



**UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”  
FACULDADE DE MEDICINA**

**Iane de Oliveira Pires Porto**

**Estrutura e variabilidade da região 3' não-traduzida  
dos genes *HLA-A*, *HLA-C* e *HLA-G* e perfil de ligação  
de microRNAs**

Tese apresentada à Faculdade de  
Medicina, Universidade Estadual  
Paulista “Júlio de Mesquita Filho”,  
Campus de Botucatu, para obtenção  
do título de Doutora em Patologia.

Orientador: Prof. Dr. Erick da Cruz Castelli

**Botucatu**

**2018**

Iane de Oliveira Pires Porto

**Estrutura e variabilidade da região 3' não-traduzida dos genes *HLA-A*, *HLA-C* e *HLA-G* e perfil de ligação de microRNAs**

Tese apresentada à Faculdade de Medicina, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Botucatu, para obtenção do título de Doutora em Patologia.

Orientador: Prof. Dr. Erick da Cruz Castelli

Botucatu  
2018

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Porto, Iane de Oliveira Pires.

Estrutura e variabilidade da região 3' não-traduzida dos genes *HLA-A*, *HLA-C* e *HLA-G* e perfil de ligação de microRNAs / Iane de Oliveira Pires Porto. - Botucatu, 2018

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu

Orientador: Erick da Cruz Castelli

Capes: 40105008

1. Antígenos *HLA*. 2. Antígenos *HLA-C*. 3. Antígenos *HLA-G*.  
4. MicroRNAs.

Palavras-chave: *HLA-A*; *HLA-C*; *HLA-G*; microRNAs.



Agradecimentos

Agradeço inicialmente (e principalmente) à minha família, pelo apoio incondicional. Sem vocês, nada disso seria possível. Esse trabalho vem do esforço dos meus pais (José e Maria Socorro) e do apoio dos meus irmãos (Bruno, Igor e Clara).

Agradeço especialmente ao meu orientador, Dr. Erick Castelli, pessoa e profissional admirável a quem eu tive a sorte e o prazer de encontrar ainda nos tempos de iniciação científica. Obrigada por acreditar em mim e por me ajudar a realizar o sonho de um dia estar do outro lado da sala de aula.

Agradeço à melhor equipe que alguém poderia desejar: Thálitta Ayala (nossa Xuxu), Andreia Souza, Michelle Paz e Emiliana Weiss. Me faltam palavras para agradecer o papel de vocês não só nesse trabalho, mas na minha vida. Obrigada!

Aos amigos de todas as horas: Sarah Muricy, Suedio Meira, Gabriela Oliveira, Karla Polo, Monica Maffei, André Jim, companheiras de moradia, e amigos cujos nomes não foram citados – obrigada por todos os momentos felizes, pelos conselhos nos momentos de dúvida, por todos os ensinamentos compartilhados comigo. Meus dias são mais felizes com vocês.

E por último, mas não menos importante: Obrigada aos membros da banca, por dedicarem tempo e conhecimento para este trabalho. A contribuição de vocês é altamente apreciada.

Obrigada!!



Resumo

## RESUMO

O complexo gênico de Antígenos Leucocitário Humanos (HLA) é a região mais variável do genoma humano. Os genes HLA de classe I são divididos em clássicos e não clássicos, a depender de suas funções primárias, níveis de expressão e polimorfismos. Assim, *HLA-A*, *HLA-B* e *HLA-C* são loci clássicos e estão associados à apresentação antigênica para células T citotóxicas, enquanto *HLA-E*, *HLA-G* e *HLA-F* são loci não clássicos e estão associados à imunomodulação e inibição de células T e NK. No entanto, o *HLA-C* também apresenta propriedades imunomodulatórias a partir da interação com o TCR de células T CD8 (ativação) e com receptores KIR de células NK (principalmente inibição). A maioria dos estudos sobre variabilidade dos genes HLA têm como foco suas regiões codificadoras e negligenciam regiões regulatórias – região promotora e região 3' não-traduzida (3'NT). Esta última apresenta um papel essencial na estabilidade do mRNA e no controle pós-transcricional mediado pela ligação de microRNAs (miRNA). Neste trabalho, nós avaliamos a estrutura e a variabilidade das regiões 3'NTs dos genes *HLA-A*, *HLA-C* e *HLA-G* por sequenciamento de nova geração além do padrão de ligação de miRNAs observado para cada gene. A região 3'NT dos genes *HLA-A* e *HLA-C* são altamente variáveis, enquanto a de *HLA-G* apresenta poucos pontos de variação. No entanto, os três genes apresentaram números similares de haplótipos de 3'NT frequentes. Vários miRNAs possuem potencial de regular *HLA-A*, *HLA-C* e *HLA-G* tanto de modo específico quanto os três loci juntos. Este fato pode estar relacionado à regulação fina da expressão de HLA em tecidos diversos e em contextos imunológicos como a gravidez, onde *HLA-C* e *HLA-G* estão presentes, mas *HLA-A* está ausente. miRNAs capazes de regular mais de um gene HLA devem ser especialmente importante em sítios imunoprivilegiados. Os miRNAs com potencial de regular os genes estudados, individualmente ou em conjunto, foram compilados neste trabalho e devem nortear estudos funcionais a respeito da regulação pós-transcricional dos níveis de HLA.



Abstract



## ABSTRACT

The Human Leucocyte Antigen (HLA) gene complex is the most variable region of the human genome. The class I HLA genes are classified as classical or non-classical depending on their primary function, expression levels and polymorphism rate. Thus, *HLA-A*, *HLA-B*, and *HLA-C* are classical loci, associated with antigen presentation to cytotoxic T cells, while *HLA-E*, *HLA-G*, and *HLA-F* are non-classical loci associated with immunomodulation inhibition of both T and NK cells. Nevertheless, *HLA-C* also presents immunomodulatory features by interacting with TCR on CD8 T (activation) cells as well as KIR receptors on NK cells (mostly inhibition). Most studies on HLA variability focus on their coding sequences and neglects their regulatory sequences – promoter and 3'untranslated region (3'UTR). The last one plays a pivotal role on mRNA stability and post-transcriptional control by microRNA (miRNA) binding. Here we evaluated 3'UTR structure and variability for *HLA-A*, *HLA-C*, and *HLA-G* by using massively parallel sequencing in a Brazilian population sample. We also report the miRNA binding pattern observed for each locus. Both *HLA-A* and *HLA-C* 3'UTR segments are highly variable, while *HLA-G* 3'UTR presents few variation sites. However, the three loci present a similar number of frequent 3'UTR haplotypes. Several miRNAs are potential regulators of *HLA-A*, *HLA-C*, and *HLA-G* on a specific manner as well as the three of them together. This fact may be associated with the fine regulation of HLA expression on different tissues and immunological contexts, such as pregnancy, where *HLA-C* and *HLA-G* are present and *HLA-A* is absent. MiRNAs that are able to regulate more than one HLA locus may be important on immunoprivileged sites. miRNAs with potential to regulate the evaluated genes, both individually or together, were compiled on the present series and may guide functional studies regarding post-transcriptional regulation of HLA levels.

## LISTA DE ILUSTRAÇÕES

### CAPÍTULO I

<b>Figura 1</b> - Organização do Complexo Principal de Histocompatibilidade em humanos.....	19
<b>Figura 2</b> - Estrutura das moléculas de MHC de classes I e II.....	20
<b>Figura 3</b> - Processamento antigênico pelo proteassomo e montagem da molécula de MHC de classe I no Retículo Endoplasmático.....	23
<b>Figura 4</b> - Esquema da nomenclatura dos alelos de HLA.....	25
<b>Figura 5</b> - Esquema da biogênese dos microRNAs.....	39

### CAPÍTULO II

<b>Figure 1</b> - Conservation level across the <i>HLA-A</i> , <i>HLA-C</i> and <i>HLA-G</i> 3'UTR segment.....	70
<b>Figure 2</b> - Number of different miRNAs binding across the HLA genes with equal or lower energy values than the cutoff.....	78
<b>Figure 3</b> - Number of miRNAs interacting with <i>HLA-A</i> , <i>HLA-C</i> or <i>HLA-G</i> ...	79
<b>Figure 4</b> - Number of miRNAs interacting specifically with either <i>HLA-A</i> , <i>HLA-C</i> and <i>HLA-G</i> across their 3'UTR.....	80
<b>Figure 5</b> - Shared miRNAs across the <i>HLA-A</i> , <i>HLA-C</i> and <i>HLA-G</i> 3'UTR....	81
<b>Figure 6</b> - Number miRNAs shared between <i>HLA-A</i> and <i>HLA-C</i> 3'UTR and their distribution along the segment.....	82
<b>Figure 7</b> - <b>A</b> ) miRNAs highly specific (not shared) to <i>HLA-A</i> (red), <i>HLA-C</i> (green) and <i>HLA-G</i> (blue) 3'UTR and their target along the UTR segment. <b>B</b> ) Number of specific miRNAs (high S score) shared between the three loci and specific to each locus.....	83

## LISTA DE TABELAS

### CAPÍTULO I

<b>Tabela 1</b> - Quantidade de alelos e proteínas já descritos para os genes HLA de classe I clássicos e não clássicos.....	24
--	----

### CAPÍTULO II

<b>Table 1</b> - <i>HLA-A</i> 3'UTR haplotypes and associated coding alleles described at the IPD-IMGT/HLA database (version 3.30.0).....	71
<b>Table 2</b> - <i>HLA-A</i> 3'UTR haplotypes and associated coding allele groups found on a Brazilian population sample.....	72
<b>Table 3</b> - <i>HLA-C</i> 3'UTR haplotypes and associated coding alleles described at the IPD-IMGT/HLA database (version 3.30.0).....	73
<b>Table 4</b> - <i>HLA-C</i> 3'UTR haplotypes and associated coding allele groups found on a Brazilian population sample.....	75

## SIGLAS E ABREVIACOES

**Clulas NK** – *Clulas Natural Killer*

**PAMPS** – *Patogen-Associated Molecular Patterns* - Padres Moleculares Associados a Patgenos

**PRRs** – *Pattern Recognition Receptors* - Receptores de Reconhecimento de Padro

**TLR** – *Toll-like Receptor* - Receptor Semelhante a Toll

**CD4/CD8** – *Cluster of differentiation*

**APC**- *Antigen-presentin Cell* – Clula apresentadora de Antgeno

**MHC** – *Major Histocompatibility Complex* – Complexo Principal de Histocompatibilidade

**HLA** – *Human Leucocyte Antigen* – Antgeno Leucocitrio Humano

**RE** – Retculo Endoplasmtico

**TAP** - *Transporter associated with Antigen Processing* - Protena transportadora associada ao processamento antignico

**KIR** - *Killer Immunoglobuline-like Receptor* – Receptor semelhante  Imunoglobulina de clulas Natural Killer

**TCR** – *T Cell Receptor* – Receptor de Clulas T

**DSA** – *Donor-specific Antibodies* – Anticorpos Especficos do Doador

**RNA** – *Ribonucleic Acid* – cido Ribonucleico

**mRNA** – *Messenger Ribonucleic Acid* – cido Ribonucleico mensageiro

**HPV** – *Human Papiloma Virus* – Papiloma Vrus Humano

**LD** – *Linkage Disequilibrium* – Desequilbrio de Ligao

**INF** - Interferon

**ISRE** - *IFN-stimulated response element* – Elemento Estimulado por Interferon

**3'NT/3'UTR** – 3' No Traduzida/3' *Untranslated Region*

**miRNA** - microRNA

**PAB** - *Poli(A) Binding Proteins* – Protenas de Ligao  Poli(A)

**RBPs** – *RNA Binding Proteins* – Protenas de Ligao ao RNA

**HER2** - *Human Epidermal growth factor Receptor-type 2* – Receptor do fator de crescimento Epidermal Humano tipo 2

**IRF-1** – *Interferon Regulatory Factor 1* – Fator Regulador de Interferon 1-

**DGCR8** - *DiGeorge syndrome chromosomal [critical] region 8* – Região 8 cromossomal [crítica] da Síndrome DiGeorge

**Xpo5** – *Exportin 5* – Exportina 5

**TRBP** - *Transactivation-responsive RNA Binding Protein* – Proteína de Ligação a RNA responsiva à transativação

**dNK** – *decidual Natural Killer* – Natural Killer decidual

**HIV** – *Human Immunodeficiency Virus* – Vírus da Imunodeficiência Human

**Ago2** – *Argonaut 2* – Argonauta 2

**RISC/miRISC** – *RNA-Induced Silencing Complex/microRNA-Induced Silencing Complex* – Complexo de Silenciamento Induzido por RNA/ Complexo de Silenciamento Induzido por microRNA

**MFE** – *Minimum Free Energy* – Energia Livre Mínima

## SUMÁRIO

### CAPÍTULO I

<b>CAPÍTULO I: REVISÃO DE LITERATURA</b>	15
<b>O Complexo Principal de Histocompatibilidade – processamento e apresentação de antígenos.....</b>	17
<b>Os genes HLA de classe I <i>HLA-A</i>, <i>HLA-C</i> e <i>HLA-G</i>.....</b>	23
<b><i>O gene HLA-A</i>.....</b>	30
<b><i>O gene HLA-C</i>.....</b>	34
<b><i>O gene HLA-G</i>.....</b>	36
<b>microRNAs e seu papel na regulação pós-transcricional do HLA de classe I.....</b>	37
<b>Métodos de detecção de alvos de microRNAs e uso do mirhunt.....</b>	41
<b>REFERÊNCIAS.....</b>	45
<b>JUSTIFICATIVA.....</b>	59
<b>OBJETIVOS.....</b>	61
<b>CAPÍTULO II: MANUSCRITO</b>	
<b>Abstract.....</b>	63
<b>Introduction.....</b>	64
<b>Methods.....</b>	66
<b>Results.....</b>	68
<b>Discussion.....</b>	83
<b>References.....</b>	90
<b>CONCLUSÃO.....</b>	97
<b>APÊNDICE.....</b>	99



# Capítulo I

## Revisão de Literatura

## CAPÍTULO I: REVISÃO DE LITERATURA

O Sistema Imunológico humano possui como função proteger o corpo contra os agentes infecciosos e o desenvolvimento de doenças ou qualquer outra ameaça capaz de causar lesão. Para isso, ele conta com as barreiras físicas, a pele e as mucosas, substâncias químicas e células capazes de reconhecer, com diferentes graus de especificidade, estruturas que não façam parte do organismo (tudo que é “não próprio”).

Com base em certos parâmetros, como tempo de ativação, células e mecanismos efetores, as respostas imunológicas podem ser divididas em duas classes: as respostas imunes inatas e as adaptativas. Enquanto as respostas inatas surgem pouco tempo após o estímulo nocivo e suas células reconhecem padrões moleculares característicos dos microrganismos, as respostas adaptativas levam mais tempo para acontecer, e apresentam um repertório de moléculas que garantem a sua especificidade ao patógeno em questão, sendo capazes de diferenciar as moléculas do hospedeiro de moléculas estranhas (Choo, 2007; Howell *et al.*, 2010; Abbas *et al.*, 2015).

Assim, a imunidade inata é tida como imediata e fornece a primeira linha de defesa do organismo, uma vez que é constituída pelas barreiras físicas e epitélios, células fagocíticas, proteínas circulantes (mediadores inflamatórios, sistema complemento, entre outros) e um tipo específico de linfócitos – as células *Natural Killer*, ou células *NK*. No caso de uma segunda exposição a determinado microrganismo, o mecanismo da resposta inata se mantém o mesmo da primeira exposição, não sendo capaz de distinguir pequenas diferenças entre microrganismos de um mesmo grupo. Por isso, ela é comumente chamada de resposta inespecífica (Murphy *et al.*, 2010; Abbas *et al.*, 2015). As células da resposta imune inata possuem em sua superfície receptores capazes de reconhecer padrões moleculares de diferentes tipos de microrganismos, conhecidos como Padrões Moleculares Associados a Patógenos (PAMPs, do inglês *Pathogen-Associated Molecular Pattern*). Por esse motivo, os receptores que reconhecem os PAMPs são chamados de Receptores de Reconhecimento Padrão (PRRs, do inglês



*Pattern Recognition Receptors*), entre os quais se destacam os receptores semelhantes a Toll (TLRs, do inglês *Toll-like Receptors*) (Murphy *et al.*, 2010).

Em contrapartida, a imunidade adaptativa é conhecida por seu alto grau de especificidade e capacidade de memória. Entre seus efetores podemos citar os linfócitos T e B e os anticorpos secretados por plasmócitos (linfócitos B ativados). A diferenciação entre as moléculas do hospedeiro e aquelas pertencentes a microrganismos que devem ser combatidos é feita por meio de moléculas especializadas de apresentação e reconhecimento de antígenos. Isso garante que mesmo microrganismos semelhantes entre si sejam reconhecidos de modo específico. As respostas adaptativas não são imediatas, mas em caso de exposição secundária, elas são rápidas e comumente exacerbadas (Murphy *et al.*, 2010; Abbas *et al.*, 2015). As respostas inatas podem ainda se dividir em celulares e humorais. As respostas celulares são mediadas pelas células T (CD4<sup>+</sup> e CD8<sup>+</sup>), enquanto a resposta humoral é mediada pelos anticorpos secretados pelas células B (Abbas *et al.*, 2015).

Para que uma resposta imune adaptativa seja deflagrada, o patógeno (quando extracelular) deve ser fagocitado por células capazes de processá-lo em peptídeos, conhecidas como células apresentadoras de antígenos (APCs, do inglês, *Antigen Presenting Cells*). Nas APCs, esses antígenos processados são acoplados a moléculas específicas – moléculas do Complexo Principal de Histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*) – que apresentarão esses antígenos aos linfócitos T, ativando-os (Abbas *et al.*, 2015). Em contrapartida, na presença de patógenos intracelulares, peptídeos antigênicos são produzidos no citosol por proteases que atuam em conjunto com o proteassoma. Esses antígenos são então transportados para o retículo endoplasmático, onde se juntam as moléculas de MHC em formação (Adiko *et al.*, 2015).

### **O Complexo Principal de Histocompatibilidade – processamento e apresentação de antígenos**

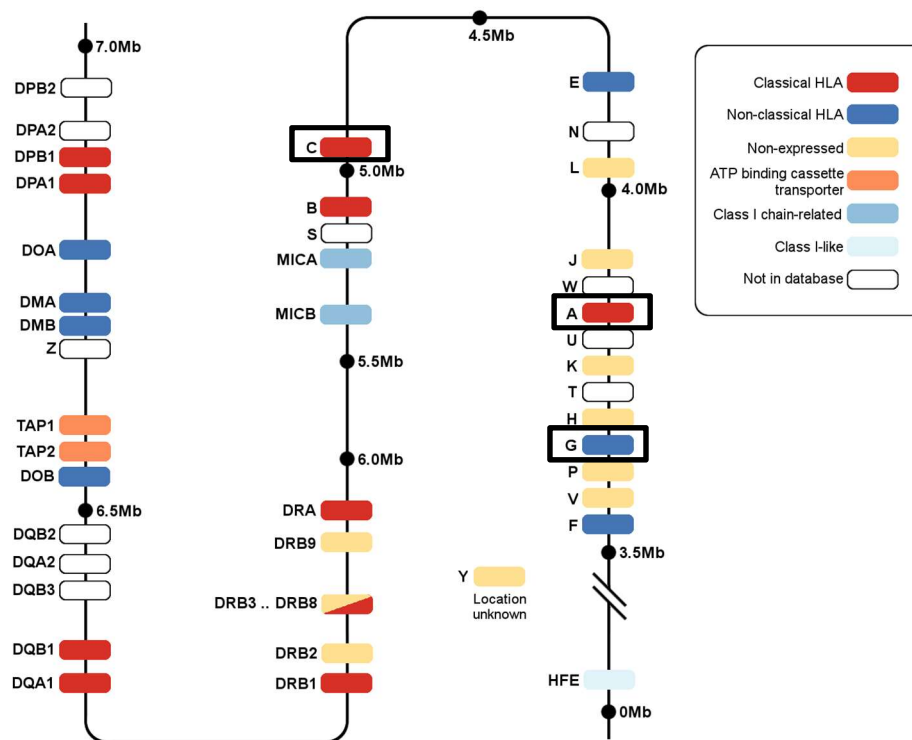
O MHC foi descoberto por meio de estudos de rejeição a transplantes em

camundongos ainda na década de 1930 e, por volta da década 1970, sabia-se que um grande complexo gênico era responsável pelo controle das respostas imunes – o Complexo Principal de Histocompatibilidade. Essa descoberta rendeu um prêmio Nobel em 1980 a George Davis Snell, Jean Baptiste Dausset e Baruj Benacerraf (Raju, 1999). Esses pesquisadores perceberam que enxertos de pele entre camundongos isogênicos não eram rejeitados, enquanto transplantes entre linhagens diferentes não eram aceitos pelo organismo do camundongo receptor.

Em humanos, essa descoberta foi feita a partir de três grupos de pacientes: (i) aqueles que haviam passado por múltiplas transfusões sanguíneas, (ii) pacientes que haviam passado por transplantes renais e (iii) mulheres múltiparas. Analisando o sangue desses pacientes, era possível encontrar, respectivamente, anticorpos que reconheciam as células sanguíneas e renais dos doadores e as células paternas. Por esse motivo, o MHC humano foi chamado de Antígenos Leucocitários Humanos (HLA, do inglês *Human Leucocyte Antigens*) (Terasaki, 2007; Thorsby, 2009; Abbas *et al.*, 2015).

Em humanos, os mais de 200 genes que compõem o MHC estão localizados no braço curto do cromossomo 6 (6p21.3), ocupando uma região que se estende por cerca de 3,5 megabases (Mb) (Figura 1) (Klein e Sato, 2000; Abbas *et al.*, 2015). Esses genes e as moléculas que eles codificam, bem como sua organização cromossômica, estão presentes em todos os vertebrados que possuem mandíbula, os gnatostomados, numa organização que vem sendo mantida conservada por mais de 500 milhões de anos (Klein e Sato, 1998). Certos genes desse complexo são particularmente variáveis e seus produtos são fundamentais ao desenvolvimento da resposta imunológica.

Com base em estrutura molecular e funções, esses genes foram convencionalmente divididos em classes I, II e III (Klein e Sato, 2000). Os genes que compõem as classes I e II também são conhecidos como Antígenos Leucocitários Humanos (HLA, do inglês *Human Leucocyte Antigens*), uma vez que atuam na apresentação antigênica a linfócitos. Os componentes do MHC de classe III incluem algumas citocinas e outros componentes das respostas imunológicas.

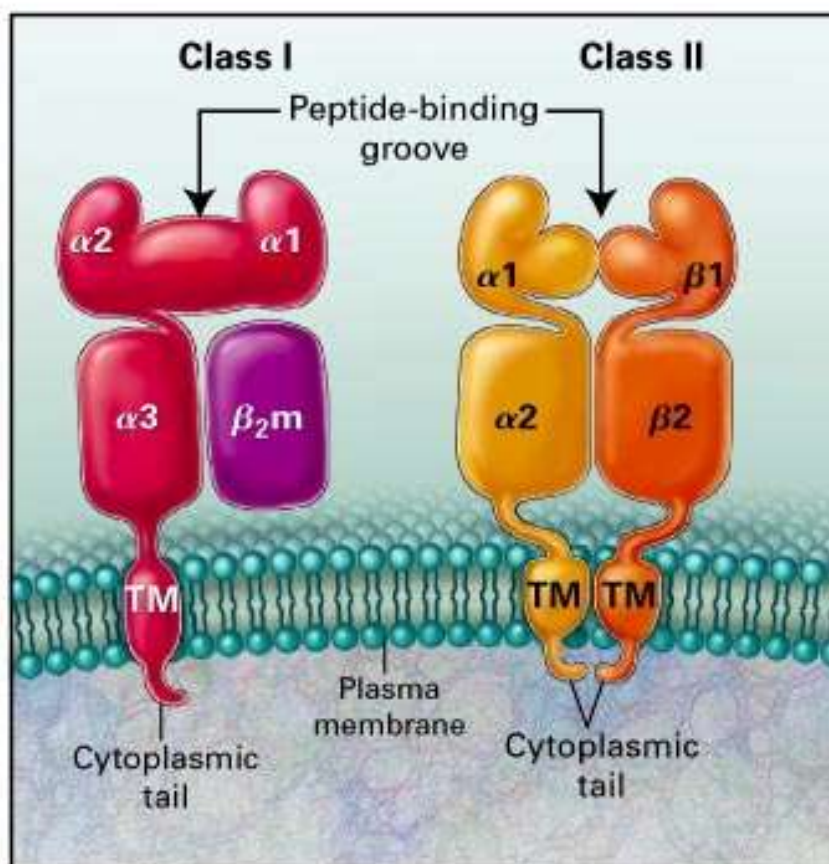


**Figura 1** – Organização do Complexo Principal de Histocompatibilidade em humanos. Os genes alvo desse trabalho estão em destaque.

Fonte: [http://hla.alleles.org/inc/images/hla\\_map\\_1000.png](http://hla.alleles.org/inc/images/hla_map_1000.png)

Os genes de classe I codificam as cadeias alfa das moléculas HLA. As cadeias alfa ligam-se não covalentemente a uma molécula de  $\beta$ 2-microglobulina, codificada por um gene localizado fora do complexo HLA, no cromossomo 15. A molécula HLA de classe I apresentam cinco domínios: dois domínios de ligação a peptídeos,  $\alpha$ 1 e  $\alpha$ 2, que juntos formam a fenda de ligação de peptídeo; um domínio de imunoglobulina (Ig),  $\alpha$ 3; um domínio transmembrana e a cauda citoplasmática (Figura 2). Dentro dessa classe, destacam-se os genes mais polimórficos do genoma humano e expressos em praticamente todas as células nucleadas – *HLA-A*, *HLA-B* e *HLA-C* (Klein e Sato, 2000; Murphy *et al.*, 2010; Abbas *et al.*, 2015). Já os genes de classe II codificam cadeias polipeptídicas  $\alpha$  e  $\beta$  e sua nomenclatura é derivada da estrutura da molécula e conta com o auxílio três letras que

representam sua classe (D), família (M, O, P, Q ou R) e o tipo de cadeia que a compõe (A ou B – cadeias  $\alpha$  ou  $\beta$ , respectivamente). Por exemplo, HLA-DRB (gene de classe II, família R, cadeia  $\beta$ ) (Klein e Sato, 2000).



**Figura 2** – Estrutura das moléculas de MHC de classes I e II.

Fonte: Adaptado de Klein & Sato, 2000

A maioria das células nucleadas humanas expressam moléculas de MHC de classe I em sua superfície associadas a antígenos próprios ou estranhos. O complexo MHC-I/peptídeo interage com o TCR de células T  $CD8^+$ , que reconhecem o antígeno como próprio ou não próprio. (Amigorena, 2016; Brutkiewicz, 2016). Caso a célula T tenha sido anteriormente apresentada ao antígeno em questão, esta interação resulta em sinais de ativação da célula T que, por sua vez, resultam na morte de células que apresentem este antígeno. Para isso,

os antígenos são mantidos na fenda de ligação a peptídeos por meio de interações não covalentes, entre a sequência do peptídeo e pontes de hidrogênio formadas entre aminoácidos conservados da molécula de HLA e aminoácidos do peptídeo (localizados nos finais da fenda) (Jensen, 2007).

A descoberta de que as moléculas do MHC de classe I e II eram necessárias para a ativação dos linfócitos – restrição por MHC – foi recompensada com um prêmio Nobel em 1996 a Peter C. Doherty e Rolf M. Zinkernagel (Adiko *et al.*, 2015). A partir de estudos nessa área, condicionou-se que as moléculas do MHC de classe I seriam responsáveis pela apresentação de antígenos de origem endógena (vírus, em especial), enquanto o MHC de classe II seria responsável por apresentar antígenos exógenos fagocitados. Porém, hoje é sabido que essa regra não é absoluta devido aos mecanismos de apresentação cruzada, onde antígenos exógenos são processados de modo a serem apresentados por moléculas de MHC de classe I e antígenos endógenos passam a ser apresentados por moléculas de MHC de classe II (Jensen, 2007; Adiko *et al.*, 2015; Brutkiewicz, 2016; Grotzke *et al.*, 2017).

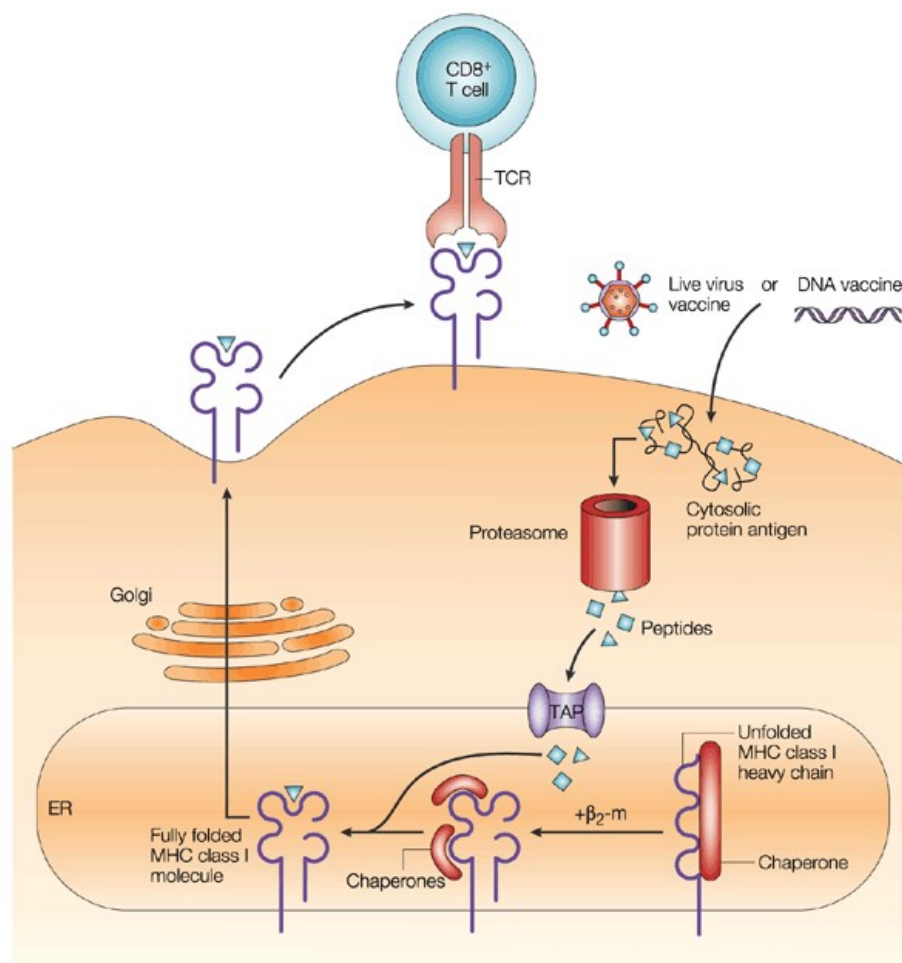
Os peptídeos que interagem com o HLA de classe I são convencionalmente gerados a partir do processamento de proteínas endógenas (normais ou alteradas) ou de patógenos (preferencialmente intracelulares) por meio da maquinaria do proteassomo e outras enzimas (Figura 3). Os proteassomos podem ser divididos a partir dos antígenos que eles processam. Quando o antígeno em questão é próprio, eles são chamados “proteassomos constitutivos”, enquanto aqueles que processam peptídeos não próprios são comumente chamados de “imunoproteassomos”. Tipos celulares diferentes possuem quantidades diferentes de cada tipo de proteassomo – células dendríticas são ricas em imunoproteassomos (Inaba e Inaba, 2005).

Após o processamento, os antígenos são transportados para o lúmen do Retículo Endoplasmático (RE) por meio da proteína transportadora associada ao processamento antigênico (TAP). Enquanto isso, no RE, a chaperona calnexina é responsável por moldar a cadeia pesada do MHC de classe I e por uni-la com a molécula de  $\beta$ 2-microglobulina. Quando a interação está estabelecida, a proteína tapasina mantém a ligação entre as duas (Figura 3). Somente quando a molécula

está completa – associada à  $\beta$ 2-microglobulina e com um antígeno em sua fenda – ela é transportada para a membrana pelo sistema de cisternas do Complexo de Golgi (Inaba e Inaba, 2005; Jensen, 2007). A montagem das moléculas de HLA corresponde a uma série de eventos finamente orquestrados que incluem um conjunto extenso de proteínas especializadas.

Apesar do MHC de classe I estar associado à apresentação de antígenos intracelulares, ocasionalmente, alguns antígenos extracelulares também podem ser apresentados aos linfócitos T CD8<sup>+</sup> por um processo chamado de apresentação cruzada. Este processo ocorre por meio de dois mecanismos distintos: um dependente de TAP e proteassomo (via citosólica) e um independente (via vacuolar) (Grotzke *et al.*, 2017). Na via citosólica, o antígeno internalizado por fagocitose precisa sair de sua vesícula e alcançar o citosol, onde ele é processado pelo proteassomo e transportado pela TAP. Esse mecanismo é mediado por uma proteína do retículo endoplasmático chamada Sec61 *translocon* (Ackerman *et al.*, 2006). Na via vacuolar, ocorre a endocitose do antígeno associado a uma proteína de membrana (como receptores Toll ou uma outra molécula de MHC), e o antígeno é degradado ainda no endossomo, sem a necessidade do proteassomo e, a protease catépsina S e o pH ácido endossomal participam desse mecanismo (Shen *et al.*, 2004).

Antígenos que se ligam ao HLA de classe II são, em geral, maiores que aqueles apresentados pelo HLA de classe I (limitados a 8-10 resíduos de aminoácidos) e, assim, as vias de processamento e acoplagem do peptídeo à fenda sofrem algumas diferenças. Como consequência disso, o processamento antigênico e sua interação com a fenda possuem um grande efeito na manutenção da estabilidade e conformação do domínio de ligação ao peptídeo (Jensen, 2007). As moléculas de MHC de classe II estão comumente associadas à apresentação de antígenos extracelulares e, por isso, sua expressão se limita às células apresentadoras de antígeno. Nessas células, os peptídeos antigênicos são gerados por proteólise nos endossomos e lisossomos.



Nature Reviews | Immunology

**Figura 3** – Processamento antigênico pelo proteassomo e montagem da molécula de MHC de classe I no Retículo Endoplasmático.

Fonte: Adaptado de Yewdell et al., 2003

### Os genes HLA de classe I *HLA-A*, *HLA-C* e *HLA-G*

Os genes que compõem o MHC de classe I em humanos podem ser classificados em clássicos e não clássicos (Beck *et al.*, 1999; Shiina *et al.*, 2009), quando levamos em consideração sua função na resposta imunológica e perfil de expressão. Os genes *HLA-A*, *HLA-B* e *HLA-C* são os chamados genes HLA de classe I clássicos. Eles são os genes mais variáveis de todo o genoma humano e juntos, contabilizam mais de 12.000 alelos descritos em um banco de dados

específico, o IPD-IMGT/HLA, disponível em <https://www.ebi.ac.uk/ipd/imgt/hla/> (Robinson *et al.*, 2015; Robinson *et al.*, 2016) (Tabela 1). As moléculas de classe I clássicas estão presentes em praticamente todas as células nucleadas do organismo e tem como função principal a apresentação antigênica aos linfócitos T CD8+. Por outro lado, os genes HLA de classe I não clássicos, *HLA-E*, *HLA-F* e *HLA-G*, são pouco variáveis, possuem expressão limitada a determinados tipos celulares (com exceção do gene *HLA-E*) e tem como função principal a imunomodulação – a apresentação antigênica aparece como uma função secundária e menos comum dessas moléculas (Wei e Orr, 1990; Geraghty *et al.*, 1992; Gobin e Van Den Elsen, 2000; Shobu *et al.*, 2006).

**Tabela 01** – Quantidade de alelos e proteínas já descritos para os genes HLA de classe I clássicos e não clássicos (dados disponíveis em IPD-IMGT/HLA Database – versão 3.34.0 <https://www.ebi.ac.uk/ipd/imgt/hla/stats.html>)

Gene	HLA-A	HLA-B	HLA-C	HLA-E	HLA-F	HLA-G
<b>Alelos</b>	4.638	5.590	4.374	27	31	61
<b>Proteínas</b>	3.172	3.923	2.920	8	6	19
<b>Alelos Nulos*</b>	224	169	171	1	0	3

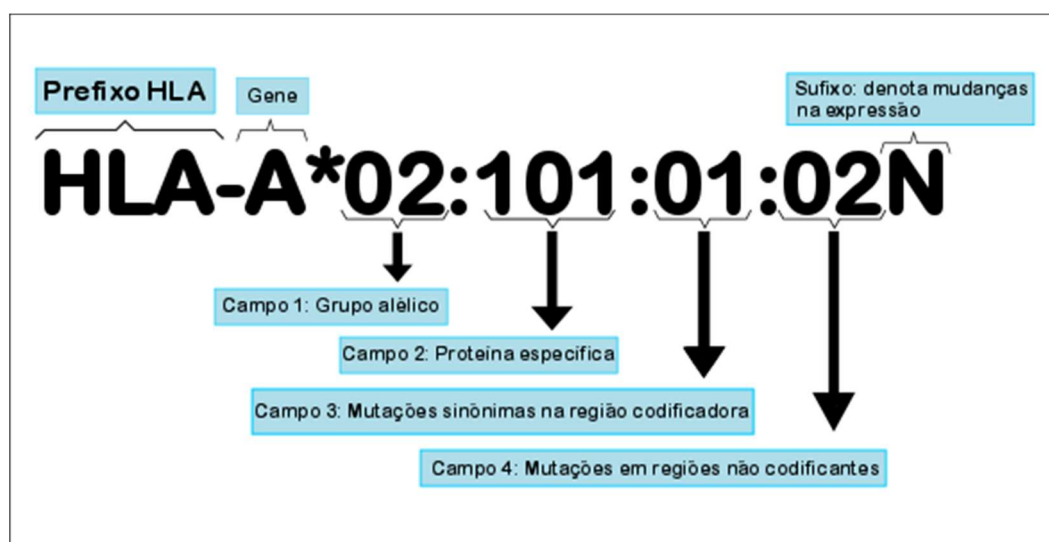
\* apresentam mutações que impedem sua expressão na superfície celular

A grande quantidade de sequências descritas, em especial para os genes clássicos de classe I, exigiram que uma nomenclatura especial fosse desenvolvida para identificar os alelos já descritos e para facilitar a nomeação de novos alelos. Cada alelo é identificado por um prefixo (HLA) e o gene correspondente (A, B, C, E, F, G para os de classe I) separados por um hífen. Depois da identificação do gene, seguem-se quatro conjuntos de dígitos separados por dois pontos (:) (Figura 4).

O primeiro campo corresponde ao antígeno sorológico, representando um grupo de alelos que codificam moléculas com motivos semelhantes e que em geral são reconhecidos pelo mesmo anticorpo em um teste sorológico. O segundo campo determina subtipos dentro do mesmo grupo sorológico, gerados por



mutações não sinônimas (mudam um ou mais aminoácidos na proteína), e a numeração é designada na ordem em que as sequências foram descritas. O terceiro campo denota mutações sinônimas em éxons, que não mudam aminoácidos da proteína. O quarto e último campo designa mudanças nos segmentos que não fazem parte da proteína final (regiões regulatórias e íntrons). Além da numeração específica, os alelos também podem receber um sufixo que representam modos alternativos de expressão – N (alelo nulo, sem expressão na superfície celular), S (solúvel ou secretado), L (níveis baixos de expressão na superfície, derivado de “*Low*”), C (para alelos que ficam retidos no citoplasma), A (aberrante, quando há dúvidas quanto a expressão da proteína) e Q (questionável). Até o momento, nenhuma sequência foi descrita como C ou A (IPD-IMGT/HLA, disponível em <http://hla.alleles.org/nomenclature/naming.html>).



**Figura 04** – Esquema da nomenclatura dos alelos de HLA.

**Fonte:** Adaptado de IPD-IMGT/HLA (<http://hla.alleles.org/nomenclature/naming.html>)

As moléculas de MHC são expressas na superfície celular em codominância, ou seja, os dois alelos herdados dos pais são expressos. Assim, a presença de duas variantes alélicas aumentariam o repertório de antígenos apresentados. Além da codominância, observam-se altos níveis de heterozigose

para esse conjunto de genes e, uma vez que variantes proteicas diferentes apresentam capacidades diferentes de acoplamento e apresentação antigênica, maximiza-se a gama de antígenos a serem apresentados aos linfócitos T (Piertney e Oliver, 2006). As altas taxas de heterozigose também estão associadas à presença de seleção balanceadora atuando nesse locus.

Um outro fator que contribui para a grande quantidade de alelos dos genes clássicos é seleção mediada por parasitas, uma vez que as moléculas de MHC atuam diretamente no combate aos patógenos, a partir da apresentação de antígenos microbianos para as células T CD8<sup>+</sup>. Essa teoria ganha força quando observamos relações bem estabelecidas entre alelos específicos de MHC I e II e a suscetibilidade ou proteção a certas doenças humanas. Como exemplo, podemos citar o efeito protetor observado em pacientes com HIV portadores do alelo *HLA-B\*57:01*. A presença desse alelo está associada a menor viremia e maior contagem de células T CD4<sup>+</sup> (Leszczyszyn-Pynka *et al.*, 2015).

Algumas evidências apontam ainda a seleção sexual como sendo um dos fatores responsáveis por moldar o MHC até o modo como o conhecemos hoje (Edwards e Hedrick, 1998; Piertney e Oliver, 2006). Por mecanismos ainda não compreendidos, receptores sensoriais olfatórios específicos reconheceriam o odor de moléculas solúveis de MHC, de modo que os indivíduos são capazes de “cheirar a variabilidade” das moléculas de MHC do parceiro (Capittini *et al.*, 2008). Em humanos, esses receptores são codificados por genes localizados na região telomérica do complexo HLA no cromossomo 6, formando um conjunto de Receptores Olfatórios ligados ao HLA (Ehlers *et al.*, 2000). Esses receptores reconhecem os sinais químicos gerados pela ligação MHC/peptídeo e assim, reconhecem e diferenciam odores de familiares e possíveis parceiros não relacionados (Ehlers *et al.*, 2000; Capittini *et al.*, 2008). Nesse contexto, parceiros seriam escolhidos com base na diversidade entre o MHC de ambos de modo a ampliar a diversidade na prole e a produzir indivíduos heterozigotos capazes de combater uma ampla variedade de patógenos (Edwards e Hedrick, 1998; Capittini *et al.*, 2008).

Alguns estudos mostram também a co-evolução entre o MHC e seus

ligantes, KIR (*Killer Immunoglobuline-like Receptor*) e TCR (*T Cell Receptor*) (Rangarajan e Mariuzza, 2014; Guethlein *et al.*, 2015). Desse modo, TCR e MHC evoluíram juntamente para garantir a habilidade de interagirem entre si, garantindo a ativação eficiente de células T (Rangarajan e Mariuzza, 2014). Por outro lado, os receptores KIR interagem com as moléculas de MHC de classe I de modo alelo específico. Acredita-se que esse fato está associado a um melhor combate a infecções virais e ao sucesso reprodutivo, uma vez que células NK possuem um papel importante nas respostas imunológicas desses dois processos.

Embora, em geral, ambos os alelos de um indivíduo são expressos (co-dominância), não está claro se há expressão alélica diferencial, i.e., se alguns alelos expressam quantidades maiores ou menores da proteína em comparação aos outros. Uma vez que as moléculas HLA de classe I possuem um papel fundamental no desenvolvimento das respostas imunológicas, em especial na ativação de células T citotóxicas, a expressão inadequada dessas moléculas está intimamente associada ao desenvolvimento de certas doenças. Como exemplo, podemos citar certas infecções virais e o desenvolvimento de neoplasias, onde o principal mecanismo de evasão ao sistema imune adaptativo é redução da expressão de moléculas de HLA na superfície celular tumoral por mecanismos diversos, que envolvem desde o comprometimento das moléculas responsáveis pela montagem das moléculas de HLA à redução dos níveis de expressão gênica (Peltenburg *et al.*, 1993; Odeberg *et al.*, 2002; Shen *et al.*, 2005; Tang *et al.*, 2009; Apps *et al.*, 2016).

Em transplantes, a compatibilidade entre o HLA do doador e do receptor é avaliada para pelo menos cinco genes, os três genes clássicos de classe I (*HLA-A*, *HLA-B* e *HLA-C*) e dois de classe II (*HLA-DQB1* e *HLA-DRB1*). A compatibilidade entre os cinco *loci* está associada a uma maior taxa de sobrevivência do enxerto e do paciente. Porém, quando ela não é possível, a incompatibilidade entre algum desses genes resulta em taxas diversas de aceitação do enxerto a depender do gene incompatível e do órgão em questão. A rejeição mediada por respostas imunes adaptativas ocorre quando antígenos do doador são apresentados às células T CD4<sup>+</sup> ou CD8<sup>+</sup> do receptor por células dendríticas deste

(reconhecimento indireto), ou mesmo do doador (reconhecimento direto). Mudanças estruturais entre as moléculas de MHC do doador e do receptor também podem estimular respostas imunes por contato direto com o TCR do doador (Moreau *et al.*, 2013). Assim, quanto maior a compatibilidade entre os dois, melhor o resultado do transplante.

Um dos genes mais considerados quando a compatibilidade completa não é possível é o *HLA-C*, uma vez que ele parece ser expresso em níveis mais baixos na superfície celular. No entanto, os resultados encontrados são conflitantes – em transplantados renais, anticorpos doador-específico (DSA) anti-*HLA-C* são comumente encontrados em pacientes cujos enxertos são rejeitados (Duquesnoy e Marrari, 2011) ao mesmo tempo que de 15 a 30% dos transplantes de células tronco hematopoiéticas realizados apresentam incompatibilidade entre os alelos de *HLA-C* do doador e do receptor (Tiercy, 2014).

A eficiência das respostas imunes e o desfecho dos processos patológicos estão intimamente associados aos níveis de expressão das moléculas HLA. O aumento da expressão dessas moléculas resulta na melhora da eficácia da apresentação antigênica e conseqüentemente das respostas imunes (Faroudi *et al.*, 2003). Duas regiões gênicas estão intimamente relacionadas com os níveis de expressão de um determinado gene – a região promotora e a região 3' não traduzida (3'NT).

Na região promotora encontram-se os motivos gênicos responsáveis pela ligação dos fatores de transcrição e outras proteínas que influenciam na ligação e funcionamento do complexo da RNA polimerase e, assim, dando início à transcrição do RNA (Juven-Gershon *et al.*, 2008). Desse modo, a região promotora é fundamental no controle dos níveis de RNA.

Os genes HLA de classe I contam com uma série de reguladores transcricionais em seus promotores que garantem sua expressão, uma vez que eles devem estar presentes na superfície da grande maioria das células do organismo. Juntamente com os genes que codificam a cadeia pesada da molécula de HLA, é preciso também que haja a expressão coordenada de proteínas associadas ao processamento antigênico e montagem da molécula HLA como um todo, como as

proteínas do proteossoma, bem como da molécula de  $\beta_2$ -microglobulina. Alguns vírus, como o Papilomavirus Humano (HPV), são capazes de modular a expressão desses genes, reduzindo a produção das moléculas de HLA e facilitando seu escape do sistema imunológico. Para isso, proteínas virais, como a proteína E7 do HPV-16 e do HPV-6a são capazes de reprimir a transcrição das cadeias pesadas de HLA e da proteína TAP por meio de interações com as regiões promotoras desses genes (Georgopoulos *et al.*, 2000).

Apesar da expressão constitutiva desses genes, algumas moléculas específicas são capazes de aumentar os níveis de HLA de classe I, mesmo em linhagens celulares onde HLA é pouco expresso. Isso pode ser observado na presença dos Interferons (IFNs) de tipo 1 ( $\alpha$  e  $\beta$ ) e tipo 2 ( $\gamma$ ). Essa citocina é capaz de estimular a transcrição tanto dos genes clássicos quanto não clássicos do HLA de classe I devido à presença de uma sequência conhecida como Elemento de Resposta Estimulado por IFN ou ISRE (do inglês, *IFN-stimulated response element*) (Chen *et al.*, 1986; Browne *et al.*, 2006) no promotor dos genes HLA. O IFN- $\gamma$  é capaz de melhorar a resposta citotóxica contra tumores mediada por células T, por estimular a expressão dos genes clássicos do MHC de classe I. No entanto, um aumento da expressão de HLA-E, por exemplo, pode resultar na inibição da citotoxicidade mediada por células NK (Browne *et al.*, 2006).

A região 3'NT, por sua vez, tem um papel muito importante na manutenção da estabilidade e estrutura secundária da molécula de mRNA (Kuersten e Goodwin, 2003). Essa região é ainda a principal fonte de sequências-alvo para microRNAs (miRNAs), responsáveis pela redução do nível proteico via inibição da tradução e, assim, atuando como um importante regulador pós-transcricional (discutido adiante) (Bartel, 2009; Pasquinelli, 2012).

Em contrapartida, a região 3'NT de uma molécula de mRNA recém-sintetizada possui um papel importante no processo de tradução. Quando uma nova molécula de mRNA é sintetizada, uma sequência rica em Adenina é adicionada ao final 3' da molécula, formando a chamada cauda poli(A). Proteínas de Ligação a Poli(A) (PAB, do inglês *Poli(A) Binding protein*) ligam-se à cauda poliA, e interagem com complexos proteicos responsáveis por recrutar as

subunidades ribossomais para a ponta 5' do mRNA como o grupo de iniciadores da tradução eIF4F (Yamashita e Takeuchi, 2017).

Desse modo, podemos concluir que a região 3'NT de uma molécula de mRNA atua de modo homólogo às regiões promotoras dos genes: fatores (especialmente proteínas de ligação a RNA, além de miRNAs) se ligam a sequências específicas e orquestram a regulação pós-transcricional, determinando a quantidade de proteína gerada por essa molécula de mRNA. Em humanos, mais 1000 proteínas de ligação a RNA (RBPs, do inglês *RNA Binding Proteins*) (Gerstberger *et al.*, 2014) e mais de 2500 miRNAs (Kozomara e Griffiths-Jones, 2014) já foram descritos. Esses dois elementos podem atuar de modo cooperativo ou competitivo, mas estudos que tentam mostrar essa relação em escala de transcriptoma ainda são escassos (Preusse *et al.*, 2015; Hafezqorani *et al.*, 2016). Nesse contexto, mudanças na sequência de 3'NT de um gene são capazes de alterar os motivos de ligação de RBPs e miRNAs, alterando os mecanismos envolvidos na estabilização ou degradação da molécula de mRNA e tradução.

### ***O gene HLA-A***

O gene *HLA-A* é o segundo mais polimórfico do genoma humano. Ele é composto por oito éxons e sete íntrons. O éxon 1 é responsável por codificar o peptídeo líder, que não faz parte da estrutura proteica final, mas comumente se liga à fenda de ligação ao peptídeo de moléculas de HLA não-clássicas, como *HLA-E*. Os éxons dois e três são responsáveis pela codificação das cadeias  $\alpha 1$  e  $\alpha 2$ , que incluem a fenda de ligação ao peptídeo e são altamente variáveis (Malissen *et al.*, 1982). A cadeia  $\alpha 3$  é codificada pelo éxon 4 e, diferente das outras duas, possui uma cadeia de aminoácidos mais conservada. Da mesma forma, as regiões transmembrana (éxon 5) e citoplasmática (éxons 6 e 7) também são mais conservadas (Tragardh *et al.*, 1979). A conservação do domínio  $\alpha 3$  pode estar associada a sua função como ligante do co-receptor CD8 e, portanto, mudanças em sua sequência seriam capazes de afetar essa interação (Salter *et al.*, 1989; Hanvesakul *et al.*, 2007).

No domínio citoplasmático, os alelos de *HLA-A* apresentam uma cisteína

não pareada (não relacionada a uma ligação dissulfeto) no resíduo 339 que parece estar envolvida na reciclagem da molécula, sua degradação e ainda no reconhecimento de receptores de células NK e formação de dímeros de MHC de classe I em exossomos (Makhadiyeva *et al.*, 2012). Assim, variações na sequência desse gene apresenta repercussões funcionais variadas, a depender da região em questão.

A região mais bem caracterizada para o gene *HLA-A* compreende os éxons dois e três, pois ambos codificam o domínio de ligação ao peptídeo. Grande parte dos estudos populacionais focaram-se nesse segmento, e pouco se sabe sobre as regiões regulatórias – promotor e região 3'NT – apesar de seu papel na manutenção dos níveis de mRNA e proteicos (Kuersten e Goodwin, 2003). Uma investigação recente mostrou o papel de polimorfismos da região 3'NT do gene *HLA-A* na ligação de certos miRNAs capazes de reduzir sua expressão em nível pós-transcricional. Por exemplo, a presença de uma Adenina na posição +2969 (rs1061235) faz com que pelo menos cinco miRNAs deixem de atuar nessa região (miR526b, miR609, miR1290, miR342-5p, miR542-5p). Entretanto, ao mesmo tempo que uma variação elimina a possibilidade de ligação de certos miRNAs, essa mesma variação cria alvos para outros (miR-520f e miR-651) (Rene *et al.*, 2015). É preciso considerar, ainda, que a expressão de HLA-A deve ocorrer de modo constitutivo em praticamente todas as células humanas e, por isso, a regulação dos seus níveis de expressão deve ser finamente regulada independentemente do microambiente celular.

A molécula codificada por esse gene interage tanto com o receptor TCR (Rock e Benacerraf, 1983; Chouaib *et al.*, 1988) quanto com alguns receptores KIR, mas esta última ocorre de um modo alelo-específico (Thielens *et al.*, 2012; Shen *et al.*, 2014). Os alelos de *HLA-A* que possuem o motivo sorológico Bw4 (alelos dos grupos *HLA-A\*03* e *HLA-A\*11*) interagem com o receptor inibitório KIR3DL2 e KIR2DL3 (Jamil e Khakoo, 2011) e alelos dos grupo *HLA-A\*11* também são capaz de interagir com os receptores ativadores KIR2DS2 e KIR2DS4 (Graef *et al.*, 2009; Liu *et al.*, 2014). Essa interação entre KIR e certos alelos de *HLA-A* parece ser a razão pela qual alelos *HLA-A* que carregam o motivo Bw4

estarem associados ao desenvolvimento de hepatite B crônica, em especial na presença de alelos de *HLA-C* do alótipo C2 (discutido adiante). Nesse contexto, KIR2DL3, ao interagir com HLA-A Bw4 estaria inibindo a resposta citotóxica mediadas por células NK (Di Bona *et al.*, 2017).

Uma vez que essas interações alelo-específicas com receptores KIR ficam à mercê da frequência desses alelos de *HLA-A* na população, a molécula de HLA-A possui como principal ligante o TCR de células T CD8<sup>+</sup>, atuando diretamente na ativação de respostas imunes adaptativas. Desse modo, a capacidade de apresentação antigênica das moléculas HLA-A possui um importante papel em infecções virais. Além disso, vários alelos de *HLA-A* associam-se à uma série de contextos patológicos, como doenças autoimunes, doenças infecciosas, neoplasias e o resultado de transplantes (Ogahara *et al.*, 1998; Silva *et al.*, 2009; Vandiedonck *et al.*, 2009; Mosaad *et al.*, 2010).

Estudos que avaliam a afinidade das moléculas de HLA aos peptídeos a serem apresentados (tanto próprios quanto não próprios), mostram a HLA-A interagindo preferencialmente a motivos conservados de proteínas humanas e de vírus que apresentam DNA como material genético (Hertz *et al.*, 2011). HLA-A parece mostrar afinidade especial a trechos conservados de proteínas virais dos *Herpesviridae* e *Adenoviridae*, e assim, sendo fundamental na formação de células T CD8<sup>+</sup> específicas. De fato, epítomos de Herpesvírus Humano 6B (HHV-6B) são capazes de gerar expansão clonal em células T CD8<sup>+</sup> e seus peptídeos são associados aos grupos alélicos *HLA-A\*02* e *HLA-A\*03* (os mais frequentes em indivíduos caucasianos e com altas frequências – superiores a 40% - em várias populações do mundo) (Iampietro *et al.*, 2014).

Em células neoplásicas, várias vias moleculares estão alteradas, e muitas vezes, o mecanismo de escape à imunovigilância se dá pela regulação negativa da expressão do HLA de classe I. Nesse contexto, Herrmann e colaboradores (2004) demonstraram que a glicoproteína HER2 (super-expressa em até 30% dos cânceres mamários) é capaz de reduzir a expressão de HLA de classe I *in vitro*, mas que esse efeito pôde ser revertido após tratamento com IFN- $\gamma$  (Herrmann *et al.*, 2004). Sabe-se que essa regulação é mediada pela via da MAPK, ativada por



HER2, mas o mecanismo exato dessa regulação ainda é desconhecido (Mimura *et al.*, 2013).

Em contraste, no carcinoma hepatocelular, não há redução da expressão de HLA de classe I e ainda mostram níveis elevados de HLA-A, mediada pelo aumento da expressão dos fatores de transcrição IRF-1 e p56, que regulam a expressão de *HLA-A* ao interagirem com os motivos ISRE e Enhancer A. Porém, não foi observada relação entre a quantidade de HLA-A e o tamanho do tumor, sua progressão e relação com a infecção por hepatite B ou C e cirrose (Shen *et al.*, 2009).

O gene *HLA-A* está cercado por genes de classe I não clássicos – *HLA-E*, *HLA-F*, *HLA-G*, além do pseudogene *HLA-H*. Acredita-se que uma possível associação entre esses genes atue na indução de alorreatividade ou de imunotolerância, uma vez que moléculas não clássicas possuem importantes papéis imunomodulatórios (Carlini *et al.*, 2016). Os estudos envolvendo *mismatches* de *HLA-A* em transplantes têm gerado resultados órgão-dependentes e alelo-específicos. Em transplantes hepáticos, a compatibilidade entre HLA-A está associada a melhores resultados com impacto na sobrevivência do enxerto (Muro *et al.*, 2012). Transplantes renais entre receptores A\*02 e doadores A\*68 ou A\*69 (sorotipo A\*28) mostram pouca produção de anticorpos e boas taxas de sobrevivência do enxerto (Dankers *et al.*, 2003). Assim, acredita-se que a associação entre *HLA-A* e outros genes não clássicos possa atuar na indução de alorreatividade ou tolerância ao órgão transplantado (Carlini *et al.*, 2016).

Anteriormente, já foram observados altos níveis de Desequilíbrio de Ligação (LD) entre alelos de *HLA-A* com alelos de *HLA-E*, de modo que alelos desses dois loci se encontram totalmente associados (*HLA-A\*01* e *HLA-E\*01:01*; *HLA-A\*23*, \*24 e \*29 e *HLA-E\*01:03*) (Geraghty *et al.*, 1992). O mesmo também já foi observado entre alelos de *HLA-A* e *HLA-G* (Ober *et al.*, 1996; Kolte *et al.*, 2010). Recentemente, alto LD foi observado entre *HLA-A*, *HLA-H* e *HLA-G* (em especial, alelos de região 3' não traduzida – 3'NT), mas não com *HLA-E*, o que é possivelmente explicado pela distância física entre este *locus* e outros genes (Carlini *et al.*, 2016).

### ***O gene HLA-C***

O gene *HLA-C*, quando comparado aos seus homólogos *-A* e *-B*, apresenta a mesma estrutura molecular e a mesma função primária de apresentação antigênica. Porém, levando os três genes em consideração, ele é o menos variável, mesmo contando com mais de 4.300 alelos, e costuma ser expresso em menores níveis (Mccutcheon *et al.*, 1995). Os mecanismos relacionados com essa menor expressão envolvem a meia vida mais curta do mRNA de *HLA-C*, responsável por menores níveis de tradução das cadeias pesadas (Mccutcheon *et al.*, 1995), a retenção da molécula no RE e sua eventual degradação (Neisig *et al.*, 1998).

Mesmo possuindo a mesma propriedade de apresentação antigênica e ser altamente polimórfico, o *HLA-C* é o único gene HLA clássico a ser amplamente expresso em tecidos placentários. Acredita-se que isso ocorra devido ao seu importante papel como modulador de células NK (King *et al.*, 2000), uma vez que o sucesso da implantação depende de uma resposta inflamatória leve (mediada pelas células dNK, pouco citotóxicas) que logo deve ser controlada e inibida (Male *et al.*, 2011). Juntamente com a *HLA-C*, as moléculas *HLA-E* e *HLA-G* (ambas com propriedades de imunomodulação e inibição de célula NK) constituem a expressão de MHC de classe I na placenta. Assim, esses três *loci* estão associados ao sucesso da gestação e à manutenção da tolerância materno-fetal (Varla-Leftherioti, 2004; Tripathi *et al.*, 2006; Betz, 2012; Klitkou *et al.*, 2015).

Uma vez que durante a gestação, as principais células de origem linfóide presentes no útero são as células NK decíduais (dNK), as moléculas *HLA-C*, *HLA-E* e *HLA-G*, possuem um papel fundamental na orquestração dos eventos necessários ao sucesso da implantação e da não rejeição materna ao feto (Parham e Guethlein, 2010). As moléculas *HLA-C* que interagem com os receptores KIR maternos podem ser divididas em dois grupos, devido a diferenças em aminoácidos fundamentais à interação – *HLA-C1*, que contém uma asparagina na posição 80 (ou 80N) e *HLA-C2*, que apresentam uma lisina na posição 80 (ou 80K) (The IPD-IMGT/HLA database, disponível em

<https://www.ebi.ac.uk/ipd/kir/ligand.html>). As moléculas HLA-C1 interagem com os receptores KIR2DL2, KIR2DL3 e KIR2DS2, sendo os dois primeiros capazes de inibir a resposta mediada por células NK, e o último capaz de ativá-las. As moléculas C2, por sua vez, são específicas para os receptores KIR2DL1, KIR2DS1 e KIR2DS2 (Vilches e Parham, 2002; Deng e Mariuzza, 2006). Além de atuar em receptores de células NK, a molécula HLA-C também desempenha um papel importante na apresentação antigênica para células T CD8<sup>+</sup> e na deflagração de respostas imunes celulares.

Por atuar tanto na resposta imune inata quanto na adaptativa e por seu papel de imunomodulação, o gene *HLA-C* tem sido bastante estudado em uma série de contextos patológicos. Entre eles podemos destacar as complicações gestacionais, como eclampsia e abortos de recorrência e espontâneos (Flores *et al.*, 2007; Yu *et al.*, 2014), e as infecções virais, com atenção especial dada ao vírus HIV (do inglês, *Human Immunodeficiency Virus*) (Qi *et al.*, 2006; Kulkarni *et al.*, 2013; Apps *et al.*, 2016). Algumas variantes de *HLA-C* têm sido associadas à infecção por HIV e desenvolvimento da doença, entre elas, uma deleção na região 3'NT que influencia o alvo do miRNA miR-148a (Kulkarni *et al.*, 2011). Níveis mais elevados de HLA-C se associam a uma maior atividade de linfócitos T citotóxicos, mas ao mesmo tempo uma maior inibição de células NK (a depender do haplótipo do hospedeiro) (Alcami e Koszinowski, 2000).

Nesse contexto, um dos mecanismos de escape usado pelo vírus HIV é a redução dos níveis de HLA-A e HLA-B (mas não de HLA-C) pela proteína Nef. Desse modo, células T CD8<sup>+</sup> deixam de ser ativadas ao mesmo tempo que células NK seriam moduladas. Em contrapartida, a proteína Vpu do HIV é capaz de reduzir a expressão de HLA-C, mantendo sob controle a resposta imune mediada por células T citotóxicas (Apps *et al.*, 2016). Altos níveis de HLA-C em infecção por HIV estão associados a maiores respostas de células T ativadas por esta molécula, porém, observam-se também maiores taxas de mutação viral, como uma tentativa do vírus de escapar à vigilância imunológica (Apps *et al.*, 2013). Assim, a regulação dos níveis de HLA-C (e outras moléculas HLA) é muito importante na determinação do resultado de uma infecção.

No que diz respeito à importância de HLA-C em transplantes, não é recomendável que o enxerto possua incompatibilidade entre nenhuma das moléculas comparadas, mas uma vez que entre os três loci clássicos o *HLA-C* é o menos expresso, alguns autores demonstram a possibilidade de transplantes bem-sucedidos mesmo quando as moléculas de HLA-C são incompatíveis (Petersdorf *et al.*, 2014; Thus *et al.*, 2014).

### ***O gene HLA-G***

O gene *HLA-G* é um HLA de classe I não clássico, pouco variável e com expressão restrita a poucos tecidos, como córnea, pâncreas e tecidos fetais (Ellis *et al.*, 1986; Le Discorde *et al.*, 2003). Até o momento, somente 61 alelos, responsáveis por codificar 19 variantes proteicas foram descritos (IPD-IMGT), porém apenas 5 delas são encontradas frequentemente (HLA-G\*01:01, G\*01:02, G\*01:03, G\*01:04 e G\*01:06) (Castelli, Ramalho, *et al.*, 2014; Castelli *et al.*, 2017). A presença de HLA-G foi primeiramente descrita no trofoblasto extraviloso e em linhagens celulares de coriocarcinoma (Ellis *et al.*, 1986). Sua estrutura molecular é semelhante à das moléculas de HLA clássicas, mas com uma redução da cauda citoplasmática devido à presença de códon de parada prematuro localizado no éxon 6 (Geraghty *et al.*, 1987).

Diferentemente dos HLA clássicos, a apresentação antigênica é uma função secundária das moléculas não-clássicas. Sua função primária é a imunomodulação, geralmente mediada por inibição da atividade de células T CD8<sup>+</sup>, via receptores ILT2 e ILT4 (receptores semelhantes à imunoglobulina transcritos 2 e 4) e de células NK. A molécula de HLA-G, em particular, interage com o receptor KIR2DL4 de células NK. Este, apesar da nomenclatura indicar ser um receptor inibitório (cauda citoplasmática longa devido à presença de motivos de tirosina inibitórios), sua interação com HLA-G resulta na indução da produção de citocinas, mas inibe respostas de citotoxicidade (Khalil-Daher *et al.*, 1999; Shiroishi *et al.*, 2003; Favier *et al.*, 2010).

Devido seu papel na imunomodulação e sua presença em tecidos fetais, o

HLA-G possui um papel importante no desenvolvimento da tolerância materno-fetal, na aceitação de tecidos transplantados e ainda no escape à vigilância imunológica por tumores (Adrian Cabestre *et al.*, 1999; Rouas-Freiss *et al.*, 1999; Carosella *et al.*, 2003).

Durante a gravidez, HLA-C, HLA-E e HLA-G participam dos fenômenos imunológicos gestacionais, da implantação ao parto (Gregori *et al.*, 2015; Hackmon *et al.*, 2017). Isso ocorre na interface materno-fetal, onde os tecidos fetais, mais especificamente o trofoblasto, estão em contato direto com leucócitos, estroma uterino e células endoteliais maternas que compreendem a decídua. A expressão de HLA-G na superfície do trofoblasto extraviloso ocorre primariamente durante o primeiro trimestre de gestação e, após este período, ela se limita ao citoplasma dessas células. No entanto, ele permanece presente nas membranas fetais amniótica e coriônica durante toda a gravidez (Hackmon *et al.*, 2017).

O *HLA-G* é o gene HLA cujas regiões regulatórias são mais bem estudadas quanto a estrutura e função (Castelli *et al.*, 2010; Donadi *et al.*, 2011; Sabbagh *et al.*, 2013; Castelli, Veiga-Castelli, *et al.*, 2014). Sua região 3'NT está associada à manutenção da estabilidade da molécula de mRNA (Rousseau *et al.*, 2003) e à regulação pós-transcricional dos níveis de HLA-G por meio da ligação de miRNAs (Tan *et al.*, 2007; Manaster *et al.*, 2012; Porto *et al.*, 2015).

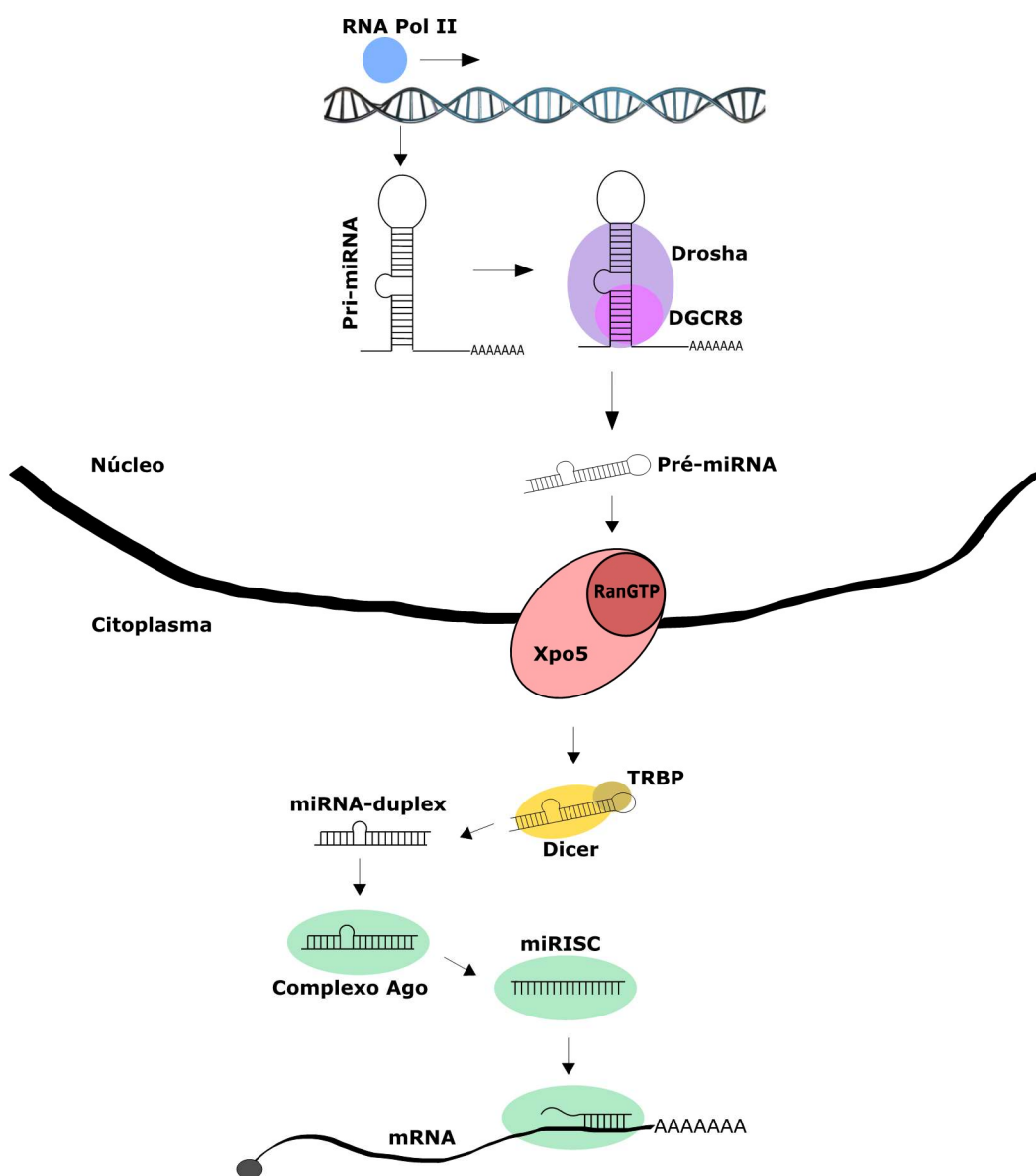
### **microRNAs e seu papel na regulação pós-transcricional do HLA de classe I**

Os miRNAs são pequenos RNA não codificantes (21-24 nucleotídeos), que atuam no mRNA alvo a partir da complementaridade de bases com sua sequência alvo. Geralmente eles são transcritos a partir de genes próprios, mas cerca de 30% deles são derivados de regiões intrônicas ou até mesmo de éxons de outros genes (Bartel, 2009). Essas pequenas moléculas de RNA já foram descritas em uma grande variedade de organismos – humanos, algas, insetos, plantas, mamíferos, protozoários, entre outros (Griffiths-Jones *et al.*, 2006; Kozomara e Griffiths-Jones, 2014). Sugere-se que miRNAs cuja transcrição ocorra dentro de “genes

hospedeiros” podem ter sua função relacionada à regulação desses mesmos genes ou genes relacionados, mas os resultados até o momento são conflitantes (Barik, 2008; Bell *et al.*, 2010).

Os miRNAs são transcritos pela RNA polimerase II e seus transcritos primários, os pri-miRNAs, podem chegar a 1kb formando estruturas de grampo. Eles podem gerar uma única molécula ou um *cluster* de miRNAs. O pri-miRNA é clivado pelo complexo proteico Drosha, com auxílio da proteína DGCR8, responsável pela ligação ao pri-miRNA e ativação da RNase III contida na Drosha. Assim, forma-se uma segunda estrutura em grampo, porém menor, contendo 70-100 nucleotídeos, o pré-miRNA. Este é transportado para o citoplasma pela proteína Exportina5 (Xpo5), que forma um complexo de transporte quando complexada à proteína nuclear Ran GTPase (RanGTP). No citoplasma, o pré-miRNA sofre uma segunda clivagem mediada pela enzima Dicer, e finalmente gera o miRNA maduro em fita dupla (Shukla *et al.*, 2011; Dong *et al.*, 2013). A precisão de clivagem da Dicer é aumentada na presença de um cofator, proteína TRBP (*Transactivation-responsive RNA Binding Protein*), que também atua na estabilização da Dicer e na determinação da fita de miRNA a interagir com a proteína Argonata, um dos componentes do RISC (Wilson *et al.*, 2015). O miRNA maduro (~ 22nt) e funcional deve ainda se soltar da sua fita complementar e se juntar ao Complexo de Silenciamento Induzido por RNA, ou RISC. O complexo miRNA/RISC interage geralmente com as regiões 3'NT do mRNA alvo, guiado pela sequência do miRNA (Dong *et al.*, 2013; Daugaard e Hansen, 2017). Um esquema da biogênese dos miRNAs está representado na Figura 5.

A interação miRNA/mRNA se baseia na complementaridade de bases entre as duas moléculas. Esta complementaridade pode ser perfeita (comumente observada em plantas) ou imperfeita. Desse modo, a sequência de RNA carregada no miRISC serve de “guia” até o mRNA alvo, no qual os resultados da interação incluem repressão da maquinaria de tradução em diferentes estágios do processo de tradução, deadenilação do mRNA e sua degradação (Bartel, 2009; Huntzinger e Izaurralde, 2011; Eichhorn *et al.*, 2014).



**Figura 5-** Esquema da biogênese dos microRNAs.

Fonte: Adaptado de Daugaard & Hansen, 2017

Em animais, o pareamento imperfeito entre miRNAs e seus alvos é suficiente para desempenhar seu papel na regulação pós-transcricional. Até pouco tempo atrás, por conta dessas diferenças entre níveis de pareamento, acreditava-se que o controle pós-transcricional por clivagem e degradação do mRNA via miRISC era um mecanismo exclusivo de plantas, enquanto em animais os miRNAs atuavam somente interferindo na maquinaria de tradução. Hoje, sabe-se

que essas duas vias de controle pós-transcricional mediada por miRNAs atuam tanto em animais quanto plantas (Huntzinger e Izaurralde, 2011). Uma das vias de inibição da tradução via miRISC tem como alvo o processo de reconhecimento do “Cap” presente na ponta 5’ do mRNA, impedindo, assim o início da tradução (Mathonnet *et al.*, 2007). A proteína Argonauta 2 (Ago2), se liga diretamente ao *cap* e reprime o estágio inicial da tradução. A degradação do mRNA costuma ocorrer por clivagem direta por meio da Ago2 ou por desestabilização da molécula via deadenilação que tem como consequência, a perda do *cap* e a degradação do mRNA (Behm-Ansmant *et al.*, 2006)

As interações miRNA/mRNA ocorrem preferencialmente na região 3’NT dos mRNAs, mas ocasionalmente elas ocorrem na região codificadora (Gaidatzis *et al.*, 2007). Nas regiões 3’NT, os miRNAs se ligam preferencialmente nas extremidades iniciais e terminais (Gaidatzis *et al.*, 2007; Didiano e Hobert, 2008), mas uma vez que a porção final dos mRNAs são compostas por sequências pouco complexas e ricas em adenina, essas interações são escassas e geralmente de baixa afinidade (Majoros e Ohler, 2007; Didiano e Hobert, 2008).

Os resultados da interação de um miRNA com seu alvo podem incluir alterações na homeostase celular, o que torna os miRNAs importantes marcadores de condições normais ou patológicas e suas funções como supressores de tumor ou “oncoMIRs” (Dong *et al.*, 2013). Sua função também já foi descrita em uma série de processos biológicos, entre eles proliferação, morte celular, desenvolvimento, diferenciação celular, entre outros (Pasquinelli e Ruvkun, 2002; Fiorenza e Barco, 2016; Roy, 2016).

Apesar do tamanho pequeno (miRNAs são os menores carreadores de informação de regulação genética tanto em animais quanto em plantas), os miRNAs representam as sequências de DNA mais conservadas já conhecidas, de modo que a família de miRNAs 854 (miR-854a-e), por exemplo, pode ser encontrada em plantas (*Arabidopsis thaliana*), no nematódeo *Caenorhabditis elegans* e em mamíferos como o camundongo e humanos (Arteaga-Vazquez *et al.*, 2006). Além da sequência, a estrutura de diversos miRNAs, incluindo *loops*, bolhas de RNA não pareado e estrutura secundária das moléculas mostram sinais



de conservação. A região mais conservada dos miRNAs é a porção final da ponta 5' (nucleotídeos 2 a 8), conhecida como *seed*. A *seed* é fundamental ao funcionamento do miRISC, de modo que o não pareamento dessa região ao mRNA alvo é suficiente para inativar os mecanismos de silenciamento, enquanto o mesmo fato não é observado para a ponta 3' do miRNA (Doench e Sharp, 2004).

No que diz respeito às regiões 3'NTs de genes HLA, sabe-se que as regiões 3'NT dos genes *HLA-A* (Rene *et al.*, 2015), *HLA-C* (Kulkarni *et al.*, 2011) e *HLA-G* (Tan *et al.*, 2007; Porto *et al.*, 2015) são alvos de miRNAs. Esses miRNAs são capazes de alterar seus níveis proteicos. Vale ressaltar que um mesmo miRNA (miR-148a-3p) foi descrito como sendo capaz de regular tanto os níveis de HLA-G e HLA-C em nível pós transcricional e, para este último, apresenta um papel importante na progressão do HIV (Kulkarni *et al.*, 2011; Manaster *et al.*, 2012; Celsi *et al.*, 2013).

### **Métodos de detecção de alvos de microRNAs e uso do miRhunt**

A detecção de alvos de miRNAs e os mecanismos que governam a interação miRNA/mRNA tornaram-se importantes áreas de estudo nas últimas décadas. Com os avanços computacionais, estudos envolvendo as características estruturais e de conservação das sequências alvos de miRNAs contribuíram bastante para o entendimento da regulação mediada por miRNAs.

Os algoritmos de predição de alvos de miRNAs usam estratégias e parâmetros variados. No entanto, um fator em comum a todos os algoritmos é a necessidade de um banco de dados com as sequências de miRNAs, obtidas em bancos de dados especializados, como o miRbase (disponível em <http://mirbase.org/>) e a sequência de mRNA alvo, geralmente uma 3'NT (Riffo-Campos *et al.*, 2016). Uma vez que o pareamento da região *seed* é fundamental à interação miRNA/mRNA, certos algoritmos, como RNAhybrid (Rehmsmeier *et al.*, 2004; Kruger e Rehmsmeier, 2006), miRanda (Enright *et al.*, 2003) e PITA (Kertesz *et al.*, 2007), usam o pareamento desse segmento do miRNA à sequência alvo de mRNA em seus cálculos. Além da *seed*, a acessibilidade ao alvo,

influenciada pela estrutura secundária do mRNA também é levada em consideração. Isso é observado em algoritmos como PITA (Kertesz *et al.*, 2007), IntaRNA (Busch *et al.*, 2008) e Vienna RNA (Lorenz *et al.*, 2011). De modo geral, os algoritmos costumam usar combinações de parâmetros (conservação de local de ligação, complementaridade à região *seed*, contribuição de regiões próximas à ligação e sua estrutura, termodinâmica, entre outros) para gerar seus resultados (Wagner *et al.*, 2014).

A ligação entre miRNA e mRNA ocorre via complementaridade de bases. No entanto, quando somente este parâmetro é avaliado, observa-se um número excessivo de falsos positivos (Bartel, 2009). Assim, os algoritmos de predição de alvos de miRNAs geralmente levam em consideração os princípios da termodinâmica que governam a interação miRNA/mRNA. Nesse contexto, eles calculam a Energia Livre Mínima (MFE, do inglês *Minimum Free Energy*) para determinar o quão forte é essa interação entre miRNA e alvo. O MFE tem como unidade o “kcal/mol” e é expresso em um valor negativo que quanto menor, melhor é a interação miRNA/mRNA (Doench e Sharp, 2004; Riffo-Campos *et al.*, 2016). Quando a acessibilidade é levada em consideração, calcula-se o “delta G” ( $\Delta G$ ), ou a diferença entre a energia necessária para romper a ligação de bases que formam uma estrutura secundária do mRNA e a energia gerada na ligação miRNA/mRNA (Kertesz *et al.*, 2007; Busch *et al.*, 2008).

Uma vez que vários algoritmos com abordagens distintas continuam a ser desenvolvidos, a seleção da melhor estratégia de predição torna-se um grande desafio. Assim, sugere-se o uso de mais de uma ferramenta, seguida da integração e comparação dos resultados. Outra dificuldade encontrada na predição *in silico* de alvos de miRNAs é determinar quais valores obtidos pelos algoritmos são relevantes – quais valores de energia seriam compatíveis à uma interação miRNA/mRNA funcional. Nesse contexto, afim de sanar a dificuldade em determinar valores de MFE ou  $\Delta G$  que sirvam de *cutoff*, um trabalho anterior do nosso grupo criou um programa que automatiza a corrida de três algoritmos distintos – RNAhybrid (Kruger e Rehmsmeier, 2006), miRanda (Enright *et al.*, 2003) e IntaRNA (Busch *et al.*, 2008) – chamado miRhunt (disponível em

[www.castelli-lab.net/apps/mirhunt](http://www.castelli-lab.net/apps/mirhunt)).

O algoritmo miRanda identifica locais de complementaridade entre miRNA e mRNA e identifica possíveis sítios de ligação. A ligação da região 5' do miRNA (*seed*) resulta em escores maiores do que aqueles cuja complementaridade ocorre somente na sua região 3'. Assim, quanto maior a complementaridade entre *seed* e alvo de ligação, melhores os resultados. Estes ainda levam em consideração estabilidade termodinâmica da ligação miRNA/mRNA e conservação evolutiva da sequência alvo (Enright *et al.*, 2003). Já o RNAhybrid permite ao usuário determinar o tamanho da *seed* a ser considerado, o uso ou não da oscilação G:U (comum em ligações RNA/RNA) e a formação de *loops* na sequência alvo, permitindo a complementaridade com regiões menos próximas à *seed*. Uma grande vantagem do seu uso é a velocidade de análise, uma vez que somente miRNAs em que há pareamento da *seed* são avaliados (mas ele permite mudanças no tamanho e na posição da *seed*) (Kruger e Rehmsmeier, 2006). O IntaRNA, por sua vez, acrescenta aos cálculos de estabilidade da interação (hibridização miRNA/mRNA) a energia necessária para tornar essa interação acessível (energia necessária para tornar o sítio alvo em uma região de fita simples), considerando que mRNAs assumem conformações secundárias. Assim como o RNAhybrid ele incorpora parâmetros relacionados à região *seed*, como o número de bases que devem ser pareadas perfeitamente e o número máximo de bases não pareadas (Busch *et al.*, 2008).

Além de automatizar as corridas dos três algoritmos, miRhunt compara seus resultados para a sequência alvo com os obtidos com milhares de sequências de 3'UTR humanas, para calcular um sistema de escores que facilita o *ranking* dos melhores resultados com base não só na afinidade ou força de ligação de um miRNA a uma região 3'NT, mas também em sua especificidade a ela (Porto *et al.*, 2015).

Para isso, o miRhunt conta com um banco de dados onde se encontram os resultados obtidos com cada um dos três algoritmos para todas as sequências de 3'NT humanas descritas até 2015 contra todos os miRNAs maduros descritos no miRbase *release* 21. Assim, ele compara a energia de ligação observada entre um

miRNA e o mRNA alvo à energia desse mesmo miRNA quando interagindo com as outras 3'NTs humanas. Para se calcular a afinidade (estabilidade), escore E, de um certo miRNA ao alvo em questão, o menor valor de MFE ou  $\Delta G$  encontrado na matriz (de cada algoritmo) é levado em consideração. Desta forma, o score E seria a razão entre o MFE ou  $\Delta G$  da interação em questão e o menor valor encontrado em toda matriz ( $E_{\text{algoritmo}} = \text{MFE}_{\text{miRNA}}$  ou  $\Delta G_{\text{miRNA}}/\text{MFE}_{\text{mínimo}}$ ). O escore E final é calculado a partir da soma dos escores dos três algoritmos ( $E_{\text{final}} = E_{\text{hybrid}} + E_{\text{miranda}} + E_{\text{intaRNA}}$ ). Quanto maior este valor, maior a probabilidade deste miRNA interagir fortemente com a região 3'NT estudada.

Para avaliar a especificidade de um miRNA à região 3'NT em questão, o MFE ou  $\Delta G$  desse miRNA é comparado aos valores deste mesmo miRNA distribuídos ao longo da matriz. O número de valores de energia menores que o encontrado para a sequência alvo é dividido pelo número total de 3'NTs disponíveis na matriz (cerca de 27 mil valores), gerando um valor (P) que determina a probabilidade empírica desse miRNA se ligar à sequência alvo estudada sem muita interferência em outros mRNAs. Calculando-se  $1 - p$ , obtém-se o escore S de especificidade e a soma do valor S dos três algoritmos gera o  $S_{\text{final}}$ . Quanto maior este valor, maior a especificidade de um dado miRNA à sequência alvo estudada (Porto *et al.*, 2015).

Esse sistema de escores minimiza o uso arbitrário de valores de MFE ou  $\Delta G$  na determinação de quais miRNAs são relevantes quanto ao seu potencial de regular um mRNA alvo. Essa estratégia é particularmente útil quando observamos que miRNAs cujos valores de MFE não são particularmente baixos, mas que comprovadamente são capazes de regular um determinado alvo. Isso é observado, por exemplo, na interação entre o miRNA *miR-148a-3p* e a região 3'NT do gene *HLA-G*. A afinidade observada entre eles é relativamente baixa devido ao alto valor de MFE dessa interação, porém, raras são as sequências de 3'NT humanas que produziriam um valor de especificidade (S) maior do que o observado para *HLA-G* e este miRNA em particular (Porto *et al.*, 2015).

Assim, miRNAs que apresentam valores de MFE ou  $\Delta G$  altos que seriam excluídos em um estudo baseado somente nesses valores, podem ser considerados

pela estratégia aqui proposta caso apresentem altos valores de especificidade. Além disso, miRNAs que apresentam alta especificidade a uma determinada sequência alvo, apresentariam pouca interferência em outras moléculas de mRNA. Contudo, estudos funcionais ainda se fazem necessários para comprovar se os miRNAs detectados pelo miRhunt como específicos ou de alta afinidade são realmente capazes de regular o mRNA alvo.

No entanto, é preciso salientar que o uso de mais de um algoritmo de predição de alvos de miRNAs não exclui a presença de falsos positivos. É preciso considerar que até o momento não se observa uma correlação entre os valores de MFE ou  $\Delta G$  e a capacidade funcional de controle pós-transcricional real de um miRNA. Em contrapartida, estes valores são boas medidas da afinidade entre um determinado miRNA e uma possível sequência alvo. Quanto maior essa afinidade, maior a possibilidade deste miRNA atuar no controle pós-transcricional do mRNA alvo.

A predição computacional de miRNAs que se ligam a um determinado mRNA alvo é necessária para nortear estudos sobre a regulação pós-transcricional. Uma vez que mais de 2000 sequências de miRNAs maduros estão disponíveis nos bancos de dados, seria totalmente inviável testá-los funcionalmente um a um. O uso de ferramentas computacionais vem então para filtrar miRNAs com alto potencial de regulação de um determinado alvo, especialmente quando múltiplos algoritmos são usados. Ao mesmo tempo, o principal obstáculo no estudo de predição de alvos de miRNAs é a falta de validação experimental das interações putativas entre miRNA e mRNA preditas pelos métodos computacionais. Nesse contexto, não existe nenhum método de predição de alvos de miRNAs que possa ser considerado “padrão-ouro” (Wagner *et al.*, 2014).



Referências

## REFERÊNCIAS

ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Imunologia Celular e Molecular** 8ª edição. Rio de Janeiro: Elsevier, 2015. 552.

ACKERMAN, A. L.; GIODINI, A.; CRESSWELL, P. A role for the endoplasmic reticulum protein retrotranslocation machinery during crosspresentation by dendritic cells. **Immunity**, v. 25, n. 4, p. 607-17, Oct 2006.

ADIKO, A. C. et al. Intracellular Transport Routes for MHC I and Their Relevance for Antigen Cross-Presentation. **Front Immunol**, v. 6, p. 335, 2015.

ADRIAN CABESTRE, F. et al. HLA-G expression in human melanoma cells: protection from NK cytotoxicity. **J Reprod Immunol**, v. 43, n. 2, p. 183-93, Jul 1999.

ALCAMI, A.; KOSZINOWSKI, U. H. Viral mechanisms of immune evasion. **Trends Microbiol**, v. 8, n. 9, p. 410-8, Sep 2000.

AMIGORENA, S. Antigen presentation: from cell biology to physiology. **Immunol Rev**, v. 272, n. 1, p. 5-7, Jul 2016.

APPS, R. et al. HIV-1 Vpu Mediates HLA-C Downregulation. **Cell Host Microbe**, v. 19, n. 5, p. 686-95, May 11 2016.

APPS, R. et al. Influence of HLA-C expression level on HIV control. **Science**, v. 340, n. 6128, p. 87-91, Apr 05 2013.

ARTEAGA-VAZQUEZ, M.; CABALLERO-PEREZ, J.; VIELLE-CALZADA, J. P. A family of microRNAs present in plants and animals. **Plant Cell**, v. 18, n. 12, p. 3355-69, Dec 2006.

BARIK, S. An intronic microRNA silences genes that are functionally antagonistic to its host gene. **Nucleic Acids Res**, v. 36, n. 16, p. 5232-41, Sep 2008.

BARTEL, D. P. MicroRNAs: target recognition and regulatory functions. **Cell**, v. 136, n. 2, p. 215-33, Jan 23 2009.

BECK, S. et al. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. **Nature**, v. 401, n. 6756, p. 921-3, Oct 28 1999.

BEHM-ANSMANT, I. et al. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. **Genes Dev**, v. 20, n. 14, p. 1885-98, Jul 15 2006.

BELL, M. L.; BUVOLI, M.; LEINWAND, L. A. Uncoupling of expression of an intronic microRNA and its myosin host gene by exon skipping. **Mol Cell Biol**, v. 30, n. 8, p. 1937-45, Apr 2010.

BETZ, A. G. Immunology: Tolerating pregnancy. **Nature**, v. 490, n. 7418, p. 47-8, Oct 4 2012.

BROWNE, S. K. et al. Differential IFN-gamma stimulation of HLA-A gene expression through CRM-1-dependent nuclear RNA export. **J Immunol**, v. 177, n. 12, p. 8612-9, Dec 15 2006.

BRUTKIEWICZ, R. R. Cell Signaling Pathways That Regulate Antigen Presentation. **J Immunol**, v. 197, n. 8, p. 2971-2979, Oct 15 2016.

BUSCH, A.; RICHTER, A. S.; BACKOFEN, R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. **Bioinformatics**, v. 24, n. 24, p. 2849-56, Dec 15 2008.

CAPITTINI, C.; MARTINETTI, M.; CUCCIA, M. MHC variation, mate choice and natural selection: the scent of evolution. **Riv Biol**, v. 101, n. 3, p. 463-80, Sep-Dec 2008.

CARLINI, F. et al. Association of HLA-A and Non-Classical HLA Class I Alleles. **PLoS One**, v. 11, n. 10, p. e0163570, 2016.

CAROSELLA, E. D. et al. HLA-G molecules: from maternal-fetal tolerance to tissue acceptance. **Adv Immunol**, v. 81, p. 199-252, 2003.

CASTELLI, E. C. et al. HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographically distinct population samples of Brazil and Cyprus. **Mol Immunol**, v. 83, p. 115-126, Mar 2017.

CASTELLI, E. C. et al. The genetic structure of 3'untranslated region of the HLA-G gene: polymorphisms and haplotypes. **Genes Immun**, v. 11, n. 2, p. 134-41, Mar 2010.

CASTELLI, E. C. et al. Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. **Front Immunol**, v. 5, p. 476, 2014.

CASTELLI, E. C. et al. Transcriptional and posttranscriptional regulations of the HLA-G gene. **J Immunol Res**, v. 2014, p. 734068, 2014.

CELSI, F. et al. HLA-G/C, miRNAs, and their role in HIV infection and replication. **Biomed Res Int**, v. 2013, p. 693643, 2013.

CHEN, E. et al. Gamma interferon and 5-azacytidine cause transcriptional elevation of class I major histocompatibility complex gene expression in K562 leukemia cells in the absence of differentiation. **Mol Cell Biol**, v. 6, n. 5, p. 1698-705, May 1986.

CHOO, S. Y. The HLA system: genetics, immunology, clinical testing, and clinical implications. **Yonsei Med J**, v. 48, n. 1, p. 11-23, Feb 28 2007.

CHOUAIB, S. et al. Allogeneic T cell activation triggering by MHC class I antigens. **J Immunol**, v. 141, n. 2, p. 423-9, Jul 15 1988.



- DANKERS, M. K. et al. Differential immunogenicity of HLA mismatches: HLA-A2 versus HLA-A28. **Transplantation**, v. 75, n. 3, p. 418-20, Feb 15 2003.
- DAUGAARD, I.; HANSEN, T. B. Biogenesis and Function of Ago-Associated RNAs. **Trends Genet**, v. 33, n. 3, p. 208-219, Mar 2017.
- DENG, L.; MARIUZZA, R. A. Structural basis for recognition of MHC and MHC-like ligands by natural killer cell receptors. **Semin Immunol**, v. 18, n. 3, p. 159-66, Jun 2006.
- DI BONA, D. et al. KIR2DL3 and the KIR ligand groups HLA-A-Bw4 and HLA-C2 predict the outcome of hepatitis B virus infection. **J Viral Hepat**, Feb 17 2017.
- DIDIANO, D.; HOBERT, O. Molecular architecture of a miRNA-regulated 3' UTR. **RNA**, v. 14, n. 7, p. 1297-317, Jul 2008.
- DOENCH, J. G.; SHARP, P. A. Specificity of microRNA target selection in translational repression. **Genes Dev**, v. 18, n. 5, p. 504-11, Mar 01 2004.
- DONADI, E. A. et al. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cell Mol Life Sci**, v. 68, n. 3, p. 369-95, Feb 2011.
- DONG, H. et al. MicroRNA: function, detection, and bioanalysis. **Chem Rev**, v. 113, n. 8, p. 6207-33, Aug 14 2013.
- DUQUESNOY, R. J.; MARRARI, M. Detection of antibodies against HLA-C epitopes in patients with rejected kidney transplants. **Transpl Immunol**, v. 24, n. 3, p. 164-71, Apr 15 2011.
- EDWARDS, S. V.; HEDRICK, P. W. Evolution and ecology of MHC molecules: from genomics to sexual selection. **Trends Ecol Evol**, v. 13, n. 8, p. 305-11, Aug 01 1998.
- EHLERS, A. et al. MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes. **Genome Res**, v. 10, n. 12, p. 1968-78, Dec 2000.
- EICHHORN, S. W. et al. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. **Mol Cell**, v. 56, n. 1, p. 104-15, Oct 02 2014.
- ELLIS, S. A. et al. Evidence for a novel HLA antigen found on human extravillous trophoblast and a choriocarcinoma cell line. **Immunology**, v. 59, p. 595-601, Jul 1986 1986.
- ENRIGHT, A. J. et al. MicroRNA targets in Drosophila. **Genome Biol**, v. 5, n. 1, p. R1, 2003.
- FAROUDI, M. et al. Lytic versus stimulatory synapse in cytotoxic T lymphocyte/target cell interaction: manifestation of a dual activation threshold. **Proc Natl Acad Sci U S A**, v. 100, n. 24, p. 14145-50, Nov 25 2003.

FAVIER, B. et al. ILT2/HLA-G interaction impairs NK-cell functions through the inhibition of the late but not the early events of the NK-cell activating synapse. **FASEB J**, v. 24, n. 3, p. 689-99, Mar 2010.

FIORENZA, A.; BARCO, A. Role of Dicer and the miRNA system in neuronal plasticity and brain function. **Neurobiol Learn Mem**, v. 135, p. 3-12, Nov 2016.

FLORES, A. C. et al. KIR receptors and HLA-C in the maintenance of pregnancy. **Tissue Antigens**, v. 69 Suppl 1, p. 112-3, Apr 2007.

GAIDATZIS, D. et al. Inference of miRNA targets using evolutionary conservation and pathway analysis. **BMC Bioinformatics**, v. 8, p. 69, Mar 1 2007.

GEORGOPOULOS, N. T.; PROFFITT, J. L.; BLAIR, G. E. Transcriptional regulation of the major histocompatibility complex (MHC) class I heavy chain, TAP1 and LMP2 genes by the human papillomavirus (HPV) type 6b, 16 and 18 E7 oncoproteins. **Oncogene**, v. 19, n. 42, p. 4930-5, Oct 5 2000.

GERAGHTY, D. E.; KOLLER, B. H.; ORR, H. T. A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. **Proc Natl Acad Sci U S A**, v. 84, n. 24, p. 9145-9, Dec 1987.

GERAGHTY, D. E. et al. Polymorphism at the HLA-E locus predates most HLA-A and -B polymorphism. **Hum Immunol**, v. 33, n. 3, p. 174-84, Mar 1992.

GERSTBERGER, S.; HAFNER, M.; TUSCHL, T. A census of human RNA-binding proteins. **Nat Rev Genet**, v. 15, n. 12, p. 829-45, Dec 2014.

GOBIN, S. J.; VAN DEN ELSEN, P. J. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F, and HLA-G. **Hum Immunol**, v. 61, n. 11, p. 1102-7, Nov 2000.

GRAEF, T. et al. KIR2DS4 is a product of gene conversion with KIR3DL2 that introduced specificity for HLA-A\*11 while diminishing avidity for HLA-C. **J Exp Med**, v. 206, n. 11, p. 2557-72, Oct 26 2009.

GREGORI, S. et al. HLA-G Orchestrates the Early Interaction of Human Trophoblasts with the Maternal Niche. **Front Immunol**, v. 6, p. 128, 2015.

GRIFFITHS-JONES, S. et al. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic Acids Res**, v. 34, n. Database issue, p. D140-4, Jan 01 2006.

GROTZKE, J. E. et al. The ongoing saga of the mechanism(s) of MHC class I-restricted cross-presentation. **Curr Opin Immunol**, v. 46, p. 89-96, May 18 2017.

GUETHLEIN, L. A. et al. Co-evolution of MHC class I and variable NK cell receptors in placental mammals. **Immunol Rev**, v. 267, n. 1, p. 259-82, Sep 2015.

HACKMON, R. et al. Definitive class I human leukocyte antigen expression in gestational placentation: HLA-F, HLA-E, HLA-C, and HLA-G in extravillous trophoblast invasion on placentation, pregnancy, and parturition. **Am J Reprod Immunol**, v. 77, n. 6, Jun 2017.

HAFEZQORANI, S. et al. Modeling the combined effect of RNA-binding proteins and microRNAs in post-transcriptional regulation. **Nucleic Acids Res**, v. 44, n. 9, p. e83, May 19 2016.

HANVESAKUL, R. et al. Indirect recognition of T-cell epitopes derived from the alpha 3 and transmembrane domain of HLA-A2. **Am J Transplant**, v. 7, n. 5, p. 1148-57, May 2007.

HERRMANN, F. et al. HER-2/neu-mediated regulation of components of the MHC class I antigen-processing pathway. **Cancer Res**, v. 64, n. 1, p. 215-20, Jan 01 2004.

HERTZ, T. et al. Mapping the landscape of host-pathogen coevolution: HLA class I binding and its relationship with evolutionary conservation in human and viral proteins. **J Virol**, v. 85, n. 3, p. 1310-21, Feb 2011.

HOWELL, W. M.; CARTER, V.; CLARK, B. The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. **J Clin Pathol**, v. 63, n. 5, p. 387-90, May 2010.

HUNTZINGER, E.; IZAURRALDE, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. **Nat Rev Genet**, v. 12, n. 2, p. 99-110, Feb 2011.

IAMPIETRO, M. et al. Human herpesvirus 6B immediate-early I protein contains functional HLA-A\*02, HLA-A\*03, and HLA-B\*07 class I restricted CD8(+) T-cell epitopes. **Eur J Immunol**, v. 44, n. 12, p. 3573-84, Dec 2014.

INABA, K.; INABA, M. Antigen recognition and presentation by dendritic cells. **Int J Hematol**, v. 81, n. 3, p. 181-7, Apr 2005.

JAMIL, K. M.; KHAKOO, S. I. KIR/HLA interactions and pathogen immunity. **J Biomed Biotechnol**, v. 2011, p. 298348, 2011.

JENSEN, P. E. Recent advances in antigen processing and presentation. **Nat Immunol**, v. 8, n. 10, p. 1041-8, Oct 2007.

JUVEN-GERSHON, T. et al. The RNA polymerase II core promoter - the gateway to transcription. **Curr Opin Cell Biol**, v. 20, n. 3, p. 253-9, Jun 2008.

KERTESZ, M. et al. The role of site accessibility in microRNA target recognition. **Nat Genet**, v. 39, n. 10, p. 1278-84, Oct 2007.

KHALIL-DAHER, I. et al. Role of HLA-G versus HLA-E on NK function: HLA-G is able to inhibit NK cytotoxicity by itself. **J Reprod Immunol**, v. 43, n. 2, p. 175-82, Jul 1999.

KING, A. et al. Surface expression of HLA-C antigen by human extravillous trophoblast. **Placenta**, v. 21, n. 4, p. 376-87, May 2000.

KLEIN, J.; SATO, A. Birth of the major histocompatibility complex. **Scand J Immunol**, v. 47, n. 3, p. 199-209, Mar 1998.

\_\_\_\_\_. The HLA system. First of two parts. **N Engl J Med**, v. 343, n. 10, p. 702-9, Sep 7 2000.

KLITKOU, L. et al. Human leukocyte antigen (HLA)-G during pregnancy part I: correlations between maternal soluble HLA-G at midterm, at term, and umbilical cord blood soluble HLA-G at term. **Hum Immunol**, v. 76, n. 4, p. 254-9, Apr 2015.

KOLTE, A. M. et al. Study of the structure and impact of human leukocyte antigen (HLA)-G-A, HLA-G-B, and HLA-G-DRB1 haplotypes in families with recurrent miscarriage. **Hum Immunol**, v. 71, n. 5, p. 482-8, May 2010.

KOZOMARA, A.; GRIFFITHS-JONES, S. miRBase: annotating high confidence microRNAs using deep sequencing data. **Nucleic Acids Res**, v. 42, n. Database issue, p. D68-73, Jan 2014.

KRUGER, J.; REHMSMEIER, M. RNAhybrid: microRNA target prediction easy, fast and flexible. **Nucleic Acids Res**, v. 34, n. Web Server issue, p. W451-4, Jul 1 2006.

KUERSTEN, S.; GOODWIN, E. B. The power of the 3' UTR: translational control and development. **Nature Reviews Genetics**, v. 4, n. 8, p. 626-637, 2003.

KULKARNI, S. et al. Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. **Proc Natl Acad Sci U S A**, v. 110, n. 51, p. 20705-10, Dec 17 2013.

KULKARNI, S. et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. **Nature**, v. 472, n. 7344, p. 495-8, Apr 28 2011.

LE DISCORDE, M. et al. Expression of HLA-G in human cornea, an immune-privileged tissue. **Hum Immunol**, v. 64, n. 11, p. 1039-44, Nov 2003.

LESZCZYSZYN-PYNKA, M. et al. Protective Effect of HLA-B\*5701 and HLA-C -35 Genetic Variants in HIV-Positive Caucasians from Northern Poland. **PLoS One**, v. 10, n. 6, p. e0127867, 2015.

LIU, J. et al. Activating killer cell immunoglobulin-like receptor 2DS2 binds to HLA-A\*11. **Proc Natl Acad Sci U S A**, v. 111, n. 7, p. 2662-7, Feb 18 2014.

LORENZ, R. et al. ViennaRNA Package 2.0. **Algorithms Mol Biol**, v. 6, p. 26, Nov 24 2011.

MAJOROS, W. H.; OHLER, U. Spatial preferences of microRNA targets in 3' untranslated regions. **BMC Genomics**, v. 8, p. 152, Jun 7 2007.

MAKHADIYEVA, D. et al. MHC class I dimer formation by alteration of the cellular redox environment and induction of apoptosis. **Immunology**, v. 135, n. 2, p. 133-9, Feb 2012.

MALE, V. et al. The effect of pregnancy on the uterine NK cell KIR repertoire. **Eur J Immunol**, v. 41, n. 10, p. 3017-27, Oct 2011.

MALISSEN, M.; MALISSEN, B.; JORDAN, B. R. Exon/intron organization and complete nucleotide sequence of an HLA gene. **Proc Natl Acad Sci U S A**, v. 79, n. 3, p. 893-7, Feb 1982.

MANASTER, I. et al. MiRNA-mediated control of HLA-G expression and function. **PLoS One**, v. 7, n. 3, p. e33395, 2012.

MATHONNET, G. et al. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. **Science**, v. 317, n. 5845, p. 1764-7, Sep 21 2007.

MCCUTCHEON, J. A. et al. Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA. **J Exp Med**, v. 181, n. 6, p. 2085-95, Jun 01 1995.

MIMURA, K. et al. The MAPK pathway is a predominant regulator of HLA-A expression in esophageal and gastric cancer. **J Immunol**, v. 191, n. 12, p. 6261-72, Dec 15 2013.

MOREAU, A. et al. Effector mechanisms of rejection. **Cold Spring Harb Perspect Med**, v. 3, n. 11, Nov 1 2013.

MOSAAD, Y. M. et al. Association of human leucocyte antigen Class I (HLA-A and HLA-B) with chronic hepatitis C virus infection in Egyptian patients. **Scand J Immunol**, v. 72, n. 6, p. 548-53, Dec 2010.

MURO, M. et al. Influence of human leukocyte antigen mismatching on rejection development and allograft survival in liver transplantation: is the relevance of HLA-A locus matching being underestimated? **Transpl Immunol**, v. 26, n. 2-3, p. 88-93, Mar 2012.

MURPHY, K.; TRAVERS, P.; WALPORT, M. **Imunobiologia de Janeway 7<sup>a</sup> edição**. Porto Alegre: Artmed, 2010. 898.

NEISIG, A.; MELIEF, C. J.; NEEFJES, J. Reduced cell surface expression of HLA-C molecules correlates with restricted peptide binding and stable TAP interaction. **J Immunol**, v. 160, n. 1, p. 171-9, Jan 01 1998.

OBER, C. et al. Population genetic studies of HLA-G: allele frequencies and linkage disequilibrium with HLA-A1. **J Reprod Immunol**, v. 32, n. 2, p. 111-23, Dec 1996.

ODEBERG, J. et al. Human cytomegalovirus (HCMV)-infected endothelial cells and macrophages are less susceptible to natural killer lysis independent of the

downregulation of classical HLA class I molecules or expression of the HCMV class I homologue, UL18. **Scand J Immunol**, v. 55, n. 2, p. 149-61, Feb 2002.

OGAHARA, S. et al. Effect of mismatched combinations of HLA-A antigens on graft survival in the transplanted kidney. **Transplant Proc**, v. 30, n. 7, p. 3500-1, Nov 1998.

PARHAM, P.; GUETHLEIN, L. A. Pregnancy immunogenetics: NK cell education in the womb? **J Clin Invest**, v. 120, n. 11, p. 3801-4, Nov 2010.

PASQUINELLI, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. **Nat Rev Genet**, v. 13, n. 4, p. 271-82, Apr 2012.

PASQUINELLI, A. E.; RUVKUN, G. Control of developmental timing by micrnas and their targets. **Annu Rev Cell Dev Biol**, v. 18, p. 495-513, 2002.

PELTENBURG, L. T.; DEE, R.; SCHRIER, P. I. Downregulation of HLA class I expression by c-myc in human melanoma is independent of enhancer A. **Nucleic Acids Res**, v. 21, n. 5, p. 1179-85, Mar 11 1993.

PETERSDORF, E. W. et al. HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation. **Blood**, v. 124, n. 26, p. 3996-4003, Dec 18 2014.

PIERTNEY, S. B.; OLIVER, M. K. The evolutionary ecology of the major histocompatibility complex. **Heredity (Edinb)**, v. 96, n. 1, p. 7-21, Jan 2006.

PORTO, I. O. et al. MicroRNAs targeting the immunomodulatory HLA-G gene: a new survey searching for microRNAs with potential to regulate HLA-G. **Mol Immunol**, v. 65, n. 2, p. 230-41, Jun 2015.

PREUSSE, M. et al. SimiRa: A tool to identify coregulation between microRNAs and RNA-binding proteins. **RNA Biol**, v. 12, n. 9, p. 998-1009, 2015.

QI, Y. et al. KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. **PLoS Pathog**, v. 2, n. 8, p. e79, Aug 2006.

RAJU, T. N. The Nobel chronicles. 1980: George Davis Snell (1903-96); Jean Baptiste Dausset (b 1916); Baruj Benacerraf (b 1920). **Lancet**, v. 354, n. 9191, p. 1738, Nov 13 1999.

RANGARAJAN, S.; MARIUZZA, R. A. T cell receptor bias for MHC: co-evolution or co-receptors? **Cell Mol Life Sci**, v. 71, n. 16, p. 3059-68, Aug 2014.

REHMSMEIER, M. et al. Fast and effective prediction of microRNA/target duplexes. **RNA**, v. 10, n. 10, p. 1507-17, Oct 2004.

RENE, C. et al. 5' and 3' untranslated regions contribute to the differential expression of specific HLA-A alleles. **Eur J Immunol**, v. 45, n. 12, p. 3454-63, Dec 2015.

- RIFFO-CAMPOS, A. L.; RIQUELME, I.; BREBI-MIEVILLE, P. Tools for Sequence-Based miRNA Target Prediction: What to Choose? **Int J Mol Sci**, v. 17, n. 12, Dec 9 2016.
- ROBINSON, J. et al. The IPD and IMGT/HLA database: allele variant databases. **Nucleic Acids Res**, v. 43, n. Database issue, p. D423-31, Jan 2015.
- ROBINSON, J. et al. The IPD-IMGT/HLA Database - New developments in reporting HLA variation. **Hum Immunol**, v. 77, n. 3, p. 233-7, Mar 2016.
- ROCK, K. L.; BENACERRAF, B. MHC-restricted T cell activation: analysis with T cell hybridomas. **Immunol Rev**, v. 76, p. 29-57, 1983.
- ROUAS-FREISS, N. et al. Role of HLA-G in Maternal–Fetal Immune Tolerance. **Transplantation Proceedings**, v. 31, p. 724-725, 1999.
- ROUSSEAU, P. et al. The 14 bp deletion-insertion polymorphism in the 3' UT region of the HLA-G gene influences HLA-G mRNA stability. **Hum Immunol**, v. 64, n. 11, p. 1005-10, Nov 2003.
- ROY, S. miRNA in Macrophage Development and Function. **Antioxid Redox Signal**, v. 25, n. 15, p. 795-804, Nov 20 2016.
- SABBAGH, A. et al. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. **Genes Immun**, Dec 19 2013.
- SALTER, R. D. et al. Polymorphism in the alpha 3 domain of HLA-A molecules affects binding to CD8. **Nature**, v. 338, n. 6213, p. 345-7, Mar 23 1989.
- SHEN, L. et al. Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo. **Immunity**, v. 21, n. 2, p. 155-65, Aug 2004.
- SHEN, Y. et al. Distribution of HLA-A, -B, and -C alleles and HLA/KIR combinations in Han population in China. **J Immunol Res**, v. 2014, p. 565296, 2014.
- SHEN, Y. et al. IRF-1 and p65 mediate upregulation of constitutive HLA-A antigen expression by hepatocellular carcinoma cells. **Mol Immunol**, v. 46, n. 10, p. 2045-53, Jun 2009.
- SHEN, Y. Q. et al. Relationship between the downregulation of HLA class I antigen and clinicopathological significance in gastric cancer. **World J Gastroenterol**, v. 11, n. 23, p. 3628-31, Jun 21 2005.
- SHIINA, T. et al. The HLA genomic loci map: expression, interaction, diversity and disease. **J Hum Genet**, v. 54, n. 1, p. 15-39, Jan 2009.
- SHIROISHI, M. et al. Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. **Proc Natl Acad Sci U S A**, v. 100, n. 15, p. 8856-61, Jul 22 2003.

- SHOBU, T. et al. The surface expression of HLA-F on decidual trophoblasts increases from mid to term gestation. **J Reprod Immunol**, v. 72, n. 1-2, p. 18-32, Dec 2006.
- SHUKLA, G. C.; SINGH, J.; BARIK, S. MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. **Mol Cell Pharmacol**, v. 3, n. 3, p. 83-92, 2011.
- SILVA, A. M. et al. Protective role of the HLA-A\*02 allele in Portuguese patients with multiple sclerosis. **Mult Scler**, v. 15, n. 6, p. 771-4, Jun 2009.
- TAN, Z. et al. Allele-specific targeting of microRNAs to HLA-G and risk of asthma. **Am J Hum Genet**, v. 81, n. 4, p. 829-34, Oct 2007.
- TANG, Q. et al. Downregulation of HLA class I molecules in primary oral squamous cell carcinomas and cell lines. **Arch Med Res**, v. 40, n. 4, p. 256-63, May 2009.
- TERASAKI, P. I. A brief history of HLA. **Immunol Res**, v. 38, n. 1-3, p. 139-48, 2007.
- THIELENS, A.; VIVIER, E.; ROMAGNE, F. NK cell MHC class I specific receptors (KIR): from biology to clinical intervention. **Curr Opin Immunol**, v. 24, n. 2, p. 239-45, Apr 2012.
- THORSBY, E. A short history of HLA. **Tissue Antigens**, v. 74, n. 2, p. 101-16, Aug 2009.
- THUS, K. A. et al. Indirectly Recognized HLA-C Mismatches and Their Potential Role in Transplant Outcome. **Front Immunol**, v. 5, p. 210, 2014.
- TIERCY, J. M. HLA-C Incompatibilities in Allogeneic Unrelated Hematopoietic Stem Cell Transplantation. **Front Immunol**, v. 5, p. 216, 2014.
- TRAGARDH, L. et al. Amino acid sequence of an immunoglobulin-like HLA antigen heavy chain domain. **Proc Natl Acad Sci U S A**, v. 76, n. 11, p. 5839-42, Nov 1979.
- TRIPATHI, P.; NAIK, S.; AGRAWAL, S. HLA-E and immunobiology of pregnancy. **Tissue Antigens**, v. 67, n. 3, p. 207-13, Mar 2006.
- VANDIEDONCK, C. et al. Association of HLA-A in autoimmune myasthenia gravis with thymoma. **J Neuroimmunol**, v. 210, n. 1-2, p. 120-3, May 29 2009.
- VARLA-LEFTHERIOTI, M. Role of a KIR/HLA-C allorecognition system in pregnancy. **J Reprod Immunol**, v. 62, n. 1-2, p. 19-27, Jun 2004.
- VILCHES, C.; PARHAM, P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. **Annu Rev Immunol**, v. 20, p. 217-51, 2002.
- WAGNER, M. et al. MicroRNA target prediction: theory and practice. **Mol Genet Genomics**, v. 289, n. 6, p. 1085-101, Dec 2014.



---

WEI, X. H.; ORR, H. T. Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue. **Hum Immunol**, v. 29, n. 2, p. 131-42, Oct 1990.

WILSON, R. C. et al. Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. **Mol Cell**, v. 57, n. 3, p. 397-407, Feb 05 2015.

YAMASHITA, A.; TAKEUCHI, O. Translational control of mRNAs by 3'-Untranslated region binding proteins. **BMB Rep**, v. 50, n. 4, p. 194-200, Apr 2017.

YU, H. et al. Interaction of parental KIR and fetal HLA-C genotypes with the risk of preeclampsia. **Hypertens Pregnancy**, v. 33, n. 4, p. 402-11, Nov 2014.



Justificativa

## JUSTIFICATIVA

Apesar do importante papel das regiões regulatórias no controle dos níveis de expressão gênica, a variabilidade dessas regiões nos genes HLA de classe I nunca foi extensamente explorada. Pouco se sabe sobre a variabilidade desses segmentos e o impacto dessas variantes no perfil de expressão desses genes. Variações na região 3' não traduzida de um gene podem causar alterações de estrutura secundária, da estabilidade da molécula de mRNA e ainda influenciar a afinidade de miRNAs, modificando os níveis proteicos.

A determinação da estrutura e dos polimorfismos das regiões 3' não traduzidas dos genes clássicos de classe I, bem como a determinação de um painel de miRNAs com afinidade por essas regiões, contribuirá para o entendimento do controle da regulação desses genes, e, ainda, norteará estudos funcionais que visam a manipulação dos níveis dessas moléculas para fins terapêuticos.

Assim, esse trabalho busca determinar a influência da estrutura da região 3' não traduzida de genes HLA (*HLA-A*, *HLA-C* e *HLA-G*) e seus polimorfismos na interação com miRNAs, determinando assim, possíveis miRNAs responsáveis pela regulação pós-transcricional desses genes isoladamente e em conjunto.



Objetivos

## 1. OBJETIVOS

Esse trabalho teve como objetivo geral avaliar a variabilidade encontrada nas regiões 3'NT dos genes HLA clássicos *HLA-A* e *HLA-C* em uma amostra da população brasileira do estado de São Paulo por meio de Sequenciamento de Nova Geração (NGS). Juntamente com a variabilidade já descrita para a região 3'NT do gene *HLA-G* avaliar os miRNAs com potencial de ligação a estes segmentos. Os objetivos específicos do trabalho são:

- Caracterizar a estrutura e a variabilidade genética das regiões 3' não traduzidas dos genes de classe I clássicos – *HLA-A* e *HLA-C*;
- Definir os haplótipos observados nestas regiões;
- Buscar computacionalmente os miRNAs com afinidade por todas as possíveis regiões 3'NT de cada gene;
- Determinar a influência dos polimorfismos encontrados na região 3'NT de cada gene na afinidade do miRNAs;
- Gerar um painel de miRNAs que possam atuar no controle pós transcricional dos genes estudados.



## Capítulo II: Manuscrito

## ***HLA-A*, *HLA-C* and *HLA-G*: a comparative study of the 3'UTR segment and microRNA binding profiles**

Iane O. P. Porto<sup>1</sup>, Thálitta H. A. Lima<sup>1</sup>, Andreia S. Sousa<sup>1</sup>, Michelle A. Paz<sup>1</sup>, Celso T. Mendes Júnior<sup>2</sup>, Eduardo A. Donadi<sup>2</sup>, Erick C. Castelli<sup>1</sup>

<sup>1</sup> São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unity, School of Medicine, Botucatu, State of São Paulo, Brazil

<sup>2</sup> Departamento de Química, Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo (USP), Ribeirão Preto, São Paulo, Brasil

<sup>3</sup> Department of Medicine, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil.

### **Abstract**

The class I Human Leucocyte Antigens (HLA) complex is the most diverse region of the human genome. The classical HLA genes (*HLA-A*, *HLA-B* and *HLA-C*) are expressed on most nucleated cells, but their levels fluctuate on a tissue-specific manner. On the other hand, non-classical class I HLA, such as *HLA-G* is only expressed on certain tissues. Differential HLA levels are also observed for the different alleles already described. One of the mechanisms related to differential expression levels is post-transcriptional regulation mediated by microRNAs (miRNAs), which bind preferentially on the 3'Untranslated Region (3'UTR) of their target mRNAs. In this context, we evaluated the variability of *HLA-A* and *HLA-C* 3'UTR on a Brazilian population sample and together with the data available for *HLA-G* 3'UTR performed a complete analysis on miRNA binding profiles observed for these segments. Both *HLA-A* and *HLA-C* 3'UTR segments are highly variable, however, they presented few frequent 3'UTR haplotypes, which is compatible with what is observed in *HLA-G* 3'UTR. These variable sites affect miRNA binding differently for each evaluated locus, i.e. *HLA-A* variable sites have little influence on miRNA binding pattern, while on *HLA-C* and *HLA-G* they seem to have greater effect. We also observed that these sequences are highly similar to each other as well as conserved through time. Thus, gene-specific sequences are associated with gene-specific miRNAs, while shared or similar sequences are usually related to shared miRNAs, such as miR-148a-3p, that may regulate *HLA-A*, *HLA-C* and *HLA-G* altogether. We believed that these 3'UTR and their regulating miRNAs evolved together in order to maintain

optimal HLA levels on the different tissues and to avoid mutations with deleterious effects on post-transcriptional regulation.

**Keywords:** HLA-A, HLA-C, HLA-G, 3' untranslated region, microRNAs, post-transcriptional regulation

## Introduction

The Human Leucocyte Antigens (HLA) complex is the most variable region throughout the human genome and their encoded molecules are associated with antigen presentation. The classical class I genes, *HLA-A*, *HLA-B* and *HLA-C*, are particularly polymorphic. These genes encode molecules that interact with T Cell Receptors (TCR) on CD8<sup>+</sup> T Cells as well as with some specific Killer Immunoglobulin-like Receptors (KIR) on Natural Killer (NK) Cells [1, 2]. In this context, class I HLA molecules play a pivotal role on immune responses.

The first HLA gene described was *HLA-A* [3]. Together with *HLA-B* and *HLA-C*, they compose the classical class I HLA locus and their main function has been associated with antigen presentation and the differentiation between self and non-self-antigens, which make these molecules extremely important on the context of grafted tissues [4, 5]. *HLA-A*, *HLA-B* and *HLA-C* are constitutively expressed on the surface of most nucleated cells. While high levels are observed for *HLA-A* and *HLA-B*, *HLA-C* levels are usually lower due to an increased mRNA degradation rate [6], as well as retention of the *HLA-C* heavy chain at the endoplasmic reticulum (ER) [7].

Nevertheless, *HLA-C* is the only classical HLA class I gene expressed placenta [8], together with *HLA-G* and *HLA-E* [9, 10], both non-classical HLA genes associated with immunomodulatory properties, including NK and T cell inhibition. In addition, *HLA-C* interacts with a series of KIR receptors, both activators (2DS and 3DS) and inhibitors (2DL and 3DL). Hence, the *HLA-C* molecule is an important regulator of NK cells activity [11].

Most studies regarding HLA variability focused on their coding regions, since the encoded molecule is important on the context of transplants and variable sites at the coding region has been associated with many diseases [12]. Thus, the information available concerning HLA regulatory regions is very limited. For most loci, the IPD-IMGT/HLA database, which is the official database that holds known HLA sequences, now on its version



3.30.0, counts with many incomplete sequences. Most of the known different HLA sequences were not characterized for the regulatory segments, including the 5' promoter and the 3'UTR. These regulatory segments account functionally for the gene expression profile. The 3'UTR segment, in particular, play an important role on mRNA stability, translational control and mRNA degradation mediated by microRNAs (miRNAs) [13-15].

The *HLA-G* gene had its regulatory regions extensively explored on the last decade [16-25] and it shows a particular pattern of linkage disequilibrium in which the promoter, coding and 3'UTR behaves essentially as a single segregation block, thus there is a straightforward association among the haplotypes at these segments. In addition, the most frequent *HLA-G* 3'UTR haplotypes detected worldwide are already established [16, 18, 26-28]. However, when it comes to other HLA loci, little is known about the variability regarding these regions, their relationship with coding sequences, their role on the regulation of expression levels and the evolutionary forces acting upon them.

Recent studies concerning *HLA-A*, show the influence of variation sites on the 3'UTR, such as rs1061235 (+2969), on miRNA binding and differential expression [29], even though only a fraction of the 3'UT segment was evaluated. In this case, specifically, the presence on a Thymine in this position hinder the interaction of several miRNAs (*miR-526b*, *miR-609*, *miR-1290*, *miR-342-5p* and *miR-542-5p*) and creates at the same time a target for *miR-520f* and *miR-651* [29]. Also, variation sites along the *HLA-A* 3'UTR may cause changes on sequences corresponding to alternative polyadenylation signals, which also participate on the regulation of translation levels [30]. Regarding *HLA-C*, the miRNA *miR-148a-3p* binds differentially depending on the 3'UTR haplotype, since the binding site lies on a polymorphic region of this segment, and is also related to HIV control based on HLA-C levels and its interaction with NK cells [31].

Both classical and non-classical HLA are greatly similar in their molecular structure, however, they present different functional properties, which is reflected on their expression patterns. Comparing HLA-A and HLA-G, for example, they usually present in a mutually exclusive expression pattern, i.e., the presence of HLA-G but not HLA-A in fetal tissues, and the opposite on most nucleated somatic cells. However, HLA-C is co-expressed with HLA-A in somatic cells, and also co-expressed with HLA-G at placenta during pregnancy. In this context, here we aimed to characterize the variability and haplotype structure of the 3'UTR segment regarding both *HLA-A* and *HLA-C* in admixed population sample such as Brazilians, comparing the results with data available on the IPD-IMGT/HLA database. Then, we

compared the differences and similarities among the 3'UTR segment of the *HLA-A*, *HLA-C* and *HLA-G* genes, highlighting which segments are most important for post-transcription regulation for each gene, which miRNAs strongly and/or specifically binds to each gene, and whether variability at the 3'UTR may influence miRNA binding. This data might provide the bases for future functional studies aiming to manipulate for clinical purposes the *HLA-A*, *HLA-C* or *HLA-G* expression pattern, by using specific miRNAs.

## Methods

*HLA-A* and *HLA-C* variability was surveyed in about 410 samples from the state of São Paulo, Brazil. All participants signed an informed consent before blood withdraw and this study protocol was reviewed and approved by the Human Research Ethics Committee from the School of Medicine of Botucatu (Unesp/Brazil). DNA samples were extracted by salting-out method [32].

*HLA-A* and *HLA-C* were amplified in their totality, encompassing a segment that includes at least 1,500 nucleotides upstream the first translated ATG, the complete coding region with all exons and introns, and the complete 3' untranslated region (3'UTR). Amplification was carried out using primers HAPR.F1 (5'-GCTACTTGGGACTGAGACAG-3') and HAUT.R1 (5'-GCCTTATGGGTACAATTTTCC-3'), for *HLA-A*, producing an amplicon of approximately 5,883 nucleotides (not considering primer sequences), between nucleotides 29,908,285 and 29,914,168 considering the sequence available for chromosome 6 (human genome assembly hg19). The *HLA-C* locus was amplified as a unique amplicon of approximately 5,708-bp using primers HCPR.F1 (5'-GTCTGAGGGATAAGGGGCA-3') and HCUT.R1 (5'-TGAAGAAGTGAACAGCAACTA-3'). Both amplifications were carried out in a 50  $\mu$ L, containing 1.25 units of PrimeSTAR® GXL DNA Polymerase (TaKaRa Bio Company ©, Otsu, Japan) and 1X PrimeSTAR GXL Buffer (Mg<sup>2+</sup> plus), 0.20mM of each dNTP and 0.30 $\mu$ M of each primer. PCR cycling conditions followed the polymerase recommendations. Amplicons were evaluated by electrophoresis on 1% agarose gel stained with GelRed® (Biotium, Inc. Hayward, CA)

Prior to library preparation, amplicons were purified using Illustra ExoProStar (GE Healthcare), quantified using Qubit dsDNA High-Sensitivity Assays (ThermoFisher Scientific, USA) and normalized to the proper concentration. For library preparation, we followed the conditions recommended by Illumina NGS protocols for amplicon quantification and sample pooling. Libraries were prepared using Nextera XT Sample Preparation Kit and

multiplexed with the Nextera XT Index Kit (Illumina, Inc., San Diego, CA). Library quantification was performed by qPCR using Kapa (Kapa Biosystems, Wilmington, MA) and the fragmentation pattern was assessed using Agilent High Sensitivity DNA Kits (Agilent Technologies, Santa Clara, CA). Finally, sequencing libraries were normalized to the recommended concentration for MiSeq Reagent Kit (V2, 500 cycles, paired-end, 2 x 250-bp). Freely available and locally developed applications were used to evaluate the paired-end sequencing data, as described below.

The mapping and genotyping procedures were previously discussed elsewhere when this same sample was evaluated for the variability on other HLA genes [16, 33, 34]. In briefly, reads were mapped by using the *hla-mapper* mapping tool (locally developed, [www.castelli-lab.net/apps/hla-mapper](http://www.castelli-lab.net/apps/hla-mapper)) with hg19 as a reference, and genotypes were inferred using The Genome Analysis Toolkit (GATK, version 3.8) HaplotypeCaller on the GVCF model with default parameters [35-37]. Then, a single VCF (variant call format) file was generated concatenating all samples together with the GATK GenotypeGVCFs. Each variable site was annotated using dbSNP version 146. The VCF file was then processed by *vcfx* (locally developed, [www.castelli-lab.net/apps/vcfx](http://www.castelli-lab.net/apps/vcfx)), function `checkpl`, with the minimum genotype likelihood set to 99.99%, which introduces missing alleles on low likelihood genotypes, thus guarantees that only high quality genotypes are passed forward to a further imputation step.

Associations between variable sites were inferred using the GATK routine ReadBackedPhasing (minimal Phase Quality Threshold of 2000 and minimum base quality set to 20). Thus, only alleles present in a same read were directly phased. Since two consecutive variation sites may be distant from each other, not all variable sites were straightforwardly phased. In addition, ReadBackedPhasing does not perform phase inference on indels and multiallelic loci. Considering these unphased sites, the PHASE algorithm [38] was used to impute the missing alleles and to get the relationships between unphased variation sites. Complete sequences were inferred using the *vcfx fasta* mode. Haplotypes and complete sequences were counted directly.

As a comparative approach, the 3'UTR sequences available on the IPD-IMGT/HLA database for both *HLA-A* and *HLA-C* were extracted from the genomic sequences available for download. To do that, the 3'UTR segment was extracted using EMBOSS *extractalign*, than aligned with *mafft* and reviewed using *Jalview*. A locally developed Perl script counted each different sequence available.

To evaluate the 3'UTR sequences regarding their potential to bind microRNAs, we used a locally developed software named mirHunt [39]. This software automates the detection of miRNAs binding sites by using three different prediction algorithms, miRanda, RNAhybrid and IntaRNA, and compares their outputs to minimize false positives ([www.castelli-lab.net/apps/mirhunt](http://www.castelli-lab.net/apps/mirhunt)). This software also scores each binding site by comparing the energy of that particular binding to the energies observed for the binding of that particular miRNA with mostly of the human 3'UTR sequences. To evaluate the miRNA binding sites, we considered the entire 3'UTR sequence plus 10 nucleotides upstream the stop codon. The 3'UTR sequences detected on our Brazilian sample for *HLA-A* and *HLA-C*, as well as the most frequent 3'UTR sequences detected for *HLA-G* in this same sample and published elsewhere [16] were used to detect the miRNA binding sites. Then, they were aligned using MAFFT for comparison purposes. Whenever possible, these sequences were also aligned with primate sequences, in order to evaluate conserved sequences. Local developed Perl scripts were used to filter data. Excel was used to generate statistical graphics to evaluate which segments are relevant for miRNA interaction.

## Results

Since the IPD-IMGT/HLA database does not consider the complete 3'UTR segment for *HLA-A*, and mostly known *HLA-A* sequences are not characterized for the 3'UTR segment, and, although the IPD-IMGT/HLA do considers the entire *HLA-C* 3'UTR segment, most of the known *HLA-C* sequences are also not characterized for this segment, here we evaluated the variability concerning the 3'UTR segments of both *HLA-A* and *HLA-C* at an admixed population sample from Brazil.

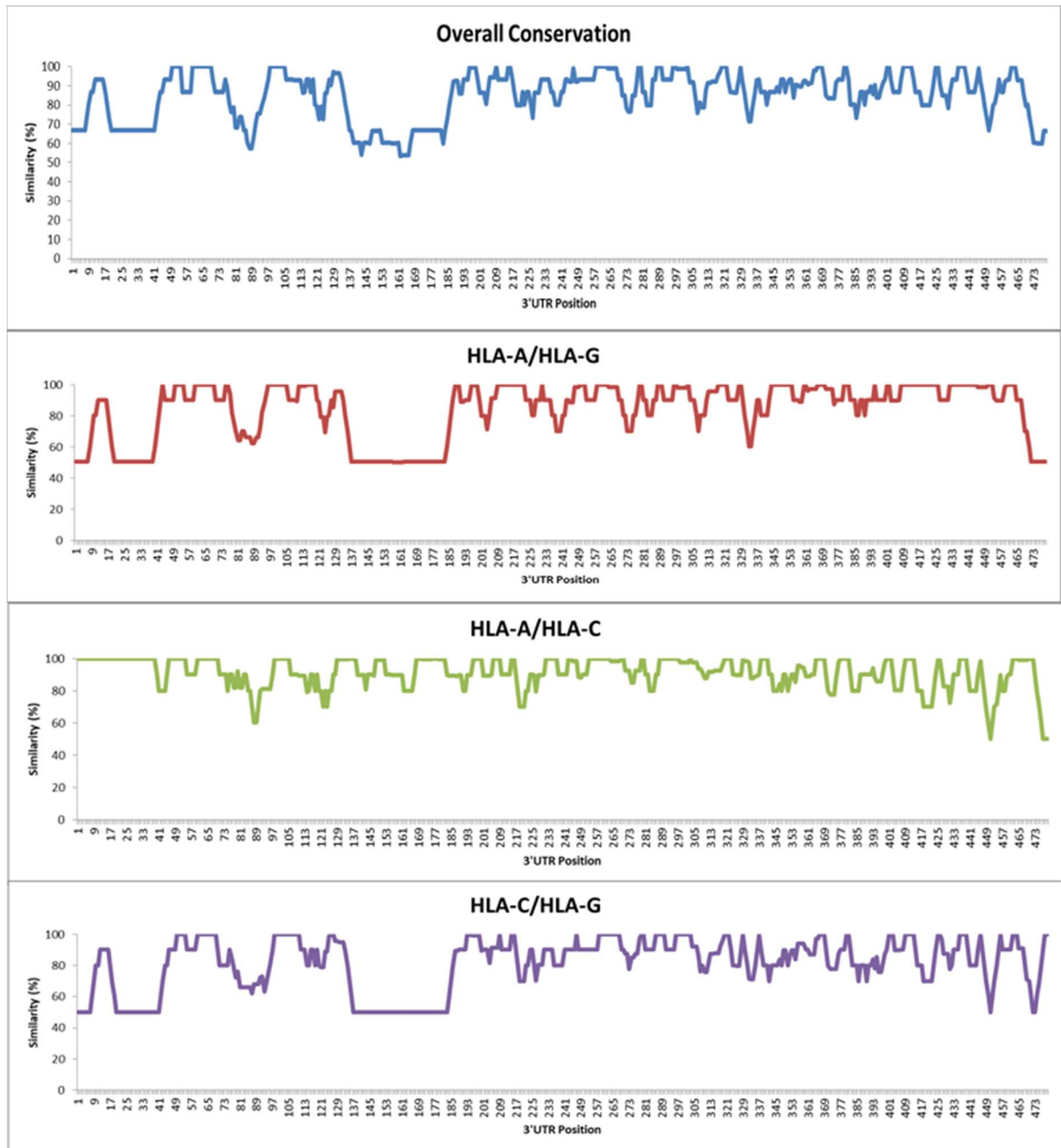
The *HLA-A* locus at the IMGT/HLA database presents 10 different 3'UTR sequences considering the 300nt available for this segment. These haplotypes are generated through 10 variation sites as described in Table 1. The small number of variation sites is explained by the fact that only the variability found on complete *HLA-A* sequences from the IMGT/HLA database were regarded, and those sequences are quite rare in the database. However, when the Brazilian data is considered, in which all the 3'UTR segment was evaluated in more than 400 individuals, 21 variation sites were detected at the *HLA-A* 3'UTR segment, and these variants were arranged in 11 distinct haplotypes as described in Table 2. Considering data from both Table 1 and Table 2, each *HLA-A* 3'UTR haplotype is associated with a single or closely related *HLA-A* coding allele groups.

The 3'UTR segment for the *HLA-C* locus at the IPD-IMGT/HLA database presented 35 variation sites, arranged into 18 different haplotypes (Table 3). These figures are compatible with the ones detected for the Brazilian population sample, i.e., 36 variation sites arranged into 20 haplotypes (Table 4). As observed for *HLA-A*, each *HLA-C* 3'UTR haplotype is associated with a specific coding allele group or closely related ones. It can be noticed many indels at the *HLA-C* 3'UTR, but not for *HLA-A*. In addition, both *HLA-A* and *HLA-C* present similar 3'UTR segments in terms of number of nucleotides, but the *HLA-C* 3'UTR is far more variable than *HLA-A*, opposing what is observed for the coding region.

To assess the similarity among the *HLA-A*, *HLA-C* and *HLA-G* 3'UTR sequences, these sequences were aligned. For *HLA-G*, we considered most frequent *HLA-G* sequences described in a previous study – UTR-1 to UTR-7, and UTR-18 [16], detected for this same sample. The aligned 3'UTR sequences are available as a supplementary file.

After alignment, the frequency of the most conserved nucleotide along the 3'UTR segment was plotted (Figure 1), comparing all genes together (overall) and as pairwise comparisons. For all comparisons, the second half of the 3'UTR segment is the most conserved segment, with the same nucleotide detected for more than 80% of the sequences. The highest similarity was observed between the classical class I genes *HLA-A* and *HLA-C* (91.39% of identity), which indicate that these loci may present a coordinated post-transcription regulation by miRNAs, while the identity of the overall sequences was 84.88%. Despite the noticeable similarities, *HLA-G* 3'UTR is the most divergent one, which, in part, may be related to its differential function and, therefore, to different miRNA-mediated regulation. In addition, some conserved DNA segments are only present for the classical class I genes, *HLA-A* and *HLA-C*, and deleted from the *HLA-G* locus (see supplementary file with the alignment).

miRNA binding sites were assessed across each gene and considering the most frequent 3'UTR sequences, as described in the methods section. Since there is no functional-evaluated cutoff for energy values to detect which miRNA binding may or may not reduce mRNA expression, here we opt to select a cutoff based on some experimental data regarding HLA genes. It has been functionally demonstrated that *miR-148a-3p* may influence both *HLA-G* and *HLA-C* expression levels, and polymorphisms might disrupt miR-148a-3p binding [24, 31, 40]. Because of that, the lowest energy values observed for miR-148a-3p considering all *HLA-C* or *HLA-G* 3'UTR sequences was selected as a cutoff to consider whether a miRNA would bind to any HLA gene and possibly interfere with its expression.



**Figure 1** – Conservation level across the *HLA-A*, *HLA-C* and *HLA-G* 3'UTR segment. Similarity levels as high as 100% indicates 3'UTR segments that present the same sequence between the compared genes. First panel (light blue): overall comparison considering *HLA-C*, *HLA-G*, and *HLA-A*. Second panel (red): comparison between the *HLA-A* and *HLA-G* 3'UTR sequences. Third panel (green): comparison between the *HLA-A* and *HLA-C* 3'UTR sequences. Forth panel (purple): comparison between the *HLA-C* and *HLA-G* 3'UTR sequences

**Table 1** – *HLA-A* 3'UTR haplotypes and associated coding alleles described at the IPD-IMGT/HLA database (version 3.30.0)

3'UTR haplotypes <sup>a</sup>	Position (Relative to the IMGT/HLA database)										Associated HLA-A coding alleles <sup>b</sup>
	2937	2969	3015	3045	3057	3082	3101	3120	3154	3186	
A*IMGT-UTR-01	G	A	G	C	C	C	C	C	T	C	A*30 (A*30:01; A*30:02), A*02 (A*02:01; A*02:48; A*02:87; A*02:661), A*03 (A*03:01; A*03:06; A*03:195; A*03:251; A*03:268; A*03:266N; A*03:272), A*68:156
A*IMGT-UTR-02	G	A	G	G	C	C	T	C	T	C	A*01
A*IMGT-UTR-03	G	A	G	C	C	T	C	C	T	C	A*02:01:01:16
A*IMGT-UTR-04	G	A	G	C	C	C	C	G	T	C	A*03:213
A*IMGT-UTR-05	G	A	A	C	C	C	C	G	T	C	A*23 and A*24
A*IMGT-UTR-06	G	A	G	C	G	C	T	C	T	C	A*11
A*IMGT-UTR-06	G	A	G	C	C	C	C	C	C	C	A*03:224
A*IMGT-UTR-07	T	A	T	C	C	C	C	C	C	C	A*29 (A*29:01; A*29:02; A*29:37; A*29:71; A*2987; A*29:90; A*29:92; A*29:94) A*32 (A*32:01; A*32:02; A*32:03; A*32:69; A*32:86; A*32:92N) and A*74 (A*74:02:01:02; A*74:01:06)
A*IMGT-UTR-08	T	T	T	C	C	C	C	C	C	C	A*31 (A*31:01; A*31:15; A*31:114; A*31:116; A*31:117; A*31:118) and A*33 (A*33:01; A*33:03; A*33:05; A*33:109; A*33:121; A:33:123N; A*125)
A*IMGT-UTR-09	T	A	T	C	C	C	C	C	C	T	A*25, A*26 and A* 66 (A*66: 01:01:01)

a: Considering that there is no official nomenclature for these haplotypes, the names used here follows a numerical sequence.

b: Allelic groups that present complete sequences (300nt) for each haplotype

**Table 2** – *HLA-A* 3'UTR haplotypes and associated coding allele groups found on a Brazilian population sample

HLA-A 3'UTR haplotypes	Frequency (2n = 816)	2937	2969	3014	3015	3036	3038	3045	3057	3101	3105	3120	3154	3172	3186	3210	3213	3225	3259	3301	3309	3310	Associated HLA-A coding alleles	IMHT/HLA haplotype
A*BRA-UTR-01	0.0980	G	A	C	G	G	G	C	C	C	A	C	T	C	C	C	T	A	G	A	G	A	A*03	A*IMGT-UTR-01
A*BRA-UTR-02	0.0613	G	A	C	G	G	G	C	G	T	A	C	T	C	C	C	G	A	G	C	C	G	A*11	A*IMGT-UTR-06
A*BRA-UTR-03	0.0968	G	A	C	G	G	G	G	C	T	A	C	T	C	C	C	G	A	G	T	C	G	A*01 and A*36	A*IMGT-UTR-02
A*BRA-UTR-04	0.0527	G	A	C	G	G	G	C	C	C	A	C	T	C	C	C	G	A	G	A	G	A	A*30	A*IMGT-UTR-01
A*BRA-UTR-05	0.0049	G	A	C	G	G	G	C	C	C	A	C	T	C	C	C	G	C	G	A	G	A	A*30:04:01	A*IMGT-UTR-01
A*BRA-UTR-06	0.1458	G	A	C	A	G	G	C	C	C	A	G	T	C	C	C	G	A	G	A	C	A	A*23 and A*24	A*IMGT-UTR-05
A*BRA-UTR-07	0.3248	G	A	C	G	G	G	C	C	C	A	C	T	C	C	C	G	A	G	A	C	A	A*02, A*68 and A*69	A*IMGT-UTR-01
A*BRA-UTR-08	0.0588	T	A	C	T	G	G	C	C	C	A	C	C	C	T	C	G	A	G	A	C	A	A*25 and A*66	A*IMGT-UTR-09
A*BRA-UTR-09	0.0821	T	A	C	T	G	G	C	C	C	A	C	C	C	C	C	G	A	G	A	C	A	A*29, A*32 and A*74	A*IMGT-UTR-07
A*BRA-UTR-10	0.0735	T	T	C	T	G	G	C	C	C	A	C	C	C	C	C	G	A	G	A	C	A	A*31 and A*33	A*IMGT-UTR-08
A*BRA-UTR-11	0.0012	G	A	G	G	A	A	C	C	C	G	C	T	T	T	A	G	A	C	A	C	A	A*80:01:01	-

White boxes refer to the reference allele according to the genome of reference (hg19); Darker shades of grey indicate variation sites that are multiallelic (+3015 and +3301)



**Table 3 – HLA-C 3’UTR haplotypes and associated coding alleles described at the IPD-IMGT/HLA Database (version 3.30.0)**

3'UTR Haplotype <sup>a</sup>	Position (Relative to the IMGT/HLA Database)																													Associated HLA-C Alleles <sup>b</sup>						
	2941	2979	2987	2988	2996	3005	3020	3028	3033	3041	3074	3095	3119	3126	3150	3153	3155	3156	3160	3161	3172	3179	3188	3193	3194	3197	3201	3218	3239		3240	3241	3250	3269	3273	3306
C*IMGT-UTR-01	C	G	A	T	T	T	G	G	G	C	C	C	A	.	C	C	T	G	C	G	A	ACCT	A	G	T	A	C	T	A	G	G	A	T	G	C	C*01 (C*01:02;C*01:03;C*01:08;C*01:30;C*01:138;C*01:139), C*04 (C*04:01;C*04:03) and C*18:02
C*IMGT-UTR-02	C	G	A	T	T	T	G	G	G	C	C	C	A	.	C	C	T	G	C	G	A	ACCT	A	G	T	A	G	T	A	G	G	A	T	G	C	C*14:02:01:01;C*14:02:01:04
C*IMGT-UTR-03	C	G	A	T	T	T	G	G	C	C	C	A	.	C	C	T	G	C	G	A	ACCT	A	G	T	A	G	T	A	G	G	A	T	G	C	C*03:02:02:01	
C*IMGT-UTR-04	C	G	A	T	T	T	G	G	C	C	C	A	.	C	C	T	G	C	G	A	ACCT	A	G	T	G	C	T	A	G	G	A	T	G	T	C*03:03 and C*03:04:01:02;	
C*IMGT-UTR-05	C	G	A	T	T	T	G	G	G	C	C	C	A	.	C	C	T	G	C	G	A	ACCT	A	G	T	G	C	T	A	G	G	A	T	G	T	C*03:04:01:01
C*IMGT-UTR-06	C	G	A	T	T	T	G	G	G	C	C	A	.	C	C	T	G	C	G	A	ACCT	A	G	T	A	C	T	A	G	G	A	T	G	C	C*04:01:01:01	
C*IMGT-UTR-07	C	G	G	T	T	T	G	G	G	C	C	C	G	G	A	C	T	G	C	A	A	ACCT	A	G	T	A	C	G	A	G	C	A	T	A	C	C*07:01
C*IMGT-UTR-08	C	G	A	C	T	T	G	G	G	C	C	C	G	G	A	C	T	G	C	A	A	ACCT	A	G	T	A	C	G	A	G	C	A	T	A	C	C*07 (C*07:02; C*07:56:02; C*07:66; C*07:67)
C*IMGT-UTR-09	C	G	A	T	T	T	G	G	G	C	C	C	G	G	A	C	T	G	C	A	A	ACCT	A	G	T	A	C	G	A	G	C	A	T	A	C	C*07:04:01:03;C*07:04:01:01
C*IMGT-UTR-10	C	G	A	T	T	C	G	G	G	C	C	C	G	G	A	C	T	G	C	A	A	ACCT	A	G	T	A	C	G	A	G	C	A	T	A	C	C*07:06
C*IMGT-UTR-11	T	G	A	T	T	T	G	G	G	C	C	C	A	.	C	T	C	.	T	A	A	ACCT	C	A	A	A	T	T	G	A	G	A	T	G	C	C*05:01, C*08 (C*08:01; C*08:02;C*08:03;C*08:27;C*08:147;C*08:22;C*08:21;C*08:20;C*08:148)
C*IMGT-UTR-12	C	G	A	T	T	T	G	G	G	C	C	C	A	.	C	T	C	.	T	A	A	ACCT	C	A	A	A	T	T	G	A	G	A	T	G	C	C*06 (C*06:02;C*06:24) C*12 (C*12:02; C*12:03), C*15 (C*15:02; C*15:04; C*15:05; C*15:17)
C*IMGT-UTR-13	C	G	A	T	T	C	G	G	G	C	C	C	A	.	C	T	C	.	T	A	A	ACCT	C	A	A	A	T	T	G	A	G	A	T	G	C	C*06:04:02
C*IMGT-UTR-14	C	G	A	T	T	T	G	G	G	C	C	C	A	.	C	T	C	.	T	A	A	ACCT	C	A	A	A	T	T	G	A	G	A	C	G	C	C*16:01:01:02

C*IMGT-UTR-15	C	G	A	T	C	T	G	G	G	C	C	C	A	.	C	T	C	.	T	A	A	ACCT	C	A	A	A	T	T	G	A	G	A	C	G	C	C*16:01:01:01;C*16:02:01;C*16:04:01:01
C*IMGT-UTR-16	C	A	A	T	T	T	G	G	G	C	C	C	A	.	C	T	C	.	T	A	A	.	C	A	A	A	T	T	G	A	G	A	T	G	C	C*02:02:02:01;C*02:129
C*IMGT-UTR-17	C	G	A	T	T	T	A	T	G	C	T	C	G	G	C	C	T	G	C	A	G	ACCT	A	G	T	A	C	G	A	G	C	G	T	A	C	C*17:03:01:02
C*IMGT-UTR-18	C	G	A	T	T	T	A	T	G	T	T	C	G	G	C	C	T	G	C	A	G	ACCT	A	G	T	A	C	G	A	G	C	G	T	A	C	C*17:01:01:02;C*17:37

a: Considering that there is no official nomenclature for these haplotypes, the names used here follows a numerical sequence.

b: Allelic groups that present complete sequences (300nt) for each haplotype

White boxes refer to the reference allele according to the IPD-IMGT/HLA; Grey boxes indicate the alternative allele; Darker shades of grey represent multiallelic sites (3201)

**Table 4 – HLA-C 3’UTR haplotypes and associated coding allele groups found on a Brazilian population sample**

3’UTR Haplotype	Freq. 2n=800	Position (Relative to IMGT/HLA Database)*																												Associated alleles	IMGT Haplotype								
		3306	3273	3269	3250	3243	3242	3240	3218	3201	3197	3194	3193	3188	3184	3172	3161	3160	3155	3154	3150	3126	3119	3074	3041	3038	3033	3028	3020			3005	2996	2988	2987	2979	2942	2941	2937
C*BRA-UTR01	0.1025	G	T	A	T	CG	G	T	C	G	T	A	C	T	GAAGT	T	T	G	CCAT	G	T	GC	C	G	G	T	C	C	C	A	A	A	C	C	A	G	T	C*07:01:01, C*07:01:02	C*IMGT-UTR07
C*BRA-UTR02	0.0850	G	T	A	T	CG	G	T	C	G	T	A	C	T	GAAGT	T	T	G	CCAT	G	T	GC	C	G	G	T	C	C	C	A	A	G	T	C	A	G	T	C*07:02:01	C*IMGT-UTR08
C*BRA-UTR03	0.0175	G	T	A	T	CG	G	T	C	G	T	A	C	T	GAAGT	T	T	G	CCAT	G	T	GC	C	G	G	T	C	C	C	G	A	A	T	C	A	G	T	C*07:18	C*IMGT-UTR10
C*BRA-UTR04	0.0113	G	T	A	T	CG	G	T	C	G	T	A	C	T	GAAGT	T	T	G	CCAT	G	T	GC	C	G	G	T	C	C	C	A	A	A	T	C	A	G	T	C*07:04:01	C*IMGT-UTR09
C*BRA-UTR05	0.0300	G	T	A	C	CG	G	T	C	G	T	A	C	T	GAAGT	C	T	G	CCAT	G	G	GC	C	A	A	T	C	A	T	A	A	A	T	C	A	G	T	C*17:01:01, C*17:03:01	C*IMGT-UTR18
C*BRA-UTR06	0.0013	G	T	A	C	CG	G	T	C	G	T	A	C	T	GAAGT	C	T	G	CCAT	G	G	GC	C	A	A	T	C	A	C	A	A	A	T	C	A	G	T	C*17:03:01	C*IMGT-UTR17
C*BRA-UTR07	0.2113	G	C	A	T	CG	C	T	A	G	T	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	C	C	C	A	A	A	T	C	A	G	T	C*01:02:01, C*04:01:01, C*04:07, C*04:09N, C*18:01, C*18:02	C*IMGT-UTR01
C*BRA-UTR08	0.0013	G	C	A	T	CG	C	T	A	G	T	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	C	C	C	A	A	A	T	C	G	G	C	C*18:01	C*IMGT-UTR01
C*BRA-UTR09	0.0013	G	C	A	T	CG	C	T	A	G	T	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	G	C	C	A	A	A	T	C	A	G	T	C*18:01	C*IMGT-UTR01
C*BRA-UTR10	0.0725	A	C	A	T	CG	C	T	A	G	C	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	G	C	C	A	A	A	T	C	A	G	T	C*03:02:02, C*03:03:01, C*03:04:01, C*03:04:02, C*03:309	C*IMGT-UTR04
C*BRA-UTR11	0.0188	A	C	A	T	CG	C	T	A	G	C	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	C	C	C	A	A	A	T	C	A	G	T	C*03:03:01, C*03:04:01	C*IMGT-UTR05
C*BRA-UTR12	0.0025	A	C	A	T	CG	C	T	A	G	C	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	G	C	C	G	A	A	T	C	A	G	T	C*03:04:01	C*IMGT-UTR04
C*BRA-UTR13	0.0350	G	C	A	T	CG	C	T	A	C	T	A	C	T	GAAGT	T	C	G	CCAT	G	G	G	T	G	G	T	C	C	C	A	A	A	T	C	A	G	T	C*14:02:01, C*14:03	C*IMGT-UTR02

<b>C* BRA- UTR14</b>	<b>0.0575</b>	G	C	G	T	C	*	TC	A	A	T	T	T	G	GAAGT	T	T	A	C	GTA	G	G	T	G	G	T	C	C	C	A	G	A	T	C	A	G	T	C*16:01:01, C*16:02:01, C*16:04:01	C*IMGT -UTR15
<b>C* BRA- UTR15</b>	<b>0.0013</b>	G	C	G	T	C	*	TC	A	A	T	T	T	G	GAAGT	T	T	A	C	GTA	G	G	T	G	G	T	C	C	C	A	A	A	T	C	A	G	T	C*16:01:01	C*IMGT -UTR14
<b>C* BRA- UTR16</b>	<b>0.1925</b>	G	C	A	T	C	*	TC	A	A	T	T	T	G	GAAGT	T	T	A	C	GTA	G	G	T	G	G	T	C	C	C	A	A	A	T	C	A	G	T	C*06:02:01, C*12:02:02, C*12:03:01, C*15:02:01, C*15:08, C*15:09, C*15:13	C*IMGT -UTR12
<b>C* BRA- UTR17</b>	<b>0.1000</b>	G	C	A	T	C	*	TC	A	A	T	T	T	G	GAAGT	T	T	A	C	GTA	G	G	T	G	G	T	C	C	C	A	A	A	T	C	A	A	T	C*05:01:01, C*08:01:01, C*08:02:01, C*08:03:01, C*08:04:01	C*IMGT -UTR11
<b>C* BRA- UTR18</b>	<b>0.0550</b>	G	C	A	T	C	*	TC	A	A	T	T	T	G	G	T	T	A	C	GTA	G	G	T	G	G	T	C	C	C	A	A	A	T	T	A	G	T	C*02:02:02, C*02:10:01, C*02:14:02	C*IMGT -UTR16
<b>C* BRA- UTR19</b>	<b>0.0025</b>	G	C	A	T	C	*	TC	A	A	T	T	T	G	G	T	T	A	C	GTA	G	G	T	G	G	TG	C	C	C	G	A	A	T	T	A	G	T	C*02:02:02	C*IMGT -UTR16
<b>C* BRA- UTR20</b>	<b>0.0013</b>	G	C	A	T	C	*	TC	A	A	T	T	T	G	G	T	T	A	C	GTA	G	G	T	G	G	TG	C	C	C	A	A	A	T	T	A	G	T	C*02:02:02	-

White boxes refer to the reference allele according to the genome of reference (hg19); Darker shades of grey indicate variation sites that are multiallelic

\* HLA-C is located on the anti-sense strand in relation to the genome reference. Thus, positions are in decrescent order.

In this matter, the number of different miRNAs binding across the HLA genes, but presenting an energy value that is similar or even lower than the cutoff here established, was assessed and plotted (Figure 2A). Since there are many different *HLA-A*, *HLA-C* or *HLA-G* sequences, only the highest number of different miRNAs binding to that particular segment was plotted. Despite the similarities observed among *HLA-A*, *HLA-C* and *HLA-G*, each gene shows a different miRNA binding pattern. Some regions seem to be a common miRNA targets for all evaluated HLA genes, while other regions seem to be gene-specific. Figure 2A demonstrates that, as expected, the first half of the 3'UTR segment is the main target for miRNAs, while the second half, which is the most conserved one (Figure 1), is not an important target for miRNAs.

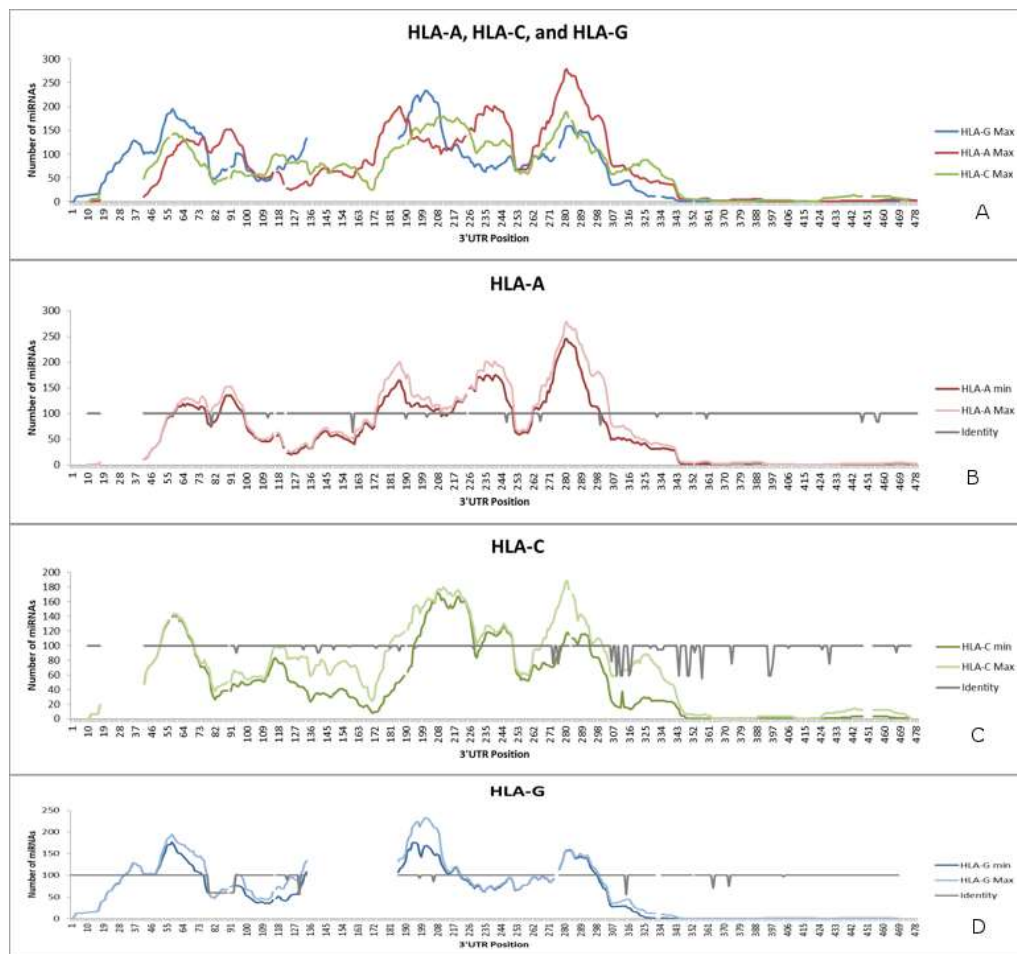
To evaluate whether variable sites may influence miRNA binding, Figures 2B-2D illustrates the minimum and maximum number of different miRNAs binding across the 3'UTR for each HLA gene here evaluated. In addition, a horizontal line (in gray) was plotted indicating the conservation (in terms of frequency of the major allele) across the 3'UTR for that particular gene. If the line is not represented at a particular segment, this is a segment that is not present at that particular gene.

For *HLA-A* (Figure 2B), there are frequent variable sites along the entire 3'UTR segment, but, apparently, they present a minor impact on miRNA binding. Most of the major targets for miRNAs are invariable segments, and frequent variable sites apparently delimit the border of important targets.

For *HLA-C* (Figure 2C), the 3'UTR is highly variable and these changes on nucleotide sequence show a great influence on miRNA binding (observed by changes on number of miRNA binding across the segment). However, frequent variable sites are mainly located on the terminal portion of the 3'UTR segment, which is not an important miRNA target.

Three different patterns were observed. For *HLA-G* (Figure 2D), it can be noticed that most of the frequent variants occur at the first half of the 3'UTR

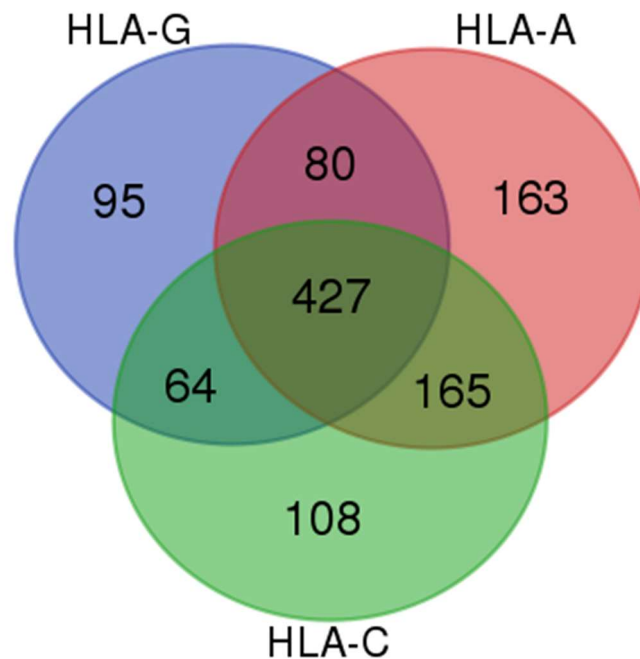
segment and they might influence miRNA binding. The variable sites that mostly influence miRNA binding are the presence or absence the 14bp, known as the 14bp polymorphism, and the variable sites at positions +3003, +3010, +3027 and +3035. However, most of the many important binding sites occur at invariable segments (Figure 2D). In addition, the 3003/3037 region is a segment binding a great number of microRNAs, and this sequence is only available for *HLA-G*. The large deletion characteristic of *HLA-G* (see alignment at the supplementary file) actually creates this important miRNA binding site.



**Figure 2** - Number of different miRNAs that bind across the HLA genes presenting energy higher than the one observed for miR-148a-3p. Panel A: the highest number of different miRNAs observed for each gene and considering all the 3'UTR sequences observed (max). Panel B: the highest (max) and lowest (min) number of different miRNAs observed for *HLA-A* considering all the *HLA-A* 3'UTR sequences. Panel C: the

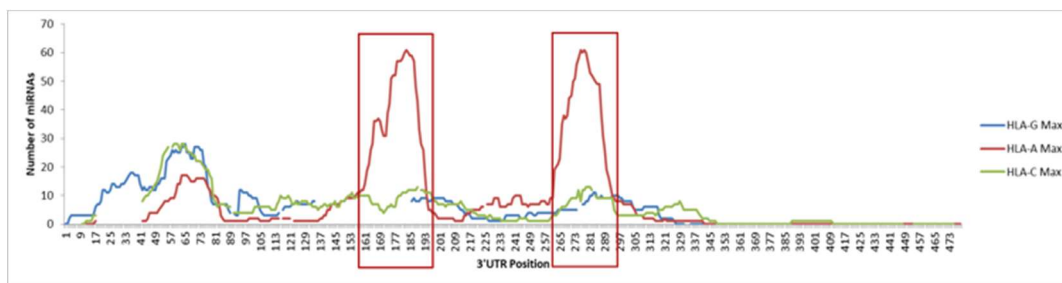
highest (max) and lowest (min) number of different miRNAs observed for *HLA-C* considering all the *HLA-C* 3'UTR sequences. Panel D: the highest (max) and lowest (min) number of different miRNAs observed for *HLA-G* considering all the *HLA-G* 3'UTR sequences. The horizontal line (identity) represents the frequency of the most common allele. This line might be discontinued in some panels because that segment is not available for that particular gene considering the alignment of all 3'UTR sequences.

Considering the number of different miRNAs that interact with each 3'UTR segment, it can be noticed that many miRNAs would regulate all the HLA genes here evaluated (Figure 3). Usually, these miRNAs target conserved segments among these genes. *HLA-A* seems to be a target for a larger number of specific miRNAs, followed by *HLA-C* and, at last, *HLA-G* (Figure 3). A large number of miRNAs (427) are able to bind to the three *loci* (Figure 3), but each gene present their own set of specific microRNAs (Table S1).



**Figure 3** – Number of different miRNAs interacting with each locus and the number of shared ones.

The number of miRNAs that bind to a unique locus were plotted across the 3'UTR segment (Figure 4). The initial portion of the 3'UTR (right after the stop codon) seem to be especially important for miRNA binding for all the three *loci*. This region is highly conserved among *HLA-A*, *HLA-C* and *HLA-G* (Supplementary alignment between nucleotides 44 and 80). For the *HLA-A locus*, for instance, this segment is highly conserved among human and different primates, and the only variable site present at this region can also be found at these primates (Figure S1). Assuming that miRNA sequences have also been conserved through evolution, this segment may be pivotal to *HLA-A*, *HLA-C* and *HLA-G* regulation since the beginning of the human race.

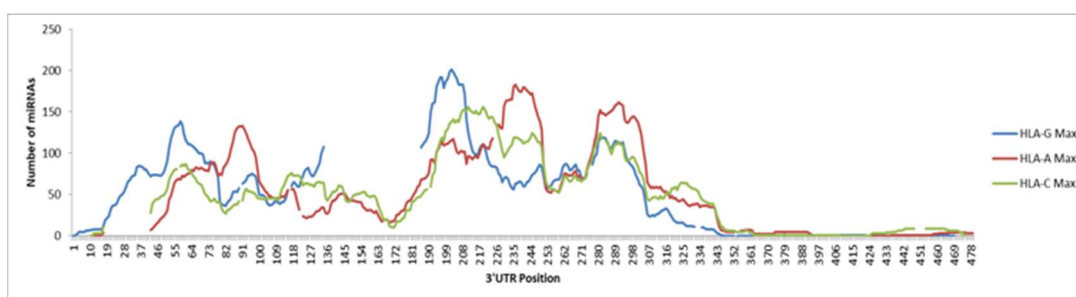


**Figure 4** – Number of different miRNAs interacting with each locus in a highly specific manner, across the 3'UTR segment. The blue line represents the highest (max) number of miRNAs observed for *HLA-G*. The red line represents the highest (max) number of miRNAs observed for *HLA-A*. The green line represents the highest (max) number of miRNAs observed for *HLA-C*. The red squares highlight two important segments for *HLA-A*.

There are two main targets for the miRNAs targeting only *HLA-A*, both in the first half of the 3'UTR (Figure 4, positions 160–185 and 260–290, marked by a red box). The first region comprehends a segment that is also present at the *HLA-C locus*, but it is deleted at *HLA-G*. In addition, this *HLA-A* segment is highly conserved among different primates (Figure S2), and the most frequent variable site detected at this segment is also present among these primates. Although present at *HLA-C*, the *HLA-A* and *HLA-C* sequences are somewhat

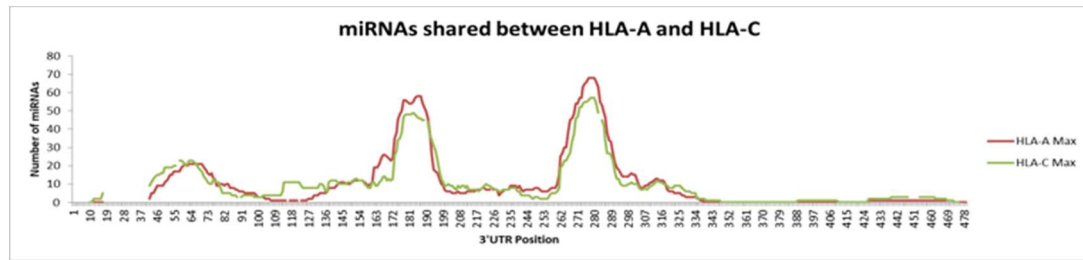


different (Supplementary alignment 1 between nucleotides 135 and 185) because of fixed mutations. These differences are important to create new binding sites for *HLA-A*, but not for *HLA-C*. However, this segment is also targeted by many shared miRNA (Figure 5). The fixed large deletion detected for *HLA-G* creates a new DNA sequence that is important for the HLA-G post-transcription regulation. As already mentioned, this sequence comprehends the variable sites +3003, +3010, +3027 and +3035, which is one of the major targets for miRNA binding (Figure 2B) and may strongly influence miRNA binding depending on the haplotype present (Figure 2B).



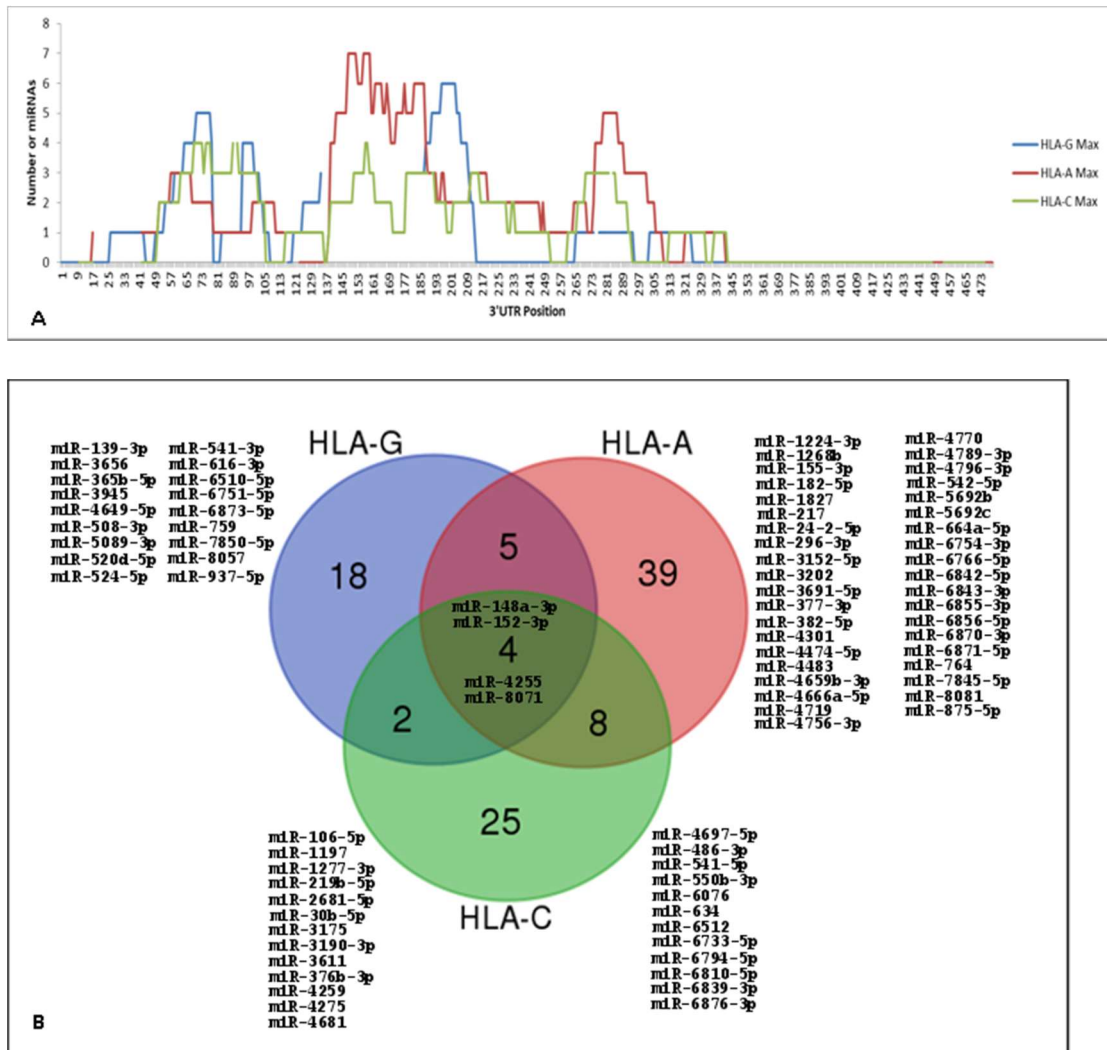
**Figure 5** – Shared miRNAs across the *HLA-A*, *HLA-C* and *HLA-G* 3'UTR. Number of different miRNAs interacting with each locus across the 3'UTR segment. Only the shared microRNAs for the three loci (427 miRNAs, Figure 3) were considered here. The blue line represents the highest (max) number of miRNAs observed for *HLA-G*. The red line represents the highest (max) number of miRNAs observed for *HLA-A*. The green line represents the highest (max) number of miRNAs observed for *HLA-C*.

The second region (260 – 290) is present at the three loci, but each locus presents its own sequence due to fixed mutations. These mutations have created many important binding sites for *HLA-A*, but only a few for *HLA-C* or *HLA-G*. Considering the *HLA-A* locus, this segment is also conserved among different primates (Figure S2). As observed for the first segment, this second one is also targeted by many shared HLA-A/C miRNAs (Figure 6). Actually, it is noteworthy that most miRNAs shared between *HLA-A* and *HLA-C* lie on those two segments (Figure 6).



**Figure 6** – Number miRNAs shared between *HLA-A* and *HLA-C* 3'UTR and their distribution along the segment. The red line represents the highest (max) number of miRNAs observed for *HLA-A*. The green line represents the highest (max) number of miRNAs observed for *HLA-C*.

miRhunt also surveys the specificity of the miRNA to the target sequence (score S) as discussed elsewhere [39]. This S score measures indirectly how many targets would better bind one specific miRNA besides the target been studied. Here we considered a locus-specific miRNA the ones presenting an S score higher than 0.9, which indicates that the MFE observed for that particular miRNA/mRNA binding is lower than at least 90% of the observed MFE values considering all the human 3'UTR sequences in this particular miRNA. The most specific miRNAs are hardly shared among the three HLA 3'UTRs, and each *loci* present some specific miRNAs (supplemental material 1 and Figures 7A-B). As potential *HLA-A*-specific regulators, we highlight miR-3152-5p, miR-217, miR-6870-3p, and miR-4474-5p; for *HLA-C*, miR-106b-5p, miR-219b-5p, miR-609, miR-6810-5p and miR-6794-5p; and for *HLA-G*, miR-365b-5p, miR-4649-5p, miR-524-5p, miR-5089-3p, and miR-759. Those miRNAs presented the highest S values considering each locus separately. An interesting finding is the specificity of miR-148a/b-3p and miR-152-3p to all 3'UTRs evaluated. When these specific miRNAs are plotted across the 3'UTR, the initial portion of the 3'UTR segment of each gene counts with a large number of specific miRNAs, since each gene presents fixed mutations at this segment. The segment present at HLA-A/-C, but not HLA-G, which in turn creates a new sequence for HLA-G, is one of the most important targets for specific miRNAs.



**Figure 7– A)** highly specific miRNAs (not shared) binding to *HLA-A* (red), *HLA-C* (green), and *HLA-G* (blue) 3'UTR sequences and estimated binding position. **B)** Number and name of specific miRNAs (presenting high S score) shared between the loci.

## Discussion

In the present series, we demonstrated that *HLA-A* and *HLA-C* 3'UTR are highly variable (Tables 2 and 4), following the same pattern observed for the coding region. This variability is also supported by the IPD-IMGT database (Tables 1 and 3). However, the majority of the *HLA-A* sequences at the

IMGT/HLA database is still incomplete and lack information on regulatory regions. On the other hand, *HLA-C* sequences at the IMGT/HLA database go beyond the end of the *HLA-C* 3'UTR segment, but many alleles described at the aforementioned database are still incomplete. Because of that, and also to evaluate the 3'UTR variability in a population-based study, getting the frequency of each variant, we surveyed *HLA-A* and *HLA-C* 3'UTR variability in a Brazilian population sample. Most variable sites here detected are already described and in agreement with the IPD-IMGT available data, but with frequency information.

We also considered the variability already described for *HLA-G* 3'UTR [16, 18, 19, 27, 41], since this segment has been extensively studied and characterized on both sequence composition and variation, and its importance for the HLA-G regulation mediated by miRNAs [24, 39, 40]. Studies addressing *HLA-A* and *HLA-C* 3'UTR structure and variability, and their role on post-transcriptional regulation, are still scarce. However, it is known that the 3'UTR of both *HLA-A* and *HLA-C* are miRNA targets and hence, participate on differential allelic expression [29, 31].

Despite the great variability detected for both *HLA-A* and *HLA-C* 3'UTRs, which contrast with the less variable and non-classical *HLA-G*, these three loci present highly similar sequences regarding the entire coding region and also the 3'UTR segments. Also, it is noteworthy that the three loci presented a small number but frequent haplotypes, (8 for *HLA-G*, 9 for *HLA-A* and 13 for *HLA-C*), in spite of the higher number of variation sites detected for *HLA-A* or *HLA-C*.

When aligned together, *HLA-A*, *HLA-C* and *HLA-G* 3'UTRs show more than 80% of identity, while in pairwise comparisons, the identity between *HLA-A* and *HLA-C* goes beyond 90%. This is attributed to the origin of HLA genes through gene duplication [42]. In this context, considering that *HLA-G* is the oldest from the loci evaluated [42], the segment that is present in *HLA-A/C* but absent in *HLA-G* (supplementary alignment 1, between nucleotides 135 to 185) may have emerged as an insertion on *HLA-A* or *HLA-C* (depending on which one is the oldest) and gave birth to an important target for specific miRNAs. On the

other hand, the original HLA-G sequence, which counts with an important series of variation sites (+3003, +3010, +3027 and +3035) retain its original sequence, which is an important allelic specific miRNA target (Figure 2B). The presence of gene-specific mutations on regions such as the one mentioned above is especially important on the binding of miRNAs with great specificity to a single locus, which minimizes possible impact on other target mRNA and its molecular pathway.

In addition to the segment present in *HLA-A/C* and absent in *HLA-G* (Figure 4, supplementary alignment 1, between nucleotides 135 - 185), a second one, present on all three loci (Figure 4, supplementary alignment 1, between nucleotides 260–290) is also an important target for miRNAs, especially for *HLA-A*. Interestingly, both segments are highly conserved on human and primate sequences (supplementary alignment 2). Thus, we believe that these segments have been influencing *HLA-A* regulation since before the separation between humans and other primates. These segments are also an important target for miRNAs with affinity to both *HLA-A* and *HLA-C* (Figure 6) and hence, they may play an important role on coordinated regulation of HLA and HLA-C levels on somatic tissues.

In addition, while *HLA-A* and *HLA-C* miRNA-mediated regulation may occur similarly on somatic tissues where both loci are expressed, *HLA-G* requires a differentiated regulation to guarantee its peculiar tissue expression [43-45]. At the same time, the expression of *HLA-C* and *HLA-G* must be finely coordinated on fetal tissues, thus, some miRNAs may play an important role on their regulation and expression levels. Despite the possibility of a single miRNA presenting multiple targets, we also observed miRNAs that are not shared between the evaluated loci. These miRNAs may be somewhat responsible for the differential expression observed for *HLA-A*, *HLA-C* and *HLA-G* on different tissues. Even though these segments are present on different HLA loci and are highly conserved throughout evolution, it should be notice that each locus present its own group of mutations. Thus, each 3'UTR segment counts with a specific sequence that gives rise to its own set of specific miRNAs (Figure 4).

In the present series, we also evaluated miRNA binding throughout the 3'UTR of the evaluated loci as well as the influence of variability on miRNA binding, which may account for differential expression profiles as already demonstrated functionally for *HLA-A* [29], *HLA-C* [31] and *HLA-G* [24, 40]. Here, we observed that variability on the 3'UTR present quite an influence on miRNA binding for both *HLA-C* and *HLA-G*, and not so much for *HLA-A* (Figure 2B – 2D). *HLA-C* and *HLA-G* perform similar immunological functions (immunomodulation) in certain tissues, such as placenta. Both genes are also quite peculiar on their expression levels, when compared with other HLA, and differential allelic expression has been observed for both [31, 46-48]. Hence, it is possible that the miRNA binding profiles are associated with the differential expression on both an allelic and tissue specific manner, and, associated with the high levels of heterozygosity, it is possible to maintain a finely tuned regulation of expression levels, avoiding both inefficient and excessive levels of these molecules, depending on the immunological context, such as pregnancy. In this context, differential allelic expression may be important to maintain their immunologic functions, avoiding under or overexpression. Thus, different UTR sequences, sensitive to a diverse set of miRNAs are also pivotal.

Considering *HLA-A* through this perspective, most miRNAs are binding preferentially on invariable sequences and little to none variation seem to interfere on or influence miRNA binding. In this scenery, the variability observed at *HLA-A* 3'UTR apparently do not influence the miRNA binding profile, thus *HLA-A* expression and maintenance of mRNA stability. Considering *HLA-A* ubiquitous expression, in tissues with different microenvironments, miRNA-mediated regulation would play a secondary role on differential allelic expression, but it would still be conditioned to tissue specific-miRNAs.

Another interesting fact is that due to the similarity observed between *HLA-A*, *HLA-C* and *HLA-G*, many miRNAs are potential regulators of all of them (Figures 3 and 5). Interestingly, *miR-148a/b-3p* and *miR-152* seem to be an important regulator for the three evaluated genes (supplemental material). This

miRNA was described acting on HLA-C [31] and HLA-G [24, 40]. In addition, the miRtarBase database also presents the interaction among both *miR-148a/b-3p* and *miR-152* and *HLA-A*, *HLA-C* and *HLA-G* 3'UTRs, reported by different assays, such as luciferase and western blot and NGS [49]. However, its influence on *HLA-A* regulation is yet to be functionally proved. All these miRNAs bind on the same region for all the 3'UTR sequences, at the beginning of mRNA second half, which is a region where miRNA binding becomes scarce and, from there on, almost inexistent. Since this specific segment (between nucleotides 280 and 305) are quite variable on both *HLA-C* and *HLA-G*, it is associated with differential expression levels mediated by miRNA on both loci [24, 31, 40]. On *HLA-A* 3'UTR there is also a variation site that lies on *miR-148a/b-3p* and *miR-152* target sequence, which may influence miRNA binding (Figure 2C). On the other hand, each gene counts with a set of miRNAs that would not interfere greatly on the regulation of other HLA genes.

Regarding HLA-A specific miRNAs, such as *miR-3152-5p*, *miR-217*, *miR-6870-3p*, and *miR-4474-5p*, it is noteworthy the expression of miR-217 on tissues such as placenta, sperm and the testis [50], and *miR-4719* on sperm, which are known to not express the HLA-A molecule [46, 51-53]. On testis, we also see the presence of *miR-5690*, which presents not only great affinity but is also highly specific for both *HLA-A* and *HLA-C* 3'UTR [50]. Interestingly, the *HLA-A* 3'UTR presented the highest number of miRNAs binding with affinity and specificity above the cutoff. *HLA-A* is constitutively expressed on most nucleated cells, and its presence must be finely tuned on the surface, since its downregulation may serve as signal for NK mediated lysis, while an overexpression may result in self-antigen presentation and, consequently, autoimmune diseases. In this context, miRNAs may play an important role on HLA-A levels on all possible microenvironments as well on tissue specific levels [53].

A previous study demonstrated the influence of a variation site on the *HLA-A* 3'UTR, (rs1061235/+2969), on miRNA binding and its differential expression [29]. It was observed both *in silico* and functionally that the presence

of a Thymine in this position hinder the interaction of several miRNAs (*miR-526b*, *miR-609*, *miR-1290*, *miR-342-5p* and *miR-542-5p*) while creates a target for *miR-520f* and *miR-651* [29]. Only *miR-609* and *miR-542-5p* were found on our evaluation, which may be explained by differences on the computational methods used. However, our methodology confirms that both *miR-609* and *miR-542-5p* present high specificity to *HLA-A*. In addition, *miR-609* is also a possible regulator for *HLA-C*.

HLA-C is the classical HLA molecule associated with lower expression levels on cell surface by a series of mechanisms [6, 7, 54]. The post-transcriptional regulation mediated by miRNAs is also known to influence HLA-C levels [31, 54]. Some miRNAs that we observed as specific to *HLA-C* 3'UTR sequences, such as *miR-106b-5p*, is expressed in all the tissues available on the miRmine database [50], including testis and placenta, where HLA-C is respectively absent [52, 53] and present [55]. Thus, we believe that *miR-106b-5p* may play an important role on HLA-C expression levels on different tissues.

According to our findings, *HLA-G* 3'UTR presents the most divergent miRNA binding profile when compared to *HLA-A* and *HLA-C*, which is explained by differences observed on its 3'UTR. *HLA-G* 3'UTR is particularly divergent from the other segments evaluated on its initial portion, which counts with a premature stop codon on exon 6 and a 14nt indel. This portion is an important target for *HLA-G*-specific miRNAs (figure 7A). Most tissues do not express *HLA-G*, but its presence is especially important on placenta and other fetal tissues, since it plays an essential role on maternal-fetal tolerance as well as in NK cells functions during invasion, placentation and the whole context of pregnancy [56, 57]. Among the miRNAs we found to be specific to *HLA-G* 3'UTR, *miR-365b-5p* and *miR-524-5p* are expressed on most tissues where HLA-G is absent, and the last is also present in placenta, where it may regulate HLA-G levels through the different stages of pregnancy [56]. *miR-524-5p* belongs to the chromosome 19 miRNA cluster (C19MC), which include a series of pregnancy-related miRNAs [58, 59]. Changes in *miR-524-5p* expression during pregnancy was related to



complications such as eclampsia, gestational hypertension, fetal growth restriction and even termination of gestation [58, 60].

Considering miRNAs with potential to regulate *HLA-A*, *HLA-C* and *HLA-G*, we found particularly interesting the presence of *miR-148a/b-3p* and *miR-152-3p*. These miRNAs have been previously described as *HLA-G* and *HLA-C* post-transcriptional regulators, and according to our results they may also regulate *HLA-A*. Further studies are still necessary to evaluate if these miRNAs are involved in the regulation other class I HLA, such as HLA-B and HLA-E. Those miRNAs have been involved with some malignances such as breast cancer [61] and hepatocellular carcinoma [62], which usually present abnormal HLA expression. They are also present in tissues with the most diverse HLA expression profile. For instance, hepatocytes, gastric epithelium present lower class I HLA levels [63] while *miR-148a/b-3p*, and *miR-152* are abundant [50]. They are also plentiful present in tissues where HLA molecules are mostly to complete absent, such as testis and sperm [52]. Therefore, these miRNAs, as well as other ones, such as miR-8071 and certain members of C19MC (*miR-513a/c-3p*, *miR-520d*) may be related to the differential HLA expression profiles observed in diverse tissues.

It should be noticed that *HLA-A*, *HLA-C* and *HLA-G* 3'UTR sequences and their variability as well as the miRNA profile acting in each one (or in more than one) must have gone through evolution forces to ensure optimal HLA levels on both tissues (guaranteeing differential tissue expression) and immunological contexts. Therefore, depending on which molecules (and how much of them) are necessary in each context, a fine-tuned regulation mechanism was developed along with HLA molecules. Thus, we believe that there is a functional/evolutionary link behind HLA post-transcriptional regulation mechanisms, and that miRNAs that may regulate these genes have been co-evolving with them, to guarantee that HLA molecules perform their functions on a tissue and allelic specific manner.

In conclusion, the present series provides an important set of data for future functional studies, which may focus on specific miRNA regulation of HLA levels, and how 3'UTR variability may influence miRNA binding and allelic specific differential levels.

## References

- [1] N. Anfossi, P. Andre, S. Guia, C.S. Falk, S. Roetynck, C.A. Stewart, V. Bresò, C. Frassati, D. Reviron, D. Middleton, F. Romagne, S. Ugolini, E. Vivier, Human NK cell education by inhibitory receptors for MHC class I, *Immunity*, 25 (2006) 331-342.
- [2] D.K. Cole, N.J. Pumphrey, J.M. Boulter, M. Sami, J.I. Bell, E. Gostick, D.A. Price, G.F. Gao, A.K. Sewell, B.K. Jakobsen, Human TCR-binding affinity is governed by MHC class restriction, *Journal of immunology*, 178 (2007) 5727-5734.
- [3] Dausset, [The HLA complex. I. Immunogenetics of the HLA system. The HLA-A, B and C allelic series], *La Nouvelle presse medicale*, 5 (1976) 1301-1304.
- [4] P. Parham, HLA, anthropology, and transplantation, *Transplantation proceedings*, 25 (1993) 159-161.
- [5] M. Muro, M.R. Lopez-Alvarez, J.A. Campillo, L. Marin, M.R. Moya-Quiles, J.M. Bolarin, C. Botella, G. Salgado, P. Martinez, F. Sanchez-Bueno, R. Lopez-Hernandez, F. Boix, A. Bosch, H. Martinez, J.M. de la Pena-Moral, N. Perez, R. Robles, A.M. Garcia-Alonso, A. Minguela, M. Miras, M.R. Alvarez-Lopez, Influence of human leukocyte antigen mismatching on rejection development and allograft survival in liver transplantation: is the relevance of HLA-A locus matching being underestimated?, *Transplant immunology*, 26 (2012) 88-93.
- [6] J.A. McCutcheon, J. Gumperz, K.D. Smith, C.T. Lutz, P. Parham, Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA, *The Journal of experimental medicine*, 181 (1995) 2085-2095.
- [7] A. Neisig, C.J. Melief, J. Neeffjes, Reduced cell surface expression of HLA-C molecules correlates with restricted peptide binding and stable TAP interaction, *Journal of immunology*, 160 (1998) 171-179.
- [8] A. King, T.D. Burrows, S.E. Hiby, J.M. Bowen, S. Joseph, S. Verma, P.B. Lim, L. Gardner, P. Le Bouteiller, A. Ziegler, B. Uchanska-Ziegler, Y.W. Loke,

Surface expression of HLA-C antigen by human extravillous trophoblast, *Placenta*, 21 (2000) 376-387.

[9] P. Tripathi, S. Naik, S. Agrawal, HLA-E and immunobiology of pregnancy, *Tissue antigens*, 67 (2006) 207-213.

[10] M. Dahl, L. Klitkou, O.B. Christiansen, S. Djurasic, Z.M. Piosik, P. Skovbo, A.M. Moller, R. Steffensen, T.V. Hviid, Human leukocyte antigen (HLA)-G during pregnancy part II: associations between maternal and fetal HLA-G genotypes and soluble HLA-G, *Human immunology*, 76 (2015) 260-271.

[11] H.G. Hilton, L.A. Guethlein, A. Goyos, N. Nemat-Gorgani, D.A. Bushnell, P.J. Norman, P. Parham, Polymorphic HLA-C Receptors Balance the Functional Characteristics of KIR Haplotypes, *Journal of immunology*, DOI 10.4049/jimmunol.1501358(2015).

[12] J. Klein, A. Sato, The HLA system. First of two parts, *The New England journal of medicine*, 343 (2000) 702-709.

[13] D. Didiano, O. Hobert, Molecular architecture of a miRNA-regulated 3' UTR, *Rna*, 14 (2008) 1297-1317.

[14] A. Yamashita, O. Takeuchi, Translational control of mRNAs by 3'-Untranslated region binding proteins, *BMB reports*, 50 (2017) 194-200.

[15] H.S. Yeh, J. Yong, Alternative Polyadenylation of mRNAs: 3'-Untranslated Region Matters in Gene Expression, *Molecules and cells*, 39 (2016) 281-285.

[16] E.C. Castelli, P. Gerasimou, M.A. Paz, J. Ramalho, I.O. Porto, T.H. Lima, A.S. Souza, L.C. Veiga-Castelli, C.V. Collares, E.A. Donadi, C.T. Mendes-Junior, P. Costeas, HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographically distinct population samples of Brazil and Cyprus, *Molecular immunology*, 83 (2017) 115-126.

[17] E.C. Castelli, L.C. Veiga-Castelli, L. Yaghi, P. Moreau, E.A. Donadi, Transcriptional and posttranscriptional regulations of the HLA-G gene, *Journal of immunology research*, 2014 (2014) 734068.

[18] E.C. Castelli, C.T. Mendes-Junior, N.H. Deghaide, R.S. de Albuquerque, Y.C. Muniz, R.T. Simoes, E.D. Carosella, P. Moreau, E.A. Donadi, The genetic structure of 3'untranslated region of the HLA-G gene: polymorphisms and haplotypes, *Genes and immunity*, 11 (2010) 134-141.

[19] E.A. Donadi, E.C. Castelli, A. Arnaiz-Villena, M. Roger, D. Rey, P. Moreau, Implications of the polymorphism of HLA-G on its function, regulation, evolution

and disease association, *Cellular and molecular life sciences : CMLS*, 68 (2011) 369-395.

[20] T.V. Hviid, R. Rizzo, L. Melchiorri, M. Stignani, O.R. Baricordi, Polymorphism in the 5' upstream regulatory and 3' untranslated regions of the HLA-G gene in relation to soluble HLA-G and IL-10 expression, *Human immunology*, 67 (2006) 53-62.

[21] C. Ober, C.L. Aldrich, HLA-G polymorphisms: neutral evolution or novel function?, *Journal of reproductive immunology*, 36 (1997) 1-21.

[22] C. Ober, C.L. Aldrich, I. Chervoneva, C. Billstrand, F. Rahimov, H.L. Gray, T. Hyslop, Variation in the HLA-G promoter region influences miscarriage rates, *Am J Hum Genet*, 72 (2003) 1425-1435.

[23] P. Rousseau, M. Le Discorde, G. Mouillot, C. Marcou, E.D. Carosella, P. Moreau, The 14 bp deletion-insertion polymorphism in the 3' UT region of the HLA-G gene influences HLA-G mRNA stability, *Human immunology*, 64 (2003) 1005-1010.

[24] Z. Tan, G. Randall, J. Fan, B. Camoretti-Mercado, R. Brockman-Schneider, L. Pan, J. Solway, J.E. Gern, R.F. Lemanske, D. Nicolae, C. Ober, Allele-specific targeting of microRNAs to HLA-G and risk of asthma, *Am J Hum Genet*, 81 (2007) 829-834.

[25] Z. Tan, A.M. Shon, C. Ober, Evidence of balancing selection at the HLA-G promoter region, *Human molecular genetics*, 14 (2005) 3619-3628.

[26] E.C. Castelli, C.T. Mendes-Junior, L.C. Veiga-Castelli, M. Roger, P. Moreau, E.A. Donadi, A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution, *Molecular biology and evolution*, 28 (2011) 3069-3086.

[27] A. Sabbagh, P. Luisi, E.C. Castelli, L. Gineau, D. Courtin, J. Milet, J.D. Massaro, H. Laayouni, P. Moreau, E.A. Donadi, A. Garcia, Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity, *Genes and immunity*, DOI 10.1038/gene.2013.67

gene201367 [pii](2013).

[28] K.E. Santos, T.H. Lima, L.P. Felicio, J.D. Massaro, G.M. Palomino, A.C. Silva, S.F. Oliveira, A. Sabbagh, A. Garcia, P. Moreau, E.A. Donadi, C.T. Mendes-Junior, E.C. Castelli, Insights on the HLA-G evolutionary history provided by a nearby Alu insertion, *Molecular biology and evolution*, 30 (2013) 2423-2434.

- [29] C. Rene, C. Lozano, M. Villalba, J.F. Eliaou, 5' and 3' untranslated regions contribute to the differential expression of specific HLA-A alleles, *European journal of immunology*, 45 (2015) 3454-3463.
- [30] S. Kulkarni, V. Ramsuran, M. Rucevic, S. Singh, A. Lied, V. Kulkarni, C. O'HUigin, S. Le Gall, M. Carrington, Posttranscriptional Regulation of HLA-A Protein Expression by Alternative Polyadenylation Signals Involving the RNA-Binding Protein Syncrip, *Journal of immunology*, 199 (2017) 3892-3899.
- [31] S. Kulkarni, R. Savan, Y. Qi, X. Gao, Y. Yuki, S.E. Bass, M.P. Martin, P. Hunt, S.G. Deeks, A. Telenti, F. Pereyra, D. Goldstein, S. Wolinsky, B. Walker, H.A. Young, M. Carrington, Differential microRNA regulation of HLA-C expression and its association with HIV control, *Nature*, 472 (2011) 495-498.
- [32] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, *Nucleic acids research*, 16 (1988) 1215.
- [33] E.C. Castelli, C.T. Mendes-Junior, A. Sabbagh, I.O. Porto, A. Garcia, J. Ramalho, T.H. Lima, J.D. Massaro, F.C. Dias, C.V. Collares, V. Jamonneau, B. Bucheton, M. Camara, E.A. Donadi, HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples, *Human immunology*, DOI 10.1016/j.humimm.2015.06.016(2015).
- [34] T.H. Lima, R.V. Buttura, E.A. Donadi, L.C. Veiga-Castelli, C.T. Mendes-Junior, E.C. Castelli, HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample, *Human immunology*, 77 (2016) 841-853.
- [35] M.A. DePristo, E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernysky, A.Y. Sivachenko, K. Cibulskis, S.B. Gabriel, D. Altshuler, M.J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data, *Nature genetics*, 43 (2011) 491-498.
- [36] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome research*, 20 (2010) 1297-1303.
- [37] G.A. Van der Auwera, M.O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.V. Garimella, D. Altshuler, S. Gabriel, M.A. DePristo, From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline,

Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.], 11 (2013) 11 10 11-11 10 33.

[38] M. Stephens, N.J. Smith, P. Donnelly, A new statistical method for haplotype reconstruction from population data, *Am J Hum Genet*, 68 (2001) 978-989.

[39] I.O. Porto, C.T. Mendes-Junior, L.P. Felicio, R.C. Georg, P. Moreau, E.A. Donadi, J.A. Chies, E.C. Castelli, MicroRNAs targeting the immunomodulatory HLA-G gene: a new survey searching for microRNAs with potential to regulate HLA-G, *Molecular immunology*, 65 (2015) 230-241.

[40] I. Manaster, D. Goldman-Wohl, C. Greenfield, D. Nachmani, P. Tsukerman, Y. Hamani, S. Yagel, O. Mandelboim, MiRNA-mediated control of HLA-G expression and function, *PloS one*, 7 (2012) e33395.

[41] E.C. Castelli, J. Ramalho, I.O. Porto, T.H. Lima, L.P. Felicio, A. Sabbagh, E.A. Donadi, C.T. Mendes-Junior, Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity, *Frontiers in immunology*, 5 (2014) 476.

[42] T. Shiina, G. Tamiya, A. Oka, N. Takishima, T. Yamagata, E. Kikkawa, K. Iwata, M. Tomizawa, N. Okuaki, Y. Kuwano, K. Watanabe, Y. Fukuzumi, S. Itakura, C. Sugawara, A. Ono, M. Yamazaki, H. Tashiro, A. Ando, T. Ikemura, E. Soeda, M. Kimura, S. Bahram, H. Inoko, Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region, *Proceedings of the National Academy of Sciences of the United States of America*, 96 (1999) 13282-13287.

[43] M. Le Discorde, P. Moreau, P. Sabatier, J.M. Legeais, E.D. Carosella, Expression of HLA-G in human cornea, an immune-privileged tissue, *Human immunology*, 64 (2003) 1039-1044.

[44] S. Lefebvre, F. Adrian, P. Moreau, L. Gourand, J. Dausset, S. Berrih-Aknin, E.D. Carosella, P. Paul, Modulation of HLA-G expression in human thymic and amniotic epithelial cells, *Human immunology*, 61 (2000) 1095-1101.

[45] Y.Q. Wang, S.L. Chen, F.Q. Xing, [Expression of HLA-G protein in trophoblast cells], *Di 1 jun yi da xue xue bao = Academic journal of the first medical college of PLA*, 25 (2005) 1488-1490.

[46] D.R. Johnson, Differential expression of human major histocompatibility class I loci: HLA-A, -B, and -C, *Human immunology*, 61 (2000) 389-396.

[47] G. Martelli-Palomino, J.A. Pancotto, Y.C. Muniz, C.T. Mendes-Junior, E.C. Castelli, J.D. Massaro, I. Krawice-Radanne, I. Poras, V. Rebmann, E.D. Carosella, N. Rouas-Freiss, P. Moreau, E.A. Donadi, Polymorphic sites at the 3' untranslated

region of the HLA-G gene are associated with differential hla-g soluble levels in the Brazilian and French population, *PloS one*, 8 (2013) e71742.

[48] X.H. Wei, H.T. Orr, Differential expression of HLA-E, HLA-F, and HLA-G transcripts in human tissue, *Human immunology*, 29 (1990) 131-142.

[49] C.H. Chou, S. Shrestha, C.D. Yang, N.W. Chang, Y.L. Lin, K.W. Liao, W.C. Huang, T.H. Sun, S.J. Tu, W.H. Lee, M.Y. Chiew, C.S. Tai, T.Y. Wei, T.R. Tsai, H.T. Huang, C.Y. Wang, H.Y. Wu, S.Y. Ho, P.R. Chen, C.H. Chuang, P.J. Hsieh, Y.S. Wu, W.L. Chen, M.J. Li, Y.C. Wu, X.Y. Huang, F.L. Ng, W. Buddhakosai, P.C. Huang, K.C. Lan, C.Y. Huang, S.L. Weng, Y.N. Cheng, C. Liang, W.L. Hsu, H.D. Huang, miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions, *Nucleic acids research*, DOI 10.1093/nar/gkx1067(2017).

[50] B. Panwar, G.S. Omenn, Y. Guan, miRmine: a database of human miRNA expression profiles, *Bioinformatics*, 33 (2017) 1554-1560.

[51] S.A. Ellis, Immune status: normal expression of MHC class I in the placenta and what is expected in clones, *Cloning Stem Cells*, 6 (2004) 121-125.

[52] D. Fiszer, M. Kurpisz, Major histocompatibility complex expression on human, male germ cells: a review, *American journal of reproductive immunology*, 40 (1998) 172-176.

[53] C. Rene, C. Lozano, J.F. Eliaou, Expression of classical HLA class I molecules: regulation and clinical impacts, *Hla*, 87 (2016) 338-349.

[54] G. Kaur, S. Gras, J.I. Mobbs, J.P. Vivian, A. Cortes, T. Barber, S.B. Kuttikkatte, L.T. Jensen, K.E. Attfield, C.A. Dendrou, M. Carrington, G. McVean, A.W. Purcell, J. Rossjohn, L. Fugger, Structural and regulatory diversity shape HLA-C protein expression levels, *Nature communications*, 8 (2017) 15924.

[55] A. King, C. Boocock, A.M. Sharkey, L. Gardner, A. Beretta, A.G. Siccardi, Y.W. Loke, Evidence for the expression of HLA-C class I mRNA and protein by human first trimester trophoblast, *Journal of immunology*, 156 (1996) 2068-2076.

[56] R. Hackmon, L. Pinnaduwege, J. Zhang, S.J. Lye, D.E. Geraghty, C.E. Dunk, Definitive class I human leukocyte antigen expression in gestational placentation: HLA-F, HLA-E, HLA-C, and HLA-G in extravillous trophoblast invasion on placentation, pregnancy, and parturition, *American journal of reproductive immunology*, 77 (2017).

[57] E.D. Carosella, P. Moreau, J. Le Maoult, M. Le Discorde, J. Dausset, N. Rouas-Freiss, HLA-G molecules: from maternal-fetal tolerance to tissue acceptance, *Advances in immunology*, 81 (2003) 199-252.

- [58] A. Lycoudi, D. Mavreli, A. Mavrou, N. Papantoniou, A. Kolialexi, miRNAs in pregnancy-related complications, *Expert review of molecular diagnostics*, 15 (2015) 999-1010.
- [59] D.M. Morales-Prieto, S. Ospina-Prieto, W. Chaiwangyen, M. Schoenleben, U.R. Markert, Pregnancy-associated miRNA-clusters, *Journal of reproductive immunology*, 97 (2013) 51-61.
- [60] I. Hromadnikova, K. Kotlabova, M. Ondrackova, P. Pirkova, A. Kestlerova, V. Novotna, L. Hympanova, L. Krofta, Expression profile of C19MC microRNAs in placental tissue in pregnancy-related complications, *DNA and cell biology*, 34 (2015) 437-457.
- [61] M.J. Chen, Y.M. Cheng, C.C. Chen, Y.C. Chen, C.J. Shen, MiR-148a and miR-152 reduce tamoxifen resistance in ER<sup>+</sup> breast cancer via downregulating ALCAM, *Biochemical and biophysical research communications*, 483 (2017) 840-846.
- [62] F. Wang, H. Ying, B. He, Y. Pan, H. Sun, S. Wang, Circulating miR-148/152 family as potential biomarkers in hepatocellular carcinoma, *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 37 (2016) 4945-4953.
- [63] A.S. Daar, S.V. Fuggle, J.W. Fabre, A. Ting, P.J. Morris, The detailed distribution of HLA-A, B, C antigens in normal human organs, *Transplantation*, 38 (1984) 287-292.





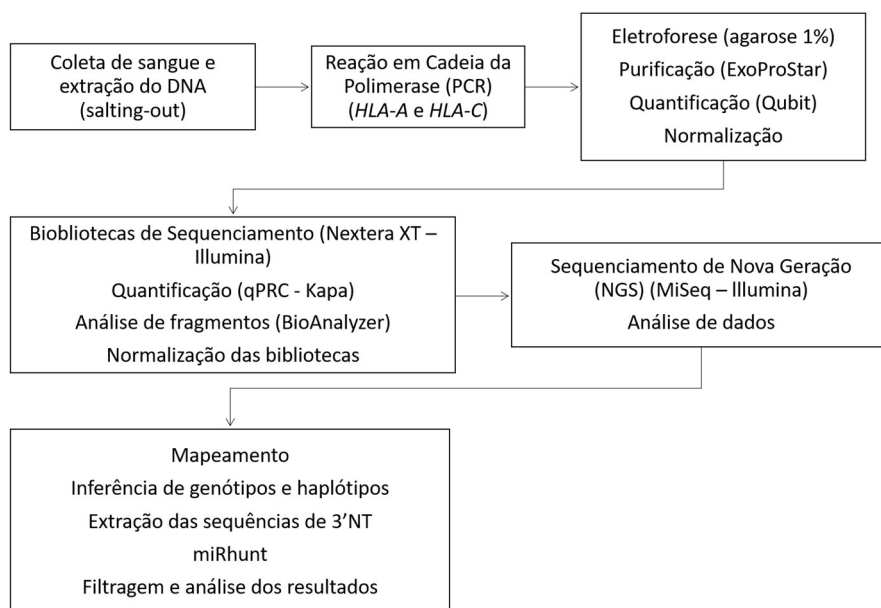
Conclusão

## CONCLUSÃO

O presente trabalho demonstrou que as regiões 3'NT dos genes *HLA-A* e *HLA-C* são altamente variáveis, assim como suas regiões codificadoras. No entanto, a quantidade de haplótipos frequentes é semelhante à observada para genes não clássicos pouco variáveis, como o *HLA-G*. Adicionalmente, esses segmentos são altamente semelhantes entre estes três genes. Esta semelhança parece ser responsável por um conjunto de miRNAs capazes de interagir com os três genes estudados, quatro deles com alto nível de especificidade. A variabilidade encontrada na região 3'NT de cada um dos genes avaliados afeta o padrão de interação de miRNAs de modos diferentes. Enquanto os pontos de variação encontrados ao longo da região 3'NT de *HLA-A* parecem ter pouca interferência na ligação de miRNAs, a interação destes com *HLA-C* e *HLA-G* é afetada pela variabilidade, de modo que alelos diferentes são regulados por miRNAs diferentes.

Finalmente, os padrões de ligação de miRNAs observados neste trabalho, poder apresentar um papel importante na expressão diferencial observada para os genes *HLA-A*, *HLA-C* e *HLA-G* em diferentes tecidos e ainda, na expressão alélica diferencial, especialmente para *HLA-C* e *HLA-G*.



**APÊNDICE A – FLUXOGRAMA DA METODOLOGIA APLICADA**

Metodologia aplicada ao desenvolvimento do trabalho

## APÊNDICE B – HAPLÓTIPOS DE HLA-G

Haplótipo de 3'NT	14-bp ins/del	Posição (relativa ao banco de dados IMGT)							
		+3003	+3010	+3027	+3035	+3142	+3187	+3196	+3227
UTR-1	del	T	G	C	C	C	G	C	G
UTR-2	ins	T	C	C	C	G	A	G	G
UTR-3	del	T	C	C	C	G	A	C	G
UTR-4	del	C	G	C	C	C	A	C	G
UTR-5	ins	T	C	C	T	G	A	C	G
UTR-6	del	T	G	C	C	C	A	C	G
UTR-7	ins	T	C	A	T	G	A	C	G
UTR-18	del	T	G	C	C	C	A	C	A

Adaptado de Castelli, 2017