

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 2020/02/03.



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

Michelle Almeida da Paz

**VARIABILIDADE DOS DOMÍNIOS ALPHA-3,
TRANSMEMBRANA E CAUDA CITOPASMÁTICA DE HLA-C E
DETECÇÃO DE VARIANTES QUE PODEM MODIFICAR SUA
FUNÇÃO**

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

Orientador: Prof. Dr. Erick da Cruz Castelli

BOTUCATU, 2018

Michelle Almeida da Paz

VARIABILIDADE DOS DOMÍNIOS ALPHA-3,
TRANSMEMBRANA E CAUDA CITOPLASMÁTICA DE
HLA-C E DETECÇÃO DE VARIANTES QUE PODEM
MODIFICAR SUA FUNÇÃO

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

Orientador: Prof. Dr. Erick da Cruz Castelli

As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de
responsabilidade dos autores e não necessariamente refletem a visão da Fapesp.

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

Paz, Michelle Almeida da.

Variabilidade dos domínios Alpha-3, transmembrana e cauda citoplasmática de HLA-C e detecção de variantes que podem modificar sua função / Michelle Almeida da Paz. - Botucatu, 2018

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu

Orientador: Erick da Cruz Castelli
Capes: 21103003

1. Antígenos HLA-C. 2. Antígenos de histocompatibilidade HLA. 3. Polimorfismo (Genética). 4. Proteínas de Membrana. 5. Imunologia celular.

Palavras-chave: HLA-C; cauda citoplasmática; domínio α3 ; domínio transmembrana; variabilidade.

Michelle Almeida da Paz

VARIABILIDADE DOS DOMÍNIOS ALPHA-3,
TRANSMEMBRANA E CAUDA CITOPLASMÁTICA DE HLA-C E
DETECÇÃO DE VARIANTES QUE PODEM MODIFICAR SUA FUNÇÃO

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

Orientador: Prof. Dr. Erick da Cruz Castelli

Prof. Dr. Erick C. Castelli
Universidade Estadual Paulista (UNESP)

Profa. Dra. Agnes Alessandra Sekijima Takeda
Universidade Estadual Paulista (UNESP)

Prof. Dr. Eduardo Antônio Donadi
Universidade de São Paulo (USP)

Botucatu, 23 de fevereiro de 2018

Agradecimientos

Primeiramente, agradeço a Deus. Agradeço ao orientador e a todos os companheiros do Laboratório de Genética Molecular e Bioinformática. À Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo FAPESP nº 2016/03622-0). Ao corpo docente e funcionários do Programa de Pós-Graduação em Patologia da Faculdade de Medicina de Botucatu. A todos amigos e familiares.

RESUMO

O Complexo Principal de Histocompatibilidade (MHC) é um complexo gênico que está intimamente envolvido com a regulação do sistema imune. Esse complexo comporta o sistema de Antígenos Leucocitários Humanos (HLA), cuja principal importância está relacionada com o reconhecimento do que é próprio ou não do organismo. *HLA-C* é o gene polimórfico menos variável dos genes HLA clássicos e o que tem menor expressão nos tecidos, exceto na interface materno-fetal, em que é o único gene clássico expresso. A molécula codificada por esse gene possui significativa função na apresentação antigênica e regulação da atividade de células NK, o que permite uma íntima associação com situações fisiológicas, como gestação, e patológicas, como doenças infecciosas, autoimunes, inflamatórias, neoplasias e rejeições a enxertos transplantados. Sua porção gênica mais estudada é a que codifica a fenda de ligação a peptídeos antigênicos, devido sua destacada importância na apresentação de antígenos a células T citotóxicas. No entanto, outras regiões do gene, que são negligenciadas nos estudos de variabilidade, também merecem destaque por influenciarem na sinalização e modulação da citotoxicidade de células efetoras, na ancoragem e estabilidade da molécula na membrana plasmática e na internalização e reciclagem da molécula HLA-C. Desta maneira, nós exploramos a variabilidade dos segmentos que codificam $\alpha 3$ (éxon 4), transmembrana (éxon 5) and cauda citoplasmática (éxon 6 and éxon 7) da molécula HLA-C em uma população miscigenada do Sudeste do Brasil, a fim de entender a influência que polimorfismos presentes nessas regiões podem impactar na estrutura e função da molécula HLA-C.

Palavras chave: *HLA-C*, Sequenciamento de Nova Geração, domínio $\alpha 3$, domínio transmembrana, cauda citoplasmática, variabilidade

ABSTRACT

The Major Histocompatibility Complex (MHC) is a gene complex closely involved in the regulation of the immune system. This complex includes the Human Leukocyte Antigen (HLA) system, whose main role is related to the recognition of self/non-self structures of humans. *HLA-C* is the least variable polymorphic gene of classical HLA genes and has the lowest expression in tissues, except at the maternal-fetal interface, where it is the only classical HLA class I expressed gene. The molecule encoded by this gene has a significant role in the antigen presentation and regulation of NK cells activities, which allows an intimate association with physiological conditions, such as pregnancy, and pathological conditions like infectious, autoimmune, and inflammatory diseases, cancer, and transplantation rejection. The most studied *HLA-C* portion is that encoding the peptide-binding groove, due to its outstanding importance in presentation of antigens to cytotoxic T cells. However, other regions of the gene, which are neglected in the variability studies, are also important in influencing the signaling and modulation of effector cell cytotoxicity, in the anchorage and stability of the molecule on the cell surface, and in the internalization and recycling of the HLA-C molecule. Here, we explore the variability of the segments encoding the $\alpha 3$ (exon 4), transmembrane (exon 5) and cytoplasmic tail (exon 6 and exon 7) domains of the HLA-C molecule in an admixed population sample from Southeastern Brazil, to understand the influence that polymorphisms present in these regions may impact the structure and function of HLA-C molecule.

Keywords: HLA-C, Next Generation Sequencing, $\alpha 3$ domain, transmembrane domain, cytoplasmic tail, variability.

SUMMARY

1.	REVISÃO DE LITERATURA	12
2.	JUSTIFICATIVA	16
3.	OBJETIVOS.....	18
4.	REFERÊNCIAS BIBLIOGRÁFICAS	18
5.	ARTICLE	21
6.	ABSTRACT	22
7.	INTRODUCTION.....	23
8.	MATERIAL AND METHODS.....	25
8.1.	DNA samples and amplification.....	25
8.2.	Library preparation and sequencing	26
8.3.	Raw data processing (mapping).....	26
8.4.	Genotype calling and processing	27
8.5.	Haplotype inference	27
8.6.	Naming process of <i>HLA-C</i> coding alleles and encoded proteins.....	28
8.7.	Template selection and molecular model construction.....	29
8.8.	Evaluation of variable regions, electrostatic potential and secondary structure prediction	29
8.9.	Other analysis	30
9.	RESULTS.....	30
9.1.	Relationship between the <i>HLA-C</i> CDS variants and encoded <i>HLA-C</i> proteins	30
9.2.	Variability in the <i>HLA-C</i> leader peptide	33
9.3.	Variability in the <i>HLA-C</i> peptide binding groove.....	35
9.4.	Variability in the <i>HLA-C</i> $\alpha 3$ domain.....	37
9.5.	Variability in the <i>HLA-C</i> transmembrane domain	40
9.6.	Variability in the <i>HLA-C</i> cytoplasmic tail	42
10.	DISCUSSION.....	43

12.	REFERENCES	46
13.	SUPPLEMENTARY DATA	52
14.	CONCLUSÃO	56

12. REFERENCES

APPS, R. et al. Influence of HLA-C Expression Level on HIV Control. **Science**, p. 1–10, 2013.

APPS, R. et al. Relative Expression Levels of the HLA Class-I Proteins in Normal and HIV-Infected Cells. **The Journal of Immunology**, v. 194, n. 8, p. 3594–3600, 2015.

ASHKENAZY, H. et al. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. **Nucleic Acids Research**, v. 38, n. SUPPL. 2, p. 529–533, 2010.

BARRETT, J. C. et al. Haploview: Analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263–265, 2005.

BERMAN, H. M. et al. The protein data bank. **Nucleic acids research**, v. 28, n. 1, p. 235–242, 2000.

BIASINI, M. et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, v. 42, n. W1, p. 252–258, 2014.

BJOKMAN, P. J. et al. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. **Nature**, v. 329, p. 512–518, 1987.

BLAIS, M. E.; DONG, T.; ROWLAND-JONES, S. HLA-C as a mediator of natural killer and T-cell activation: Spectator or key player? **Immunology**, v. 133, n. 1, p. 1–7, 2011.

BRANDT, D. Y. C. et al. Mapping Bias Overestimates Reference Allele Frequencies at the *HLA* Genes in the 1000 Genomes Project Phase I Data. **G3: Genes|Genomes|Genetics**, 2015.

CASTELLI, E. C. et al. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. **Human Immunology**, v. 76, n. 12, p. 945–953, 2015.

CELIK, A. A. et al. The diversity of the HLA-E-restricted peptide repertoire explains the immunological impact of the Arg107Gly mismatch. **Immunogenetics**, v. 68, n. 1, p. 29–41, 2016.

CEREB, N.; HUGHES, A. L.; YANG, S. Y. Cw*1701 a new HLA-C allelic lineage with an unusual transmembrane domain. v. 115, p. 252–255, 1997.

CHAZARA, O.; XIONG, S.; MOFFETT, A. Maternal KIR and fetal HLA-C: a fine balance. **Journal of Leukocyte Biology**, p. 703–716, 2011.

CHOO, J. A. L. et al. The Immunodominant Influenza A Virus M158-66 Cytotoxic T Lymphocyte Epitope Exhibits Degenerate Class I Major Histocompatibility Complex Restriction in Humans. **Journal of Virology**, v. 88, n. 18, p. 10613–10623, 2014.

CINGOLANI, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. **Fly**, 2012.

CONNOLLY, J. M. et al. Recognition by CD8 on cytotoxic T lymphocytes is ablated by several substitutions in the class I alpha 3 domain: CD8 and the T-cell receptor recognize the same class I molecule. **Proceedings of the National Academy of Sciences of the United States of America**, v. 87, n. 6, p. 2137–41, 1990.

DAVIS, D. M. et al. The transmembrane sequence of human histocompatibility leukocyte antigen (HLA)-C as a determinant in inhibition of a subset of natural killer cells. **The Journal of experimental medicine**, v. 189, n. 8, p. 1265–74, 1999.

DI MARCO, M. et al. Unveiling the Peptide Motifs of HLA-C and HLA-G from Naturally Presented Peptides and Generation of Binding Prediction Matrices. **The Journal of Immunology**, p. ji1700938, 2017.

DOSZTÁNYI, Z. et al. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. **Bioinformatics**, v. 21, n. 16, p. 3433–3434, 2005.

FISCHER, J. C. et al. Relevance of C1 and C2 epitopes for hemopoietic stem cell transplantation: role for sequential acquisition of HLA-C-specific inhibitory killer Ig-like receptor. **Journal of immunology** (Baltimore, Md. : 1950), v. 178, n. 6, p. 3918–23, 2007.

FISCHER, J. C. et al. The impact of HLA-C matching depends on the C1/C2 KIR ligand status in unrelated hematopoietic stem cell transplantation. **Immunogenetics**, v. 64, n. 12, p. 879–885, 2012.

GASTEIGER, E. et al. Protein Identification and Analysis Tools on the ExPASy Server. **The Proteomics Protocols Handbook**, p. 571–607, 2005.

GRUDA, R. et al. Intracellular cysteine residues in the tail of MHC class I proteins are crucial for extracellular recognition by leukocyte Ig-like receptor 1. **Journal of immunology** (Baltimore, Md. : 1950), v. 179, n. 6, p. 3655–61, 2007.

HANSON, J. et al. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. **Bioinformatics**, v. 33, n. 5, p. 685–692, 2017.

HARRIS, K.; NIELSEN, R. Error-prone polymerase activity causes multinucleotide mutations in humans. **Genome Research**, v. 24, n. 9, p. 1445–1454, 2014.

HIBY, S. E. et al. Combinations of Maternal KIR and Fetal HLA-C Genes Influence the Risk of Preeclampsia and Reproductive Success. **The Journal of Experimental Medicine**, v. 200, n. 8, p. 957–965, 2004.

JONES, D. T. Protein secondary structure prediction based on position-specific scoring matrices. **J. Mol. Biol.**, v. 292, p. 195–202, 1999.

JONES, D. T.; COZZETTO, D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. **Bioinformatics**, v. 31, n. 6, p. 857–863, 2015.

KÄLL, L.; KROGH, A.; SONNHAMMER, E. L. L. A combined transmembrane topology and signal peptide prediction method. **Journal of Molecular Biology**, v. 338, n. 5, p. 1027–1036, 2004.

KÄLL, L.; KROGH, A.; SONNHAMMER, E. L. L. An HMM posterior decoder for sequence feature prediction that includes homology information. **Bioinformatics**, v. 21, n. SUPPL. 1, p. 251–257, 2005.

KAUR, G. et al. Structural and regulatory diversity shape HLA-C protein expression levels. **Nature Communications**, v. 8, n. May, p. 15924, 2017.

KLEIN, J.; SATO, A. The HLA system Second of Two Parts. **The New England Journal of Medicine**, p. 782–786, 2000.

KRAEMER, T. et al. HLA-E: Presentation of a Broader Peptide Repertoire Impacts the Cellular Immune Response — Implications on HSCT Outcome. **Stem Cells International**, v. 2015, p. 1–5, 2015.

KRAEMER, T.; BLASCZYK, R.; BADE-DOEDING, C. HLA-E: A novel player for histocompatibility. **Journal of Immunology Research**, v. 2014, p. 1–6, 2014.

KROGH, A. et al. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. **Journal of Molecular Biology**, v. 305, n. 3, p. 567–580, 2001.

KULPA, D. A.; COLLINS, K. L. **The emerging role of HLA-C in HIV-1 infection** *Immunology*, 2011.

KUMAR, S. et al. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. **Briefings in Bioinformatics**, v. 9, n. 4, p. 299–306, 2008.

LOVELL, S. C. et al. Structure validation by C alpha geometry: phi,psi and C beta deviation. **Proteins-Structure Function and Genetics**, v. 50, n. August 2002, p. 437–450, 2003.

MARSH, S. G. E. et al. Nomenclature for factors of the HLA system, 2010. **Tissue Antigens**, p. 291–455, 2010.

MARTÍNEZ-LOSADA, C. et al. Patients lacking a KIR-ligand of HLA group C1 or C2 have a better outcome after umbilical cord blood transplantation. **Frontiers in Immunology**, v. 8, n. JUL, p. 1–8, 2017.

MCKENNA, A. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, p. 1297–1303, 2010.

MIRDITA, M. et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. **Nucleic Acids Research**, v. 45, n. D1, p. D170–D176, 2017.

MORRISON, B. A. et al. Multiple sclerosis risk markers in HLA-DRA, HLA-C, and IFNG genes are associated with sex-specific childhood leukemia risk. **Autoimmunity**, v. 43, n. 8, p. 690–7, 2010.

MOZER-LISEWSKA, I. et al. Genetic (KIR, HLA-C) and Some Clinical Parameters Influencing the Level of Liver Enzymes and Early Virologic Response in Patients with Chronic Hepatitis C. **Archivum Immunologiae et Therapiae Experimentalis**, v. 64, n. 1, p. 65–73, 2016.

PARHAM, P. MHC class I molecules and kirs in human history, health and survival. **Nature Reviews Immunology**, p. 201–214, 2005.

PETERSDORF, E. W. Optimal HLA matching in hematopoietic cell transplantation. **Current Opinion in Immunology**, 2008.

PETTERSEN, E. F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 2004.

RAMALHO, J. et al. HLA-E regulatory and coding region variability and haplotypes in a Brazilian population sample. **Molecular Immunology**, v. 91, n. September, p. 173–184, 2017.

RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: The European Molecular Biology Open Software Suite. **The European Molecular Biology Open Software Suite**, v. 16, n. 6, p. 276–277, 2000.

ROBINSON, J. et al. The IPD and IMGT/HLA database: Allele variant

databases. **Nucleic Acids Research**, v. 43, n. D1, p. D423–D431, 2015.

SANCHEZ-MAZAS, A. **An apportionment of human HLA diversity.** Tissue Antigens, 2007

SCHAEFER, M. R. et al. A novel trafficking signal within the HLA-C cytoplasmic tail allows regulated expression upon differentiation of macrophages. **Journal of immunology** (Baltimore, Md. : 1950), v. 180, n. 12, p. 7804–17, 2008.

SÖDING, J.; BIEGERT, A.; LUPAS, A. N. The HHpred interactive server for protein homology detection and structure prediction. **Nucleic Acids Research**, v. 33, n. SUPPL. 2, 2005.

STEPHENS, M.; SMITH, N. J.; DONNELLY, P. A New Statistical Method for Haplotype Reconstruction from Population Data. **The American Journal of Human Genetics**, p. 978–989, 2001.

VAN DEUTEKOM, H. W. M.; KEŞMİR, C. Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? **Immunogenetics**, v. 67, n. 8, p. 425–436, 2015.

VARLA-LEFTHERIOTI, M. Role of a KIR/HLA-C allorecognition system in pregnancy. **Journal of Reproductive Immunology**, v. 62, n. 1–2, p. 19–27, 2004.

WANG, S. et al. Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. **Scientific Reports**, v. 6, n. January, p. 1–11, 2016.

WANG, Z. C. et al. Molecular characterization of the HLA-Cw*0409N Allele. **American Journal Of Veterinary Research**, v. 8859, n. 2, p. 1457–1461, 2002.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. **Curr Protoc Bioinformatics**, v. 54, p. 1–55, 2017.

WESLEY, P. K. et al. The CD8 coreceptor interaction with the $\alpha 3$ domain of HLA class I is critical to the differentiation of human cytotoxic t-lymphocytes specific for HLA-A2 and HLA-Cw4. **Human Immunology**, v. 36, n. 3, p. 149–155, 1993.

YAN, R. et al. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. **Scientific Reports**, v. 3, 2013.

ZHI, D. et al. Killer cell immunoglobulin-like receptor along with HLA-C ligand genes are associated with type 1 diabetes in Chinese Han population. **Diabetes/Metabolism Research and Reviews**, v. 32, n. 30, p. 13–23, 2014.

ZHOU, Y. et al. SPIDER2: A Package to Predict Secondary Structure, Accessible Surface Area, and Main-Chain Torsional Angles by Deep Neural Networks. [s.l: s.n.]. v. 1484

ZIMMERMANN, L. et al. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. **Journal of Molecular Biology**, 2017.

14. CONCLUSÃO

Até mesmo considerando uma população miscigenada e heterogênea como a brasileira, os domínios $\alpha 3$, transmembrana e citoplasmático das proteínas HLA-C apresentam certo nível de conservação. As proteínas diferentes encontradas compartilham alguns segmentos entre si e isso pode ser consequência da conversão de genes, dando origem a divergentes sequências de HLA-C que apresentam altas frequências. Ainda que alguns resíduos se apresentam variáveis, geralmente a troca ocorre entre aminoácidos de perfis quimicamente semelhantes. De regra, as variações de resíduos que foram encontradas nos domínios estudados não alteram a estrutura secundária da proteína. Mesmo os resíduos, que potencialmente podem alterar a função da molécula HLA-C por estarem localizados em regiões críticas de interação com receptores, estão em baixa frequência ou são raros. Essa conservação característica nos os domínios $\alpha 3$, transmembrana e citoplasmático mostra-se importante para sustentar toda a função imunológica da molécula HLA-C e sua importância clínica.