

AVALIAÇÃO DA ACURÁCIA DA INFERÊNCIA DE ALELOS HLA UTILIZANDO ALGORITMO DE MAXIMIZAÇÃO DE EXPECTATIVA

Isadora Laguila Altoé (PIBIC/FA), Kayo Thiago Ribeiro Perroni, Matheus Braga (Coorientador), Larissa Danielle Bahls Pinto (Orientador). E-mail: ldbpinto2@uem.br

Universidade Estadual de Maringá, Departamento de Ciências Básicas da Saúde, Maringá, PR.

Imunologia /Imunogenética

Palavras-chave: HLA; Inferência estatística; Diversidade genética.

RESUMO

O estudo investigou a acurácia da plataforma HaploStats para inferência estatística de alelos HLA em alta resolução na população do noroeste do Paraná, a partir de resultados obtidos por tipificação HLA de baixa/média resolução. O sistema HLA é altamente polimórfico e fundamental na resposta imune, sendo relevante para transplantes, doenças e genética populacional. A imputação de alta resolução, embora mais econômica que métodos laboratoriais, pode apresentar viés quando bancos de dados de referência não representam populações miscigenadas como a brasileira. Foram analisadas 56 amostras de doadores voluntários cadastrados no REDOME. A imputação pelo HaploStats utilizou como referência a população caucasiana e os resultados foram comparados com tipificações reais em alta resolução. Houve uma acurácia de 94,11%, sendo que os erros ocorreram mais em HLA de classe II que em classe I. Parte dos alelos incorretamente imputados eram comuns no Brasil, indicando lacunas de representatividade nos bancos de referência. Apesar dessas limitações, o desempenho foi satisfatório, demonstrando que o HaploStats é uma ferramenta viável, de baixo custo e aplicável na prática clínica e em estudos de associação HLA-doença, podendo otimizar a seleção de doadores e reduzir custos em tipagens de alta resolução. Contudo, recomenda-se ampliar a inclusão de perfis genéticos de populações miscigenadas nos bancos de dados para aumentar a precisão da imputação.

INTRODUÇÃO

O Complexo Principal de Histocompatibilidade (MHC), denominado sistema *HLA* nos humanos, é formado por moléculas de superfície celular essenciais para a imunidade adaptativa e a homeostasia da imunidade inata. Sua variabilidade influencia respostas a tumores, infecções, transplantes, vacinas e hipersensibilidades. Os genes *HLA* localizam-se no braço curto do cromossomo 6, organizados em três classes: I (*HLA-A*, *-B*, *-C*), II (*HLA-DR*, *-DQ*, *-DP*) e III, que codifica moléculas não relacionadas à histocompatibilidade, como componentes do complemento e TNF (1).

A região *HLA* é altamente polimórfica e sua distribuição alélica varia conforme a ancestralidade, permitindo pesquisas sobre transplantes e suscetibilidade a doenças devido ao papel central das moléculas *HLA* na resposta imune (2). A tipificação *HLA* pode ser realizada por métodos de baixa/média resolução (PCR-SSP, PCR-SSO) ou alta resolução (sequenciamento Sanger, NGS) (3). Embora a alta resolução seja mais precisa para transplantes e estudos de associação, é mais cara e demorada, motivando o uso de métodos de imputação baseados em dados públicos. A imputação depende de fatores como tamanho e diversidade do painel de referência, qualidade dos dados e correspondência étnica entre painel e população estudada. A limitação é agravada em populações miscigenadas, como a brasileira, que são sub-representadas nos bancos públicos (4). Neste contexto, o estudo propõe avaliar a acurácia da plataforma HaploStats (NMDP), que utiliza bancos de dados de diferentes etnias, para identificar o painel mais adequado à população paranaense em estudos *HLA*-doença.

MATERIAIS E MÉTODOS

Seleção de amostras

O estudo utilizou resultados de tipificação *HLA* de doadores cadastrados no REDOME e convocados para exames confirmatórios entre janeiro de 2015 e janeiro de 2025, todos com TCLE assinado e dados sigilosos conforme a Política de Privacidade. Foram selecionadas 56 amostras com código REDOME, o que permitiu acessar e organizar eficientemente as informações para a pesquisa.

Imputação dos dados

Os resultados de tipificação *HLA* em baixa/média e alta resolução foram organizados em tabelas no Excel. Para avaliar a capacidade da plataforma HaploStats em identificar alelos específicos, os dados em baixa resolução foram inseridos no sistema, adotando a população caucasiana como referência, e registrado o alelo de alta resolução mais frequente. Em seguida, comparou-se esse resultado com a tipificação real em alta resolução para calcular a acurácia da imputação, considerando a etnia do banco de referência (optou-se por utilizar como referência a população caucasiana) e o haplótipo utilizado.

RESULTADOS E DISCUSSÃO

As amostras englobaram indivíduos de ambos os sexos e de diferentes etnias — brancos, pardos, negros, amarelos —, sendo que houve uma predominância de brancos na casuística, o que é esperado devido à maior contribuição da população europeia na ancestralidade dos paranaenses.. A idade dos participantes variou de 18 a 46 anos, com uma média de 24,7 anos (desvio-padrão de 7,05).

Os alelos imputados (*HLA-A*, *-B*, *-C*, *-DRB1* e *-DQB1*) foram comparados com as tipificações reais obtidas por metodologia de alta resolução. O estudo demonstrou que a plataforma Haplostats obteve um desempenho consistente na imputação da região *HLA* na população examinada, sendo que foram registrados 527 acertos e 33

erros nas imputações realizadas. Dessa forma, a acurácia alcançada foi de 94,11%, enquanto a taxa de erro foi de 5,89%, sendo um resultado considerado satisfatório para apoiar tanto a avaliação quanto a seleção de doadores de medula óssea, aumentando, assim, a eficácia do procedimento de transplante nas comunidades envolvidas.

A respeito dos alelos imputados equivocadamente, tiveram 11 erros de HLA de classe I, contra 19 erros de HLA de classe II. A análise dos erros revelou que 9 são classificados como comuns, 5 intermediários e 1 bem documentado pelo catálogo internacional CIWD v. 3.0.0. Já o catálogo CIWD Brasil aponta 14 como comuns e apenas 1 como intermediário (Tabela 2), (5, 6).

De forma relevante, a ferramenta errou a imputação de um mesmo alelo comum até 7 vezes (DQB1*02:02), refletindo a não representatividade do haplótipo segregado existente no Brasil no banco de dados. Os resultados indicam a necessidade de maior representatividade de populações miscigenadas, como a brasileira, nos bancos de dados que servem como referência para a ferramenta a fim de aumentar sua acurácia. Todavia, o HaploStats demonstrou uma baixa taxa de erro, sendo assim uma ferramenta viável e válida, de excelente custo-benefício para inferência de resultados HLA em alta resolução, otimizando a seleção de doadores e ampliando o acesso a dados de alta resolução na prática clínica.

Tabela 1: Classificação dos alelos imputados erroneamente

Alelo	Fa	CIWD internacional	CIWD Brasil	Frequência dos alelos
<i>C*02:10</i>	0,11	C	C	Comum
<i>DRB1*14:54</i>	0,18	C	C	Comum
<i>DQB1*03:19</i>	0,07	C	C	Comum
<i>B*40:129</i>	0,04	WD	C	Descrito em 3 populações, sendo 2 brasileiras
<i>B*35:11</i>	0,07	I	C	Raro no Brasil, comum nos EUA
<i>B*35:03</i>	0,07	C	C	Comum
<i>DQB1*02:02</i>	0,32	C	C	Comum
<i>DRB1*04:05</i>	0,07	C	C	Comum
<i>DQB1*03:02</i>	0,29	C	C	Comum
<i>DQB1*04:02</i>	0,29	C	C	Comum
<i>A*31:15</i>	0,04	I	C	Bem documentado no Brasil
<i>B*15:05</i>	0,04	I	C	Comum
<i>DRB1*04:11</i>	0,14	C	C	Comum
<i>B*39:13</i>	0,04	WD	C	Raro
<i>B*39:10</i>	0,04	I	C	Comum
<i>DQB1*06:05</i>	0,04	WD	não consta	Comum

Fa – Frequência alélica na população do estudo, C – comum, I – intermediário, WD – well documented (bem documentado)

CONCLUSÕES

Os resultados evidenciam que a imputação HLA por meio do HaploStats apresenta alta acurácia e bom custo-benefício, sendo uma alternativa promissora frente às limitações de custo e tempo das metodologias de alta resolução. Entretanto, a ocorrência de erros associados à falta de representatividade de haplótipos característicos de populações miscigenadas, como a brasileira, ressalta a importância de ampliar e diversificar os bancos de dados de referência. Esse investimento pode aumentar a precisão na compatibilidade doador/receptor e impulsionar estudos de associação HLA-doença.

AGRADECIMENTOS

Agradeço à Fundação Araucária pelo financiamento do PIBIC e a Profa. Dra. Larissa Danielle Bahls Pinto e Matheus Braga por todo apoio e assistência.

REFERÊNCIAS (Arial 12, negrito, alinhado à esquerda, deixe uma linha em branco)

- 1) BODMER, J. G. et al. Nomenclature for factors of the HLA system, 1998. Tissue Antigens 53: 407-443, 1999.
- 2) SANCHEZ-MAZAS A. et al. Immunogenetics as a tool in anthropological studies. Immunol 2011; 133: 143–64
- 3) RIBEIRO, B. G. C. Estudo e proposta de sistema para o processo de tipificação HLA praticado pelo Laboratório de HLA da UERJ. 2017. 128 f. Dissertação (Mestrado em Ciências Computacionais) - Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2017.
- 4) ZHENG, X. et al. HIBAG—HLA genotype imputation with attribute bagging. Pharmacogenomics J 14, 192–200 (2014). <https://doi.org/10.1038/tpj.2013.18>
- 5) HURLEY, C. K, et al. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. HLA. 2020; 95: 516–531. <https://doi.org/10.1111/tan.13811>
- 6) DA SILVA, J.S., et al. Common, Intermediate and Well-Documented HLA Alleles in the Brazilian Population: An Analysis of the Brazilian Bone Marrow Donor Registry (REDOME). HLA. 2025;105(2):e70051. doi:10.1111/tan.70051