria dos ESTs são obtidos a partir de bibliotecas de diferentes indivíduos e tecidos, e com isto, o alinhamento de sequências sobrepostas para a mesma região pode levar à identificação de novos SNPs e, consequentemente, novos alelos (Picoult-Newberg et al. 1999). Portanto, a análise de ESTs parece ser uma boa alternativa para o estudo de SNPs em vários genes, especialmente os genes HLA classe I, genes que possuem acentuados polimorfismos os quais estão associados à rejeição de transplantes clínicos, suscetibilidade a doenças infecciosas, doenças autoimunes e predisposição a um amplo espectro de doenças crônicas não infecciosas.

Com isto, esta dissertação teve como principal objetivo predizer e validar SNPs presentes nas regiões codificantes dos genes HLA classe I, utilizando o banco de dados de ESTs humano.

CAPÍTULO I

1. REVISÃO BIBLIOGRÁFICA

1.1 Estrutura e função do sistema HLA

Os genes HLA (Antígenos Leucocitários Humanos) estão localizados no MHC (Complexo Principal de Histocompatibilidade Humano) que localiza-se no braço curto do cromossomo 6 (6p21.3), uma das regiões com os maiores níveis de polimorfismo do genoma e alta concentração gênica. Cerca de 40% dos genes localizados no MHC estão envolvidos em algum processo imunológico. Os genes HLA são herdados em blocos ou séries chamados haplótipos e expressos codominantemente em cada indivíduo. Influenciam na rejeição de transplantes clínicos, suscetibilidade a doenças infecciosas e predisposição a um amplo espectro de doenças crônicas não infecciosas (Shiina, 2009).

O MHC apresenta três regiões: classe I, II e III (Figura 1). A região de classe II, a mais centromérica do complexo, estende-se por aproximadamente um megabase (Mb) e contém os genes HLA de classe II. A região de classe III também se estende por um Mb e apresenta genes envolvidos na resposta imune, porém não possui genes HLA (Shiina, 2009). A região de classe I é a mais telomérica, estende-se por dois Mb e contém os genes HLA-A, HLA-B e HLA–C, cada um dos quais com 1.729, 2.329 e 1.291 alelos descritos, respectivamente (IMGT/HLA Database http://www.ebi.ac.uk/imgt/hla acesso em: 29/12/2011).

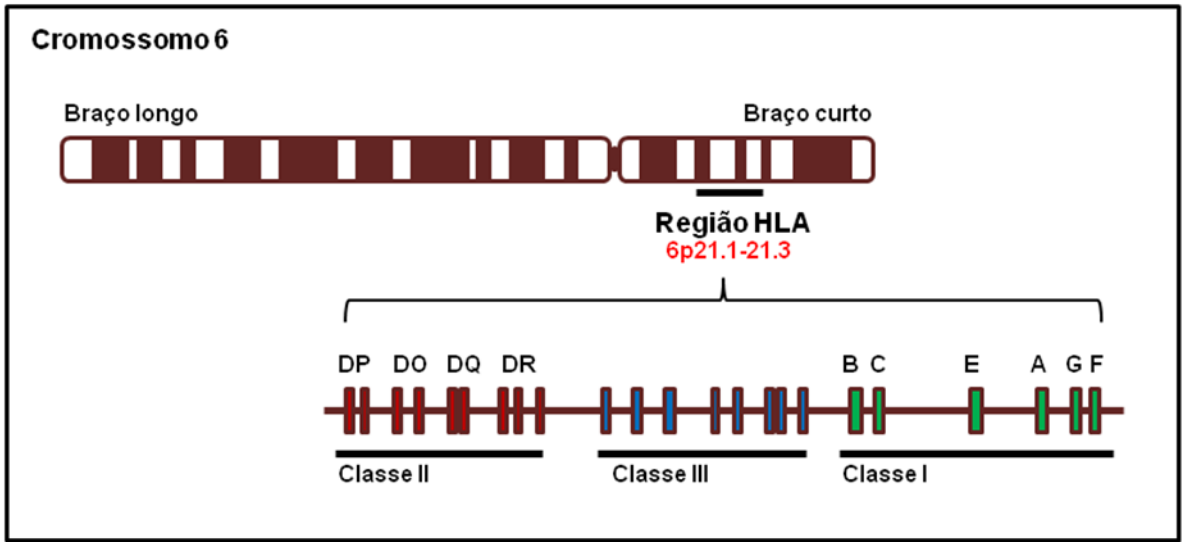


Figura 1. Estrutura esquemática do cromossomo 6, com representação da localização relativa dos principais genes HLA.

Dentro dos *loci* de classe II estão genes que codificam diversas proteínas que exercem papéis essenciais no processamento do antígeno, como as proteínas de classe II clássicas – HLA-DR, HLA-DP e HLA-DQ, expressas na superfície das células apresentadoras de antígenos (APCs). O MHC de classe II também codifica proteínas de classe II não clássicas, como HLA-DM e HLA-DO (Sanchez-Mazas et al. 2011). A região MHC de Classe III contém numerosos genes que codificam proteínas relacionadas ao sistema imune, como proteínas do sistema complemento, citocinas TNF-α e β (fatores α e β de necrose tumoral-dependente) e proteínas de choque térmico (*heat shock*) – além de outras não relacionadas – como enzimas requeridas para a síntese de esteróides, e muitas outras proteínas ainda não identificadas (Vandiedonck & Knight, 2009).

As moléculas HLA-A, HLA-B e HLA-C, também conhecidas como moléculas HLA de classe I clássicas, são as principais responsáveis pela apresentação de peptídeos citoplasmáticos. Essas moléculas são constituídas por duas subunidades: a cadeia α e a β_2-microglobulina. A cadeia α é codificada pelos genes HLA de classe I clássicos, HLA-A, -B e –C, e apresenta três domínios extracelulares: α1, α2 e α3. Os domínios α1 e α2 formam a fenda apresentadora de peptídeos, um sulco flanqueado por duas α-hélices dispostas sobre um assoalho de folhas β-pregueadas (Figura 2) (Vandiedonck & Knight, 2009).

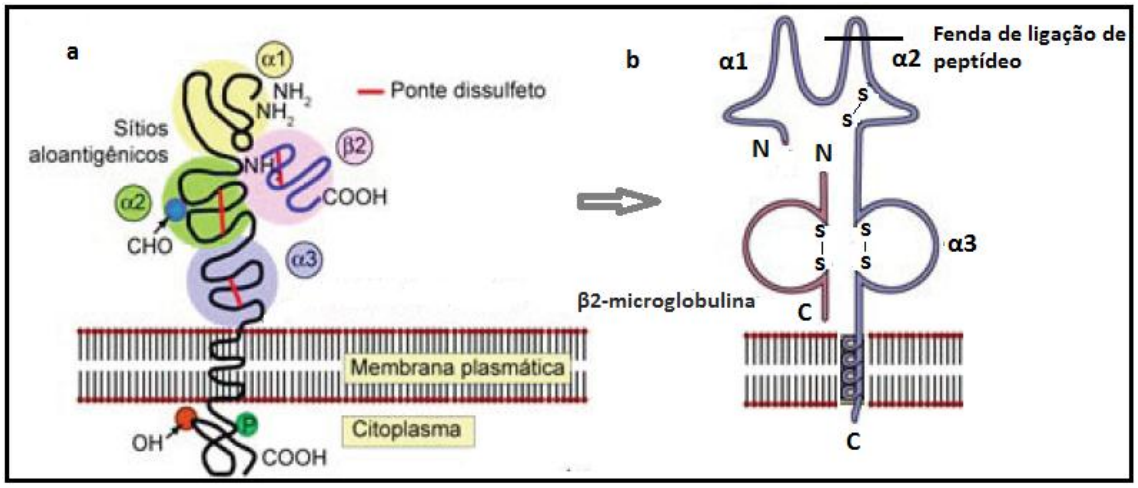


Figura 2. Estrutura esquemática de uma molécula MHC de classe I**.** As moléculas de classe I são compostas por uma cadeia peptídica α (pesada) ligada não covalentemente a uma cadeia leve não polimórfica, chamada β2-microglobulina.(b) O diagrama mostra os três domínios globulares da cadeia pesada: α1 (amarelo), α2 (verde) e α3 (azul). O domínio α3 está intimamente associado ao peptídeo β2-microglobulina (rosa) (a). Essa associação é estabilizada por pontes dissulfeto (vermelho). O sítio aloantigênico (fenda de ligação de peptídeo) é encontrado nos domínios α1 e α2. O domínio α2 é glicosilado (azul – CHO). Fonte: modificado a partir de Abbas et al. (2005).

Na ocorrência de uma infecção viral, bacteriana ou alteração de proteínas próprias como ocorre no câncer, peptídeos das proteínas estranhas são clivados no citoplasma da célula e bombeados para o retículo endoplasmático. As moléculas HLA de classe I tem a importante função de apresentar estes peptídeos endógenos/exógenos às células T CD8 positivas. Primeiramente, os peptídeos são gerados no citosol pela ação de proteassomas (LPM 2 e 7) e então, transportados ao retículo endoplasmático. Em paralelo, também no retículo endoplasmático, ocorre a síntese da cadeia pesada α e da β2-microglobulina das moléculas de classe I. O sítio de ligação do peptídeo, uma fenda profunda, é formado lateralmente pelas α hélices dos domínios α1 e α2, e o assoalho, por oito (08) fitas β, pregueadas. Após a formação deste heterodímero, o peptídeo se acomoda na fenda, estabilizando a molécula de classe I, então, o complexo HLA + peptídeo via complexo de Golgi é exteriorizado e expresso nas superfícies celulares. Linfócitos T $CD8^+$ reconhecem esse complexo e desencadeiam uma resposta que culmina na eliminação da célula apresentadora (Brodsky et al. 1996; Parham, 2005).

1.2 Polimorfismos dos genes HLA

Desde a descoberta do primeiro polimorfismo biológico humano no ano 1900 por Landsteiner (Gröger, 2000), a notável natureza da variação genética humana tem sido revelada; desde as diferenças de um único nucleotídeo à variação genômicas estruturais de milhões de pares de bases. Estes incluem polimorfismos de nucleotídeo único (SNPs), as substituições e os polimorfismos de inserção-deleção (DIPs), bem como a variação envolvendo mais de uma base, como *short tandem repeats* (STRs) (microssatélites), grandes deleções ou inserções (INDELs), inversões, entre outros (Scherer et al. 2007).

Como marcadores genéticos, STRs e SNPs revelaram-se, particularmente, informativos. As variações são encontradas em quase todos os genes, mas, nenhuma parte do genoma pode competir com o polimorfismo extremo encontrado nos genes HLA (Vandiedonck & Knight, 2009).

Os genes HLA classe I apresentam centenas de alelos em cada *locus* e as combinações ao acaso permitem uma acentuada variabilidade genotípica (polialelismo). Uma das possíveis explicações é de que o acentuado polimorfismo existe para garantir a sobrevivência das espécies (Turner, 2004).

O número de alelos para o *locus* HLA-B dobraram nos últimos anos, atualmente, apresenta 2.329 alelos, sendo considerado como um dos genes mais polimórfico do

genoma humano. O locus HLA-A possui 1.729 alelos e o HLA-C tem 1.291 alelos descritos, totalizando 5.349 alelos HLA classe I (IMGT/HLA Database http://www.ebi.ac.uk/imgt/hla acesso em: 29/12/2011). No mês de maio/2011 o mesmo banco apresentava 2.125 alelos HLA-B, 1.601 alelos HLA-A e 1.112 HLA-C depositados, o que representa um aumento de mais de 500 novos alelos em seis meses.

Na figura 3 encontramos o número de alelos nomeado por ano a partir de 1987 a março de 2011, representando o número crescente de alelos descritos durante os anos.

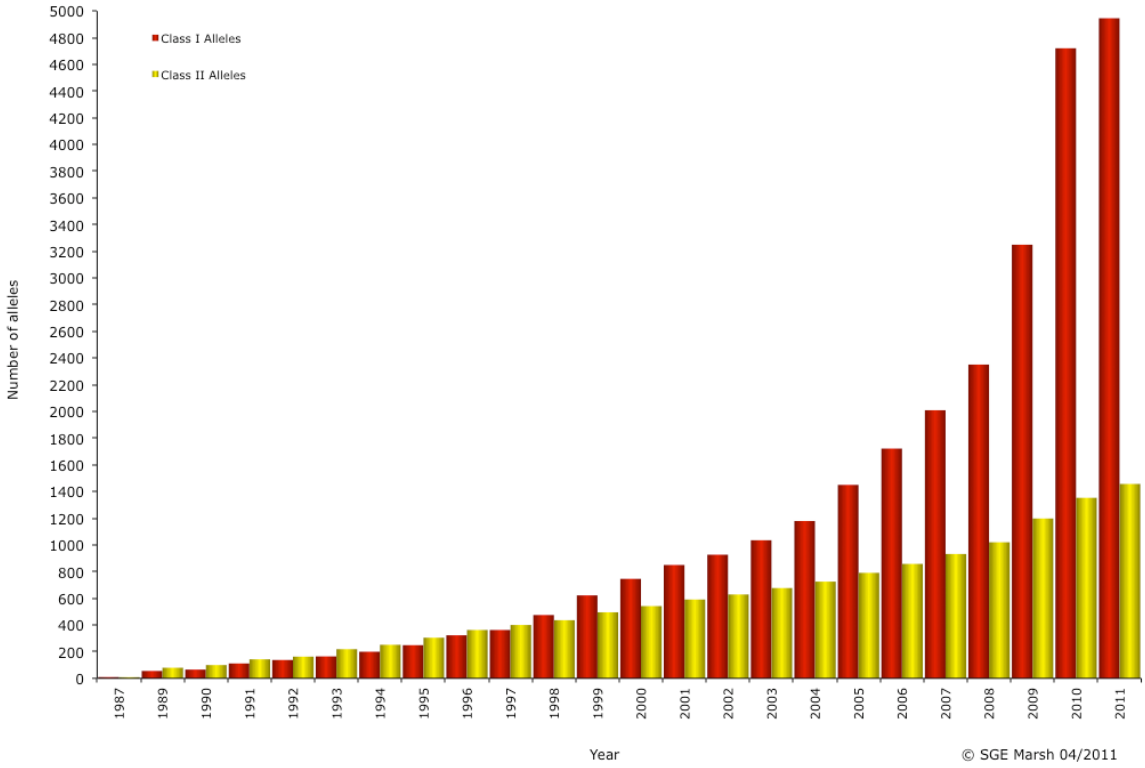


Figura 3. Gráfico mostrando o número de alelos nomeado por ano a partir de 1987 a março de 2011. Colunas em vermelho: alelos HLA classe I, coluna em amarelo: alelos HLA classe II (Robinson et al. 2011).

Os polimorfismos dos genes HLA clássicos estão predominantemente localizados nos exons 2 e 3 dos genes de classe I, os quais codificam para as estruturas moleculares alfa 1 e alfa 2 do receptor membranar para peptídeos antigênicos. Com isto, as substituições não-sinônimas de um único nucleotídeo levam à mudanças de aminoácidos nos domínios envolvidos na ligação de peptídeos antigênicos (Vandiedonck & Knight, 2009).

A sequência de aminoácidos das moléculas HLA classe I determina as propriedades das bolsas dos receptores dentro das quais as cadeias laterais dos peptídeos

serão ligadas e, consequentemente, a especificidade ao peptídeo antigênico. Os resíduos polimórficos que ocupam a fenda de ligação e em particular as bolsas determinam as propriedades de ligação ao peptídeo de diferentes moléculas HLA. Todos os sítios nas moléculas HLA que interagem com cadeias laterais do peptídeo têm contribuição de resíduos polimórficos, o que pode levar a mudanças na especificidade de ligação ao peptídeo entre os diferentes alelos HLA. Devido à especificidade de ligação, podemos concluir que somente certos patógenos podem se ligar a uma dada molécula HLA, e aqueles que não se ligam não são imunogênicos. Analogamente, um patógeno pode escapar da detecção por sofrer mutações que eliminam de suas proteínas todos os peptídeos capazes de se ligar a moléculas HLA. Esse processo de evasão da resposta imune é muito mais difícil para um patógeno potencial devido à existência de diferentes *loci* HLA que codificam diferentes moléculas HLA (Lechler & Warrens, 2000). Fugir da resposta imune torna-se ainda mais difícil para um patógeno porque o polimorfismo para cada *locus* pode dobrar o número de diferentes moléculas HLA expressas por um indivíduo, pois trata-se de genes codominantes, e a maioria dos indivíduos são heterozigotos (Sanchez-Mazas et al. 2011).

Há vários estudos em busca de uma explicação em relação à evolução do sistema HLA. A partir da observação dos níveis de diversidade dos genes HLA o modelo de evolução neutra torna-se incompatível, implicando que o modelo alternativo de seleção balanceadora seja mais plausível. Vários dos polimorfismos moleculares atribuídos a esse tipo de seleção ocorrem em proteínas celulares que interagem com microorganismos. Nas moléculas HLA, praticamente toda a variação localiza-se nos domínios α1 e α2, nos sítios constituintes da fenda apresentadora de peptídeos. Portanto, as regiões com maior evidência da atuação da seleção balanceadora são as que interagem diretamente com os antígenos, indicando que os microorganismos patogênicos sejam a fonte da seleção atuante nos genes HLA. Essa hipótese evolutiva é conhecida como seleção dirigida por patógenos (Garrigan & Hedrick, 2003).

Estudos de associação com doenças infecciosas corroboram a hipótese de seleção dirigida por patógenos. Por exemplo, foram descritas associações entre alelos HLA e resistência à malária em populações do oeste africano (Hill et al. 1997). Alelos de HLA-B foram associados a um maior ou menor progressão à infecção por HIV (Carrington et al. 1999). Se realmente foi a seleção dirigida por patógenos que moldou o polimorfismo dos genes HLA, é plausível supor que regiões geográficas com alta diversidade patogênica apresentem um maior sinal seletivo. Prugnolle e colaboradores (2005)

mostraram que há uma correlação positiva entre a carga de patógenos de uma determinada região geográfica e os níveis de diversidade nos *loci* HLA, especialmente em HLA-B. Todos esses resultados sustentam a hipótese de seleção dirigida por patógenos, mas ainda há debate a respeito dos mecanismos específicos de como a interação com os microorganismos moldou a variação dos genes HLA.

Borghans e colaboradores (2004) realizaram simulações matemáticas do modelo de co-evolução entre patógenos e sistema imunológico. De acordo com esse modelo, as mudanças nos genes dos patógenos e genes imunológicos ocorreriam paralelamente: os patógenos tenderiam a acumular mutações de escape, isto é, mutações que evitassem a apresentação pelas moléculas HLA mais comuns na população de hospedeiros e, por outro lado, novas moléculas HLA que apresentassem de forma eficiente essas mutações de escape, seriam adaptativas e aumentariam de frequência na população de hospedeiros, recomeçando o ciclo. Esse modelo de co-evolução entre patógenos e hospedeiros se mostrou mais adequado que o de sobredominância simples, explicando de maneira satisfatória várias das características dos dados reais tal como o grande número de alelos, altas taxas de heterozigose e grande persistência dos alelos nas populações ao longo das gerações.

1.3 Associações entre o sistema HLA e diversas doenças

A diversidade e polimorfismo dos genes do sistema HLA e sua participação na resposta imune fazem com que ele desempenhe um papel importante na patogenia de várias doenças de distintas etiologias, incluindo as autoimunes, as infecciosas, as neoplásicas e as idiopáticas (Doan et al. 2008).

As mutações pontuais, deleções e outras variantes dentro do sistema HLA são responsáveis por doenças monogênicas que apresentam herança mendeliana, tais como hemocromatose ou hiperplasia congênita de adrenal. Porém, há um papel mais significativo desta região em doenças multifatoriais (Smith et al. 2006). Desde o início dos anos 70, testes sorológicos já associavam antígenos HLA específicos e doenças autoimunes, tais como espodilite anquilosante, psoríase e doença celíaca. A importância do MHC foi posteriormente demonstrada através dos microssatélites, polimorfismos de nucleotídeo único (SNPs) e outros marcadores genéticos para um número amplo de doenças, principalmente, nos estudos de associação de genoma completo (GWAS) (The Wellcome Trust Case Control Consortium, 2007). São as doenças autoimunes que estão associadas, principalmente, com as variações genéticas do sistema HLA, entre elas,

estão incluídos os principais *loci* de risco para diabetes tipo 1 (Grayson et al. 2010), artrite reumatóide (Weyand & Goronzy, 2000) e lupus eritematoso sistêmico (Logar et al. 2002). Estudos também mostram associações com doenças infecciosas e inflamatórias, bem como com o câncer (Pardoll, 2003). Para a maioria destas associações do sistema HLA com doenças tem o haplótico como causador, ao contrário de variantes específicas (Vandiedonck & Knight, 2009).

Os antígenos ou alelos do sistema HLA são associados e estudados em relação a várias doenças, normalmente observados em estudos populacionais e familiares. A obtenção dos resultados é através do cálculo do risco relativo, que indica a força da associação do HLA com a doença. É o risco de desenvolver a doença quando o antígeno está presente comparado com o risco de não desenvolver a doença quando o antígeno está ausente. A fração etiológica estima em porcentagem a susceptibilidade da doença, sendo calculada nos casos em que o risco relativo é maior que 1. A fração preventiva estima a porcentagem de proteção contra a doença, sendo calculada nos casos em que o risco relativo é menor que 1. Por fim, o risco absoluto é o risco real do portador de determinado marcador vir a desenvolver a doença (Donadi, 2000).

Para entender o mecanismo de associação entre antígenos HLA e determinadas doenças, algumas hipóteses tem sido formuladas. Entre estas, destacam-se: 1) As moléculas de histocompatibilidade são glicoproteínas que podem funcionar facilitando ou impedindo a entrada de alguns agentes etiológicos nas células (Donadi, 2000; Lechler & Warrens, 2000). 2) Ocorrência de mimetismo molecular entre os antígenos HLA e os agentes etiológicos. A hipótese de que o mimetismo molecular compõe o mecanismo patogênico de doenças, como febre reumática, artrite associada à doença inflamatória intestinal, síndrome de Reiter e artrite reumatóide, é defendida por vários autores. Esse mecanismo pode ser definido como a ocorrência de epítopos comuns entre um agente infeccioso e um tecido do organismo, a qual pode provocar resposta imune celular ou humoral contra tecidos próprios. O desenvolvimento da doença cardíaca reumática, por exemplo, se dá por mimetismo molecular entre proteínas estreptocócicas e proteínas do tecido cardíaco nas quais estão incluídas moléculas HLA (Donadi, 2000; Lechler & Warrens, 2000; Stanevicha et al. 2003). 3) Indução aberrante da expressão de moléculas HLA de classe II. Células teciduais que normalmente não expressam moléculas HLA de classe II podem ser estimuladas a apresentá-las na presença de citocinas, derivadas, por exemplo, de células infectadas por vírus. Assim, essas células podem apresentar aos linfócitos T antígenos derivados da degradação do próprio tecido, gerando uma resposta

auto-imune (Donadi, 2000; Trabace, 2000). 4) Participação de outros genes do MHC, ou mesmo de fora do MHC, que estejam em desequilíbrio de ligação com os genes de histocompatibilidade. Alguns fatores do sistema complemento, os fatores de necrose tumoral, os peptídeos processadores e transportadores de determinantes antigênicos, as proteínas do choque térmico, todos tem algum papel na resposta imune, sendo codificados por genes situados dentro do MHC e podem ser herdados em haplótipos em desequilibrio de ligação (Donadi, 2000; Lechler & Warrens, 2000).

1.4 Bioinformática

Na segunda metade da década de 80, com o surgimento dos sequenciadores automáticos, houve uma explosão na quantidade de sequências gênicas e protéicas, oriundas de diferentes grupos e instituições de pesquisa, o que exigiu recursos computacionais cada vez mais eficientes. A computação contribuiu não somente para o processamento e armazenamento de dados, mas também com métodos matematicamente sofisticados, a exemplo disto, temos o método de alinhamentos múltiplos, utilizado no desenvolvimento deste trabalho. Com isto, a união entre biologia e ciência da computação criou um novo campo chamado Bioinformática (Pasternak, 2007; Lesk, 2008). Segundo Lesk (2008) "*Hoje, a bioinformática é uma ciência aplicada. Nós utilizamos programas de computador para fazer inferências a partir de dados obtidos da biologia moderna, para fazer conexões entre eles e para derivar predições importantes e relevantes*".

Há vários *softwares* já disponíveis e outros que vem sendo criados para facilitar o processamento de dados biológicos. Dentre eles há o CLC Combined Workbench ® (CLCbio, Aarhus, Dinamarca), *software* que permite a análise de DNA, RNA e proteínas através de alinhamentos, desenho de primers, estudo protéico 3D, construção de cladogramas, busca de seqüências no NCBI, entre outras ferramentas (http://www.clcbio.com/).

O CLC Combined Workbench tem como principal função realizar alinhamentos múltiplos em curto período de tempo. O alinhamento múltiplo de sequências biológicas consiste na coleção de três ou mais sequências que são parcialmente (alinhamento local) ou completamente alinhadas (alinhamento global) simultaneamente. O método utilizado pelo BLAST (ferramenta de busca de alinhamento local) verifica regiões de identidade local. Para cada entrada em um banco de dados, ele busca por pequenas regiões

contíguas que pareiam com pequenas regiões contíguas da sequência-alvo. A saída do BLAST é um conjunto de segmentos com pareamentos locais (Lesk, 2008; Smith; Waterman, 1981). Há várias variáveis do programa BLAST, a variável utilizada neste trabalho foi a BLASTn, compara sequências-alvo de nucleotídeo diretamente com banco de dados de nucleotídeos.

Nos estudos biológicos é fundamental que os alinhamentos indiquem uma real similaridade entre as sequências analisadas. Para isto, na análise dos alinhamentos alguns valores são fornecidos como resultado para cada alinhamento gerado. A porcentagem de identidade quantifica o grau de similaridade entre as sequências comparadas. Enquanto o *e-value* corresponde à probabilidade de se obter, com outra sequência aleatória de mesmo tamanho e composição de nucleotídeos, outro alinhamento com s*core* igual ou superior. Isso ocorre devido ao processo de alinhamento ser de forma aleatória, e quando o algoritmo compara fragmentos com diferentes tamanhos, sequências que não têm relação alguma podem ser alinhadas. Desta forma, o *e-value*, mais confiável tende a zero (Altschul et al. 1990).

1.5 Identificação de SNPs utilizando Banco de Dados de ESTs

Os SNPs (polimorfismos de base única) são uma das mais comuns fontes de variações genéticas. Essas variações podem ser mudanças de base classificadas como transições ou transversões, inserções ou deleções de bases. Os SNPs podem ser localizados em regiões codificadoras (cSNPs) ou não codificadoras (nSNPs), distribuídos aleatoriamente ou formando *clusters* em alguns genes. Os cSNPs podem modificar a sequência de aminoácidos do produto através de substituições não-sinônimas que influenciam na estrutura e na função protéica (Aouacheria et al. 2006). Já os nSNPs presentes em regiões não traduzidas do gene (UTR-SNPs) podem ter efeitos na expressão gênica por afetar elementos reguladores, ou por influenciar na estabilidade do RNAm (Goto et al. 2001).

Os estudos dos SNPs tem muitas aplicações na caracterização da estrutura dos genes, na história de grupos populacionais e na identificação de *loci* responsáveis pela variação fenotípica de origem multifatorial, incluindo suscetibilidade a doenças. Apesar de terem surgido por mutações, muitas das posições que contêm SNPs apresentam baixas taxas de mutações, e podem ser utilizadas como marcadores moleculares estáveis no mapeamento de genes (Lesk, 2008).

Há duas principais formas de detectar SNPs; através de procedimento experimental e por método *in silico*. O método experimental em geral se refere à triagem do DNA feita por meio de sequenciamento. O método *in silico* se refere à triagem feita por análise computacional em sequências provenientes de diferentes indivíduos, armazenadas em bancos de dados (Useche et al. 2001).

Para a identificação de novos SNPs por método *in silico*, utilizam-se diversos tipos de ácidos nucléicos, dentre eles as seqüências parciais de cDNA provenientes de diferentes tecidos depositadas em bancos de dados públicos, que são os ESTs (etiquetas de sequências expressas). Os projetos ESTs envolvem, inicialmente, a preparação de bibliotecas de cDNA, isolamento e sequenciamento de clones, análise dos dados gerados e, finalmente, a submissão destes dados ao Genbank, o repositório mundial de sequências de organismos. Uma grande vantagem de projetos ESTs é o acesso mais fácil às informações de espécies que possuem um genoma complexo (Vetorre et al. 2003).

Já que a maioria dessas bibliotecas são obtidas de diferentes indivíduos, o alinhamento de sequências sobrepostas para a mesma região pode levar à identificação de novos SNPs. Além disso, o uso de ESTs em estudos de identificação de SNPs tem a vantagem de facilitar o achado de polimorfismos diretamente ligados a regiões codificantes do gene (Picoult-Newberg et al. 1999).

A descoberta experimental de SNPs consiste de uma série de etapas trabalhosas que tornam este processo complexo e caro. A análise *in silico* tem sido proposta como um método de descoberta alternativo que utiliza e tira proveito de grandes conjuntos de dados com informações potenciais de SNPs que foram gerados com outros fins e que não tenham sido utilizados como fonte de informação de SNPs (Useche et al. 2001). Para ser capaz de analisar milhares de sequências, a descoberta automatizada de SNPs a partir de análise *in silico* será a abordagem predominante no futuro.

Por tudo isso, a análise de ESTs parece ser uma boa alternativa para o estudo de SNPs em vários genes, especialmente dos genes HLA classe I, genes que possuem um acentuado polimorfismo e estão relacionados na rejeição a transplantes clínicos, suscetibilidade a doenças infecciosas, doenças autoimunes e predisposição a um amplo espectro de doenças crônicas não infecciosas.

REFERÊNCIAS BIBLIOGRÁFICAS

Referências Bibliográficas

Abbas A.; Lichtman A. (2005) Cellular and Molecular Immunology. 5. ed. Philadelphia: Saunders.

Altschul S. F. et al. (1990). Basic local alignment tool. *Journal of Molecular Biology*. 215:403-410.

Aouacheria A. (2006). In silico whole-genome scanning of cancer-associated nonsynonymous SNPs and molecular characterization of a dynein light chain tumour variant. *Oncogene*. 24 (40):6133-6142**.**

Borghans J., Beltram J.B., De Boer R.J. (2004). MHC polymorphism under host-pathogen coevolution. *Immunogenetics*. 55(11):732-739.

Brodsky F. M; Lem L; Bresnahan P.A. (1996). Antigen processing and presentation. *Tissue Antigens.* 47: 464-471.

Carrington M. et al. (1999). HLA and HIV-1: heterozygote advantage and B*35- Cw*04 disadvantage. *Science.* 283 (5408):1748-1752.

Doan T. et al. (2008). Imunologia ilustrada. Porto Alegre, RS: Artmed,

Donadi E.A. (2000). Como entender a nomenclatura e os mecanismos de associação entre os antígenos e os alelos de histocompatibilidade com as doenças. *Medicina*. 33:7-18.

Garrigan D.; Hedrick P. (2003). Perpective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*. 57:1707-1722.

Goto Y. et al. (2001). A novel single-nucleotide polymorphism in the 3'-untranslatedregion of the human dihydrofolate reductase gene with enhanced expression. *Clinical Cancer Research***.** 7(7):1952-1956.

Grayson B. L. et al (2010). Genome-Wide Analysis of Copy Number Variation in Type 1 Diabetes. *Plos One.* 5 (11).

Gröger H. (2000) Karl Landsteiner and medical science in Vienna around 1900. The significance of laboratory medicine for clinical medicine. *Vox Sang*. 78:3-6.

Hill A. et al. (1997). Genetic analysis of host-parasite coevolution in human malaria. *Philosophical Transactions of the Royal Society*.352(1359):1317-1325.

IMGT/HLA Database. Disponível em: < http://www.ebi.ac.uk/imgt/hla > Acessos em: 11/05/2011 e 29/12/2011.

Lechler R.; Warrens A. (2000). HLA in Health and Disease. 2 ed. London: Academic press.

Lesk A.M. (2008). Introdução à bioinformática. In: Organização e Evolução de Genomas. 2 ed. Porto Alegre: Artmed. cap. 2, p. 87 -132.

Logar D. et al. (2002). The contribution of HLA-DQB1 coding and QBP promoter alleles to anti-Ro alone autoantibodyresponse in systemic lupus erythematosus. *Rheumatology*. 41: 305-311.

Parham P. (2005). MHC class I molecules and KIRs in human history, health and survival. *Nature Reviews Immunoly*. 5 (3): 201-214.

Pardoll D. (2003). Does the Immune System see Tumors as Foreign or Self? *Ann Rev of Immun.* 21:807-839.

Pasternak J.J. (2007). Uma introdução à genética molecular humana: Mecanismos das doenças hereditárias. 2 ed. Guanabara koogan, cap. 8:149 -168.

Picoult-Newberg L. et al. (1999) Mining SNPs from EST databases. *Genome Research.*. 9:167–174.

Prugnolle F. et al (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology*. 15: 1022-1027.

Robinson J. et al. (2011). The IMGT/HLA Database. *Nucleic Acids Research*. 39: (Database issue):D1171-6.

Shiina T. et al. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *Journal Human Genetics*. 54:15-39.

Sanchez-Mazas A. et al. (2011). Immunogenetics as a tool in anthropological studies. *Immunology*. 133: 143-164.

Scherer S.W. et al. (2007). Challenges and standards in integrating surveys of structural variation. *Nature Genetics*. 39:S7–15.

Smith W. P. et al. (2006). Toward understanding MHC disease associations: partial resequencing of 46 distinct HLA haplotypes. *Genomics*. 87:561–571.

Smith T. F.; Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*. 147:195-197.

Stanevicha V. et al. (2003). HLA class II associations with rheumatic heart disease among clinically homogeneous patients in children in Latvia. *Arthritis Research & Therapy*. 5(6):340-346.

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 447:661–678.

Trabace S. (2000). HLA and disease association. *The Journal of Headache and Pain*. 1(2), S109-S113.

Turner D. (2004). The human leucocyte antigen (HLA) system. *Vox Sanguinis*. 87(1): 587-590.

Useche F.J. et al. (2001) High-throughput identification, database storage and analysis of SNPs in ESTs sequences. *Genome Informatics*, 12,194.

Vandiedonck C.; Knight C.J. (2009). The human major histocompatibility complex as a paradigm in genomics research. *Briefings in Functional Genomics and Proteomics*. 8(5): 379–394.

Vetorre A.L. et al (2003). Analysis and functional annotation of na expressed sequence tag collection for tropical crop sugarcane. *Genome Research*. 13:2725-2735.

Weyand C.M.; Goronzy J.J. (2000). Association of MHC and rheumatoid arthritis: HLA polymorphisms in phenotypic variants of rheumatoid arthritis. *Arthritis Research & Teraphy*. 2(3):212-6.

CAPÍTULO II

International Journal of Immunogenetics **(Fator de impacto/2010: 1.62)**

Using ESTs database to validate and predict single polymorphisms at the HLA system.

INTERNATIONAL JOURNAL OF
IMMUNOGENETICS

Short Communication

doi: 10.1111/j.1744-313X.2011.01076.x

Using ESTs database to predict and validate single polymorphisms at the HLA system

T. C. Figueiredo*,† & J. R. M. de Oliveira*,†,‡

Summary

We propose a bioinformatics pipeline in which we use an ESTs database to predict and validate single-nucleotide polymorphisms (SNPs) directly linked to gene-coding regions at the HLA class I genes (HLA-A, HLA-B and HLA-C). Annotation originated from our analysis revealed various classes of possible new variations that may indicate possible new alleles. Thus, bioinformatics pipelines seem to be useful approaches to help screening for novel genetic variations at the HLA panel, and further analysis will foster this aim to provide celerity at the massive analysis of data currently generated in large-scale high-throughput experiments.

Introduction

The human leucocyte antigen system (HLA), covering about 4 Mb, is located on the short arm of chromosome 6 and is divided into three regions: class I, class II and class III, according to the structure and function of their genes. The class I region contains the HLA-A, HLA-B and HLA-C *loci*, while the class II region encodes the HLA-DR, HLA-DQ and HLA-DP *loci*. Class III genes encode, among others, complement factors, tumour necrosis factor and the enzyme 21-hydroxylase (Marsh *et al*., 2002).

Since the discovery of the first human biological polymorphism, the remarkable nature of human genetic variation has been revealed, ranging from single-nucleotide differences to large structural genomic variation, spanning thousands of base pairs. These include single-nucleotide polymorphisms (SNPs), substitutions, deletion and insertion polymorphisms as well as variation involving more than one base such as short tandem repeats (STRs) (microsatellites), large insertion or deletions, inversions and other repeats. While such variants have been identified in almost all genes, nowhere in the genome can yet compete with the number of polymorphism found in HLA genes (Scherer *et al*., 2007; Vandiedonck & Knight, 2009).

The number of alleles for the HLA-B locus has doubled over the last years, with more than 2329 alleles so far, considered as the most polymorphic gene in the human genome. The HLA-A locus has 1729 alleles, and the HLA-C has 1291 alleles described, in a total of more than 5400 HLA class I alleles (IMGT/HLA Database, October 2011).

Several decades of intensive research has defined the remarkable genomic environment of the major histocompatibility complex (MHC) and how genetic variation within this region plays a key role in susceptibility to autoimmune, infectious and other diseases (Lie & Thorsby, 2005). More significant, however, is the role played by this region of the human genome in many common multifactorial diseases. The importance of the MHC for a range of diseases was robustly demonstrated using microsatellites, single-nucleotide polymorphisms (SNPs) and other markers of genetic variability and continues today in the era of genome-wide association studies (The Wellcome Trust Case Control Consortium, 2007).

Many researchers have studied the relationships between disease and biological variations such as single-nucleotide polymorphisms (SNPs), copy number variation, sequence repeats and genetic rearrangement (Bae *et al*., 2008). Studies on genetic variation associated with diseases become intense as many genetic variations are thought to affect the structure and function of proteins, as a result of amino acid substitutions. Significantly, SNPs, which report over 90% of genetic variation in the human genome, can have a major impact on how humans respond to disease, to drugs and to other therapies (Kim *et al*., 2008).

The main methods for detecting SNPs are experimental methods and analysis *in silico*. There are several experimental techniques to discover and detect SNPs. The experimental method in general refers to the screening performed by DNA sequencing. The method *in silico* refers to the screening performed by

* Keizo Asami Laboratory (LIKA), Federal University of Pernambuco, Recife, PE, Brazil, † Biological Sciences Graduate Program, Federal University of Pernambuco, Recife, PE, Brazil and ‡ Department of Neuropsychiatry, Federal University of Pernambuco, Recife, PE, Brazil

Received 25 September 2011; revised NA; accepted 29 November 2011

Correspondence: João Ricardo Mendes de Oliveira, Department of Neuropsychiatry, Federal University of Pernambuco, Recife, PE 50670-901, Brazil. Tel: +55 81 21268539; Fax: +55 81 21268485; E-mail:joao.ricardo@ufpe.br

computer analysis of sequences from different individuals, stored in databases and has been proposed as an alternative discovery method that uses large data sets with information of potential SNPs that were generated for other purposes and have not been used as a source of information of SNPs (Useche *et al.*, 2001).

Several types and sources of nucleic acids can be used for the *in silico* identification of new SNPs, including the partial sequences of cDNA from different tissues, these are known as Expressed Sequence Tags (ESTs). The identification of ESTs has proceeded rapidly, and over 52 million ESTs are currently available in public databases (http://www.ncbi.nlm.nih.gov/). As most of these libraries are derived from different individuals, the alignment of overlapping sequences for the same region may lead to identification of new SNPs. In addition, the use of ESTs in studies to identify SNPs has the advantage of facilitating the discovery of polymorphisms directly linked to the gene-coding regions (Balcluniene *et al.*, 2001).

Motivated by the former interest of our group in genetic polymorphisms involved with the immune system, we selected five genes from a previous expression microarrays study of hippocampal cornu ammonis (CA1) (Colangelo *et al.*, 2002) area of Alzheimer's disease subjects (AD) and using the methodology of searching for polymorphisms in the database of ESTs, it predicted the existence of deletions ranging from 1 to 10 bp in the inflammatory pathway genes related to Alzheimer's disease. A virtual validation confirmed that some of the variations identified had been reported previously and had been confirmed in DNA samples, showing that this method is a feasible way to detect genetic variations that merit further exploration in genetic risk factor association studies (Colangelo *et al.*, 2002; Lemos *et al.*, 2009). A similar approach was used for major depressive disorder (anterior cingulated cortex), bipolar affective disorder (left dorsolateral prefrontal cortex) and sporadic Creutzfeldt–Jakob disease (prefrontal cortex) (Souza *et al.*, 2010).

At the present moment, our group is screening DNA samples to validate the variations predicted from the database of ESTs, and confirmation of our findings will support the current concept that microdeletions and other rare variations that differ from SNPs might be important culprits as genetic risk factor for common neurological and behavioural conditions, as well as for other complex traits (Walsh *et al.*, 2008).

Material and methods

Here, we use a ESTs database to predict and validate SNPs directly linked to gene-coding regions of HLA class I genes (HLA-A, HLA-B and HLA-C) that may be associated with different diseases.

In this analysis, we select the coding regions (CDS) of HLA class I genes to predict and validate single-nucleotide polymorphisms using ESTs database. The software CLCbio Workbench Combined® (http://www.clcbio.com) version 5.7.1 was used during all the following steps. The main function of CLC Combined Workbench is to perform multiple alignments using large data sets in a short period of time, initially to build spliced ESTs and mRNA (cDNA) files retrieved, respectively, from the Goldenpath (http://www.genome.ucsc.edu) and NCBI (http://www.ncbi.nlm.nih.gov) databases and later to perform multiples batches of Smith–Waterman BLASTn alignments (Fig. 1).

An initial number of 18.029 ESTs related to the genes previously selected were screened to decrease the chances of handling sequencing artefacts and to select possible SNPs, selecting alignments with identity

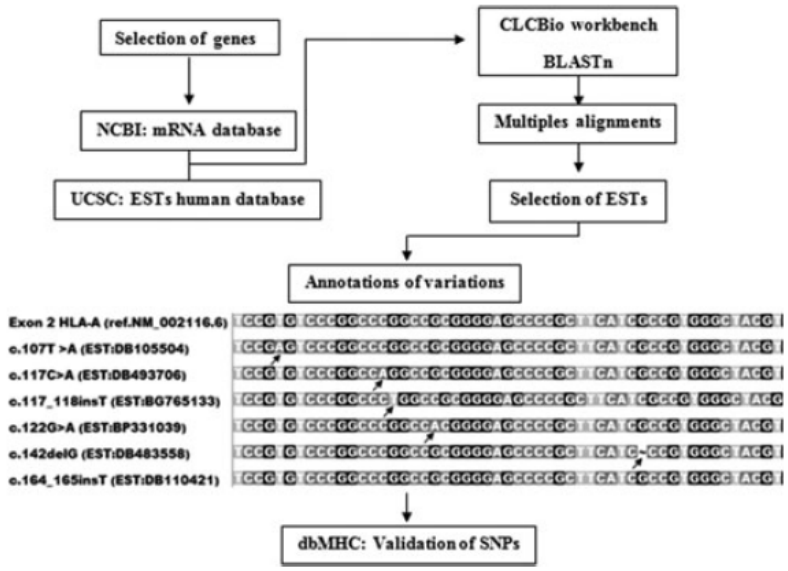


Figure 1. Sequential steps of the proposed Bioinformatics pipeline for screening candidate variations from ESTs database.

between 97 and 99% and *e*-value index under 10^{-20}. The percentage of identify quantifies the degree of similarity between the sequences compared. The expect value (*e*-value) corresponds to the probability of obtaining any alignment score compared to another random sequence with the same size and nucleotides content. For example, an alignment obtaining an *e*-value of 0.05 means that there are 5 in 100 chances for this overlapping to occur by chance, meaning that a lower *e*-value is associated with a higher overlapping significance (Altschul *et al.*, 1990).

A Database of Major Histocompatibility Complex (dbMHC) available at NCBI was used to validate the SNPs found. This validation was conducted on 6 April 2011. The whole process was accelerated by an additional piece of hardware, named CUBE® version 1.06 -CLC bio LLC, Cambridge, MA, USA. The Cube takes out the hassle of high-performance computing, by offering a powerful, flexible and integrated hardware-accelerated bioinformatics solution, which is accessed with ease through the USB port and a graphical user interface. This hardware ensures a performance per accelerator unit of – at least – 50 times that of a fast 3 GHz/4 GB RAM desktop computer, when running the Smith–Waterman algorithm (http://www.clcbio.com).

Results and discussion

From the initial number of 18.029 ESTs that were aligned with the mRNAs of genes, 3.287 ESTs were obtained after the application of the parameters to decrease the chances of handling sequencing artefact and to select possible SNPs. Annotation from our analysis revealed various classes of variations: 94 transversions, 100 transitions, 10 deletions and 17 insertions, scoring a total of 221 potential SNPs. Of these SNPs that were found, 101 were already described in dbMHC (Tables S1–S3 in Supporting Information) and 120 are potential new polymorphisms (Tables S4–S6 in Supporting Information).

Some of the SNPs were found in up to 210 different ESTs (Electronic Supplementary Material). These ESTs that were aligned with the exons and selected for annotation of possible SNPs presented *e*-value $<10^{-100}$ and identity above 97% to decrease the chance for spurious association because of sequencing artefacts.

In Figure 2 are different polymorphisms found in exon 2 of HLA-A and not described in the database. Several other polymorphisms were found in this region (Electronic supplementary material).

The HLA-A, HLA-B and HLA-C genes contain eight exons: the analysis identified potential new SNPs in exons 1, 2, 3, 4 and 5. In general, the annotations of the variations revealed a higher number of SNPs in exons 2 (67 SNPs) and 3 (61 SNPs). These variations were observed mainly in HLA-B. These results corroborate with data reported for HLA class I genes. Polymorphisms in genes of HLA class I are located mainly in exons 2 and 3, which encode alpha 1 and alpha 2 domains of the molecule that contain the binding site for antigen, and the rate of not synonymous substitutions is higher in regions binding antigenic peptides (Hughes & Nei, 1988). This feature of HLA genes was also observed in the results of our validation. After a virtual translation of the aligned sequences, a higher number of not synonymous mutations (156) were found when compared to synonymous mutations (38) and frameshift mutations (25) (Supporting information). Most of the data generated with this analysis are potentially new and are not yet deposited in public databases, such as the MHC database (NCBI) and IMGT/HLA Database (Tables S4–S6 in Supporting Information).

The regions with the greatest evidence of the role of balancing selection are those that interact directly with antigen, indicating that pathogenic microorganisms are the source of selection acting on HLA genes (Garrigan & Hedrick, 2003).

There is still no consensus among researchers who study the evolution of the HLA system. However, there is a dominant evolutionary hypothesis, known as pathogen-driven selection. Borghans *et al.* (2004) performed mathematical simulations of the co-evolution model between pathogens and immune system. According to this model, changes in the genes of pathogens and immune genes occur in parallel: the pathogens tend to accumulate escape mutations (mutations that avoid presentation by common HLA molecules in the host population) and new HLA molecules that present more efficiently these peptides of escape tend to arise. This mechanism explains satisfactorily many of the characteristics of HLA class I genes as the large number of alleles, high heterozygosity rates and

Figure 2. Sequences alignment of ESTs with exon 2 of gene HLA-A, not described in the HLA system database. Underlined letters indicate positions of virtual variations for nucleotide and amino acids (frame +3).

great persistence of alleles in populations over generations. Thus, it is evident the importance of studying polymorphisms of the HLA genes, mainly SNPs located in exons 2 and 3. The methodology used for this work was able to find several candidate SNPs in these regions that cause changes in protein structure and may indicate possible new alleles.

Another important factor was the repetition of SNPs in different ESTs. The repetition of SNPs throughout the analysis is an important finding, because the presence of the same SNP in different EST could also indicate that these are potential real SNPs (Balcluniene *et al.*, 2001).

The advantage of using existing EST data has following advantages: lower costs, less time for the SNP discovery process and without specialized equipment, as the primary sequence data are available. This approach involves a large amount of data and may indicate potential new SNPs (Navratil *et al.*, 2008). Furthermore, with the reference number of the EST, it is possible to analyse in which tissue was found the EST, the type of the study performed and the title of the study. These data can assist in experimental screening studies.

In conclusion, the approach used in this work was able to analyse and handle a large amount of data from different genomic libraries. Annotation from our analysis revealed various classes of possible new variations that can help genotyping studies. Most of these variations were found mainly in exons 2 and 3 which encode the molecular structures of the alpha 1 and alpha 2 domains of the class I membrane receptor, which form the antigen binding site. These structures are critical to critical numerous immune functions.

Thus, bioinformatics pipelines seem to be useful approaches to help screening for novel genetic variations at the HLA panel, and further analysis will foster this aim to provide celerity at the massive analysis of data currently generated in large-scale high-throughput experiments.

Acknowledgements

We are greatly indebted to Henrique Castelletti for technical support. This study received financial support from the following Brazilian funding agencies and academic bureaus: LIKA- UFPE, PROPESQ-UFPE and CAPES.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment tool. *Journal of Molecular Biology*, **215**, 403.

Bae, J.S., Cheong, H.S., Kim, J.O., Lee, S.O., Kim, E.M., Lee, H.W., Kim, S., Kim, J.W., Cui, T. & Inoue, I. (2008) Identification of SNP markers for common CNV regions and association analysis of risk of subarachnoid aneurysmal hemorrhage in Japanese population. *Biochemical and Biophysical Research Communications*, **373**, 593.

Balcluniene, J., Syvänen, A.C., McLeod, H.L., Pettersson, U. & Jazin, E.E. (2001) The Geographic distribution of Monoamine Oxidase haplotypes supports a Bottleneck during the dispersion of Modern Humans from Africa. *Journal Molecular Evolution*, **52**, 157.

Borghans, J.A., Beltman, J.B. & De Boer, R.J. (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, **55**, 732.

Colangelo, V., Schurr, J., Ball, M.J., Pelaez, R.P., Bazan, N.G. & Lukiw, W.J. (2002) Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and proinflammatory signaling. *Journal of Neuroscience Research*, **70**, 462.

Garrigan, D. & Hedrick, P. (2003) Perpective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707.

Hughes, A. & Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167.

Kim, B.C., Kim, W.Y., Park, D., Chung, W.H., Shin, K.S. & Bhak, J. (2008) SNP@Promoter: a database of human SNPs (single nucleotide polymorphisms) within the putative promoter regions. *BMC Bioinformatics*, **9**, 1.

Lemos, R.R., Castelletti, C.H., Lima Filho, J.L., Marques, E.T. & Oliveira, J.R. (2009) In silico identification of new genetic variations as potential risk factors for Alzheimer's Disease in a Microarray oriented simulation. *Journal of Molecular Neuroscience*, **39**, 244.

Lie, B.A. & Thorsby, E. (2005) Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Current Opinion in Immunology*, **17**, 529.

Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B. & Erlich, H.A. (2002) Nomenclature for factors of the HLA system. *Tissue Antigens*, **60**, 460.

Navratil, V., Penel, S., Delmonte, S., Mouchiroud, D., Gautier, C. & Aouacheria, A. (2008) DigiPINS: a database for vertebrate exonic single nucleotide polymorphisms and its application to cancer association studies. *Biochimie*, **90**, 567.

Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E. & Feuk, L. (2007) Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, **39**, 15.

Souza, M.B., Lemos, R.R., Cunha, J.E., Lima Filho, J.L. & Oliveira, J.R. (2010) Searching for new genetic risk factors for neuropsychiatric disorders in expression databases. *Journal of Molecular Neuroscience*, **41**, 193.

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661.

Useche, F.J., Gao, G., Harafey, M. & Rafalski, A. (2001) High-throughput identification, database storage and analysis of SNPs in ESTs sequences. *Genome Informatics*, **12**, 194.

Vandiedonck, C. & Knight, C.J. (2009) The human major histocompatibility complex as a paradigm in genomics research. *Briefings in Functional Genomics and Proteomics*, **8**, 379.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M. *et al.* (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, **320**, 539.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 HLA – A: previously described polymorphisms.

Table S2 HLA – B: previously described polymorphisms.

Table S3 HLA – C: previously described polymorphisms.

Table S4 HLA – A: possible new polymorphisms.

Table S5 HLA – B: possible new polymorphisms.

Table S6 HLA – C: possible new polymorphisms.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary Material

Using ESTs database to validate and predict single polymorphisms at the HLA system

International Journal of Immunogenetics

Thalita Cristina Figueiredo[1,2] and João Ricardo Mendes de Oliveira[1,2,3]

[1]*Keizo Asami Laboratory (LIKA), Federal University of Pernambuco, Recife, PE, Brazil.*
[2]*Biological Sciences Graduate Program, Federal University of Pernambuco, Recife, PE, Brazil.*
[3]*Department of Neuropsychiatry, Federal University of Pernambuco, 50670-901,Recife, PE, Brazil.*

Correspondig author: J.R.M. Oliveira. Phone: +55 (81) 21268539. E-mail: joao.ricardo@ufpe.br.

S1 - HLA – A: previously described polymorphisms

Localization	Variations Nucleotide	Variations Amino acids	refSNPs	ESTs
Exon 1	c.28C > G	Val3Leu	rs1143146	DB268746 DB250117 DB160195 DB114455 DB258930 DB264247 DA835878 DA946046 DA945638 DA951841 DA945169 DA938437 DA938025 DA944499 DA950544 DA937572 DA943839 DA936525 DA936314 DA941122 DA934015 DA940404 DA940144 DA940102 DB008539 DA849606 DA933999 DA933886 DA959389 DA998521 DA808596 DA931741 DA886399 DA958458 DA931407 DA846457 DA852313 DA957099 DA845240 DA949729 DA947318 DA540482 DA484235 DA363558 DA083124 DA139255 DA139060 DA137926 DA129325 DA132369 DA130075 DA038869 DA038731 DA038015 BP306838 BP303889 BP302833 BP302831 BP302828 BP302762 BP302751 BP302676 BP302469 BP300956 BP300615 BP300425 BP300223 BP298536 BP298219 BP298209 BP297996 BP297407 BP297310 BP295011 BP257013 BP256937 BP256399 52226270 52223971 52172243 52172167 BP256374 BP256339 BP256164 BP256159 BP256040 BP255772 BP255760 BP255732 BP255676 BP255581 BP255538 BP255328 BP255017 BP255011 BP254931 BP254930 BP254919 BP254835 BP254486 BP254183 BP254165 BP254143 BP254123 BP254091 BP254030 BP253962 BP253948 BP253898 BP253872 BP253845 BP253640 BP253600 BP253320 BP253256 BP253195 BP252950 BP252949 BP252786 BP252712 BP252600 BP252538 BP252474 BP252443 BP252416 BP252286 BP252236 BP252199 BP252124 BP252097 BP210168 BP194823 CB150738 CB149619 CB149607 CB140984 CB140980 CB140960 CB138698 CB136457 CB121546 CB116871 CB116840 CB116443 CB113803 CB113301 CB113285 CB113020 CB113017 CB112935 CB112842 CB112928 CB112496 CB112488 CB109052 CB108949 DC361390 DC407248 DC380395 DC380391 DC406553 DC390186 DC406349 DC404608 DC407337 DC413787 DC422785 DC421018 DC428366 DC380025 DC406845 DC405592 DC393415 DC393391 DC393385 DC393382 DC393369 DC393359 DC362360 DB466126 DB501532 DC315843 DC325662 DC386824 DC396046 DC421443 DC424697 DC413408 DC295210 DC295116 BJ994553 DT218592 DB178044 DB190860 DB190841 DB190247 DB260233 DC327401 DC373127 DC298320 DT216982 DT215921 DB112272 DB151287 DB190946 DB190936 DB190880 DB190701 DB177310 DB261609
Exon 1	c.41C > T	Ser14Leu	rs2230954	DB103173 DB181436 DB205326 DB159587 DB204372 DB210052 DB125907 DB118873 DB118153 DB130437 DB163416 DB117600 DB117200 DB188898 DB115894 DB187725 DB271503 DB114746 DB114650 DB120598 DB120525 DB113170 DB152578 DB106652 DA952441 DA951942 DB018735 DA943306 DA936645 DB016990 DA817670 DB015931 DB008493 DA822540 DB008111 DB047414 DA933057 DB007956 DB007250 DA959651 DA932339 DB006870 DA814385 DB006477 DB006425 DB012070 DA931293 DA813636 DA957689 DA956026 DA948058 DA540665 DA681912 DA429897 DA429635 DA429433 DA429161 DA429106 DA428743 DA428245 DA427711 DA550991 DA426564 DA426382 DA426272 DA426221 DA432513 DA432251 DA432218 DA425785 DA425589 DA425185 DA424975 DA541149 DA200670 DA260030 DA199759 DA008721 DA008436 DA008345 DA008160 BP305988 BP304573 BP303612 BP302811 BP302743 BP302734 BP302334 BP301678 BP301205 BP300399 BP300315 BP298074 BP297894 BP297065 BP296403 BP295619 BP295561 BP295508 BP289597 BP264296 BP264039 BP263399 BP262704 BP262647 BP262224 BP262070 BP262019 BP262003 BP261980 BP261973 BP261156 BP261152 BP261058 BP260764 BP260713 BU541672 BM046946 BM043730 BI831726 DC344750 DC405946 DC390287 DC397833 DC413167 DC405082 DC412946 DB479900 DB478529 DB476075 DB483558 DB151810 DB256811 DB105660 DB105504 DB111579 DB183871 DB183190 DB182822 DB444064 DB493706 DB488964 DB181098 BP367301 BP306937 BP303692 BP303221 BP301933 BP301729 BP301640 BP301360 BP301240

Exon 1	c.47C>T	Ala16Val	rs41554816	DA821981
Exon 2	c.97T>A	Phe33Ile	rs1136659	DB205326 DB204372 DB210052 DB271503 DA008160 BP303801 BP303692 BP303612 BP303602 BP303595 BP303492 BP303396 BP303221 BP302811 BP302743 BP302734 BP302712 BP302520 BP302453 BP302334 BP302123 BP301937 BP301933 BP301899 BP301755 BP301741 BP301729 BP301678 BP301640 BP301480 BP301360 BP301243 BP301240 BP301205 BP301147 BP301010 BP300851 BP300738 BP300729 BP300701 BP300605 BP300588 BP300558 BP300556 BP300411 BP300399 BP300338 BP300315 BP300276 BP300136 BP300082 BP300027 BP299985 BP299882 BP299859 BP299784 BP299742 BP299703 BP299660 BP299655 BP299644 BP299636 BP299625 BP299588 BP299400 BP299322 BP299255 BP299128 BP298956 BP298888 BP298867 BP298749 BP298616 BP298423 BP298404 BP298352 BP298340 BP298335 BP298248 BP298192 BP298074 BP298059 BP298042 BP297946 BP297913 BP297471 BP297203 BP297065 BP297021 BP296979 BP296925 BP296895 BP296875 BP296796 BP296781 BP296762 BP296746 BP296618 BP296559 BP296509 BP296504 BP296501 BP296434 BP296392 BP296367 BP296321 BP296255 BP296196 BP296000 BP295957 BP295830 BP295568 BP295497 BP295485 BP295479 BP295291 BP295284 BP295010 BP294803 BP294653 BP271238 BP270802 BP264320 BP264313 BP264296 BP264165 BP264039 BP264019 BP263994 BP263868 BP263514 BP263258 BP263174 BP262911 BP262884 DB479900 DB478529 DB476075 DB455738 DB448923 DB445935 DB503261 DB256811 BP299463
Exon 2	c.98T > C	Phe33Tyr	rs2075684	DB205326 DB204372 DB210052 DB271503 DA008160 BP303801 BP303692 BP303612 BP303602 BP303595 BP303492 BP303396 BP303221 BP302811 BP302743 BP302734 BP302712 BP302520 BP302453 BP302334 BP302123 BP301937 BP301933 BP301899 BP301755 BP301741 BP301729 BP301678 BP301640 BP301480 BP301360 BP301243 BP301240 BP301205 BP301147 BP301010 BP300851 BP300738 BP300729 BP300701 BP300605 BP300588 BP300558 BP300556 BP300411 BP300399 BP300338 BP300315 BP300276 BP300136 BP300082 BP300027 BP299985 BP299882 BP299859 BP299784 BP299742 BP299703 BP299660 BP299655 BP299644 BP299636 BP299625 BP299588 BP299400 BP299322 BP299255 BP299128 BP298956 BP298888 BP298867 BP298749 BP298616 BP298423 BP298404 BP298352 BP298340 BP298335 BP298248 BP298192 BP298074 BP298059 BP298042 BP297946 BP297913 BP297471 BP297203 BP297065 BP297021 BP296979 BP296925 BP296895 BP296875 BP296796 BP296781 BP296762 BP296746 BP296618 BP296559 BP296509 BP296504 BP296501 BP296434 BP296392 BP296367 BP296321 BP296255 BP296196 BP296000 BP295957 BP295830 BP295568 BP295497 BP295485 BP295479 BP295291 BP295284 BP295010 BP294803 BP294653 BP271238 BP270802 BP264320 BP264313 BP264296 BP264165 BP264039 BP264019 BP263994 BP263868 BP263514 BP263258 BP263174 BP262911 BP262884 DB479900 DB478529 DB476075 DB455738 DB448923 DB445935 DB503261 DB256811 BP299463
Exon 2	c.257A > G	Gln86Arg	rs1064588	DC344750 DC412946 DB183871 DB183190 DB181098
Exon 2	c.259G > C	Glu87Gln	rs2230991	DC344750 DC412946 DB183871 DB183190 DB181098
Exon 2	c.253G > C	Asp85His	rs11539963	BG475386
Exon 2	c.98T > A	Phe33Tyr	rs2075684	DB009832 DA358826 DA432971 DA432899 DA008715 DA008594 CX758695 CX758340 CX752735 CV814296 BP331555 BP331346 BP330204 BP329811 BP329579 BP329346 BP328645 BP328298 BP328195 AL541870 CF132090 CF131062 CF130927 CF128829 CF128197 CF128064 CF127655 CF127135 BQ053341 CF126960 CF126934 CF126871 CF126784 CF126486 CF126330 CF126167 CF126124 CD250861 CB216533 BQ961694 BQ054199 BQ053007 BQ052973 BQ052954 BQ052528 BQ052505 BQ052311 BQ051831 BQ051738 BQ051688 BM926448 BM917981 BM917659 BM917450 BM917202 BM838759 BM837450 BM829375 BM785669 BM564459 BG820628 BG766705 BG763613 BG762259 BG761529 BF530676 BF529683 BE906971 CB265480
Exon 2	c.271G > A	Val91Met	rs41564215	DB162753 DB122551 DB161447 DB120602 DB152660 DB257132 DA819721 DA706675 DA815873 DA728153 DB010798 DA566146 DA552557 DA011056 DA012885 DN995449 BP368264 BP367036 BP340306 BP339915 BP339543 BP339423 BP339259 BP339172 BP339095 BP338410 BP338098 BP327192 BP286372 BP285287 CV025826 AL521301 CB216937 CB216568 CB215714 BU186051 BQ892932 BQ881181 BQ675465 BQ674389 BQ673413 BQ672490 BQ672085 BQ670249 BQ670081 BQ668693 BM820527 BM050851 BM008840 BG824362 BG751854 BG386952 BG386532 BF725135 BF342244 BF340444 BE314635 DC406185 DT215168 DB256953
Exon 2	c.282G > C	Gln94His	rs1059463	DB162753 DB122551 DB161447 DB120602 DB152660 DB257132 DA819721 DA706675 DA815873 DA728153 DB010798 DA566146 DA552557 DA011056 DA012885 DN995449 BP368264 BP367036 BP340306 BP339915 BP339543 BP339423 BP339259 BP339172 BP339095 BP338410 BP338098 BP327192 BP286372 BP285287 CV025826 AL521301 CB216937 CB216568 CB215714 BU186051 BQ892932 BQ881181 BQ675465 BQ674389 BQ673413 BQ672490 BQ672085 BQ670249 BQ670081 BQ668693 BM820527 BM050851 BM008840 BG824362 BG751854 BG386952 BG386532 BF725135 BF342244 BF340444 BE314635 DC406185 DT215168 DB256953
Exon 2	c.78C > T	Synonymous	rs1136657	DC327401 DT216982 DT215921 BY796012 BP321903 DB112272 B151287 DB190946 DB190936 DB190880 DB190866 DB190701 DB177310 DB294516 DB261609 DB190392 DC392799
Exon 3	c.363A > G	Ile121Met	rs1136695	CA8433545 BQ690576 BG78626 BP375178 BP301941 BP300949 BP297812 BP295834 BP271293 BP270156 BP264370 BP262003 CB161399 CA847986 BU552434 BU543585 BQ712618 BQ708589 BQ707368 BM048775 BM047999 BM043730 BI831726 BI712717 BI544885 BG773029 BG425785 BG755570 BG341668 BF975696 BM045637 BG775632 BY996721 CA943281 BU197141 BU541673 BU541672 CA843118 BQ891746 BM045720 BI599289 BM007756 BP302755 AL541196 BQ711460 BM048843 BE816326 BM008489 BU185790 AL547313 BU541409 BI548344 BG704883 BM046946 BU153564 BQ957160 BQ941596 BQ933852 BQ924953 BQ231645

				BQ228747 BQ228320 BQ212719 BQ935323 BQ957434 BG437791 BQ960744 BG437329 DC345225 DB182822 DC344750 DB455738 DB445935 DB483558 DB478529 DB448923 DC412946
Exon 3	c.413G > A	Arg138Gln	rs3173420	CA843354 BQ690576 BG78626 BP375178 BP301941 BP300949 BP297812 BP295834 BP271293 BP270156 BP264370 BP262003 CB161399 CA847986 BU552434 BU543585 BQ712618 BQ708589 BQ707368 BM048775 BM047999 BM043730 BI831726 BI712717 BI544885 BG773029 BG425785 BM045637 BG775632 BU541673 BU541672 BQ891746 BM045720 BI599289 BM007756 BP302755 AL541196 BQ711460 BM048843 BM008489
Exon 3	c.502A > C	Lys168Gln	rs1059517	CA843354 BQ690576 BG78626 BP375178 BP301941 BP300949 BP297812 BP295834 BP271293 BP270156 BP264370 BP262003 CB161399 CA847986 BU552434 BU543585 BQ712618 BQ708589 BQ707368 BM048775 BM047999 BM043730 BI831726 BI712717 BI544885 BG773029 BG425785 BG755570 BG341668 BF975696 BM045637 BG775632 BY996721 CA943281 BU197141 BU541673 BU541672 CA843118 BQ891746 BM045720 BI599289 BM007756 BP302755 AL541196 BQ711460 BM048843 BE816326 BM008489 BQ890082 BQ708456 BQ707630 BQ707324 BQ706803 BQ706608 BQ581402 BM007402 BG745507 BG745595 BM007997 BI468489 BQ705901 BQ270389 BM007757 CR998363 CT002386 CR996952 CR993463 CR992423 CR991337 CR990993 CR989653 CR981742 BX426221 BU196368 BU191105 BU168815 BU163573 BU157846 BU151560 BU149459 BQ894316 BQ892628 BQ880652 BQ880525 BQ880250 BQ878220 BQ691441 BQ690745 BQ687688 BQ687395 BQ686406 BQ685668 BQ685392 BQ433938 BM557737 BM006837 BF828600 BQ889579 BI917235 BQ689755 BG390745 CR998012 AL543167 BM006835 AL549738 BQ889398 AL532717 BU185790 AL547313 BU541409 BI548344 BG704883 BM046946 DR422466 BP254523 CN415945 CB159209 CB123178 CB122442 BU171620 BU160995 BQ937637 BQ229590 BI599464 BF879988 BF529029 AU135983 BE410912 BE385415 BU172559 BG911485 BI602583 BP303571 BP294989 BU188476 BI015313 DC404608 BP296432 DA379865 BP258534 BQ876578 BQ717172 BQ716269 BG753944 BG684084 BE260855 BG754839 AL525554 BX403863 DC345225 DC390186 DC393382 DB183871 DC406845 DB182822 DC393391 DC362360 DC406349 DC421018 DC393412 DC422785 DC406553 DC405592 DC393415
Exon 3	c.524A > G	His175Arg	rs1059536	CA843354 BQ690576 BG78626 BP375178 BP301941 BP300949 BP297812 BP295834 BP271293 BP270156 BP264370 BP262003 CB161399 CA847986 BU552434 BU543585 BQ712618 BQ708589 BQ707368 BM048775 BM047999 BM043730 BI831726 BI712717 BI544885 BG773029 BG425785 BG755570 BY996721 CA943281 BU197141 BU541673 BU541672 CA843118 BQ891746 BP302755 AL541196 CR998363 CT002386 CR996952 CR993463 CR992423 CR991337 CR990993 CR989653 CR981742 BX426221 BU196368 BU191105 BU168815 BU163573 BU157846 BU151560 BU149459 BQ894316 BQ892628 BQ880652 BQ880525 BQ880250 BQ878220 BQ691441 BQ690745 BQ687688 BQ687395 BQ686406 BQ685668 BQ685392 BQ433938 BM557737 BM006837 BF828600 BQ889579 BQ689755 CR998012 AL543167 BM006835 AL549738 BQ889398 AL532717 BU185790 AL547313 BU541409 BM046946 DB183871 DB182822
Exon 3	c.555T > G	Asp185Glu	rs1059542	CA843354 BQ690576 BG78626 BP375178 BP301941 BP300949 BP297812 BP295834 BP271293 BP270156 BP264370 BP262003 CB161399 CA847986 BU552434 BU543585 BQ712618 BQ708589 BQ707368 BM048775 BM047999 BM043730 BI831726 BI712717 BI544885 BG773029 BG425785 BG755570 BY996721 BU197141 BU541672 BQ891746 BM045720 BI599289 BM007756 BP302755 AL541196 BQ711460 BM008489
Exon 4	c.650C > G	Pro217Arg	rs45562634	CK003226

S2 - HLA – A: possible new polymorphisms

Localization	Variations Nucleotide	Variations Amino acids	ESTs
Exon 1	c.38T > C	Leu13Pro	DA927219
Exon 1	c.68G > C	Trp23Stop	DA916693
Exon 1	c.58A > G	Thr20Ala	DA921806
Exon 1	c.24C > T	Thr8Ile	DA921017
Exon 1	c.11T > C	Met4Thr	DA432808
Exon 1	c.20G > T	Arg7Leu	DA067510
Exon 1	c.14C > G	Ala5Gly	DB483883

Exon 1	c.9C > A	Synonymous	DB487371
Exon 2	c.117C >A	Synonymous	DB493706
Exon 2	c.142delG	Ala47del fs	DB483558
Exon 2	c.122G > A	Arg41His	BP331039
Exon 2	c.117_118insT	Gly40del fs	BG765133
Exon 2	c.321delC	Tyr108del fs	BG830531 BF568512 BG423505 BI911418
Exon2	c.164_165insT	Gly56del fs	DB110421
Exon 2	c.107T >A	Val36Glu	DB105504
Exon 3	c.546_547insC	Tyr183del fs	BU541673
Exon 3	c.473_474insC	Ala159del fs	BM048843
Exon 3	c.535delC	Glu178del fs	AL559934 AL541209 AL541172
Exon 4	c.654C > T	Pro219Leu	BF809421
Exon 4	c.784_785insC	Trp228del fs	W60665
Exon 4	c.792G > A	Gly231Ser	AL568783 CN415948 AL573721 BQ227788 BM687466 BG167153 BG167012 BF970707 BF771086 BF770493 AV691800 CV361575 CV361517 AL574382 BE160372 BE160290
Exon 4	c.672C > A	Thr224Asn	BF852692
Exon 4	c.716T > G	Leu239Arg	BG746185
Exon 4	c.729G > A	Arg243Gln	CD620640 BQ689949 BQ669609 BM016527 AL542160 CV338925
Exon 4	c.742C > A	Asp248Glu	CR994104 CR989368 CR986250 CD742827 CR982557 CD742770 CD512828 CD368629 CD367343 CD366495 CD365760 CD365012 CB529818 CB529146 CB528498 CA436948 CA308900 BQ710156 BM923524 BM705335 BG744685 BF975616 AV752612 AV708859
Exon 5	c.950T > A	Leu317His	AL569142
Exon 5	c.956G > A	Gly319Glu	AW276632
Exon 5	c.987delC	330Stop	BE889148
Exon 5	c.906C > G	Synonymous	BQ329879
Exon 5	c.935T > A	Ile312Asn	BM477309
Exon 5	c.935T > C	Ile312Thr	BF913065
Exon 5	c.943delC	Leu315del fs	BE736535
Exon 5	c.928delG	Gly310del fs	BF760613 BG954531
Exon 5	c.936T > G	Ile312Met	AL568790

Exon 5	c.910C > G	Pro304Ala	BF760459

S3 - HLA – B: previously described polymorphisms

Localization	Variations Nucleotide	Variations Amino acids	refSNPs	ESTs
Exon 1	c.5T > G	Leu2Arg	rs9266206	DC389103 DC427657 DA541904 DA930711 DA924571 DA931653 DA936636 DA939321 BM836944 DA542552 DA928609 AL702287 DA042581 CB151350
Exon 1	c.11T > G	Met4Thr	rs1050458	DC389103 DC427657 DA541904 DA930711 DA924571 DA931653 DA939321 BM836944 DA542552 DA928609 AL702287 DA042581 CB151403 DB485391
Exon 1	c.15G > A	Synonymous	rs1050459	DC389103 DC427657 DA541904 DA930711 DA924571 DA931653 DA939321 BM836944 DA542552 DA928609 DA947352 DA946897 DC391606
Exon 1	c.25G > C	Val9Leu	rs1050462	DB485391
Exon 2	c.103T > G	Ser35Ala	rs1131170	DC305825 DC338806 CB268492 BQ229697 BQ072921 BM922688 BM472134 BI912479 BI909824 BI906865 BI906311 BI905837 BI767121 BI518587 BG685270 BG391494 BG257308 DC365452 DC365423 DC365404 DC365381 DC365366 DC365360 DC365342 CV023913 BQ050308 BM807044 BM478193 BM477065 BM476664 BM476615 BM464730 BM464722 BM455947 BM455142 BM455022 BM454348 BG176695 BG106129 BG024549 BF796283 BF791685 BM467474 DC338211 DC298756 DA010226 DA011557 BG677736 BG389471 DC365763 DC409940 DC406029 DC405412 DC395062 DC393679 DC393663 DC405179 DC394702 DC393615 DC393553 DC392590 DC338192 DC405578 DC392382 DC413465 DC413460 DC413456 DC413425 DC366542 DB144277 DB150656 DC366091 DB101323 DB113845 DA704719 DB011522 DA696465 DA689681 BX439601 CD558337 CD517041 CD515795 CD515151 CD359493 CD252240 BQ889065 BG697915 BG681249 BG681244 BG679402 BG679396 BG678756 BG676167 BG676106 BG661353 BG661350 BF526497 BF342367 DA727676 BG740716 DC393700 DC393624 DC397309 BX403678 AL545647 BF338633 DB189226 AL541507 BF339971 BF339884 BF338329 BF338100 BF796353 DA682168 BF338602 BF795890 DC354688 DC312687 DC312327 DC312227 DA369727 DA139483 DA136249 DA129621 DA135951 DA140655 CD579538 CB114884 BQ432426 BM765029 BM753340 BM752535 BG900409 BF972746 BE888511 DC312890 DC311020 DA137129 BX458605 BX417652 BX417959 AL553252 AL542771 AL541548 BM849937 BM849370 BX458291 BX439035 BX365082 AL540445 AL556116 BX366315
Exon 2	c.97T > G	Tyr33Asp	rs2596492	DB150656 DC366091 DB101323 DB113845 DA704719 DB011522 DA696465 DA689681 BX439601 CD558337 CD517041 CD515795 CD515151 CD359493 CD252240 BQ889065 BG697915 BG681249 BG681244 BG679402 BG679396 BG678756 BG676167 BG676106 BG661353 BG661350 BF526497 BF342367 DA727676 DC393700 DC393624 DC397309 BX403678 AL545647 BF338633 DB189226 AL541507 BF339971 BF339884 BF338329 BF338100 DC311020 DA137129 BX458605 BX417652 BX417959 AL553252 AL542771 AL541548 BM849937 BM849370 BX458291 BX439035 BX365082 AL540445 AL556116
Exon 2	c.106G > A	Val36Met	rs1050486	DB150656 DC366091 DB101323 DB113845 DA704719 DB011522 DA696465 DA689681 BX439601 CD558337 CD517041 CD515795 CD515151 CD359493 CD252240 BQ889065 BG697915 BG681249 BG681244 BG679402 BG679396 BG678756 BG676167 BG676106 BG661353 BG661350 BF526497 BF342367 DA727676 DC393700 DC393624 DC397309 BX403678 AL545647 BF338633 DB189226 AL541507 BF339971 BF339884 BF338329 BF338100 DB189226 AL541507 BF339971 BF339884 BF338329 BF338100 BF796353 DA682168 BF338602 BF795890 DC354688 C312687 DC312327 DC312227 DA369727 DA139483 DA136249 DA129621 DA135951 DA140655 CD579538 CB114884 BQ432426 BM765029 BM753340 BM752535 BG900409 BF972746 BE888511 DC312890
Exon 2	c.145G > A	Val49Met	rs41564319	AW402383
Exon 2	c.97 T > C	Tyr33His	rs2596492	DC393749 DC393741 DC393715 DC393655 DC393607 DC393604 DC393580 DC393535 DC393534 DC393697 DC393657 DB009313 DB010004 DC393613 DC354847
Exon 2	c.117C > T	Synonymous	rs3177891	DC393749 DC393741 DC393715 DC393655 DC393607 DC393604 DC393580 DC393535 DC393534 DC393697 DC393657 DB009313
Exon 2	c.193G > A	Ala65Thr	rs1050529	DC380770
Exon 2	c.203G > C	Arg68Thr	rs45512291	AL541507
Exon 2	c.171C > A	Phe57Leu	rs41546213	DA831578

Exon 2	c.246G > A	Synonymous	rs1050564	EL955197 DC380938 DC380892 DC380781 DC380403 DC380321 DC387862 DC398033 DC387583 DC388698 DC380186 DC380143 DC380081 DC380036 DC388520 DC388397 CV571297 CD620610 CD620595 BX480500 BX480274 BX480182 BU663606 BM754711 BM754266 BG742683 BG697797 BG682240 BG682013 BG681967 BG681905 BG681812 BG680913 BG680893 BG680043 BG678159 BG676285 AW630067 DC305825 DC338806 CB268492 BQ229697 BQ072921 BM922688 BM472134 BI912479 BI909824 BI906865 BI906311 BI905837 BI767121 BI518587 BG685270 BG391494 BG257308 DC380770 DC427722 DC427721 DC427720 DC427718 DC427717 DC427711 DC427710 DC427700 DC427673 DC427668 DC427667 DC427665 DC427663 DC427674 DA828462 DA831578 CD620601 BM749909 BG675726
Exon 2	c.272A > G	Tyr91Cys	rs1071816	EL955197 DC380938 DC380892 DC380781 DC380403 DC380321 DC387862 DC398033 DC387583 DC388698 DC380186 DC380143 DC380081 DC380036 DC388520 DC388397 CV571297 CD620610 CD620595 BX480500 BX480274 BX480182 BU663606 BM754711 BM754266 BG742683 BG697797 BG682240 BG682013 BG681967 BG681905 BG681812 BG680913 BG680893 BG680043 BG678159 BG676285 AW630067 DC305825 DC338806 CB268492 BQ229697 BQ072921 BM922688 BM472134 BI912479 BI909824 BI906865 BI906311 BI905837 BI767121 BI518587 BG685270 BG391494 BG257308 DC380770 DA828462 DA831578 CD620601 BM749909 BG675726 DC365452 DC365423 DC365404 DC365381 DC365366 DC365360 DC365342 CV023913 BQ050308 BM807044 BM478193 BM477065 BM476664 BM476615 BM464730 BM464722 BM455947 BM455142 BM455022 BM454348 BG176695 BG106129 BG024549 BF796283 BF791685 BM467474 DC338211 BG389471
Exon 2	c.277G > A	Ala93Thr	rs1131204	EL955197 DC380938 DC380892 DC380781 DC380403 DC380321 DC387862 DC398033 DC387583 DC388698 DC380186 DC380143 DC380081 DC380036 DC388520 DC388397 CV571297 CD620610 CD620595 BX480500 BX480274 BX480182 BU663606 BM754711 BM754266 BG742683 BG697797 BG682240 BG682013 BG681967 BG681905 BG681812 BG680913 BG680893 BG680043 BG678159 BG676285 AW630067 DC305825 DC338806 CB268492 BQ229697 BQ072921 BM922688 BM472134 BI912479 BI909824 BI906865 BI906311 BI905837 BI767121 BI518587 BG685270 BG391494 BG257308 DC380770 BM556597 DA828462 DA831578 CD620601 BM749909 BG675726 DC365452 DC365423 DC365404 DC365381 DC365366 DC365360 DC365342 CV023913 BQ050308 BM807044 BM478193 BM477065 BM476664 BM476615 BM464730 BM464722 BM455947 BM455142 BM455022 BM454348 BG176695 BG106129 BG024549 BF796283 BF791685 BM467474 DC338211 BG389471 DC365763 DC409940 DC406029 DC405412 DC395062 DC393679 DC393663 DC405179 DC394702 DC393615 DC393553 DC392590 DC338192 DC405578 DC392382 DC413465 DC413460 DC413456 DC413425 DC366542 DB144277 DB150656 DB101323 DB113845 DA704719 DB011522 DA696465 DA689681 BX439601 CD558337 CD517041 CD515795 CD515151 CD359493 CD252240 BQ889065 BG697915 BG681249 BG681244 BG679402 BG679396 BG678756 BG676167 BG676106 BG661353 BG661350 BF526497 BF342367 DA727676 DC393700 DC393624 BX403678 AL545647 BF338633 AL541507 AL556116 BX366315 DC296858 DC424119 DA469485 DA646697 CB152414 CB151614 CB151549 CB151403 CB151350 CB151150 CB151132 CB151069 CB141751 CB139133 CB139104 CB138528 CB137681 CB137654 CB131731 CB131713 CB131697 CB130653 CB129451 CB128505 CB127979 CB127943 CB127875 CB127387 CB127195 CB127039 CB127021 CB126854 CB125963 CB125502 CB122314 CB115359 CB113352 BQ083366 BM856651 BM769702 BM769608 BM769587 BM769406 BM147094 BE242740 CB127742 CB128545 CB126018
Exon 2	c.292G > T	Asp98Tyr	rs1131215	DC365452 DC365423 DC365404 DC365381 DC365366 DC365360 DC365342 CV023913 BQ050308 BM807044 BM478193 BM477065 BM476664 BM476615 BM464730 BM464722 BM455947 BM455142 BM455022 BM454348 BG176695 BG106129 BG024549 BF796283 BF791685 BM467474 AL556116 BX366315 DC296858 DC424119 DA469485 DA646697 CB152414 CB151614 CB151549 CB151403 CB151350 CB151150 CB151132 CB151069 CB141751 CB139133 CB139104 CB138528 CB137681 CB137654 CB131731 CB131713 CB131697 CB130653 CB129451 CB128505 CB127979 CB127943 CB127875 CB127387 CB127195 CB127039 CB127021 CB126854 CB125963 CB125502 CB122314 CB115359 CB113352 BQ083366 BM856651 BM769702 BM769608 BM769587 BM769406 BM147094 BE242740 CB127742 CB128545 CB126018
Exon 2	c.272A> T	Tyr91Phe	rs1071816	DC365763 DC409940 DC406029 DC405412 DC395062 DC393679 DC393663 DC405179 DC394702 DC393615 DC393553 DC392590 DC338192 DC405578 DC392382 DC413465 DC413460 DC413456 DC413425 DC366542 DB144277 DB150656 DB101323 DB113845 DA704719 DB011522 DA696465 DA689681 BX439601 CD558337 CD517041 CD515795 CD515151 CD359493 CD252240 BQ889065 BG697915 BG681249 BG681244 BG679402 BG679396 BG678756 BG676167 BG676106 BG661353 BG661350 BF526497 BF342367
Exon 2	c.272A > C	Tyr91Ser	rs1071816	AL556116 BX366315 DC296858 DC424119 DA469485 DA646697 CB152414 CB151614 CB151549 CB151403 CB151350 CB151150 CB151132 CB151069 CB141751 CB139133 CB139104 CB138528 CB137681 CB137654 CB131731 CB131713 CB131697 CB130653 CB129451 CB128505 CB127979 CB127943 CB127875 CB127387 CB127195 CB127039 CB127021 CB126854 CB125963 CB125502 CB122314 CB115359 CB113352 BQ083366 BM856651 BM769702 BM769608 BM769587 BM769406 BM147094 BE242740 CB127742 CB128545 CB126018
Exon 3	c.412G > A	Asp138Asn	rs709055	CB151614 CB150500 CB127943 CB126854 DA570835 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BM819582 BI915180 BE772772 CA394168 AL553546 AL540353 BG432538 CD620611 AL542472 AL542062 AL541166 AL540609 BG767029 DW407943 DW407913 DW407653 DC424119 DW407942 BM820957 CD620592 BU172833 DR423347 BG757531 BY994189 BY995604 BM738128 BF129139 BG340735 BX458762 BX356955 DA933708 CB125963 CD517041 CD515795 CD515151 CD515135 CD359493 BQ889065 BQ720157 BQ717938 BM686499 BI083969 BG682289 BG682210 BG681760 BG681610 BG680237 BG680080 BG679402 BG679396 BG676227 BF526958 BF344619 DA647205 DA938353 DA936857 DC381097 DC413460 DC413460 DC393553 DB190886 DB190609 DB190921 DB111564 DB105462 DB190557 DC380892 DC398033 DC388397 DC392590 DC380794 DC380770 DC380698 DC391599 DC296858 DB151134 DB105737 DC338192 DC406029 DC413425 DC427711 DC380143 DC427674 DC427663 DC380596 DC391967 DC381124 DC427667 DC381156 DC427720 DC427721 DC427665 DC380781 DC380403 DC392609 DC358384 DC365114

				DC393624 DC380321 DC427718 DC427700 DC427717 DC427668 DC380877 DC427710 DC380081 DC388520 DC389788 DB104271 DC427673 DC380186 DC394702 DC405578 DC392382 DC405179 DT218920 DC393663 DC393615 DC380273 DC391909 DC389103 DC390111 DC388963 DC395062 DC388698 DC380036 DC380799
Exon 3	c.363C > G	Ser121Arg	rs1140412	DA570835 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BM819582 BI915180 BE772772 CA394168 AL553546 AL540353 BG432538 CD620611 AL542472 AL542062 AL541166 AL540609 BG767029 BX458762 BX356955 DA933708 CV023913 BM807044 BM478193 BM477065 BM476615 BM464730 BM464722 BM455947 BM455022 BG106129 BG025650 BM476664 BM454348 BI222491 AL556116 BF796353 BF791685 BE546563 DB241833 DR423143 DR422645 AL533106 CB215655 CB156048 BG547097 AV704243 BI918872 BX444755 BF343837 BF342882 BF341207 BF339741 BX403678 DB194513 CD558337 BG740716 BG676167 BX439601 BG697915 BG681249 BG676838 BG676106 BG675796 BF342367 BF339884 BF338100 DA560696 BG681244 BG678756 BF339971 BF338633 BG681649 BM467474 BF796283 BQ050308 BG024549 BG176695 DA593857 DA585729 DA585660 BI668141 AW966655 DA647205 DA938353 DA936857 AL541062 AL537919 BU150172 BU146868 BU146488 BU145515 BQ898329 BQ893633 BQ882365 BQ881346 BQ880956 BQ878893 BQ878404 BG112930 BG027684 BI668378 BI549502 BI461185 BY993828 BF514040 DN831915 BI560067 DB247267 DA586135 DT215163 DB194365 DA592832 CB216333 CB215936 BU146933 BI868229 BQ898420 CN415964 DT215118 DT219750 BE839103 DT216564 DT219820 DB190610 DB110092 DC381097 DB110894 DC365366 DC365342 DC358386 DB190886 DB190609 DB190921 DB111564 DB105462 DB190557 DB254672 DB110976 DC380892 DC398033 DC388397 DC365423 DC365360 DC380794 DC380770 DC380698 DB151134 DB105737 DC365381 DC427711 DC380143 DC397168 DB104400 DC427674 DC427663 DC380596 DC391967 DC381124 DC427667 DC381156 DC427720 DC427721 DC427665 DC380781 DC380403 DC358384 DC380321 DC427718 DC427700 DC427717 DC427668 DC380877 DC427710 DC380081 DC388520 DC389788 DB104271 DC427673 DC380186 DB058238 DB255312 DC427722 DC365452 DT216515 DC380273 BJ996422 DC391909 DC389103 DC390111 DC388698 DB473343 DB484104 DC380036 DB505977 DT214947 DB495877 DB503377 DB249153 DC380799 DB503530 DB493538 DB462213 DB479272
Exon 3	c.419A > C	Tyr140Ser	rs4997052	BI222491 AL556116 BE546563 DB241833 DR423143 DR422645 AL533106 CB215655 CB156048 BG547097 AV704243 BI918872 BX444755 DA593857 DA833094 DA829718 DA585729 DA585660 BI668141 AW966655 DA831601 DA647205 DA938353 DA936857 AW603500 AL541062 AL537919 BU150172 BU146868 BU146488 BU145515 BQ898329 BQ893633 BQ882365 BQ881346 BQ880956 BQ878893 BQ878404 BG112930 BG027684 BI668378 BI549502 BI461185 BY993828 BF514040 DN831915 BI560067 DA373215 CT002994 CR998615 CR997350 DV183563 DB110092 DB111564 DB104400 DC358384 DT218274 DB104271 DB058238 DB55312 DC381190 DT216515 DB473343 DB484108 DB505977 DT214947 DB495877 DB503377 DB493538 DB462213 DB479272
Exon 3	c.418T > A	Tyr140Asn	rs9266150	BX458762
Exon 3	c.361A > T	Ser121Cys	rs41556417	DA373215 CT002994 CR998615 CR997350 CR997045 CR995537 CR988851 CR976965 CR999834 CT001573 CR982627 CR977139 DB247267 CA849102 BM787963 DB183563 DB182047 DT218274
Exon 3	c.369C > T	Synonymous	rs3179865	DN831915 DA373215 CT002994 CR998615 CR997350 DB183563 DB110894 DB182047
Exon 3	c.379G > C	Val127Leu	rs1131112	AL541062 AL537919 BU150172 BU146868 BU146488 BU145515 BQ898329 BQ893633 BQ882365 BQ881346 BQ880956 BQ878893 BQ878404 BG112930 BG027684 BI668378 BI549502 BI461185 BY993828 BF514040 DN831915 BI560067 DA373215 CT002994 CR998615 CR997350 DB183563 DB190886 DB190609 DB190921 DB182047 DB190557 DT218274 DB473343 DB484104 DB503377
Exon 3	c.419A > T	Tyr140Phe	rs4997052	DA647205 DC381097 DB190886 DB190609 DB190921 DC380892 DC398033 DC388397 DC380794 DC380770 DC380698 DC380143 DC427674 DC427663 DC380596 DC391967 DC381124 DC427667 DC381156 DC427720 DC427721 DC427665 DC380781 DC380403 DT218274 DC380321 DC427718 DC427700 DC427717 DC427668 DC380877 DC427710 DC380081 DC388520 DC389788 DC427673 DC380186 DC427722 DC380273 DC388698 DC380036 DC380799
Exon 3	c.368A > T	Tyr123Phe	rs41562013	DB190886 DB190609 DB190921 DB190557 DC427711
Exon 3	c.387G > C	Synonymous	rs12721836	DB183563 DB110894 DB182047 DT218274 DC345263 DB503377
Exon 3	c.499A > T	Thr167Ser	rs41541519	CB151614 CB150500 CB127943 CB126854 DA570835 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BM819582 BI915180 BE772772 CA394168 AL553546 AL540353 BG432538 CD620611 AL542472 AL542062 AL541166 AL540609 BG767029 DW407943 DW407913 DW407653 DC424119 DW407942 BM820957 CD620592 BU172833 DA938353 DA936857 CD137654 DB111564 DB105462 DC296858 DB151134 DB105737
Exon 3	c.409C > T	His137Tyr	rs1050379	DB183563 DB110894 DB058238 DB255312 BJ996422 DB473343 DB484104 DB503377
Exon 3	c.353C > T	Thr118Ile	rs12721827	DT216564
Exon 3	c.355C > A	Leu119Ile	rs12721829	DT216564 DB190886 DB190609 DB190921 DB182047 DB190557 BJ996422

Exon 3	c.487C > G	Synonymous	rs709054	DA829718 DA585729 DA585660 BI668141 AW966655 DA831601 DA647205 DA586135 DT215163 DT212118 BE839103 DC393679 DC413465 DT219820 DC413456 DC366542 DB110092 DC381097 DC365366 DC365342 DC358386 DC413460 DC413460 DC393553 DB254672 DB110976 DC380892 DC398033 DC388397 DC365423 DC365360 DC392590 DC380794 DC380770 DC380698 DC365381 DC338192 DC406029 DC413425 DC427711 DC380143 DC397168 DB104400 DC427674 DC427663 DC380596 DC391967 DC381124 DC427667 DC381156 DC427720 DC427721 DC427665 DC380781 DC380403 DC392609 DC358384 DC365114 DC393624 DC380321 DC427718 DC427700 DC427717 DC427668 DC380877 DC427710 DC380081 DC388520 DC389788 AL541062 AL537919 BU150172 BU146868 BU146488 BU145515 BQ898329 BQ893633 BQ882365 BQ881346 BQ880956 BQ878893 BQ878404 BG112930 BG027684
Exon 3	c.503A > G	Gln168Arg	rs41548613	DC385754
Exon 3	c.512G > T	Trp171Leu	rs41551018	CB151614 CB150500 CB127943 CB126854 DA570835 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BM819582 BI915180 BE772772 CA394168 AL553546 AL540353 BG432538 CD620611 AL542472 AL542062 AL541166 AL540609 BG767029 DW407943 DW407913 DW407653 DC424119 DW407942 BM820957 CD620592 BU172833 BX458762 BX356955 DA933708 CB125963 DA938353 DA936857
Exon 3	c.527A > T	Glu176Val	rs707912	CB151614 CB150500 CB127943 CB126854 DA570835 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BM819582 BI915180 BE772772 CA394168 AL553546 AL540353 AL542472 AL542062 AL541166 AL540609 BG767029 DW407943 DW407913 DW407653 DC424119 DW407942 BM820957 CB125963 CD517041 CD515795 CD515151 CD515135 CD359493 BQ889065 BQ720157 BQ717938 BM686499 BI083969 BG682289 BG682210 BG681760 BG681610 BG680237 BG680080 BG679402 BG679396 BG676227 BF526958 BF344619 BF343837 BF342882 BF341207 BF339741 BX403678 DB194513 CD558337
Exon 3	c.534delG	Glu178del fs	rs41559517	AL553546 BX356955
Exon 3	c.535delC	Gln179del fs	rs41557920	AL540353
Exon 3	c.537C > T	Arg180Trp	rs9266144	DR423143 DR422645 AL533106 CB215655 CB156048 BG547097 AV704243 BI918872 DA829718 DA585729 DA585660 DA831601 AL541062 AL537919 BU150172 BU146868 BU146488 BU145515 BQ898329 BQ893633 BQ882365 BQ881346 BQ880956 BQ878893 BQ878404 BG112930 BG027684 BI668378 BI549502 BI461185 BY993828 BF514040 DN831915 DA373215 CT002994 CR998615 CR997350
Exon 3	c.559G > C	Glu187Val	rs2308466	DR422645 AL533106 CB215655 CB156048 BG547097 AV704243 BI918872
Exon 3	c.560A > T	Glu187Gln	rs2308466	DR422645 AL533106 CB215655 CB156048 BG547097 AV704243 BI918872
Exon 4	c.652A > G	Ile218Val	rs1050341	DA069779 CN415997 CF126316 CB266188 BQ925999 BQ880081 BQ710321 BQ422681 BM853136 BI018970 BG989423 BG545602 BG108318 BF902064 AW351882 BF829089 AL541512 BX424871 BF899345 BF901021 BX438817 BU781641 AL550551 CF127683 BQ950074 BF902061 BG989174 BF901026 AL559899 BI052604 BF902071 BX425095 AL542511 BU686992 AL547121 AL533479 BX417794 H02678 BX457496 AL572196 BG987981 BF901008 BF901006 AA188398 BQ706551 BX403477 AA327267 BG954002 BF891843 BG745774 BX446371 AW851064 BF899340 AA209510 AW793284 BX412479 BG989169
Exon 4	c.668C > T	Ala223Val	rs1050723	CR999293 CR991971 CR991697 CR988546 CR987134 CR985485 CR984502 CR981602 CR977778 BU191227 BQ712697 BQ711915 BQ709263 BQ689985 BQ685972 BQ685475 BM917765 BI012501 BI010713 BG818166 BG285893 BG252007 BF976374 BF975283 AI114563 CA397572 AI110593 BM706717 BM007917 CR979770 BU145622 CT003703 BX442956 BQ685256
Exon 4	c.814G > A	Val272Met	rs41558116	BF360173 BE832416
Exon 4	c.786T > C	Synonymous	rs1050823	BG681139 BG990319 BG989496 BG739920 BG989247 BF230033 BI003224
Exon 5	c.916G > A	Val306Ile	rs1131500	BF841817 CN415997 BX457496 BX434074 BX378969 BX378968 AL576306 AL572196 AL568958 AL568940 AL565135 AL533105 CD620600 CD620599 CD620593 CF131910 CF127683 CD523222 CD521585 CD515072 BX412479 BX412216 BX412135 AL569612 CB145685 CB121138 CA439443 CA439162 CA439083 CA431842 CA414834 CA309330 CA306114 BU943852 BU861341 BU838588 BU788986 BU786851 BU627799 BQ883939 BQ880081 BQ639241 BQ638176 BQ632152 BQ422681 BQ229678 BM991229 BM989982 BM853136 BM851280 BM851220 BM851046 BM848838 BM819412 BM763284 BI021515 BI018970 BI005632 BI001076 BG990299 BG990027 BG989494 BG987787 BG984021 BG954002 BG572872 BG484360 BG179533 BG108318 BG026446 BF934577 BF915549 BF915010 BF906089 BF901344 BF847158 BF847083 BF832755 BF832740 BF760100 BF378654 BF378653 BF378562 BF330970 BF330966 BF083225 BE832466 BE831217 BE828417 BE826308 BE619806 BE270974 BE149772 AW947364 AW945862 AW835500 AW800141 AW799957 AW799201 AW799199 AW799044 AW798321 AW797150 AW796820 AW795874 AW795868 AW795850 AW795847 AW794434 AW603462 AW376417 AW372940 AW366805 AW366802 AW366701 AW366694 AW366693 AW366691 AW366268 AW352359 AW352306 AA209510 AA075426 CF551975 CA431478 BI064193 BG989162 AW795876 AW376424 AA573819 BX366314 AW361438 CB268915 CB266188 BQ925999 BG978852 AV701958 AW945923 AL567443 BX365081 BE927007 AW842147 AW351882 AA360691 AA340073 AL568414 BG113110 AW795872 BF934431 BF933173 BF933155 BF755848 BF087863 AW799976 AW795873 AW795852 AW794662 AW604957 AW404877 AW366267 AA361528 AA360829 AA327536 AA327131 AA319983 AA300921 H60666

Exon 5	c.985G > A	Ala329Thr	rs1051488	BG983930 BE695922 DB197834 DA761792 DA896317 BP424289 BP417392 AV734316 AV691664 CN415997 BX457496 BX434074 BX378969 BX378968 AL576306 AL572196 AL568958 AL568940 AL565135 AL533105 CD620600 CD620599 CD620593 CF131910 CF127683 CD523222 CD521585 CD515072 BX412479 BX412216 BX412135 AL569612 CB145685 CB121138 CA439443 CA439162 CA439083 CA431842 CA414834 CA309330 CA306114 BU943852 BU861341 BU838588 BU788986 BU786851 BU627799 BQ883939 BQ880081 BQ639241 BQ638176 BQ632152 BQ422681 BQ229678 BM991229 BM989982 BM853136 BM851280 BM851220 BM851046 BM848838 BM819412 BM763284 BI021515 BI018970 BI005632 BI001076 BG990299 BG990027 BG989494 BG987787 BG984021 BG954002 BG572872 BG484360 BG179533 BG108318 BG026446 BF934577 BF915549 BF915010 BF906089 BF901344 BF847158 BF847083 BF832755 BF832740 BF760100 BF378654 BF378653 BF378562 BF330970 BF330966 BF083225 BE832466 BE831217 BE828417 BE826308 BE619806 BE270974 BE149772 AW947364 AW945862 AW835500 AW800141 AW799957 AW799201 AW799199 AW799044 AW798321 AW797150 AW796820 AW795874 AW795868 AW795850 AW795847 AW794434 AW603462 AW376417 AW372940 AW366802 AW366701 AW366694 AW366693 AW366691 AW366268 AW352359 AW352306 AA209510 AA075426 CF551975 CA431478 BI064193 BG989162 AW795876 AW376424 AA573819 BX366314 AW361438 CB268915 CB266188 BU786781 BQ925999 BG978852 AV701958 AL567443 BX365081 BE927007 AW842147 AW351882 AA360691 AA340073 AL568414 BG113110 AW795872 BF934431 BF933173 BF933155 BF755848 BF087863 AW799976 AW795873 AW795852 AW794662 AW604957 AW404877 AW366267 AA361528 AA360829 AA327536

S4 - HLA – B: possible new polymorphisms

Localization	Variations Nucleotide	Variations Amino acids	ESTs
Exon 1	14_15insT	Pro6del fs	DC345206 DC345208
Exon 1	c.61A > G	Glu21Gly	DC385754
Exon 1	c.9C > T	Synonymous	DB010004
Exon 1	c.58A > G	Thr60Val	DC392567
Exon 2	c.99C > T	Synonymous	DC380596
Exon 2	c.88_89insT	Met29del fs	BG740716
Exon 2	c.155T > C	Val51Ala	DA839967
Exon 2	c.106delG	Val35del fs	BF527705
Exon 2	c.123C > A	Synonymous	BP299155
Exon 2	c.175A > G	Arg59Gly	DA828462
Exon 2	c.327C > G	Tyr109Stop	CD620601
Exon 2	c.321delC	Tyr108del fs	BG675726
Exon 2	c.186C > T	Synonymous	BM749909
Exon 2	c.250delT	Trp84del fs	BF339971 BF339884 BF338602
Exon 2	c.259A > G	Asp87Glu	AL556116 BX366315 DC296858 DC424119 DA469485 DA646697 CB152414 CB151614 CB151549 CB151403 CB151350 CB151150 CB151132 CB151069 CB141751 CB139133 CB139104 CB138528 CB137681 CB137654 CB131731 CB131713 CB131697 CB130653 CB129451 CB128505 CB127979 CB127943 CB127875 CB127387 CB127195 CB127039 CB127021 CB126854 CB125963 CB125502 CB122314 CB115359 CB113352 BQ083366 BM856651 BM769702 BM769608 BM769587 BM769406 BM147094 BE242740 CB127742 CB128545 CB126018 CB127563 CB129357 CB151176 DA466342
Exon 3	c.411T > A	His137Gln	BX458762

Exon 3	c.382G > A	Gly126Arg	DA560696
Exon 3	c.377A > T	Asp126Val	DC380321
Exon 3	c.482_483insA	Asp161del fs	BF527705
Exon 3	c.432delC	Lys143del fs	BG327758
Exon 3	c.467C > A	Arg155Ser	DR423143 DR422645 AL533106 CB215655 CB156048 BG547097 AV704243 BI918872 BX444755 DA829718 DA585729 DA585660 BI668141 AW966655 DA831601 DA647205 DA586135 DT215163 DT215118 DB110092 DC381097 DB110894 DC365366 DC365342 DC358386 DB254672 DB110976 DC380892 DC398033 DC388397 DC365423 DC365360 DC380794 DC380770 DC380698 DC365381 DC427711 DC380143 DB104400 DC427674 DC427663 DC380596 DC391967 DC381124 DC427667 DC381156 DC427720 DC427721 DC427665 DC380781 DC380403 DC358384 DC380321 DC427718 DC427700 DC427717 DC427668 DC380877 DC427710 DC380081 DC388520 DC389788 DB104271
Exon 3	c.440A > G	Tyr145Cys	EL594759
Exon 3	c.432_433insC	Lys143del fs	BG740716
Exon 3	c.491delC	Ala161del fs	BG678756 DB105462
Exon 3	c.459C > T	Synonymous	DA831601
Exon 3	c.523delC	Arg175del fs	BG432538 CD620611 BG685579 BF796353 BE546563 BG697915 BG681249 BG676838 BG676106 BG675796 BF342367 BF339884 BF339971 BF338633 BG024549 BI560067
Exon 3	c.524_525insC	Arg175del fs	BU172833
Exon 3	c.540G > T	Arg180Leu	CB151614 CB150500 CB127943 CB126854 DA570835 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BM819582 BI915180 BE772772 CA394168 BG767029 DW407943 DW407913 DW407653 DC424119 DW407942 BM820957 DR423347 BG757531 BY994189 BY995604 BM738128 BF129139 BX458762 DA933708 CV023913 BM807044 BM478193 BM477065 BM476615 BM464730 BM464722 BM455947 BM455022 BG106129 BG025650 BM476664 BM454348 BI222491
Exon 4	c.600G > A	Synonymous	AL545775
Exon 4	c.651delC	Ile218del fs	BG680237 BG675796
Exon 4	c.640A > G	His214Arg	BQ229678
Exon 4	c.635delA	His212del fs	BM830341
Exon 4	c.645C >A	His215Gln	BF343837 BG682289
Exon 4	c.648C > G	His216Gln	BF343837 BG682289
Exon 4	c.671C > A	Thr224Asn	BG745122
Exon 4	c.686C > G	Ala229Gly	BG745122
Exon 4	c.628A > T	Synonymous	BF901006
Exon 4	c.658G > A	Asp220Asn	AA327267
Exon 4	c.642C > G	Thr213Ser	BG954002
Exon 4	c.718A > G	Ala239Thr	BF899345
Exon 4	c.730delG	Asp243del fs	BI083969
Exon 4	c.773G > C	Arg257Thr	CA397572
Exon 4	c.818T > A	Val273Glu	AJ713801
Exon 4	c.860A > T	Gln287Leu	AL547121
Exon 4	c.863G > A	Glu289Lys	BF230033
Exon 5	c.902C > G	Ser301Cys	CV337985

Exon 5	c.920C > A	Pro301His	AA682542
Exon 5	c.949G > A	Val316Ile	CD620614
Exon 5	c.938C > G	Ala308Gly	BI010534
Exon 5	c.939T > A	Synonymous	AA553776
Exon 5	c.914C > G	Thr305Ser	BI010534
Exon 5	c.976G > A	Val322Met	R23714
Exon 5	c.984T > C	Val325Ala	AV761909
Exon 5	c.978G > A	Ala315Thr	CA431478
Exon 5	c.965T > G	Ile310Met	AE376424
Exon 5	c.975T > C	Ile322Thr	AW794662
Exon 5	c.955G > T	Ala319Ser	AW404877
Exon 5	c.941G > C	Gly314Ala	AW795873
Exon 5	c.998G > C	Arg335Thr	BQ721967 BG680080 BF337806
Exon 5	c.1001G > C	Arg334Thr	BQ2681441
Exon 5	c.982G > C	Ala328Pro	CF551975
Exon 5	c.989T > C	Val330Ala	BF934431
Exon 5	c.980T > A	Val327Asp	AW3662671
Exon 5	c.990G > A	Synonymous	AA361528 AA360829 AA327536 AA327131 AA319983 AA300921
Exon 5	c.1001G > T	Arg334Met	H60666

S5 - HLA – C: previously described polymorphisms

Localization	Variations Nucleotide	Variations Amino acids	refSNPs	ESTs
Exon 1	c.22G > A	Ala8Thr	rs2308525	DB252759 DB266655 DA920150 DA952406 DA952089 DA938276 DA943008 DB021398 DA940174 DA926125 DA931547 DA957863 DA957083 DA923870 DA962132 DA916532 DA877394 DA876729 DA954464 DA359170 BX488697 BM856204 BM856160 BM839114 BM837683 BM837007 BM828491 BM786396 BM785679 BM769118 BM767151 BI668378 BI603776 BI598423 BI549502 BI461185 BG398919 BG330385 DT219378 DT219333 DB454481 DB473960 DB473343 DB450563 DB443145 DB503377 DB485099 DB484104 BJ996422 DT218274 DT218298 DB498295 DB497752 DB497453 DB496762 DB453344 DB492148 DB491970 DB489038 DB487180 DB502371 DB501996 DB463509 DB501478 DB184473 DB144716 DB183563 DB110660 DB143773 DB182047 EL582477 EL582473 EL582472 EL582403 EL582402 DT216564 DB184620 DB190921 DB190886 DB190610 DB190609 DB190557 DB190016 DB183740 DB110894 DB058238 T52219
Exon 1	c.47G > C	Gly16Ala	rs1050451	DB252759 DB266655 DA920150 DA952406 DA952089 DA938276 DA943008 DB021398 DA940174 DA926125 DA931547 DA957863 DA957083 DA923870 DA962132 DA916532 DA877394 DA876729 DA954464 DA359170 BX488697 BM856204 BM856160 BM839114 BM837683 BM837007 BM828491 BM786396 BM785679 BM769118 BM767151 BI668378 BI603776 BI598423 BI549502 BI461185 DB119711 DA136341 BG398919 BG330385 DT219378 DT219333 DB454481 DB473960 DB473343 DB450563 DB443145 DB503377 DB485099 DB484104 BJ996422 DT218274 DT218298 DB498295 DB497453 DB496762 DB453344 DB492148 DB491970 DB489038 DB487180 DB502371 DB501996 DB463509 DB501478 DB184473 DB144716 DB183563 DB110660 DB143773 DB182047 EL582477 EL582473 EL582472 EL582403 EL582402 DT216564 DB184620 DB190921 DB190886 DB190610 DB190609 DB190557 DB190016 DB183740 DB110894 T52219
Exon 1	c.28C > A	Leu10Ile	rs2308527	DB119711 DA136341 BJ996422 DT218274 DT218298 DB498295 DB497453 DB496762 DB492148 DB491970 DB487180 DB502371 DB501996 DB463509 DB501478 DB184473 DB144716 DB183563 DB110660 DB143773 DB182047 EL582477 EL582473 EL582472 EL582403 EL582402 DT216564 DB184620 DB190921 DB190886 DB190610 DB190609 DB190557 DB190016 DB183740 DB110894
Exon 2	c.97G > T	Asp33Tyr	rs9264668	DB492148 DB491970 DB11066 BJ996422 DB498295 DB496762 DB495637 DB487180 DB502371 DB463509 DB144716 DB501478 DT216564 DB184620 DB183740 DB110894 DB454481 DB473960 DB453344 DB473343 DB450563 DB503377 DB484104 DB501996 DB058238 DB489038 DT219378 DT219333 DB497752 DB485099 DB497453 DB443145 DB143773 EL582477 EL582472 EL582402
Exon 2	c.134G > A	Arg45His	rs34483041	DB502371 DT216564 DB184620 DB183740 DB110894 DB454481 DB473960 DB453344 DB473343 DB450563 DB503377 DB484104 DB501996 DB058238 DB489038 DT219378 DT219333 DB497752 DB485099 DB443145 EL582477 EL582472 EL582402

Exon 2	c.126A > G	Synonymous	rs41542719	DT216564 DB184620 DB183740 DB110894 DB190921 DB190886 DB190610 DB190609 DB190016 DB190557 EL582477 EL582472 EL582402
Exon 2	c.118G > A	Gly40Ser	rs41551915	DB454481 DB473960 DB453344 DB473343 DB450563 DB503377 DB484104 DB501996 DB058238 DB489038 DB497752 DB485099 DB443145
Exon 2	c.103G > T	Ala35Ser	rs1050445	DB190921 DB190886 DB190610 DB190609 DB190016 DB190557
Exon 2	c.112C > T	Arg38Trp	rs3177892	DB190921 DB190886 DB190610 DB190609 DB190016 DB190557
Exon 2	c.127G > A	Glu43Lys	rs1050438	DB497453
Exon 2	c.222G > A	Synonymous	rs1050556	DB152050
Exon 2	c.201G > A	Synonymous	rs41545020	DB492148 DB491970 DB110660 BJ996422 DB498295 DB496762 DB495637 DB487180 DB502371 DB463509 DB144716 DB501478 DB454481 DB473960 DB453344 DB473343 DB450563 DB503377 DB484104 DB501996 DB489038 DT219378 DT219333 DB497752 DB485099 DB497453 DB190921 DB190886 DB190610 DB190609 DB190016 DB190557 DB443145 DB143773
Exon 2	c.204A > G	Synonymous	rs9266179	DB144716
Exon 2	c.193G > A	Ala65Thr	rs1050529	DB443145
Exon 2	c.263C > T	Thr88Ile	rs41559414	DA122782
Exon 3	c.361A > G	Arg121Trp	rs1131118	CT005179 CT004160 CT002868 CT002485 CT001941 CT001866 CT001837 CT000916 CT000643 CT000527 CR999137 CR998193 CR997902 CR997416 CR997320 CR996457 CR996345 CR996135 CR995534 CR995260 CR994958 CR994433 CR994349 CR994152 CR994011 CR993729 CR993078 CR992937 CR990488 CR990281 CR990203 CR990055 CR990001 CR989309 CR989133 CR988694 CR988020 CR987787 CR985461 CR985317 CR984166 CR984080 CR984059 CR983496 CR983428 CR982326 CR982013 CR981859 CR981104 CR980794 CR980472 CR980358 CR980260 CR980162 CR979687 CR978812 CR978295 CR978173 CR978153 CR977723 CR977271 CR975955 CR975700 CR975363 CR975168 DN990076 CD519504 CD519193 CD518815 CD251428 CD245343 BU542060 BU541239 BU186006 BU073289 BQ948870 BQ948342 BQ945500 BQ928128 BQ887521 BQ887454 BQ670074 BQ669246 BQ638172 BM046290 BM043546 BF339889 CT001573 CR988300 CR985767 CR984618 DA379871 CX753219 CX753133 BQ417227 BQ417031 CA849102 AL556076 CT002994 CR998615 CR997350 CR997045 CR995537 CR991022 CR988851 CR978748 CR976965 CR984463 CN415991 CN415970 AL524367 CB269028 CB265394 BU164745 BQ962558 BQ961803 BQ430599 BQ230617 BQ219799 BI667827 BI258703 BG829779 BG756407 BG720748 BG715933 BG709186 BG707998 BG704635 BG684727 BF975588 BE908353 BE907330 BE905416 DA373215 CR982627 CR977139 CT004504 CT003877 CT003729 CT002048 CR996921 CR987641 CR976441 CR979090 CR999834 BF337407 CR981869 BG538939 BG398177 BE879626 CT003357 CR998501 CR996103 CR988259 CR979157 BU158366 BG707963 BU527415 BU190906 BQ691183 BQ689765 BQ688513 BQ686181 BQ651721 BQ647414 BQ646739 BQ053218 BQ053106 BM561917 BG758194 BF912738 BE268116 BF974934 BE396551 BU157829 BG341496 BI225923 AL560013 BF724266 CR998001 AL560087 BP393775 BP392316 BQ932702 BM453070 BE877208 AL552699 AL541111 BX463710 AL541377 BG254852 AL532778
Exon 3	c.368C > A	Ser123Tyr	rs1131115	CT005179 CT004160 CT002868 CT002485 CT001941 CT001866 CT001837 CT000916 CT000643 CT000527 CR999137 CR998193 CR997902 CR997416 CR997320 CR996457 CR996345 CR996135 CR995534 CR995260 CR994958 CR994433 CR994349 CR994152 CR994011 CR993729 CR993078 CR992937 CR990488 CR990281 CR990203 CR990055 CR990001 CR989309 CR989133 CR988694 CR988020 CR987787 CR985461 CR985317 CR984166 CR984080 CR984059 CR983496 CR983428 CR982326 CR982013 CR981859 CR981104 CR980794 CR980472 CR980358 CR980260 CR980162 CR979687 CR978812 CR978295 CR978173 CR978153 CR977723 CR977271 CR976028 CR975955 CR975700 CR975363 CR975168 DN990076 CD519504 CD519193 CD518815 CD251428 CD245343 BU542060 BU541239 BU186006 BU073289 BQ948870 BQ948342 BQ945500 BQ928128 BQ887521 BQ887454 BQ670074 BQ669246 BQ638172 BM046290 BM043546 BF339889 CT001573 CR988300 CR985767 CR984618 DN831915 AL556076 CT002994 CR998615 CR997350 CR997045 CR995537 CR991022 CR988851 CR978748 CR976965 CR984463 CN415991 CN415970 AL524367 CB269028 CB265394 BU164745 BQ962558 BQ961803 BQ430599 BQ230617 BQ219799 BI667827 BI258703 BG829779 BG720748 BG715933 BG709186 BG707998 BG704635 BG684727 BF975588 BE908353 BE907330 BE905416 DA373215 CR982627 CR977139 CT004504 CT003877 CT003729 CT002048 CR996921 CR987641 CR976441 CR979090 CR999834 BF337407 CR981869 BG538939 BG398177 BE879626 CT003357 CR998501 CR996103 CR988259 CR979157 BU158366 BG707963 BU527415 BU190906 BQ691183 BQ689765 BQ688513 BQ686181 BQ651721 BQ647414 BQ646739 BQ053218 BQ053106 BM561917 BG758194 BF912738 BE268116 BF974934 BE396551 BU157829 BG341496 BI225923 AL560013 BF724266 CR998001 AL560087
Exon 3	c.368C > T	Ser123Phe	rs1131115	DA379871 CX753219 CX753133 BQ417227 BQ417031 CA849102 DR159599 CX787704 CX787205 CX786377 CX165617 CN416002 CN415999 CN415995 CN415981 CF127886 CF127825 CF127614 CF127070 CF126577 CB216947 CB215143 CB150250 CB141388 BU151654 BQ899278 BQ712729 BQ712046 BQ712027 BQ711932 BQ711420 BQ711173 BQ709786 BQ709328 BQ709195 BQ708837 BQ708485 BQ708033 BQ706911 BQ706447 BM904367 BM904337 BM802883 BM477276 BM476190 BM458847 BM007796 BG823761 BG745468 BG745347 CB215718 DB194365 BQ890668 BM837769
Exon 3	c.123C > G	Ser123Cys	rs1131115	BP393775 BP392316 BQ932702 BM453070 BE877208 AL552699 AL541111 BX463710 AL541377 BG254852 AL532778
Exon 3	c.419C > A	Ser140Tyr	rs713032	AL541377 BG254852 AL532778
Exon 3	c.419C > T	Ser140Phe	rs713032	DR159599 CX787704 CX787205 CX786377 CX165617 CN416002 CN415999 CN415995 CN415981 CF127886 CF127825 CF127614 CF127070 CF126577 CB216947 CB215143 CB150250 CB141388 BU151654 BQ899278 BQ712729 BQ712046 BQ712027 BQ711932 BQ711420 BQ711173 BQ709786 BQ709328 BQ709195 BQ708837 BQ708485 BQ708033 BQ706911 BQ706447 BM904367 BM904337 BM802883 BM477276 BM476190 BM458847 BM007796 BG823761 BG745468 BG745347 CB215718 DB194365 BQ890668 BM837769 CN415988 CN415978 CN415968 CN415963

				BQ677414 CN415977 CN415973 CN415983 CT004408 CT004325 CT004321 CT004286 CT004228 CT003950 CT003847 CT003811 CT003555 CT003483 CT003356 CT003354 CT003200 CT002822 CT002703 CT002401 CT002331 CT002294 CT001925 CT001721 CT001197 CT001080 CT000923 CT000861 CT000407 CT000361 CT000181 CT000117 CR999703 CR999518 CR998934 CR998881 CR998537 CR998507 CR998265 CR997988 CR997898 CR995762 CR995443 CR995406 CR994830 CR993307 CR993173 CR992773 CR992720 CR992424 CR991655 CR991218 CR990884 CR990861 CR990680 CR990510 CR989920 CR989237 CR988963 CR988735 CR987930 CR987667 CR987018 CR986167 CR985935 CR984862 CR984397
Exon 3	c.412G > A	Asp138Asn	rs2308575	DR159599 CX787704 CX787205 CX786377 CX165617 CN416002 CN415999 CN415995 CN415981 CF127886 CF127825 CF127614 CF127070 CF126577 CB216947 CB215143 CB150250 CB141388 BU151654 BQ899278 BQ712729 BQ712046 BQ712027 BQ711932 BQ711420 BQ711173 BQ709786 BQ709328 BQ709195 BQ708837 BQ708485 BQ708033 BQ706911 BQ706447 BM904367 BM904337 BM802883 BM477276 BM476190 BM458847 BM007796 BG823761 BG745468 BG745347 CB215718 DB194365 BQ890668 BM837769 CX753051 CB150512 BU150719 BU146279 BQ898807 BQ897775 BQ897527 BQ893651 BQ881776 BQ878007 BQ720774 BQ717834 BQ716908 BM767176 DA646247 CR999764 CR999703 CR999518 CR998934 CR998881 CR998537 CR998507 CR998265 CR997988 CR997898 CR995762 CR995443 CR995406 CR994830 CR993307 CR993173 CR992773 CR992720 CR992424 CR991655 CR991218 CR990884 CR990861 CR990680 CR990510 CR989920 CR989237 CR988963 CR988735 CR987930 CR987667 CR987018 CR986167 CR985935 CR984862 CR984397
Exon 3	c.387C>G	Synonymous	rs1050384	DR159599 CX787704 CX787205 CX786377 CX165617 CN416002 CN415999 CN415995 CN415981 CF127886 CF127825 CF127614 CF127070 CF126577 CB216947 CB215143 CB150250 CB141388 BU151654 BQ899278 BQ712729 BQ712046 BQ712027 BQ711932 BQ711420 BQ711173 BQ709786 BQ709328 BQ709195 BQ708837 BQ708485 BQ708033 BQ706911 BQ706447 BM904367 BM904337 BM802883 BM477276 BM476190 BM458847 BM007796 BG823761 BG745468 BG745347 CB215718 DB194365 BQ890668 BM837769
Exon 3	c.355C > T	Leu119Phe	rs1071649	CN415988 CN415978 CN415968 CN415963 BQ677414 CN415977 CN415973 CN415983
Exon 3	c.512T > G	Leu171Trp	rs1050366	CT005179 CT004160 CT002868 CT002485 CT001941 CT001866 CT001837 CT000916 CT000643 CT000527 CR999137 CR998193 CR997902 CR997416 CR997320 CR996457 CR996345 CR996135 CR995534 CR995260 CR994958 CR994433 CR994349 CR994152 CR994011 CR993729 CR993078 CR992937 CR990488 CR990281 CR990203 CR990055 CR990001 CR989309 CR989133 CR988694 CR988020 CR987787 CR985461 CR985317 CR984166 CR984080 CR984059 CR983496 CR983428 CR982326 CR982013 CR981859 CR981104 CR980794 CR980472 CR980358 CR980260 CR980162 CR979687 CR978812 CR978295 CR978173 CR978153 CR977723 CR977271 CR976028 CR975955 CR975700 CR975363 CR975168 DN990076 CD519504 CD519193 CD518815 CD251428 CD245343 BU542060 BU541239 BU186006 BU073289 BQ948870 BQ948342 BQ945500 BQ928128 BQ887521 BQ887454 BQ670074 BQ669246 BQ638172 BM046290 BM043546 BF339889 CT001573 CR988300 CR985767 CR984618 DN831915 AL541062 AL537919 BU150172 BU146868 BU146488 BU145515 BQ898329 BQ893633 BQ882365 BQ881346 BQ880956 BQ878893 BQ878404 BG112930 BG027684 BU150332 BQ723649 BI668378 BI549502 BI461185 DA379871 CX753219 CX753133 BQ417227 BQ417031 CA849102 CT002994 CR998615 CR997350 CR997045 CR995537 CR991022 CR988851 CR978748 CR976965 CR984463 CN415991 CN415970 AL524367 CB269028 CB265394 BU164745 BQ962558 BQ961803 BQ430599 BQ230617 BQ219799 BI667827 BI258703 BG829779
Exon 3	c.474C > T	Synonymous	rs41553316	AL556076 CN415991 CN415970 AL524367 CB269028 CB265394 BU164745 BQ962558 BQ961803 BQ430599 BQ230617 BQ219799 BI667827 BI258703 BG829779 BG756407 BG720748 BG715933 BG709186 BG707998 BG704635 BG684727 BF975588 BE908353 BE907330 BE905416
Exon 3	c.477C > T	Synonymous	rs41550715	BQ881455 BG109133
Exon 3	c.459C > T	Synonymous	rs1050371	DR159599 CX787704 CX787205 CX786377 CX165617 CN416002 CN415999 CN415995 CN415981 CF127886 CF127825 CF127614 CF127070 CF126577 CB216947 CB215143 CB150250 CB141388 BU151654 BQ899278 BQ712729 BQ712046 BQ712027 BQ711932 BQ711420 BQ711173 BQ709786 BQ709328 BQ709195 BQ708837 BQ708485 BQ708033 BQ706911 BQ706447 BM904367 BM904337 BM802883 BM477276 BM476190 BM458847 BM007796 BG823761 BG745468 BG745347 CB215718 DB194365 BQ890668 BM837769
Exon 3	c.485C > A	Thr162Asn	rs2308584	CT004408 CT004325 CT004321 CT004286 CT004228 CT003950 CT003847 CT003811 CT003555 CT003483 CT003356 CT003354 CT003200 CT002822 CT002703 CT002401 CT002331 CT002294 CT001925 CT001721 CT001197 CT001080 CT000923 CT000861 CT000407 CT000361 CT000181 CT000117
Exon 3	c.527C > A	Ala176Glu	rs2308590	CT005179 CT004160 CT002868 CT002485 CT001941 CT001866 CT001837 CT000916 CT000643 CT000527 CR999137 CR998193 CR997902 CR997416 CR997320 CR996457 CR996345 CR996135 CR995534 CR995260 CR994958 CR994433 CR994349 CR994152 CR994011 CR993729 CR993078 CR992937 CR990488 CR990281 CR990203 CR990055 CR990001 CR989309 CR989133 CR988694
Exon 3	c.526G > A	Ala176Thr	rs41552817	CX753051 CB150512 BU150719 BU146279 BQ898807 BQ897775 BQ897527 BQ893651 BQ881776 BQ878007 BQ720774 BQ717834 BQ716908 BM767176 DA646247
Exon 4	c.703G>A	Ala235Thr	rs41562012	BF854586
Exon 4	c.735G > C	Synonymous	rs1050326	CV570738 BQ690144 BM808258 BF854586
Exon 4	c.744G > A	Synonymous	rs1050320	CV570738 BQ690144 BM808258 BF854586
Exon 4	c.747C > T	Synonymous	rs1050317	CV570738 BQ690144 BM808258 BF854586

S6 - HLA – C: possible new polymorphisms

Localization	Variations Nucleotide	Variations Amino acids	ESTs
Exon 1	c.27C > T	Synonymous	DA083967
Exon 1	c.35T > C	Leu12Pro	DA727821
Exon 1	c.37C > T	Leu13Phe	DA565403
Exon 1	c.49C > A	Leu17Met	DB497752
Exon 2	c.122G > T	Gly41Arg	DA873578
Exon 2	c.140T > A	Ile47Asn	DB009627
Exon 2	c.135C > T	Synonymous	DB153681
Exon 2	c.112 C > A	Synonymous	DB498295 DB496762 DB487180
Exon 2	c.162C > G	Asp53Glu	DA431649
Exon 2	c.193G > A	Ala64Thr	DA727821
Exon 2	c.189C > T	Synonymous	DA964724
Exon 2	c.184A > G	Ser61Gly	DB009826
Exon 2	c.203G > A	Arg68Lys	DA073694
Exon 2	c.160G > A	Asp53Asn	DB497453 DB501478
Exon 2	c.255C > T	Synonymous	DA426786
Exon 2	c.249T > C	Synonymous	BF338032
Exon 2	c.267G > T	Gln88His	BM704428
Exon 3	c.495G > A	Synonymous	CN415983
Exon 3	c.539T > A	Leu180Gln	BU527415 BU190906 BQ691183 BQ689765 BQ688513 BQ686181 BQ651721 BQ647414 BQ646739 BQ053218 BQ053106 BM561917 BG758194 BF912738 BE268116
Exon 4	c.629G > A	Synonymous	AL574842 AL526608 AL563985 AL562495 AL569206 AL569394
Exon 4	c.764T > C	Val254Ala	BF808068

CAPÍTULO III

Journal of Molecular Neuroscience **(Fator de impacto/2010: 2.992)**

Reconsidering the Association Between the Major Histocompatibility Complex and Bipolar Disorder.

J Mol Neurosci
DOI 10.1007/s12031-011-9656-6

Reconsidering the Association Between the Major Histocompatibility Complex and Bipolar Disorder

Thalita Cristina Figueiredo · João Ricardo Mendes de Oliveira

Received: 13 September 2011 / Accepted: 20 September 2011

Abstract Bipolar disorder (BD) is a cyclical and chronic affective disorder, globally recognized as an important public health problem and characterized by mood changes with recurring phases such as mania and depression. It is considered a complex disease, depending on the interaction of genetic and environmental triggers (stressors factors), but with a poorly known pathogenesis. Recent studies have implicated immune factors in the pathogenesis of BD and more particularly associated with different human major histocompatibility complex (MHC) regions. A major consortium study have recently linked BD to hundreds of variations with stronger associations in the MHC region, such as the rs3130297 SNP, located in the NOTCH4 gene, with an additional overlapping association with schizophrenia. This short review focuses on studies that investigated the association between bipolar disorder and the MHC, and the involvement of the immune system in the pathogenesis of the disease, in order to provide further information for additional diagnostic and therapeutic strategies. Fully understanding the etiology and pathophysiology of BD is extremely important to define new approaches for intervention and prevention, maybe through the modulation of the immune system.

Keywords Bipolar disorder · MHC region · Immune system · Polymorphisms

T. C. Figueiredo · J. R. M. de Oliveira (✉)
Keizo Asami Laboratory (LIKA),
Federal University of Pernambuco,
Recife, Pernambuco, Brazil
e-mail: joao.ricardo@ufpe.br

T. C. Figueiredo · J. R. M. de Oliveira
Biological Sciences Graduate Program,
Federal University of Pernambuco,
Recife, Pernambuco, Brazil

J. R. M. de Oliveira
Department of Neuropsychiatry,
Federal University of Pernambuco,
50670-901, Recife, Pernambuco, Brazil

Published online: 11 October 2011

Introduction

Bipolar disorder (BD) is a cyclical, recurrent, and chronic affective disorder, characterized by mood changes with recurring phases such as mania and depression and globally recognized as an important public health problem. Understanding the etiology and pathophysiology of this disorder in a multifactorial manner is a major challenge for effective treatments (Müller-Oerlinghausen et al. 2002).

BD is considered a complex disease, and multiple etiological models tried to explain the emergence and manifestation of the symptoms based on biochemical, genetic, and immunological paradigms (Machado-Vieira et al. 2004). Many studies have been dedicated to determining the genetic predispositions for bipolar disorder, and the possible candidate genes for this disorder have been presumed to be located in different chromosome regions. It was in the 1970s when different studies began to relate the etiology of BD with the human major histocompatibility complex (MHC) on chromosome 6 (Shapiro et al. 1976, 1977; Johnson 1978; Targum et al. 1979). However, the hypothesis that BD could be caused by infectious agents and the involvement of the immune system was first formulated in the 19th century in the *American Journal of Insanity* (Yolken and Torrey 1995).

The human MHC, also commonly named as human leukocyte antigens (HLA), is located on the short arm of

chromosome 6, with about 4 Mb, and is divided into three regions: class I, class II, and class III (Shiina et al. 2009). The class I and class II regions contain the classical HLA genes which encode glycoprotein molecules expressed at the cell surface where they present antigenic peptides to CD8-positive and CD4-positive T cells, respectively. The class III region is involved in the regulation of the humoral immune response and in the inflammatory reaction through genes encoding the tumor necrosis factor (TNF), heat shock proteins, or components of the complement cascade (Bayley et al. 2004; Vandiedonck et al. 2004; Vandiedonck and Knight 2009).

The MHC region is one of the most polymorphic regions of the whole human genome, and many of the polymorphisms found in classical HLA genes are associated with clinical organ transplantation rejection, susceptibility to infectious diseases, autoimmune diseases, and susceptibility to a broad spectrum of chronic non-infectious conditions (Lie and Thorsby 2005).

Previous studies started to conjecture the involvement of immunological disturbances in mood disorders due to the clinical overlapping between part of the "sickness behavior" stereotype in the infectious process and some of the most common affective symptoms (Licinio and Frost 2000; Fertuzinhos et al. 2004). Later, the MHC became an interesting and more specific biological target for mood disorders, which has led to a rapid expansion in the number of investigations associating the HLA alleles with bipolar disorder, but with inconsistent and conflicting findings (Wentzel et al. 1982).

MHC Class I and Class II Regions and Bipolar Disorder

Most association studies linking the MHC region with BD involve the classical HLA genes. For example, the HLA-B16 was associated with mood disorders in general, including the forms with mania or presenting depression as the only recurrent symptom (Ozcan et al. 1996). Both HLA-A29 and HLA-B21 antigens were detected more frequently in bipolar patients than in controls, with statistically significant differences (Ventura et al. 1990). Modrego and Ferrández (2000) observed an increased frequency of HLA-DR2 and HLA-DR3 in patients with bipolar disorder. In another study, HLA class I and class II allelic frequencies were assessed in 87 bipolar patients and compared with 206 normal controls in the Korean population with a weak positive association, maybe due to a small sample size and the clinical heterogeneity of BD (Jun et al. 2002). Another study in Caucasian Turkish patients also showed no association with HLA antigens (Ucok et al. 2005). Both publications suggest consecutive studies with a larger subject cohort and more advanced methods to clarify the genetic influence of HLA on bipolar disorder.

Most studies on the genetic relationship between specific HLA alleles and bipolar disorder were conducted based on the frequency of the HLA antigen investigated serologically. However, the serological method has a limitation in specificity, so that it is difficult to detect differences in the frequencies of specific alleles, even if there is no cross-reaction. Therefore, distinguishing similar subtypes of antigenicity by only serological HLA typing has difficulties in elucidating the exact relationship between a precise HLA subtype and a specific mood disorder (Jun et al. 2002; Vandiedonck and Knight 2009).

With the advent of advances in biotechnology, statistics, population genetics, and psychiatry, together with the ability to acquire and study larger samples of patients, the possible association between MHC region and bipolar disorder came to be present in the studies. Data from Genome-Wide Association Studies are beginning to provide strong support for genetic risk for bipolar disorder (Williams et al. 2011).

Additional approaches with microarrays allowed studies such as the one run by Nakatani et al. (2006), developing a genome-wide expression analysis and examining the expression levels of more than 12,000 genes in Brodmann's Area and 46 genes of the dorsolateral prefrontal cortex from postmortem brains in BD patients and controls using Affymetrix GeneChips. The results showed that HLA-DRA was decreased in bipolar patients in both microarray and quantitative RT-PCR analyses. The expression of HLA-DRA (inflammatory response) was suggested to be probably controlled by X-box binding protein 1 (XBP1), binding to the HLA-DRA promoter (Gomez et al. 2005). Interestingly, a pathway analysis detected an association between XBP1, a component of network 2 (nervous system development and function) and bipolar disorder (Kakiuchi et al. 2003).

In 2009, the International Schizophrenia Consortium released the analysis from a genome-wide association study of 3,322 individuals with schizophrenia, 2,784 with bipolar disorder, and 8,020 controls. They provided molecular genetic evidence for a substantial polygenic component to the risk of schizophrenia and bipolar disorder involving thousands of common alleles of very small effect at the MHC. The researchers detected a hotspot located on chromosome 6p22, the same site of several histone genes and the HLA alleles. Thus, they seem to be involved in the controlling of the transcription and DNA repair, in antimicrobial defense mechanisms and for part of HLA class II, suggesting that autoimmune mechanisms may be involved in these psychiatric illnesses. They found more than 450 variations with stronger associations in the MHC region, but the best inputted SNP (rs3130297) was located in the NOTCH4 gene, reaching genome-wide significance ($P=4.79\times10^{-8}$, T allele odds ratio=0.747, and minor

allele frequency=0.114) (The international schizophrenia consortium et al. 2009).

NOTCH4 lies at the centromeric end of the HLA class III region, approximately 335 kb telomeric to the DRB1 locus (Matsuzaka et al. 2001) and is a gene with previously reported associations with schizophrenia and bipolar disorder (Wei and Hemmings 2000; Swift-Scanlan et al. 2002). The statistical simulations confirmed that these findings are not caused by a handful of genetic variants with large effects or rare genetic variants. Individually, the effects of these common variants have no statistical difference but maybe a cumulative relevance.

The researchers mentioned that the involvement of many of these common variations suggests that schizophrenia and bipolar disorder can be triggered by different processes in different patients. Another important point of this study was that this variation contributes to the risk of bipolar disorder and schizophrenia, but not to several non-psychiatric conditions, compared as additional internal control groups.

Other genome-wide studies involving the overlap of schizophrenia and bipolar disorder with the MHC region were performed. Most are pointing to the class I and class II regions (Lichtenstein et al. 2009; Moskvina et al. 2009; Shi et al. 2009). Williams et al. (2011) performed a study upon variants that show genome-wide evidence for association in the largest publicly available schizophrenia and bipolar disorder data sets.

In general, the association to the MHC was detected genome wide with levels of significance for bipolar disorder and schizophrenia. Kaminsky et al. (2011) performed an epigenome-wide scan in postmortem brain simples from individuals affected with bipolar disorder and identified lower DNA methylation in bipolar disease patients compared with controls at an extended HCG9 region (P=0.026). The analysis of nearly 40,000 CpGs revealed complex relationships between DNA methylation and age, medication as well as DNA sequence variation in patients.

Another approach combined the use of expression with genotyping arrays in BD samples and identified 45 SNPs associated with differentially expressed genes, such as 15 cis SNPs associated with decreased expression of HLA-DBP1 gene in the prefrontal cortex of bipolar patients (Choi et al. 2011).

There are several hypotheses of how the classical HLA molecules may be associated with susceptibility of various diseases in general. HLA molecules can function as receptors for infectious agents, occurrence of molecular mimicry between HLA antigens and certain microorganisms, aberrant expression of HLA class II molecules triggering autoimmune mechanisms, can induce vigorous T cell response and aberrant expression of cytokines among other hypotheses (Lechler and Warrens 2000).

The functional interpretations are necessarily speculative, since the true functional variants are still unknown; however, the associations between bipolar disorder and the extended MHC locus potentially lend support to long-held hypotheses of the importance of either infective agents and/or autoimmunity in the etiopathogenesis this disease (Williams et al. 2011).

MHC Class III, Cytokines, and Bipolar Disorder

Many association studies between the MHC region and bipolar disorder have been performed, but there are fewer studies about the role of changes in the immune system in pathogenesis of disease. Some studies have shown that cytokines can interfere in the metabolism of neurotransmitter systems, neuroendocrine and neuronal activity, and regulation of the growth and proliferation of glial cells (Silverman et al. 2005; Steiner et al. 2008). Cytokines interact with the neuroendocrine system, e.g., the hypothalamic–pituitary–adrenocortical system, the autonomic system, and the neurotransmitter system (dopamine, serotonin, and glutamate) (Kim et al. 2007).

Several studies suggest the involvement of TNF-α cytokine in the pathogenesis of BD (Clerici et al. 2009; Guloksuz et al. 2010; Drexhage et al. 2010). TNF-α is a 157-amino acid cytokine encoded by a gene located in the MHC class III region (Bayley et al. 2004) and produced in response to injury and inflammatory or infectious stimulus by macrophages, lymphocytes, neutrophils, and structural cells, including fibroblast, smooth muscle cells (Balakumar and Singh 2006), astrocytes, and microglia (Kronfol and Remick 2000).

Some studies reported increased proinflammatory cytokines and hyperactivity of T helper cell 1 in BD, with significantly higher TNF-α levels in bipolar patients during manic and depressive episodes, while others reported increased production of interleukin-6 and TNF-α during mania when compared with nonbipolar controls (O'Brien et al. 2006; Kim et al. 2007; Czerski et al. 2008). TNF-α has been found to significantly upregulate the activity of the serotonin transporter, an effect that can determine reduced function of serotonergic transmission by the reduction of synaptic availability of serotonin (Irwin and Miller 2007).

Additional studies have described impairments in neuroplasticity and neuronal survival among BD patients, suggesting that several cytokines, particularly TNF-α, might play a critical role in neuronal survival, cell resilience, neuroplasticity, and autoimmunity of neuronal cell as the main neuropathological and immunepathological correlates of bipolar disorder (Brietzke and Kapczinski 2008; Schloesser et al. 2008).

Bipolar patients might also display higher levels of circulating autoantibodies and proinflammatory profile of cytokines, regardless of stage of disease or the use of medication when compared with healthy controls. This may

suggest the participation of immune mechanisms and/or inflammation in the pathophysiology of BD. Following this line of evidence, it is interesting to highlight recent studies investigating anti-inflammatory drugs as a new therapeutic strategy for bipolar patients. Several conventional mood stabilizers have been also considered to have anti-inflammatory properties, and the cyclooxygenase-2-selective anti-inflammatories such as Celecoxib are one of the attempts to find immunemodulators with antidepressant effects. (Nery et al. 2008; Goldstein et al. 2009).

Conclusions and Perspectives

Despite the ramping number of studies and research groups focusing on BD, there is a major debate over the actual etiopathogenesis. The extensive polymorphism of the MHC region, the few studies with different ethnic groups, and the often discordant results point to the need for more research on the participation of MHC and the immune system in the pathogenesis of this common mood disorder.

The recent biotechnological advances and the possibility of pooling larger studies in a major consortium are significant factors for finding consistent results involving the MHC region, immune system, and BD in the near future. This knowledge might also be useful for estimating genetic risk for BD and hopefully also allow preventive and therapeutic interventions in a proper manner.

Acknowledgments This study received financial support from the following Brazilian funding agencies and academic bureaus: LIKA-UFPE, PROPESQ-UFPE, and CAPES.

References

Balakumar P, Singh M (2006) Anti-tumor necrosis factor-α therapy in heart failure: future directions. Pharmacol Toxicol 99:391–397

Bayley JP, Ottenhoff TH, Verweij CL (2004) Is there a future for TNF promoter polymorphisms? Genes Immun 5:315–329

Brietzke E, Kapczinski F (2008) TNF-α as a molecular target in bipolar disorder. Prog Neuropsychopharmacol Biol Psychiatry 32:1355–1361

Choi KH, Higgs BW, Wendland JR, Song J, McMahon FJ, Webster MJ (2011) Gene Expression and Genetic Variation Data Implicate PCLO in Bipolar Disorder. Biol Psychiatry 69:353–359

Clerici M, Arosio B, Mundo E et al (2009) Cytokine polymorphisms in the pathophysiology of mood disorders. CNS Spectr 14:419–425

Czerski PM, Rybakowski F, Kapelski P, Rybakowski JK, Dmitrzak-Weglarz M, Leszczyńska-Rodziewicz A (2008) Association of tumor necrosis factor -308G/A promoter polymorphism with schizophrenia and bipolar affective disorder in a Polish population. Neuropsychobiology 57:88–94

Drexhage RC, Knijff EM, Padmos RC et al (2010) The mononuclear phagocyte system and its cytokine inflammatory networks in schizophrenia and bipolar dirsorder. Expert Ver Neurother 10:59–76

Fertuzinhos SMM, Oliveira JRM, Nishimura AL et al (2004) Analysis of IL-1α, IL-1 β, and IL-RA Polymorphisms in Dysthymia. J Mol Neurosci 22:251–256

Goldstein BI, Kemp DE, Soczynska JK, McIntyre RS (2009) Inflammation and the phenomenology, pathophysiology, comorbidity, and treatment of bipolar disorder: a systematic review of the literature. J Clin Psychiatry 70:1078–1090

Gomez JA, Majumder P, Nagarajan UM, Boss JM (2005) X box-like sequences in the MHC class II region maintain regulatory function. J Immunol 175:1030–1040

Guloksuz S, Cetin EA, Cetin T, Deniz G, Oral ET, Nutt DJ (2010) Cytokine levels in euthymic bipolar patients. J Affect Disord 126:458–462

Irwin M, Miller A (2007) Depressive disorders and immunity: 20 years of progress and discovery. Brain Behav Immun 21:374–383

Johnson GF (1978) HLA antigens and manic–depressive disorders. Biol Psychiatry 13:409–412

Jun TY, Pae CU, Chae JH, Pyo CW, Han H (2002) Human leukocyte antigen alleles in patients with bipolar disorder in the Korean population. Psychiatry Clin Neurosci 56:453–457

Kakiuchi C, Iwamoto K, Ishiwata M et al (2003) Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. Nat Genet 35:171–175

Kaminsky Z, Tochigi M, Jia P. et al (2011) A multi-tissue analysis identifies HLA complex group 9 gene methylation differences in bipolar disorder. Mol Psychiatry

Kim Y, Jung H, Mynt A, Kim H, Park S (2007) Imbalance between proinflammatory cytokines in bipolar disorder. J Affect Disord 104:91–95

Kronfol Z, Remick D (2000) Cytokines and the brain: implications for clinical psychiatry. Am J Psychiatry 157:683–94

Lechler R, Warrens A (2000) HLA in Health and Disease, 2nd edn. Academic press, London

Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 373:234–239

Licinio J, Frost P (2000) The neuroimmune-endocrine axis: pathophysiological implications for the central nervous system cytokines and hypothalamus-pituitary-adrenal hormone dynamics. Braz J Med Res 33:1141–1148

Lie BA, Thorsby E (2005) Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. Curr Opin Immunol 17:529

Machado-Vieira R, Kapczinski F, Soares JC (2004) Perspectives for the Development of New Animal Models of Bipolar Disorder. Prog Neuropsychopharmacol Biol Psychiatry 28:209–224

Matsuzaka Y, Makino S, Nakajima K et al (2001) New polymorphic microsatellite markers in the human MHC class III region. Tissue Antigens 57:397–404

Modrego PJ, Ferrández J (2000) Familial multiple sclerosis with repetitive relapses of manic psychosis in two patients (mother and daughter). Behav Neurol 12:175–179

Moskvina V, Craddock N, Holmans P et al (2009) Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. Mol Psychiatry 14:252–260

Müller-Oerlinghausen B, Berghöfer A, Bauer M (2002) Bipolar Disorder. Lancet 359:241–247

Nakatani N, Hattori E, Ohnishi T et al (2006) Genome-wide expression analysis detects eight gene with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation. Hum Mol Genet 15:1949–1962

Nery FG, Monkul ES, Hatch JP et al (2008) Celecoxib as na adjunct in the treatment of depressive or mixed episodes of bipolar

disorder: a doubleblind, randomized, placebo-controlled study. Hum Psychopharmacol 23:87–94

O'Brien SM, Scully P, Scott LV, Dinan TG (2006) Cytokine profiles in bipolar affective disorder: focus on acutely ill patients. J Affect Disord 90:263–267

Ozcan ME, Taskin R, Banoglu R, Babacan M, Tuncer E (1996) HLA antigens in schizophrenia and mood disorders. Biol Psychiatry 39:891–895

Schloesser R, Huang J, Klein P, Manji H (2008) Cellular plasticity cascades in the pathophysiology and treatment of bipolar disorder. Neuropsychopharmacology 33:100–133

Shapiro RW, Bock E, Rafaelsen OJ, Ryder LP, Svejgaard A (1976) Histocompatibility antigens and manic-depressive disorders. Arch Gen Psychiatry 33:823–825

Shapiro RW, Ryder LP, Svejgaard A, Rafaelsen OJ (1977) HLA antigens and manic-depressive disorders: further evidence of na association. Psychol Med 7:387–396

Shi J, Levinson DF, Duan J et al (2009) Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 460:753–757

Shiina T, Hosomichi K, Inoko H, Kulski JK (2009) The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 54:15–39

Silverman MN, Pearce BD, Biron CA, Miller AH (2005) Immune modulation of the hypothalamicpituitary- adrenal (HPA) axis during viral infection. Viral Immunol 18:41–78

Steiner J, Bielau H, Brisch R et al (2008) Immunilogical aspects in the neurobiology of suicide: elevated microglial density in schizophrenia and depression is associated with suicide. J Psychiatry Res 42:151–157

Swift-Scanlan T, Lan TH, Fallin MD et al (2002) Genetic analysis of the (CTG)n NOTCH4 polymorphism in 65 multiplex bipolar pedigrees. Psychiatr Genet 12:43–47

Targum SD, Gershon ES, Van Eerdewegh M, Rogentine N (1979) Human leukocyte antigen system not closely linked to or associated with bipolar manic–depressive illness. Biol Psychiatry 14:615–636

The International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P (2009) Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder. Nature 460:748–752

Ucok A, Akar U, Polat A, Yazici O (2005) Human leukocyte antigen alleles in patients with bipolar disorder in Turkey. Eur Psychiatry 20:83

Vandiedonck C, Knight CJ (2009) The human major histocompatibility complex as a paradigm in genomics research. Brief Funct Genomics Proteomics 8:379–394

Vandiedonck C, Beaurain G, Giraud M et al (2004) Pleiotropic effects of the 8.1 HLA haplotype in patients with autoimmune myasthenia gravis and thymus hyperplasia. Proc Natl Acad Sci 101:15464–15469

Ventura T, Lobo A, Marco JC (1990) HLA antigens in bipolar affective patients. Actas Luso Esp Neurol Psiquiatric Cienc Afines 18:339–343

Wei J, Hemmings GP (2000) The NOTCH4 locus is associated with susceptibility to schizophrenia. Nature Genet 25:376–377

Wentzel J, Roberts DF, Whalley LJ (1982) HLA in manic–depressive psychosis. Psychol Med 12:275–278

Williams HJ, Craddock N, Hamshere ML et al (2011) Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boumdaries. Hum Mol Genet 20:387–391

Yolken RH, Torrey EF (1995) Viruses, Schizophrenia and Bipolar Disorder. Clin Microbiol Rev 8:131–145

CAPÍTULO IV

CONCLUSÃO

Após utilizar uma metodologia baseada em diferentes ferramentas de bioinformática foi possível verificar que o estudo de sequências de ESTs para se identificar novas variações em grandes conjuntos de dados é uma abordagem alternativa promissora e acessível.

A partir da validação virtual verificamos que a técnica foi capaz de identificar 101 variações previamente anotadas e confirmadas em amostras de DNA, além de 120 possíveis novos SNPs encontrados nas regiões codificantes dos genes HLA classe I, sendo a maioria, mutações não sinônimas que se repetem em ESTs provenientes de diferentes bibliotecas. Cada possível novo SNP que foi encontrado pode representar um novo alelo.

Portanto, a análise de ESTs parece ser uma boa alternativa para o estudo de SNPs em vários genes, especialmente nos genes HLA classe I, genes que possuem acentuados polimorfismos os quais estão associados à rejeição de transplantes clínicos, suscetibilidade às doenças infecciosas, doenças autoimunes e predisposição a um amplo espectro de doenças crônicas não infecciosas.

ANEXOS

COMPROVANTE DO ACEITE DO ARTIGO: "Using ESTs database to predict and validate single polymoorphisms at the HLA system".

Decision Letter (IJIG-Sep-11-0167)

From: paul.travers@ed.ac.uk

To: thalita.figueiredo87@gmail.com

CC:

Subject: International Journal of Immunogenetics - Decision on Manuscript ID IJIG-Sep-11-0167

Body: @@date to be populated upon sending@@

Dear Ms Figueiredo

It is a pleasure to accept your manuscript entitled "Using ESTs database to predict and validate single polymorphisms at the HLA system" in its current form for publication in the International Journal of Immunogenetics.

COMPROVANTE DO ACEITE DO ARTIGO: "*Reconsidering the Association Between the Major Histocompatibility Complex and Bipolar Disorder*".

Date: Sep 20, 2011
To: "João Oliveira" joao.ricardo@ufpe.br
From: "Journal of Molecular Neuroscience" Lorievic.Hayag@springer.com
Subject: Decision on your manuscript #JOMN-882R1

Dear Dr João Oliveira:

We are pleased to inform you that your manuscript, "Reconsidering the association between the MHC region and Bipolar Disorder" has been accepted for publication in Journal of Molecular Neuroscience.

For queries regarding your accepted paper, please click the following link http://www.springer.com/12031; then click on "Contacts", and then "Production Editor", complete the query form and click "Submit".

Please remember to always include your manuscript number, #JOMN-882R1 , whenever inquiring about your manuscript. Thank you.

Best regards,

The Editorial Office
Journal of Molecular Neuroscience

METODOLOGIA

A figura a seguir sumariza a metodologia utilizada e está detalhada em seguida.

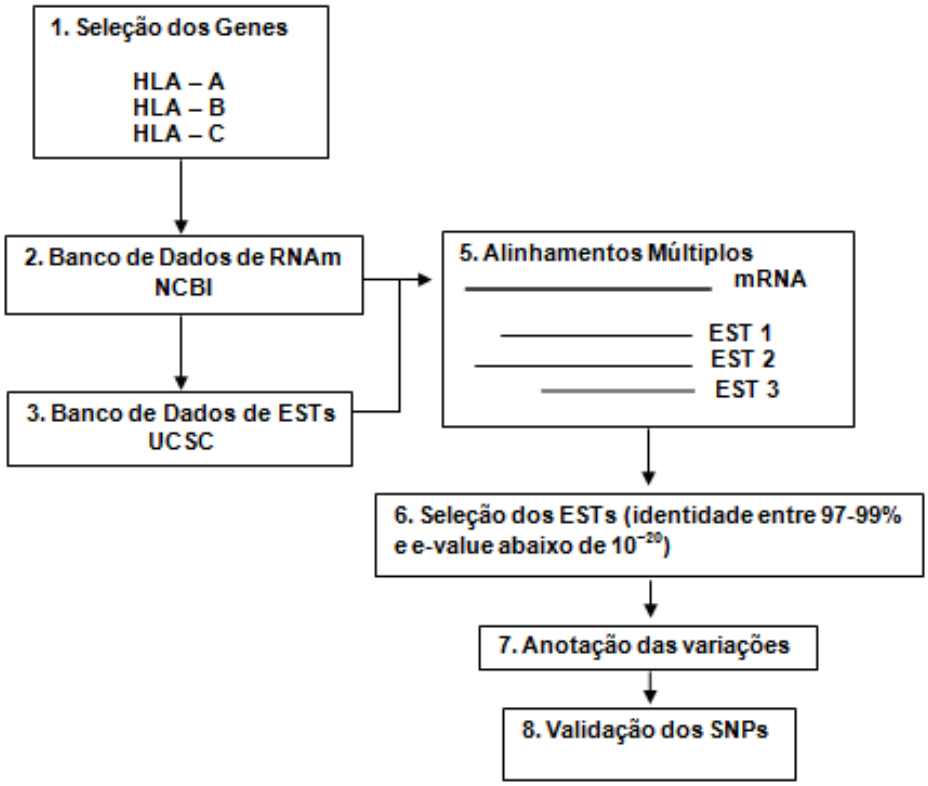


Figura 4. Esquema sumarizado da Metodologia.

O banco de dados de ESTs

As sequências de ESTs spliced foram provenientes do banco de dados público Golden Path – UCSC (http://www.genome.ucsc.edu).

Sequências de RNAm

Foram feitos downloads das sequências referências de RNAm dos genes HLA classe I, a partir do site NCBI (http://www.nlm.nih.gov).

BLASTn (Smith-Waterman)

O BLASTn foi realizado entre as sequências de cada éxon, separadamente, e todos os ESTs pelo software CLCbio Workbench Combined® versão 5.7.1 (http://www.clcbio.com), com auxílio do CUBO® versão 1.0.6, hardware com função de acelerar o processamento do BLASTn.

Alinhamento entre as sequências de exons e ESTs

Cada éxon dos genes abordados apresentou *scores* de similaridade com vários ESTs, como critério de seleção foram usados dois filtros de exclusão: Sequências com *e-value* > 10^{-20} e identidade % > 99.

Identificação de Inserções, Deleções, Substituições, Mutações Sinônimas e Não sinônimas

As identificações de deleções (quando uma base nitrogenada é deletada da sequência), inserções (quando há inserção de uma base nitrogenada), transições (quando ocorre uma troca de uma base purina (A ou T) por outra purina ou troca de uma base pirimidina (C ou T) por outra pirimidina) e transversões (troca de uma base purina por uma pirimidina e vice-versa) (figura 5) foram feitas com as sequências de nucleotídeos selecionadas pelos dois filtros. Posteriormente, as sequências foram traduzidas em aminoácidos para identificação de mutações sinônimas (não há mudança do aminoácido correspondente ao códon) e não sinônimas (há mudança do aminoácido a partir da mudança no códon).

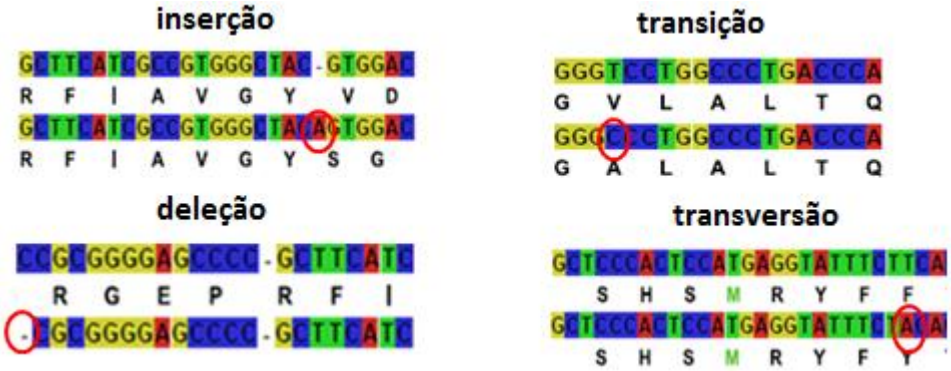


Figura 5. Tipo de mutações anotadas durante a análise de SNPs

Validação das variações

A validação foi feita utilizando a ferramenta dbMHC Sequence Alignment Viewer (NCBI) que tem como opção marcar todas as posições que possuem SNPs nas sequências dos éxons dos genes HLA. Assim, foram verificadas as variações já depositadas nos principais bancos de dados de SNPs do sistema HLA. Na figura 6 está a interface da ferramenta. Os nucleotídeos marcados em verde representam que há polimorfismos naquela posição. Ao clicar em cima do nucleotídeo marcado, é possível verificar detalhes de cada SNP (figura 7).

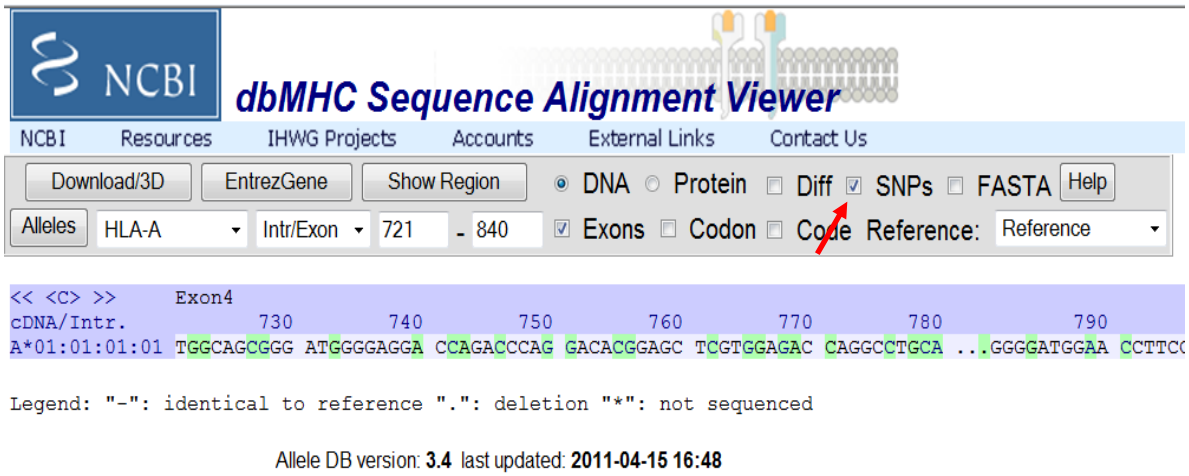


Figura 6. Ferramenta utilizada para validação dos SNPs encontrados.

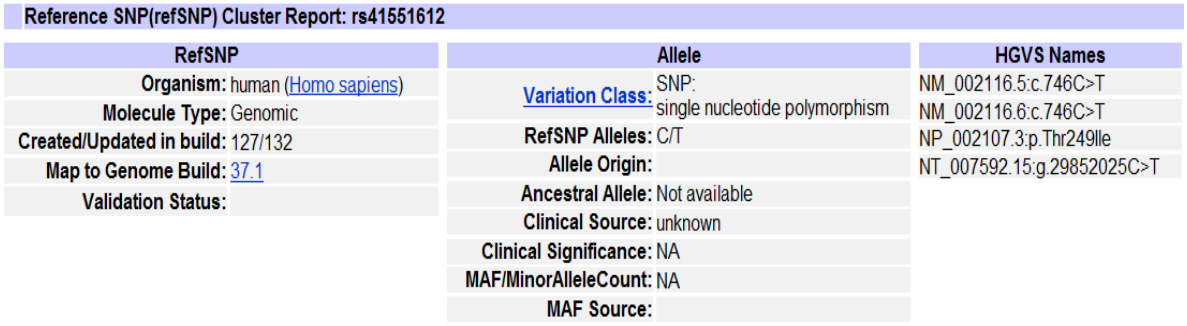


Figura 7. Detalhes dos SNPs a partir do uso da ferramenta do dbMHC.

RESULTADOS COMPLEMENTARES

A técnica de *screening in silico* de SNPs a partir de ESTs vem se tornando uma metodologia rápida e de baixo custo (Useche et al. 2001). Esse tipo de abordagem envolve uma grande quantidade de dados que pode indicar potenciais novos SNPs, portanto, ponteciais novos alelos e regiões *hotspots* que poderão ser alvos de pesquisas de genotipagem experimental na população.

No gráfico abaixo (figura 8) está representando o número de SNPs encontrados por éxon de cada gene utilizando a metodologia de *screening in silico* a partir de ESTs. Os polimorfismos dos genes HLA clássicos estão predominantemente localizados nos éxons 2 e 3 dos genes de classe I, os quais codificam para as estruturas moleculares alfa 1 e alfa 2 do receptor membranar para peptídeos antigênicos (Hughes & Nei, 1988).

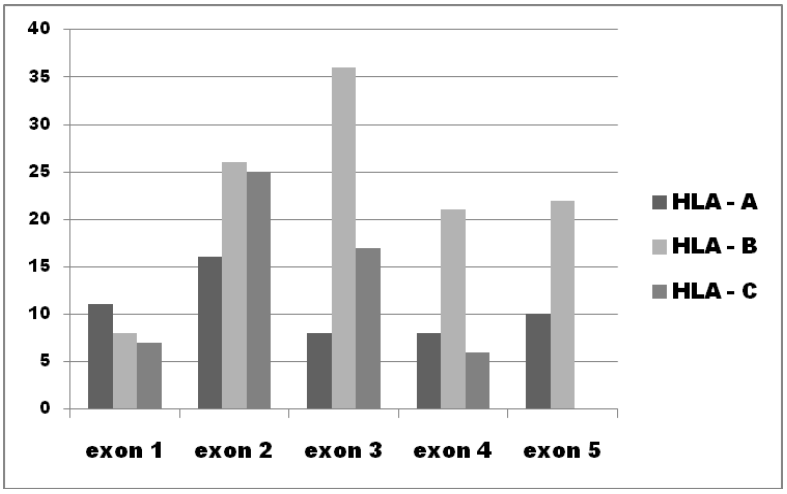


Figura 8. Número de SNPs encontrados por exon de cada gene HLA classe I.

As sequências de nucleotídeos das moléculas HLA classe I localizadas, principalmente, nos exons 2 e 3, determinam as propriedades das bolsas dos receptores dentro das quais as cadeias laterais dos peptídeos serão ligadas e, consequentemente, a especificidade ao peptídeo antigênico (Hughes & Nei, 1988). Portanto, a diversidade nestas regiões tem papel fundamental na resposta imune e muitas vezes estão relacionadas na patogenia de várias doenças de distintas etiologias.

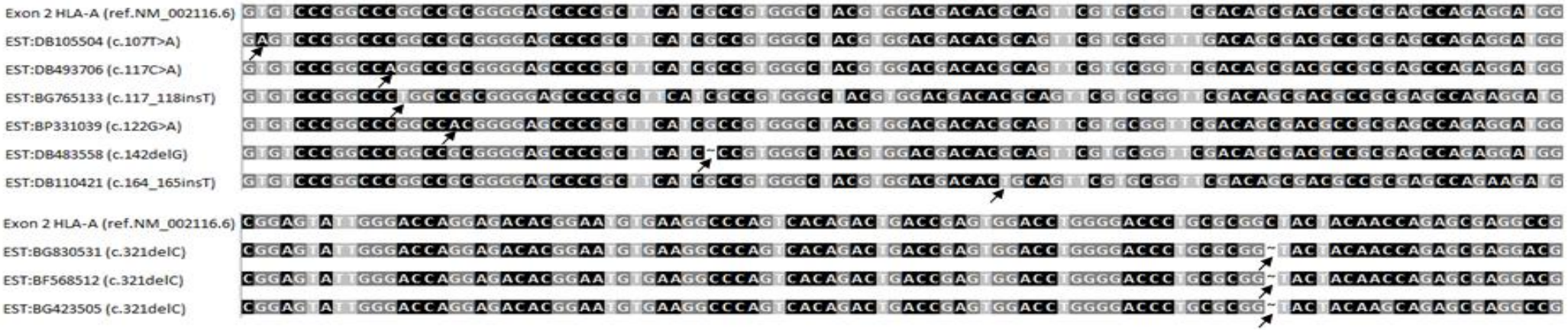
O processo de validação virtual com busca dos SNPs no banco de dados do NCBI confirmou algumas das variações encontradas e nos revelou possíveis novas mutações causadas por mudanças de base única (SNPs) (Figura 9). A maioria das mutações foram não sinônimas, isto colabora com a hipótese de co-evolução entre patógenos e sistema imunológica, e assim, a importância da diversidade destes genes. Diante do exposto, a metodologia proposta por este trabalho chegou a um dado experimentalmente já comprovado previamente em amostras de DNA genômico e revelou possíveis novos polimorfismos.

Figura 9. Mutações descritas (a) e não descritas (b).

Como as regiões dos exons 2 e 3 estão diretamente relacionadas na resposta imune, tornam-se regiões de maior interesse. Sendo assim, novos alinhamentos foram realizados a partir dos ESTs selecionados que apresentaram SNPs não depositados no banco de dados do NCBI, também para obter maiores informações sobre os ESTs alinhados.

Nos quadros apresentados logo em seguida estão detalhes do alinhamento, bem como, dos ESTs, e, apontados por uma seta estão os possíveis novos SNPs. Estes dados podem auxiliar pesquisas de genotipagem experimental para confirmação destes SNPs.

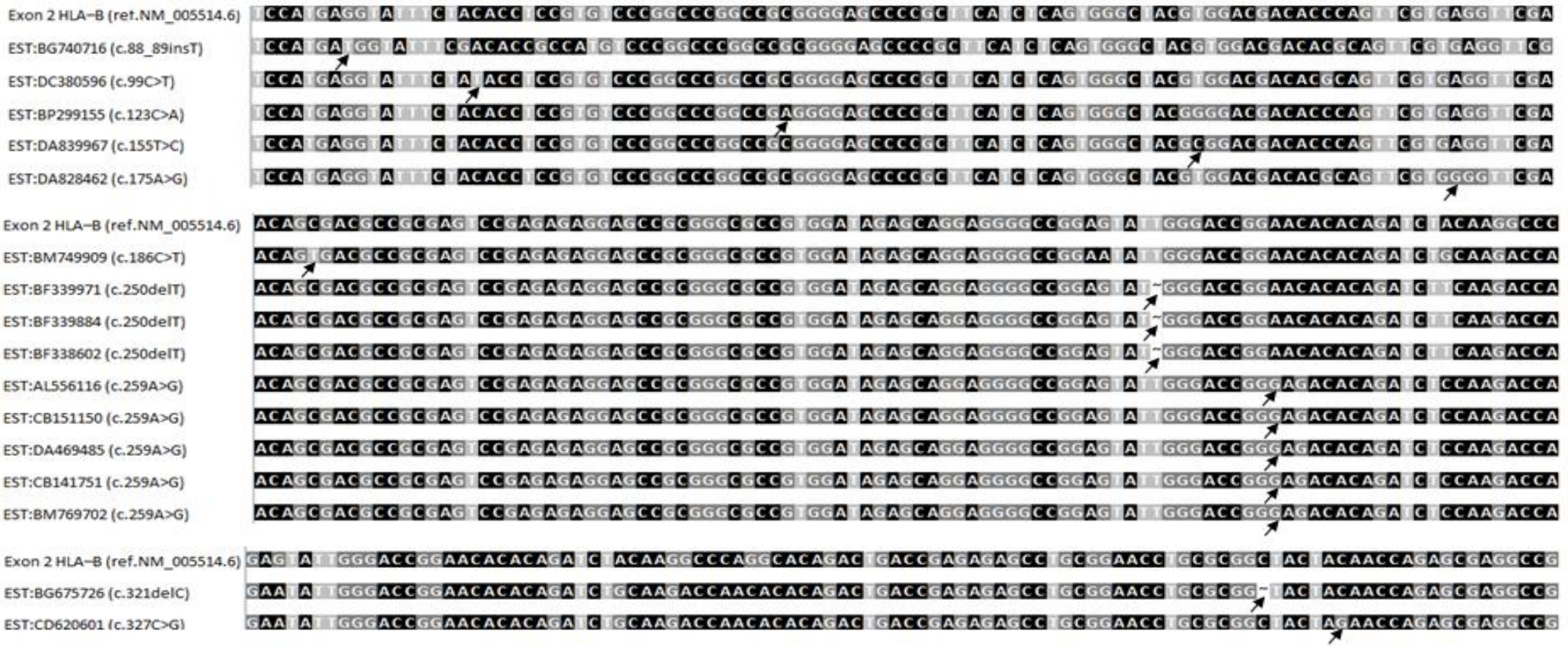
HLA – A (Exon 2)						
Mutation nucleotide	**Mutation amino acids**	**EST**	**Library name**	**Tissue type**	**Study**	**Blastn**
c.107T > A	Val36Glu	DB105504	LIBEST_018519 THYMU2	Thymus	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 8e-120; **Identity:** 95%
c.117C > A	Synonymous	DB493706	LIBEST_019377	Hypothalamus	RIKEN full-length enriched human cDNA library	**E-value:** 1e-127; **Identity:** 97%
c.117_118insT	Gly40del fs	BG765133	LIBEST_007270 NIH_MGC_49	Skin - melanotic melanoma	NIH-MGC EST Sequencing Project	**E-value:** 1e-132; **Identity:** 98%
c.122G > A	Arg41His	BP331039	LIBEST_016425	Rectum	Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions	**E-value:** 1e-132; **Identity:** 98%
c.142delG	Ala47del fs	DB483558	LIBEST_019375	Hippocampus	RIKEN full-length enriched human cDNA library	**E-value:** 2e-125; **Identity:** 97%
c.164_165insT	Gly56del fs	DB110421	LIBEST_018519 THYMU2	Thymus	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 3e-128; **Identity:** 97%
c.321delC	Tyr108del fs	BG830531	LIBEST_007245 NIH_MGC_42	Pancreas- epithelioid carcinoma	NIH-MGC EST Sequencing Project	**E-value:** 4e-132; **Identity:** 98%
		BF568512	LIBEST_007245 NIH_MGC_42	Pancreas- epithelioid carcinoma	NIH-MGC EST Sequencing Project	**E-value:** 4e-132; **Identity:** 98%
		BG423505	LIBEST_004068 NIH_MGC_14	Kidney - renal cell adenocarcinoma	NIH-MGC EST Sequencing Project	**E-value:** 2e-135; **Identity:** 99%



HLA – A (Exon 3)						
Mutation nucleotide	**Mutation amino acids**	**EST**	**Library name**	**Tissue type**	**Study**	**Blastn**
c.535delC	Glu178del fs	AL559934	LIBEST_013050	B Cells	Full-length cDNA libraries and normalization	**E-value:** 1e-140; **Identity:** 99%
		AL541209	LIBEST_013037	Placenta	Full-length cDNA libraries and normalization	**E-value:** 3e-139; **Identity:** 99%
		AL541172	LIBEST_013037	Placenta	Full-length cDNA libraries and normalization	**E-value:** 1e-140; **Identity:** 99%
c.546_547insC	Tyr183del fs	BU541673	LIBEST_008834 NIH_MGC_40	Prostate - carcinoma	NIH-MGC EST Sequencing Project	**E-value:** 5e-129; **Identity:** 96%

HLA – B (Exon 2)						
Mutation nucleotide	**Mutation amino acids**	**EST**	**Library name**	**Tissue type**	**Study**	**Blastn**
c.88_89insT	Met29del fs	BG740716	LIBEST_008848 NCI_CGAP_Skn3	Skin	NIH-MGC EST Sequencing Project	**E-value:** 1e-122; **Identity:** 96%
c.99C > T	Synonymous	DC380596	LIBEST_018468 PLACE6	Placenta	NEDO human cDNA sequencing project	**E-value:** 2e-127; **Identity:** 97%
c.123C > A	Synonymous	BP299155	LIBEST_016419	Macrophage	Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions	**E-value:** 6e-131; **Identity:** 98%
c.155T > C	Val51Ala	DA839967	LIBEST_018468 PLACE6	Placenta	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-136; **Identity:** 99%
c.175A>G	Arg59Gly	DA828462	LIBEST_006999 PLACE1	Placenta	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-126; **Identity:** 97%
c.186C > T	Synonymous	BM749909	LIBEST_010310	Stomach - lymphoblast	Transcriptome analysis of human gastric	**E-value:** 6e-126;

			S10SNU1			**Identity:** 97%
c.250delT	Trp84del fs	BF339971	LIBEST_007138 NCI_CGAP_Brn64	Brain - glioblastoma	NIH-MGC EST Sequencing Project	**E-value:** 2e-120; **Identity:** 95%
		BF339884	LIBEST_007138 NCI_CGAP_Brn64	Brain - glioblastoma	NIH-MGC EST Sequencing Project	**E-value:** 2e-120; **Identity:** 95%
		BF338602	LIBEST_007138 NCI_CGAP_Brn64	Brain - glioblastoma	NIH-MGC EST Sequencing Project	**E-value: 2e-120; Identity: 95%**
c.259A > G	Asp87Glu	AL556116	LIBEST_013024	Hela cells cot 25-normalized	Full-length cDNA libraries and normalization	**E-value: 6e-121; Identity: 95%**
		CB151150	LIBEST_012528 C1SNU17	Uterine - epithelial	Identification of intrahepatic cholangiocarcinoma related genes by comparison with normal liver tissues using expressed sequence tags	**E-value: 6e-126; Identity: 97%**
		DA469485	LIBEST_018366 D9OST2	CD34+ cells	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value: 6e-126; Identity: 97%**
		CB141751	LIBEST_012555 L7N800102	Liver	Identification of intrahepatic cholangiocarcinoma related genes by comparison with normal liver tissues using expressed sequence tags	**E-value: 6e-126; Identity: 97%**
		BM769702	LIBEST_010316 S14K402	Stomach	Transcriptome analysis of human gastric cancer	**E-value: 6e-126; Identity: 97%**
		BE242740	LIBEST_005410	Leukopheresis - myeloid cell	Leukemia cDNA Sequencing Project	**E-value: 6e-126; Identity: 97%**
		DC424119	LIBEST_006982 MAMMA1	Mammary gland	HRI human cDNA project	**E-value: 3e-124; Identity: 96%**
		BM856651	LIBEST_010316 S14K402	Stomach	Transcriptome analysis of human gastric cancer	**E-value: 2e-127; Identity: 97%**
		BM147094	LIBEST_005410	Leukopheresis - myeloid cell	Pediatric Leukemia cDNA Sequencing Project	**E-value: 6e-126; Identity: 96%**
c.321delC	Tyr108del fs	BG675726	LIBEST_008849 NCI_CGAP_Skn4	Skin - squamous cell carcinoma	NIH-MGC EST Sequencing Project	**E-value: 2e-125; Identity: 97%**
c.327C > G	Tyr109Stop	CD620601	LIBEST_013869	LIBEST_013869	Circular rapid amplification of cDNA ends for high-throughput extension cloning of partial genes	**E-value: 6e-126; Identity: 97%**



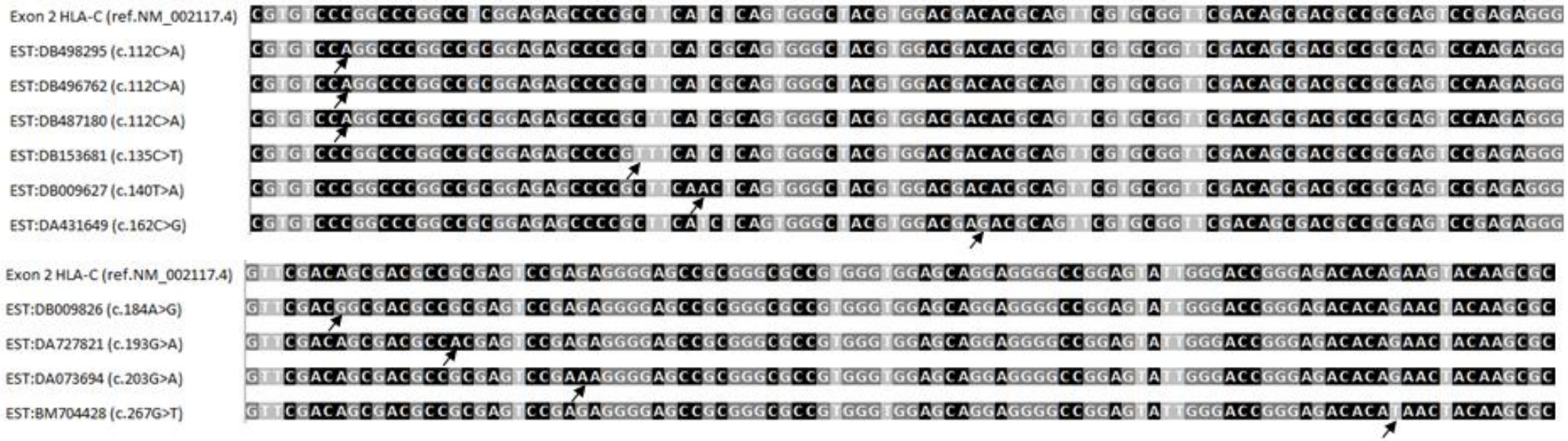
HLA – B (Exon 3)						
Mutation nucleotide	**Mutation amino acids**	**EST**	**Library name**	**Tissue type**	**Study**	**Blastn**
c.377A > T	Asp126Val	DC380321	LIBEST_018468 PLACE6	Placenta	NEDO human cDNA sequencing project	**E-value:**1e-83; **Identity:** 95%
c.382G > A	Gly126Arg	DA560696	LIBEST_018410 HEART2	Heart	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 2e-121; **Identity:** 95%
c.411T > A	His137Gln	BX458762	LIBEST_013037	Placenta	Full-length cDNA libraries and normalization	**E-value:** 6e-131; **Identity:** 97%

c.432delC	Lys143del fs	BG327758	LIBEST_004068 NIH_MGC_14	Kidney - renal cell adenocarcinoma	NIH-MGC EST Sequencing Project	**E-value:** 1e-127; **Identity:** 97%
c.432_433insC	Lys143del fs	BG740716	LIBEST_008848 NCI_CGAP_Skn3	Skin	NIH-MGC EST Sequencing Project	**E-value:** 3e-123; **Identity:** 96%
c.440A > G	Tyr145Cys	EL594759	LIBEST_004068 NIH_MGC_14	Skin – epidermis – keratinocyte	Large-scale identification of human genes implicated in epidermal barrier function	**E-value:** 4e-137; **Identity:** 98%
c.459C > T	Synonymous	DA831601	LIBEST_006999 PLACE1	Placenta	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-121; **Identity:** 95%
c.482_483insA	Asp161del fs	BF527705	LIBEST_007139 NCI_CGAP_Brn67	Brain - anaplastic oligodendroglioma	NIH-MGC EST Sequencing Project	**E-value:** 3e-139; **Identity:** 99%
c.491delC	Ala161del fs	BG678756	LIBEST_008849 NCI_CGAP_Skn4	Skin - squamous cell carcinoma	NIH-MGC EST Sequencing Project	**E-value:** 3e-123; **Identity:** 96%
c.523delC	Arg175del fs	BG432538	LIBEST_006833 NIH_MGC_75	Kidney	NIH-MGC EST Sequencing Project	**E-value:** 2e-130; **Identity:** 97%
		CD620611	LIBEST_013869 FLP		Circular rapid amplification of cDNA ends for high-throughput extension cloning of partial genes	**E-value:** 7e-130; **Identity:** 97%
		BE546563	LIBEST_004011 NIH_MGC_12	Cervical - carcinoma cell line	NIH-MGC EST Sequencing Project	**E-value:** 1e-122; **Identity:** 95%
		BG697915	LIBEST_008848 NCI_CGAP_Skn3	Skin	NIH-MGC EST Sequencing Project	**E-value:** 1e-122; **Identity:** 96%
		BF342367	LIBEST_007138 NCI_CGAP_Brn64	Brain – glioblastoma	NIH-MGC EST Sequencing Project	**E-value:** 1e-122; **Identity:** 96%
c.524_525insC	Arg175del fs	BU172833	LIBEST_005606 NIH_MGC_67	Eye - retinoblastoma	NIH-MGC EST Sequencing Project	**E-value:** 1e-127; **Identity:** 96%
*c.540G > T	Arg180Leu	CB151614	LIBEST_012528 C1SNU17	Uterine – ephitelial cell	Identification of intrahepatic cholangiocarcinoma related genes by comparison with normal liver tissues using expressed sequence tags	**E-value:** 3e-134; **Identity:** 98%
		DA570835	LIBEST_006980 HEMBA1	Whole embryo, mainly head	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 1e-132; **Identity:** 97%
		BM819582	LIBEST_010322 S19N665307	Stomach	Transcriptome analysis of human gastric cancer	**E-value:** 1e-132;

						Identity: 97%
		DR423347	LIBEST_017582	Eye - pterygium	NEIBank analysis of Human pterygium	**E-value:**9e-129; **Identity:** 98%
		BG106129	LIBEST_007318 NIH_MGC_85	Lymphoma, cell line	NIH-MGC EST Sequencing Project	**E-value:** 9e-124; **Identity:** 96%

* CB150500 CB127943 CB126854 AL542797 CD108637 CD107454 CD106942 CA395240 BM821619 BI915180 BE772772 CA394168 BG767029 DW407943 DW407913 DW407653 DC424119 DW407942 BM820957 BG757531 BY994189 BY995604 BM738128 BF129139 BX458762 DA933708 CV023913 BM807044 BM478193 BM477065 BM476615 BM464730 BM464722 BM455947 BM455022 BG025650 BM476664 BM454348 BI222491

HLA – C (Exon 2)						
Mutation nucleotide	**Mutation amino acids**	**EST**	**Library name**	**Tissue type**	**Study**	**Blastn**
c.112 C > A	Synonymous	DB498295	LIBEST_019377	Hypothalamus	RIKEN full-length enriched human cDNA library	**E-value:** 6e-131; **Identity:** 98%
		DB496762	LIBEST_019377	Hypothalamus	RIKEN full-length enriched human cDNA library	**E-value:** 6e-131; **Identity:** 98%
		DB487180	LIBEST_019377	Hypothalamus	RIKEN full-length enriched human cDNA library	**E-value:** 6e-131; **Identity:** 98%
c.135C > T	Synonymous	DB153681	LIBEST_018520 THYMU3	Thymus	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-136; **Identity:** 99%
c.140T > A	Ile47Asn	DB009627	LIBEST_018489 TCOLN2	Colon - tumor tissue	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-136; **Identity:** 99%
c.162C > G	Asp53Glu	DA431649	LIBEST_018350 COLON2	Colon	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 2e-137; **Identity:** 99%
c.184A > G	Ser61Gly	DB009826	LIBEST_018489 TCOLN2	Colon - tumor tissue	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-136; **Identity:** 99%
c.193G > A	Ala64Thr	DA727821	LIBEST_006990 NT2RM1	Teratocarcinoma	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 6e-136; **Identity:** 99%
c.203G > A	Arg68Lys	DA073694	LIBEST_018310 BRACE2	Cerebellum	Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes	**E-value:** 7e-135; **Identity:** 98%
c.267G > T	Gln88His	BM704428	LIBEST_010285 UI-E-CI1	Eye - RPE and Choroid	Normalization and subtraction: two approaches to facilitate discovery	**E-value:** 6e-136; **Identity:** 99%



HLA – C (Exon 3)						
Mutation nucleotide	**Mutation amino acids**	**EST**	**Library name**	**Tissue type**	**Study**	**Blastn**
c.495G > A	Synonymous	CN415983	LIBEST_014185 GRN_PREHEP	Embryonic stem cells	Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation	**E-value:** 6e-126; **Identity:** 97%

RESUMOS DE CONGRESSOS

XVIII Simpósio Sobre o Cérebro

ANÁLISE *IN SILICO* DE NOVOS POTENCIAIS POLIMORFISMOS GENÉTICOS DO SISTEMA HLA ASSOCIADOS AO RISCO DE TRANSTORNO BIPOLAR DO HUMOR EM BANCOS DE DADOS DE ESTs

Thalita Cristina Figueiredo Cunha[1]; João Ricardo Mendes de Oliveira[1,2]

[1]Pós-Graduação em Ciências Biológicas – Universidade Federal de Pernambuco (UFPE).[2]Pós-Graduação em Neuropsiquiatria e Ciências do Comportamento – Universidade Federal de Pernambuco (UFPE).

O Transtorno Bipolar do Humor (TBH) é um dos mais graves tipos de doença mental e caracteriza-se pela presença de episódios alternados de humor. O TBH se insere entre as doenças geneticamente complexas, cuja manifestação depende da presença de um conjunto de genes que interagem entre si resultando em uma fisiopatologia também complexa, até o momento pouco definida. Fatores genéticos, imunológicos e ambientais estão envolvidos na patogênese de várias doenças psiquiátricas, a exemplo, o TBH. O sistema HLA se destaca pelo seu polimorfismo e capacidade de conferir susceptibilidade ou resistência a vários distúrbios. Além disto, as doenças psiquiátricas tem uma herdabilidade relativamente alta. Isto vem conduzindo um aumento nos estudos de associação dos genes HLA com o TBH. Sendo assim, o objetivo desse estudo é definir os principais polimorfismos do sistema HLA que podem estar envolvidos no TBH, através de técnicas de Bioinformática. Para obter os resultados optou-se por utilizar o software CLCbio Workbench Combined® versão 5.7.1, para revelar variações novas, a partir de sequências públicas de *Expressed Site Tags* (ESTs). Como resultados um total de 6.257 ESTs foram alinhados para o gene HLA-A, sendo 1.080 sequências de ESTs selecionadas depois da aplicação de parâmetros apropriados de estringência utilizados para minimizar erros de alinhamentos. A anotação revelou 90 polimorfismos de base única (SNPs), sendo 28 já descritas e 62 novos potenciais fatores de risco genético a serem selecionados para análise como fatores de risco para o TBH. Dentre os SNPs identificados, 44 foram encontrados em pelo menos dois ESTs provenientes de bibliotecas diferentes. A repetição de SNPs na análise é um dado relevante, pois a presença do mesmo SNP em ESTs diferentes também é um indicativo de SNPs reais. O próximo passo é validar e triar os novos polimorfismos encontrados, em amostras de DNA de pacientes com TBH.

VII Escola Latino Americana de Genética Humana e Médica

Using ESTs database to validate and predict single polymorphisms at the HLA system

Thalita Cristina Figueiredo[1,2] and João Ricardo Mendes de Oliveira[1,2,3]

[1]*Keizo Asami Laboratory (LIKA), Federal University of Pernambuco, Recife, PE, Brazil.*
[2]*Biological Sciences Graduate Program, Federal University of Pernambuco, Recife, PE, Brazil.*
[3]*Department of Neuropsychiatry, Federal University of Pernambuco, 50670-901,Recife, PE, Brazil.*

Abstract The human leukocyte antigen system (HLA) is highly polymorphic and presents an associated ability to confer susceptibility to a large number of infections and autoimmune diseases. In this study, we propose a Bioinformatics pipeline in which we use ESTs database to validate and predict Single Nucleotide Polymorphisms (SNPs) directly linked to gene coding regions at the HLA class I genes (HLA-A, HLA-B and HLA-C) that might be further explored as potentially pathogenic genetic risk factors. The CLCbio Workbench Combined® version 5.7.1 was initially used to build expression sites tags (ESTs) and mRNA files, retrieved respectively from the Goldenpath (UCSC) and NCBI databases and latter to perform multiple batches of Smith–Waterman alignments. Annotation originated from our analysis revealed various classes of possible new variations. Most of these variations were found mainly in exons 2 and 3, which encode the molecular structures for alpha 1 and alpha 2 membrane receptor where is the antigen binding site. These structures are key to critical numerous immune functions. Therefore, the knowledge about the HLA system has great medical interest because it is directly related to the process of organs and tissues screened for transplants, with susceptibility to pathogens, as well as individual variability in susceptibility to disorders of autoimmune etiology. Bioinformatics pipelines seems to be useful approaches to help screening for novel genetic variations at the HLA panel and further analysis will foster this aim to provide celerity at the massive analysis of data currently generated in large scale high throughput experiments.

Keywords: *HLA system; Bioinformatics; ESTs; SNPs; Variations.*

57º CONGRESSO BRASILEIRO DE GENÉTICA

Potential new polymorphisms in exons 2 and 3 of the HLA class I genes found in the database of ESTs

Figueiredo, TC[1,2] and Oliveira, JRM[1,2,3].

[1]Keizo Asami Laboratory (LIKA), Federal University of Pernambuco, Recife, PE, Brazil.
[2]Biological Sciences Graduate Program, Federal University of Pernambuco, Recife, PE, Brazil.
[3]Department of Neuropsychiatry, Federal University of Pernambuco, 50670-901,Recife, PE, Brazil.

thalita.figueiredo87@gmail.com

Keywords: HLA system; Bioinformatics; ESTs; SNPs; Variations.

The human leukocyte antigen system (HLA) is highly polymorphic and presents an associated ability to confer susceptibility to a large number of infections and autoimmune diseases. In addition, is responsible for the rejection of organ transplants and tissue. In this study, using the bioinformatics pipeline was possible to identify polymorphisms in exons 2 and 3 of HLA class I genes that are not described in the main databases of the HLA system. The CLCbio Workbench Combined® version 5.7.1 was initially used to build expression sites tags (ESTs) and mRNA files, retrieved respectively from the Goldenpath (UCSC) and NCBI databases and latter to perform multiple batches of Smith–Waterman alignments. From the initial number of 18.029 ESTs that were aligned with the mRNAs of genes, 3.287 ESTs were obtained after the application of the parameters to decrease the chances of handling sequencing artifact and to select possible SNPs. Annotation originated from our analysis revealed various classes of possible new variations. Most of these variations were found mainly in exons 2 and 3, which encode the molecular structures for alpha 1 and alpha 2 membrane receptor where is the antigen binding site. These membrane receptors interact with T lymphocytes (CD8 +) that detect expression changes in cells, these changes may occur due to infections or the growth of tumor cells. Thus, T lymphocytes (CD8 +) recognize cells that should attack to combat the spread of a particular infection or a tumor. Therefore, the knowledge about the HLA system has great medical interest because it is directly related to the process of organs and tissues screened for transplants, with susceptibility to pathogens, as well as individual variability in susceptibility to disorders of autoimmune etiology. Bioinformatics pipelines seems to be useful approaches to help screening for novel genetic variations at the HLA panel and further analysis will foster this aim to provide celerity at the massive analysis of data currently generated in large scale high throughput experiments.

Financial Support: LIKA- UFPE, PROPESQ-UFPE and CAPES.