Predicted epitope binding core motifs and common amino acid polymorphisms distinguish protective- from susceptibility-associated HLA alleles in leishmaniasis

Nicky de Vrij [1,2], Pieter Meysman [1,3], Sofie Gielis [1], Wim Adriaensen [2], Kris Laukens [1,3,*] and Bart Cuypers [1,3,4,*]

[1] Adrem Data Lab, Department of Mathematics and Computer Science, University of Antwerp, Middelheim 1, 2020 Antwerp, Belgium.
[2] Clinical Immunology Unit, Department of Clinical Sciences, Institute of Tropical Medicine, Nationalestraat 122, 2000 Antwerp, Belgium.
[3] Biomedical Informatics Network Antwerpen (biomina), University of Antwerp, Antwerp, Belgium
[4] Unit of Molecular Parasitology, Department of Biomedical Sciences, Institute of Tropical Medicine, Nationalestraat 122, 2000 Antwerp, Belgium
* Authors to whom correspondence should be addressed. E-mail: kris.laukens@uantwerpen.be; bart.cuypers@uantwerpen.be

Abstract: Susceptibility for leishmaniasis is largely dependent on genetic- and immune factors of the host. Despite the previously described association of human leukocyte antigen (HLA) gene cluster variants as genetic susceptibility factors, little is known on the mechanisms that mediate these associations. To characterize the functionality underpinning these associations between HLA and disease, we predicted the epitope binding repertoires for all known leishmaniasis-associated HLA variants collected in a thorough literature review. We identified several amino acid polymorphisms in the HLA sequences that distinguished protective- from risk-associated HLA-DRB1 alleles. Proteome-wide and multi-species T cell epitope binding predictions were carried out across these alleles, enabling us to map the effects on the epitope binding repertoires. The protective-associated HLA-DRB1 alleles were characterized by common binding core motifs, which map to the identified amino acid polymorphisms. These results strongly suggest that polymorphism in the HLA region, resulting in differential antigen presentation, affects the association between HLA and leishmaniasis disease development. Finally, we established a valuable open-access resource of putative epitopes, of which a set of 14 HLA-unrestricted strong-binding epitopes, conserved across species, were prioritized for further epitope discovery in the search for novel subunit-based vaccines.

Keywords: *Leishmania*; immunoinformatics; antigen presentation; HLA association; leishmaniasis; vaccine candidates; immunogenetics; human leukocyte antigen.

1. Introduction

Leishmaniasis is a vector-borne disease, transmitted through sandflies, caused by protozoan parasites of the genus *Leishmania*. Leishmaniasis imposes a substantial global burden, affecting 1.5 million people annually [1]. It is one of the 20 "neglected tropical diseases", predominantly affecting the poorest populations in low- to middle-income countries [2]. Leishmaniasis is endemic in 87 countries, spanning multiple continents and predominantly affecting countries with a warm and tropical climate. Indeed, Afghanistan, Algeria, Brazil, Colombia, Pakistan and the Syrian Arab Republic represent more than 70% of all cutaneous leishmaniasis (CL) cases globally. Brazil, India, South Sudan and Sudan host 78% of all visceral leishmaniasis (VL) cases. However, due to climate change, the geographical distribution might expand towards cooler regions [3-5].

Leishmaniasis features a wide spectrum of clinical manifestations, ranging from the skin disease cutaneous leishmaniasis to visceral leishmaniasis, which affects the visceral organs and is lethal without treatment. CL clinical presentations can also range from the often self-healing localized CL (LCL) to diffuse CL (DCL) which imposes numerous diffuse skin lesions, and mucocutaneous CL (MCL) where infected individuals develop disfiguring mucosal lesions. CL can be caused by around ten distinct species of *Leishmania,* including *L. major, L. braziliensis* and *L. mexicana,* while VL is predominantly caused by *L. donovani* and *L. infantum* [6]. The majority of infected individuals do not develop symptomatic disease [7-9]. Yet, knowledge on why some infected individuals remain asymptomatic while others develop disease is scarce. The development of the different clinical manifestations is primarily dependent on the infecting species of *Leishmania,* but genetic factors and the immune response of the host play a central role as well [10].

Several major histocompatibility complex (MHC) genes of the human leukocyte antigen (HLA) gene cluster have been identified as genetic susceptibility factors for leishmaniasis, with variants affecting disease outcome both

positively and negatively [11, 12]. The HLA gene complex is located within the human chromosome 6p21 region and is highly polymorphic [13]. An individual carries several HLA gene variants, or alleles, on a chromosome. The entirety of an individuals HLA genes on a single chromosome is known as a haplotype. The varying HLA genes code for distinct MHC glycoproteins and these MHC molecules present antigens to T cells, eliciting a T cell-mediated immune response upon activation. This T cell-mediated immunity dictates the antileishmanial immune response in both animal models and human disease [14]. However, findings *in vitro* or in animal models may not translate to protective or detrimental responses in humans [14].

Despite the previous identification of specific genes of the HLA loci as genetic susceptibility factors, little is known about the properties (e.g. common amino acid variants and their physicochemical properties, binding affinity and epitope dominance hierarchy) that underpin these associations. The identification of HLA variants as genetic susceptibility factors and the importance of T cell-mediated immunity suggests that antigen presentation plays a pivotal role in disease development and progression. Yet, several complications hamper research into antigen presentation in leishmaniasis. Firstly, *Leishmania* is an eukaryotic parasite with complex cellular mechanisms, such as immune modulation, and several life stages. Secondly, *Leishmania* exhibits a multitude of possible epitopes due to its relatively large proteome, coding for roughly 8000 proteins. As a consequence, relatively few epitopes have been characterized for murine leishmaniasis and even less so for human leishmaniasis [15]. Yet, identifying and characterizing these epitopes and their binding properties are crucial to understanding and studying the complex host-pathogen interaction process, but also to guide the development of safe and effective multi-epitope-based vaccines.

Traditionally, epitope discovery was centered around long and low-throughput epitope screenings. In recent times, immunoinformatics-based T cell epitope predictions, using computational prediction algorithms, allow for high-throughput *in silico* epitope identification and prioritization [16]. These prediction tools have thus enabled reverse-vaccinology approaches that can expedite traditional epitope-based vaccine discovery. However, the usability of these immunoinformatic predictions is not limited to reverse-vaccinology, as these tools have been used to uncover links between disease pathogenesis and antigen presentation as well [17]. In *Leishmania* research, these strategies have been widely used to predict potential vaccine candidates [18-21]. As most of these approaches require relatively high computing power, they started from only a subset of *Leishmania* proteins, mostly limited to those that are known to be immunogenic in animal models. This strategy is, however, hampered by the observed discordance between animal models and humans [14]. In addition, these immunoinformatic tools have not been fully wheeled to flesh out associations between leishmaniasis pathogenesis and antigen presentation..

Therefore, we performed unprecedented multi-species and proteome-wide T cell epitope binding predictions to A) explore the properties underpinning the known class I and class II HLA associations in human leishmaniasis and B) establish a comprehensive resource of promising epitopes linked to broad-class protection and risk for both CL and VL. Furthermore, we provide an exploratory analysis of the physicochemical- and binding properties underpinning the observed variation in protection- and risk-associated HLA profiles.

2. Materials and Methods

2.1 Literature search for leishmaniasis-associated HLA alleles

We conducted a literature review searching for HLA alleles that are significantly associated with genetic susceptibility for any type of leishmaniasis. Specifically, PubMed and Web of Science were screened for relevant publications using the following keywords in all fields: [("*Leishmania*" OR "Leishmaniasis" or "Kala-azar") AND ("HLA" OR "Human leukocyte antigen" OR "Histocompatibility")]. EndNote X9 was used to filter out any duplicate publications found in the resulting list. The abstracts of the resulting publications were screened for relevance, and only relevant studies on human hosts were included for review. Following this, a complete screening of the full texts was performed to identify publications suitable for inclusion. HLA alleles associated with genetic susceptibility to leishmaniasis were extracted from the resulting publications. The identified HLA alleles were labeled as either 'Risk' for susceptibility-associated loci or 'Protective' for protection-associated loci (as defined by authors). The demographic information of the human study populations was also collected. The PRISMA statement guidelines were followed for article selection [22].

2.2 Multiple sequence alignment of leishmaniasis-associated HLA alleles

The protein sequences of leishmaniasis-associated HLA-DRB1 alleles were acquired through the IPD-IMGT/HLA database (release 3.42.0, 2020-10-15) [13]. The HLA-DRB1 locus was specifically chosen for a sequence-based comparison, in order to unravel common amino acid variants that underpin the disease association status of an allele. The diverse spectrum of leishmaniasis-associated alleles in the HLA-DRB1 locus enabled us to filter out nat-

urally occurring variation in the amino acid sequences, thus allowing identification of substitutions likely to be causal. The first 29 amino acid residues were trimmed off to fit a position numbering system that confers structural equivalence across the different alleles [23]. This approach allows the identification of functionally important positions, such as peptide-binding or T cell receptor contact positions. The resulting trimmed sequences were aligned using Clustal Omega with default parameters, and subsequently exported to Jalview (version 2.11.1.3) for visualisation [24, 25]. Amino acids were coloured using the Zappo colour scheme, which describes their physicochemical properties and colours only non-identical positions.

2.3 Prediction of epitope binding to leishmaniasis-associated HLA alleles

The proteomes of the different *Leishmania* species with known HLA allele associations were downloaded from TriTrypDB (release 46), except for *L. donovani,* for which we used the PacBio assembly by Dumetz et al. [26, 27]. Pseudogenes were excluded because they are unlikely to be transcribed into proteins, and thus to be processed into epitopes. Epitope binding predictions across protective- and risk-associated HLA alleles were restricted to the *Leishmania* species known to be associated with a particular HLA allele.
Epitope binding prediction requires specific 4-digit resolution HLA alleles as input. Thus, in case only a 2-digit allele group of HLA was identified in the literature review, the "classical allele freq" search tool of the Allele Frequencies Net Database (accessed on 11/02/2020) was used to find the most common alleles in the population studied [28]. All alleles were subsequently cross-checked with the wide list of available HLA proteins included in the tools to narrow down the chosen allele sequences to only those for which predictions were possible using the chosen tools.
Proteome-wide prediction of epitope binding to leishmaniasis-associated HLA alleles was carried out using the NetMHCpan v4.1b and NetMHCIIpan v4.0 tools [29]. The NetMHCpan peptide fragment length was set to 9mer, which deviates from the default range of fragment lengths (between 8-11mer). For NetMHCIIpan the default peptide fragment length of 15mer was chosen. For both tools, a sliding window of 1 was used. 9mer and 15mer fragment length peptides are the most commonly eluted ligands for MHC-I and MHC-II, respectively [30, 31]. Predicted epitopes, their binding cores, their binding affinity (BA) to a specific HLA allele, the eluted ligand score percentage rank (%Rank), and the protein(s) each epitope is derived from, were extracted from the output of the epitope binding prediction tools. The %Rank is the percentile rank of the epitopes, as defined by the probability of it being an eluted ligand, when compared to a set of 100.000 random naturally occurring peptides.

2.4 Standardization of epitope binding affinity

The predicted affinity values should be standardized according to a common reference point to enable the comparison of binding preference across alleles, analogous to the method described by Meysman et al [17]. In brief, a collection of 107 highly expressed human proteins was used to calculate the relative binding affinity of each epitope to a particular HLA allele. The binding affinity of the best scoring predicted epitope of the human dataset was then divided by the binding affinity of a predicted epitope of *Leishmania* spp. to determine a relative affinity for that particular *Leishmania* spp. epitope. This is based on the hypothesis that human proteins are the most likely competitors for binding to HLA alleles and are thus best suited as reference points.

To screen the resulting predicted epitopes for those with a maximal likelihood of being presented in vivo, two filters were applied. Firstly, only leishmanial epitopes with a %Rank lower than or equal to 2 were considered to be suitable for further analysis. Secondly, only those leishmanial epitopes with a relative score above 1 (i.e. those that bind better than the best binding human epitope) were considered likely to outperform competitor self-peptides and thus get presented. Predicted epitopes that pass both these criteria are further referred to as strong-binding epitopes.

Leishmania genomes feature a lot of tandemly repeated genes that are known to vary between strains of the same species [32]. Together with conserved gene families, this can lead to the same epitopes being included in many proteins identical in sequence and function, varying only in gene identifier. For these duplicate epitopes, only one was included for further analysis for each HLA allele.

The described standardization and filtering steps have been integrated into a bioinformatic pipeline, available at https://github.com/BioinformaNicks/LeisHLA.

2.5 Characterization of epitope binding repertoires

Firstly, the relative affinity distributions of all epitopes for each HLA allele were plotted on a protective-associated against risk-associated basis. Subsequently, linear mixed models were implemented, using the lmerTest package in the programming language R, to identify differences in the relative affinity distributions of protective- and

risk-associated HLA alleles [33]. Here, the HLA association status served as a fixed effect while the specific HLA alleles served as a random effect. This model enables us to explore the difference between disease association status of the HLA alleles, while the intrinsic variability between HLA alleles is controlled for.

Finally, common amino acid variants within the sequences of the different leishmaniasis-associated HLA profiles might translate to functional variability within the peptide binding groove. In order to explore this, the *in silico* predicted epitopes were directly wheeled for the creation of consensus sequence logos of the 9mer binding cores. The predicted 9mer binding cores were used to create these consensus sequence logos for each HLA allele, using the ggseqlogo R package [34].

2.6 Epitope prioritization for future discovery

Potential strong-binding epitopes were prioritized based on two features. The overlap of these different properties was used to conclude a prioritized list of potential epitopes that are most suitable candidates for future epitope discovery.
The first feature is the conservation of predicted epitopes between species. Epitopes that are conserved between different species are more interesting candidates as they (1) increase the coverage of a vaccine therefore protecting against multiple species and (2) are less likely to mutate over the course of evolution as conserved sequences tend to be part of essential functional domains of a protein.
The second feature of interest is the promiscuity of epitopes. Promiscuous epitopes that bind across different alleles are more suitable for discovery because they tend to be immunodominant epitopes towards which the bulk of the immune response is skewed [35]. Moreover, this prevents confounding of immunological evaluation by HLA type restrictions.

3. Results

3.1. HLA alleles associated with protection or risk

Following the outlined search strategy of leishmaniasis-associated HLA literature, 752 publications were identified for further screening. After duplicate filtering, 451 were retained and 33 were selected after title and abstract screening. Based on the full texts and predefined exclusion criteria, 16 were deemed highly relevant for inclusion. Seven publications were further excluded because they contained either no associations or non-significant associations after multiple testing correction. This workflow is outlined in the PRISMA Flow Diagram (Supplementary figure 1).

The literature review resulted in a list of HLA alleles that were either associated with increased susceptibility or protection for infection by *Leishmania spp.* (Table 1). This list consists of 12 high-resolution 4-digit alleles, 5 broad haplotypes, and 11 low-resolution 2-digit alleles. Most association studies conducted case/control frequency enrichment analysis to define protection- or risk-associated alleles. However, Fakiola et al. and Lara et al. performed a familial segregation analysis in addition to a case/control frequency enrichment, improving statistical confidence [36-38]. Protection-associated alleles were generally defined as alleles that occur significantly more in healthy endemic controls than in patients with diagnosed clinical symptoms, and vice versa for risk-associated alleles. It is important to note that several association studies selectively typed only either class I or class II HLA.

Table 1. Leishmaniasis-associated HLA alleles extracted from the literature were divided into those that confer protection against contracting the disease (Protective) and those that increase the susceptibility for the disease (Risk). CL: Cutaneous leishmaniasis; LCL: Localized cutaneous leishmaniasis; DCL: Diffuse cutaneous leishmaniasis; MCL: Mucocutaneous leishmaniasis; VL: Visceral leishmaniasis. † = *L. donovani* does not typically cause CL, but this species has been reported to cause both VL and CL in the Sri Lankan population [39]. * = Included patients with both LCL and DCL.

Type	Reference	Protective alleles	Risk alleles	Sample size	Species	Study Location
VL	Singh et al., 2018 [40]	DRB1*0101 DRB1*1501 DRB1*1502 DRB1*1602	DRB1*11 DRB1*1404 DRB1*1301 DRB1*1302	*Patients N = 889* *Control N = 977*	*L. donovani* *L. infantum*	India Brazil

Type	Reference	Protective alleles	Risk alleles	Sample size	Species	Study Location
VL	Fakiola et al., 2013 [36]	DRB1*01 DRB1*15 DRB1*16	DRB1*11 DRB1*13 DRB1*14	*Three cohorts. Total Patients N = 2287 Total Control N = 3692*	*L. donovani* *L. infantum*	India Brazil
VL	Faghiri et al., 1995 [41]		A*26	*Patients N = 52 Control N = 222*	*L. donovani*	Iran
MCL	Petz-Erler et al., 1991 [42]	DRB1*15 DRB1*16	DQB1*03	*Patients N = 43 Control N = 111*	*L. braziliensis*	Brazil
LCL†	Samaranayake et al., 2016 [12]	DRB1*15-DQB1*06	DRB1*15	*Patients N = 140 Control N = 140*	*L. donovani*†	Sri Lanka†
LCL	Ribas-Silva et al., 2015 [43]		C*04	*Patients N = 186 Control N = 278*	*L. braziliensis*	Brazil
LCL	Olivo-Díaz et al., 2004 [44]	DRB1*15 DRB1*16 DPB1*0401	DRB1*0407 DQA1*3011 DPA1*0401 DPB1*0101 DRB1*0407-DQA1*3011-DQB1*0302 DRB1*0407-DQA1*3011-DQB1*0301	*Patients N = 65 Control N = 100*	*L. mexicana*	Mexico
CL*	El-Mogy et al., 1993 [45]	-	A*11 B*05 B*07	*Patients N = 27 Control N = 200*	*L. major*	Egypt
LCL	Lara et al., 1991 [46]	B*15	A*28 Bw22 DQw3 Bw22-CF31 Bw22-DR*11-DQ*7	*24 families. Patients N = 127 Control N = 160*	*L. guyanensis* *L. braziliensis*	Venezuela

The associations in Table 1 were cross-checked with a list of HLA alleles available in the chosen prediction tools. Consequently, a subset of these HLA alleles or haplotypes were excluded from analysis due to unavailability, namely: Bw22-CF31, Bw22-DR*11-DQ*7, DRB1*0407-DQA1*3011-DQB1*0301, DRB1*0407-DQA1*3011-DQB1*0302 and DRB1*15-DQB1*06. Moreover, the HLA allele groups A*28, B*5, Bw22, and DQw3 are 'broad antigen serotypes', defined as a category of allele groups that are difficult to discern from each other with older serotyping methods [47]. Although newer techniques have enabled the distinction of the 'split antigens' within these broad antigen serotypes, it is hard to distinguish which split antigen was truly associated with leishmaniasis at the time several of these studies were carried out. Therefore, these broad antigens and associated haplotypes (Bw22-DR*11-DQ7 and Bw22-CF31) were excluded to minimize noise. The remaining associated haplotypes were also excluded for predictions due to unavailability in the prediction tools. Concludingly, the list of HLA alleles included for further analysis was narrowed down to 12 high-resolution class II HLA alleles, 2 low-resolution class II HLA alleles, and 5 low-resolution class I HLA alleles.

3.2 Sequence-based comparison of leishmaniasis-associated HLA-DRB1 alleles

Only a small number of HLA class I alleles and an even smaller number of non-HLA-DRB1 class II alleles are associated with leishmaniasis. Due to this small number, variation in the sequences of these alleles is not a reliable proxy to infer the correlation of amino acid substitutions with protective-or risk-associated effects. In contrast, HLA-DRB1 alleles are widely associated with leishmaniasis. Thus, the amino acid sequences of the different leishmaniasis-associated HLA-DRB1 alleles were aligned to identify distinct substitutions in the HLA profile that explain the disease association status. This multiple sequence alignment, shown in Figure 1, identified several key amino acid substitutions that are completely (n=3) or largely (n=10) shared in either protective- or risk-associated HLA alleles. Interestingly, some of these are located in the functionally important HLA-DRB1 binding pocket at positions 1, 4, 6 and 9 [23, 48-50]. These residues and their functional positions are shown in Table 2.

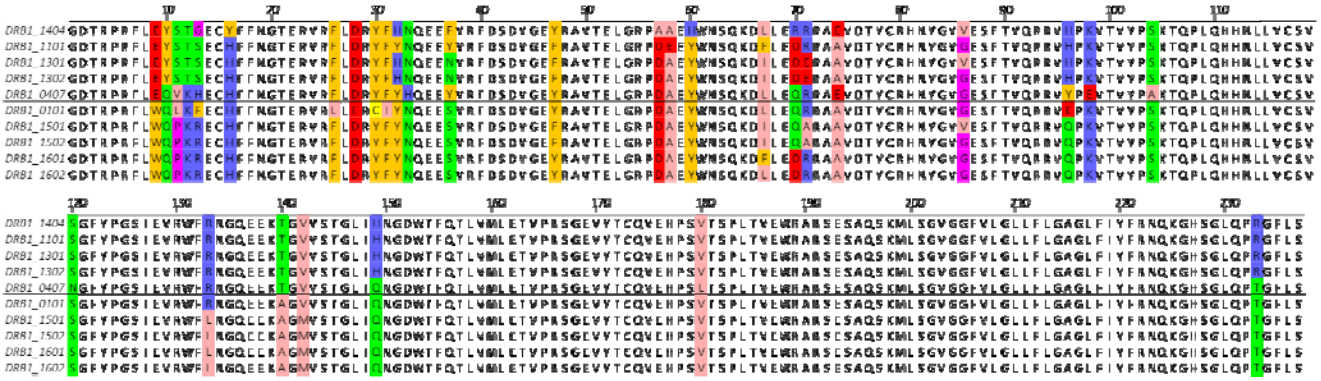


Figure 1. A multiple sequence alignment of the risk-associated HLA-DRB1 alleles (top 5) versus the protective-associated HLA-DRB1 alleles (bottom 5). Non-similar positions were colored using the physicochemical properties of the respective amino acids. Rose = hydrophobic/aliphatic; Orange = aromatic; Blue = positively charged; Red = negatively charged; Green = hydrophilic; Purple = conformationally special; Yellow = Cysteine.

Table 2. Functionally important observed amino acid substitutions are shown between protective (top 5, white) HLA alleles and risk (bottom 5, greyscale) alleles. Bolded residues denote amino acid variants shared across several alleles, and variants shared across remaining other alleles are underlined. The respective epitope binding pocket positions are denoted with Px. A = Alanine; C = Cysteine; D = Aspartic acid; E = Glutamic acid; F = Phenylalanine; G = Glycine; H = Histidine; I = Isoleucine; K = Lysine; L = Leucine; M = Methionine; N = Asparagine; P = Proline; Q = Glutamine; R = Arginine; S = Serine; T= Threonine; V = Valine; W = Tryptophan; Y = Tyrosine.

Sequence residue	9	10	11	12	13	26	28	30	37	67	70	71	74	86	96	133	140	142	149	233
DRB1*15:01	W	Q	P	K	R	F	D	Y	S	I	Q	A	A	V	Q	L	A	V	Q	T
DRB1*15:02	W	Q	P	K	R	F	D	Y	S	I	Q	A	A	G	Q	L	A	V	Q	T
DRB1*16:01	W	Q	P	K	R	F	D	Y	S	*F*	D	R	A	G	Q	L	A	V	Q	T
DRB1*16:02	W	Q	P	K	R	F	D	Y	S	L	D	R	A	G	Q	L	A	V	Q	T
DRB1*01:01	W	Q	L	K	F	L	E	C	S	L	Q	R	A	G	E	R	A	V	Q	T
DRB1*04:07	E	Q	V	K	H	F	D	Y	Y	L	Q	R	E	G	Y	R	T	V	Q	T
DRB1*11:01	E	Y	S	T	S	F	D	Y	Y	*F*	D	R	A	G	H	R	T	M	H	R
DRB1*13:01	E	Y	S	T	S	F	D	Y	N	I	D	*E*	A	V	H	R	T	M	H	R
DRB1*13:02	E	Y	S	T	S	F	D	Y	N	I	D	*E*	A	G	H	R	T	M	H	R
DRB1*14:04	E	Y	S	T	G	F	D	Y	F	L	R	R	E	V	H	R	T	M	H	R
Functional position	P9		P6		P4	P4	P4/7	P6	P9	P7/TCR	P4	P4/7	P4	P1		CD4 contact	CD4 contact	CD4 contact	CD4 contact	

The leishmaniasis protective- and risk-associated HLA-DRB1 alleles can be completely distinguished at residues 9, 37 and 140. Indeed, the protective-associated alleles share a tryptophan at residue 9, a serine on residue 37 and an alanine at residue 140. In sharp contrast, the risk-associated alleles share a glutamic acid at residue 9, a tyrosine at residue 140 and exhibit a range of different amino acids at residue 37. These substitutions confer distinct physicochemical properties on these functionally important residues. The shared tryptophan at residue 9 of the protective alleles is an aromatic and hydrophobic amino acid, while the glutamatic acid, shared across the risk alleles, is negatively charged. Similarly, the serine at residue 37 of the protective alleles confers a hydrophilic char-

acteristic. The risk alleles exhibit aromatic amino acids at this position, with the exception of DRB1*13:01 and DRB13:02 containing the hydrophilic asparagine. Finally, a hydrophobic alanine is shared across residue 140 of the protective alleles, while the risk alleles share a hydrophilic threonine. Residue 140 is postulated to interact with the CD4 co-receptor, and the distinct physicochemical properties at this residue may influence CD4 co-receptor binding and subsequent TCR-pMHC interactions [23]. Consequently, the identified amino acid substitutions might alter the binding affinity and stability of a peptide by only enabling specific amino acids to bind at these binding pocket positions.

3.3 Prediction of binding epitopes and their relative affinity

To explore the potential functional effects of the identified variants, and to provide a valuable resource of epitopes for future *Leishmania* immune research, binding epitopes were predicted across the above-defined protective- and risk-associated HLA profiles and the species corresponding to these alleles.

Predictions of 9mer HLA class I epitopes were performed for the complete *L. braziliensis, L. donovani* and *L. major* proteomes. Using a sliding window of 1 in the NetMHCpan tool, a total of 4,580,567, 5,226,608 and 5,326,280 HLA class I 9mer peptides were respectively derived from these proteomes. Even though HLA associations were also identified for *L. guyanensis,* this species was excluded from HLA class I predictions due to the unavailability of a high-quality reference proteome. With regards to HLA class II epitope predictions, 15mer peptides were derived from the proteomes of *L. braziliensis, L. donovani, L. infantum* and *L. mexicana*. A total of 4,535,501, 5,175,104, 5,144,590 and 4,996,408 HLA class II 15mer epitopes were derived from the respective proteomes, using a sliding window of 1 in the NetMHCIIpan tool.

We performed a series of subsequent filtering steps to narrow down these peptides to only those strong-binding epitopes that are likely to get presented in *vivo* with sufficiently high affinity to displace competitor self-peptides (see method section). Moreover, only those that were unique to each HLA allele (i.e. contained only one epitope of tandemly repeated gene duplicates), were considered to be strong-binding epitopes. The resulting strong-binding epitopes and their properties are listed in in a spreadsheet provided at https://github.com/BioinformaNicks/LeisHLA.

The relative affinity distributions of the strong-binding HLA class II epitopes were plotted for each species (Figure S2-3).

Using linear mixed effect models, no significant difference ($p < 0.05$) was observed between the relative affinity distributions of protective- and risk-associated alleles. Therefore, a general alteration in binding affinity for *Leishmania* epitopes conferred by amino acid substitutions in the HLA-DRB1 profile, probably does not explain the differences between the protective or risk status of the associated HLA alleles.

3.4 Relationship between amino acid substitutions in the HLA repertoire and epitope binding cores

Since the strength or affinity of the peptide-MHC-II binding complex is not correlated with the disease association status, we hypothesize that the amino acid substitutions across the HLA-DRB1 profiles might alter the epitope binding repertoires of the different HLA alleles. To explore this hypothesis, the binding cores of strong-binding epitopes were visualized, using consensus sequence logos, to identify binding motifs within the protective- or risk-associated HLA-II alleles.

Interestingly, epitopes that bind to protective-associated class II HLA-DRB1 alleles exhibit a common combination of motifs at binding core positions 4, 6 and 9 (Figure 2). In sharp contrast, this combination of motifs is not seen in risk-associated HLA-DRB1 alleles, which display a more promiscuous nature across the binding core positions. These findings overlap with the previously identified amino acid substitutions that correspond to these binding core positions, in the protective- and risk-associated HLA profiles, confirming a relationship between these variants and the (predicted) epitope binding repertoires.

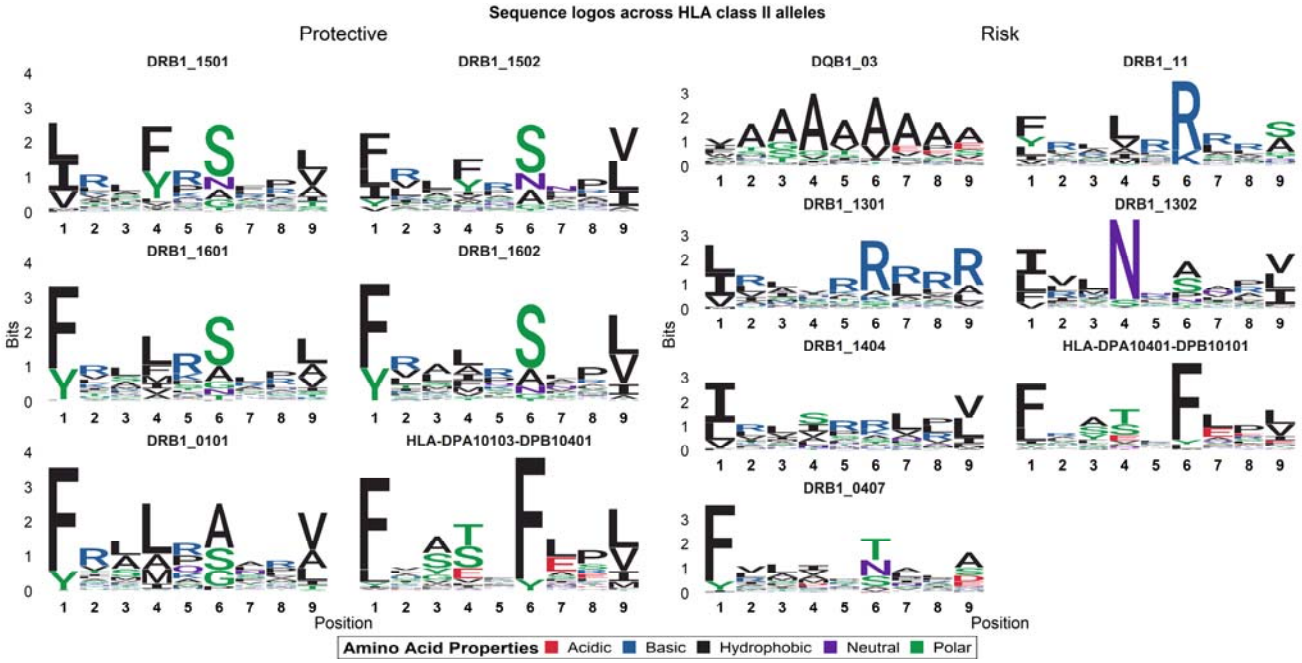


Figure 2. Consensus sequence logos of the 9mer binding cores of predicted epitopes across HLA class II alleles. This plot shows the sequence conservation level (in bits) of an amino acid occurring at a certain position in the binding core. Protective and risk-associated HLA alleles have been separated in a 2-column matrix.

Position 4 and 9 across protective-associated DRB1 alleles consist predominantly of similar hydrophobic and polar amino acids, while position 6 consist mainly of the SA(N)G motif. While the risk-associated DRB1*13:02 shares the combination of a SANG motif at position 6 and a hydrophobic motif at position 9, it does not share the hydrophobic motif on position 4. Within the protective-associated alleles, position 4 displays a unique Y motif for DRB1*15:01 and DRB1*15:02, which maps to a R71A amino acid substitution. This R71A amino acid variant, however, subsequently maps to the LAMI motif in binding core position 4 of the other protective-associated alleles. Similarly, DRB1*01:01 displays a SAG motif instead of the SANG motif seen across the rest of the protective-associated alleles. This lack of asparagine binding in the binding core of DRB1*01:01 maps to the P11L amino acid substitution.

Although position 1 of the protective-associated DRB1 alleles shares a consistent hydrophobic motif, this residue is found across all risk-associated DRB1 alleles as well. This finding potentially reflects the previously reported limited binding repertoire of the HLA-DRB1 pocket 1 position, generally enabling the binding of the following amino acids: FYLVM [49]. Moreover, the occurrence of the FY amino acids in the binding cores perfectly correlates with the G86V dimorphism in the HLA-DRB1 sequence.

As such, our data suggests that, at least for DRB1, protective-associated class II HLA alleles have similar binding core properties that discern them from risk-associated HLA alleles.

HLA class II epitope conservation across Leishmania species and HLA allotypes

The finding that protective-associated HLA-DRB1 alleles are consistently associated with leishmaniasis, and share common binding motifs, highlights that epitopes that bind to these HLA alleles are the most likely to confer a disease-specific immune response. Thus, these epitopes may be interesting candidates for inclusion in future epitope discovery studies.

To further prioritize the in silico predicted epitopes for future discovery, we set out to identify whether any epitopes were conserved across *Leishmania* spp. (Figure 3), and whether these epitopes bind across the different protective-associated HLA alleles (Figure 4-5).

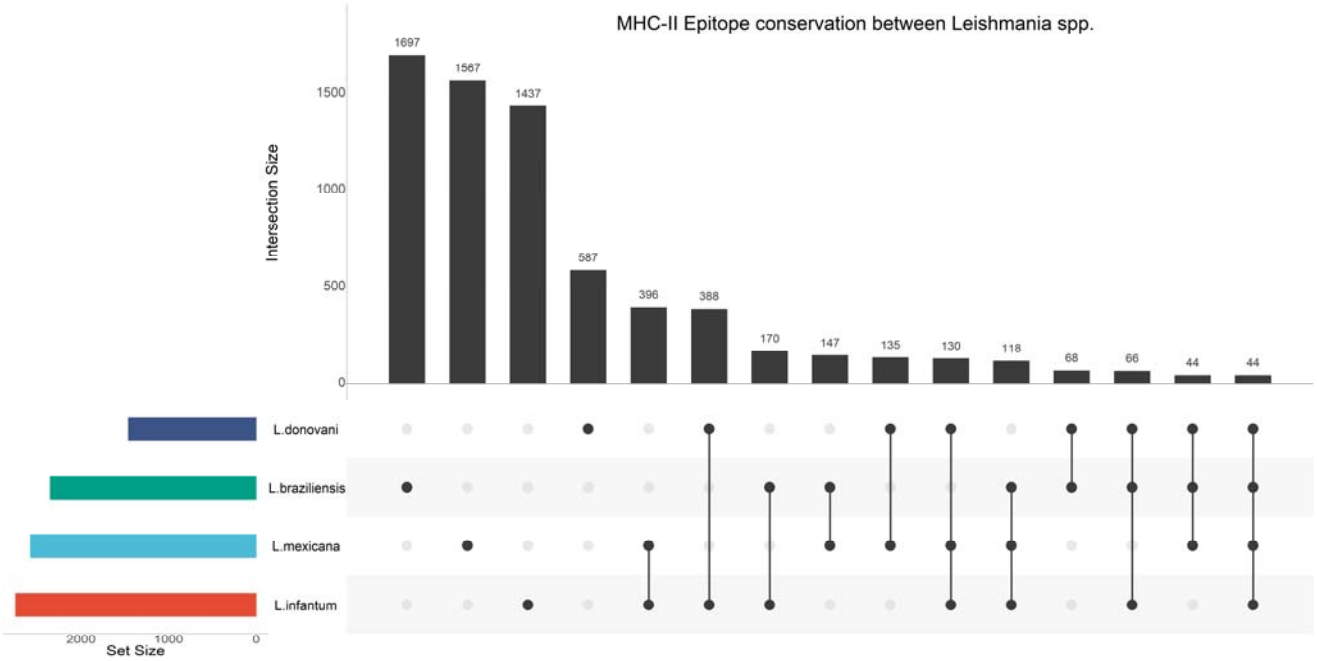


Figure 3. Upset Plot of strong-binding epitopes that are 1) unique to protective-associated alleles and 2) conserved across two major VL- (*L. infantum* and *L. donovani*) and two major CL-causing *Leishmania* species (*L. braziliensis* and *L. mexicana*). The set size denotes the total number of strong-binding epitopes predicted for each species, while the intersection size denotes the number of strong-binding epitopes shared across the species.

A total of 44 protective-associated strong-binding epitopes are conserved across *L. braziliensis, L. donovani, L. infantum* and *L. mexicana*. These epitopes and their properties are listed in an excel file provided at https://github.com/BioinformaNicks/LeisHLA.

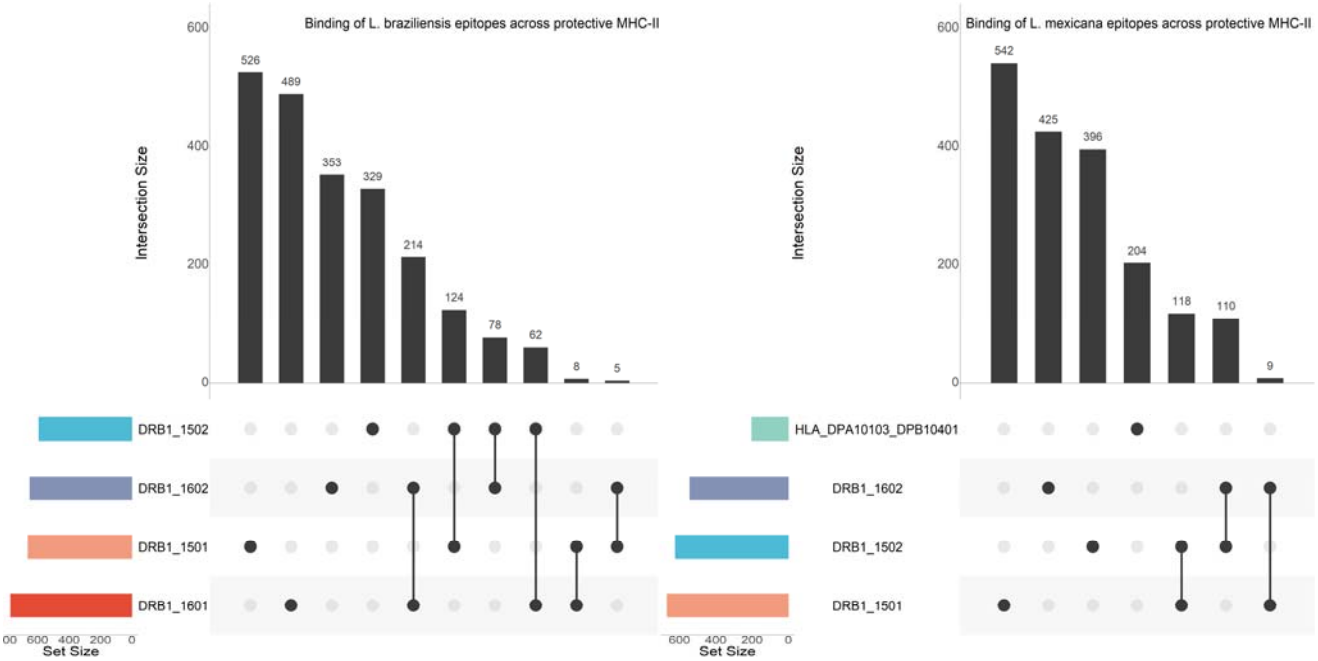


Figure 4. Upset Plots showing unique strong-binding epitopes of the CL-causing species (*L. braziliensis* and *L. mexicana*) that bind across the protective-associated alleles known for these species. The set size denotes the total number of strong-binding epitopes predicted for each species, while the intersection size denotes the number of strong-binding epitopes shared across the species.

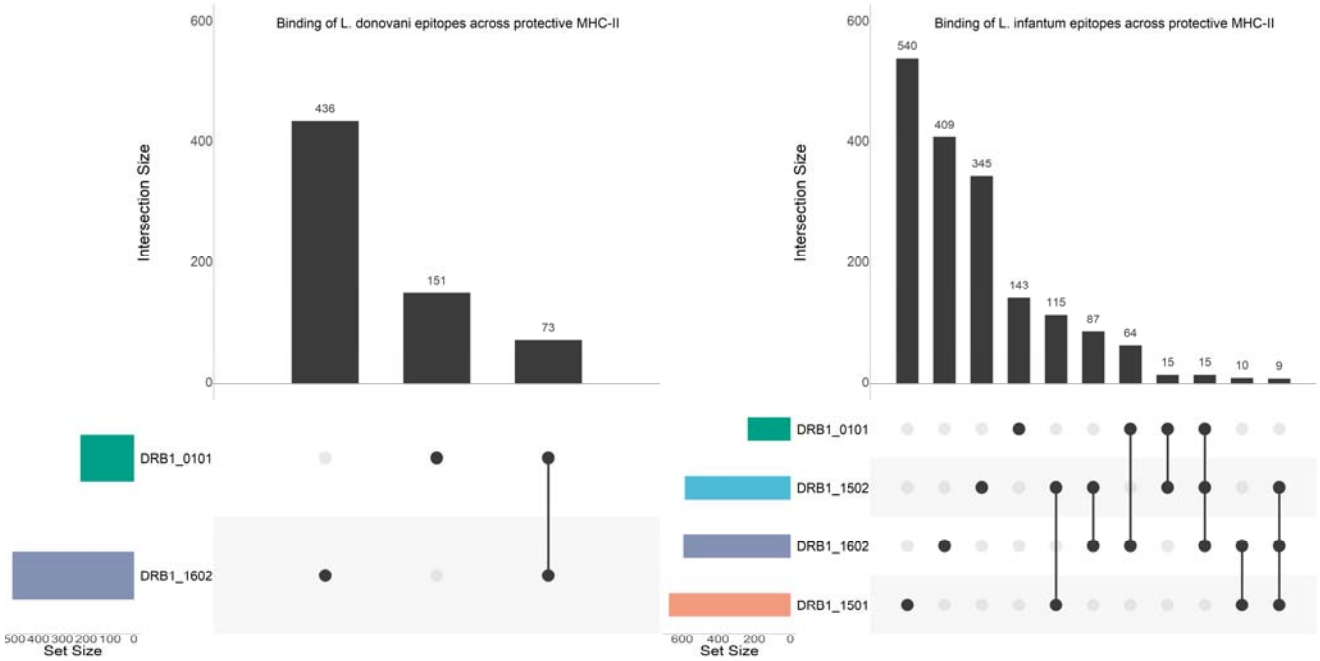


Figure 5. Upset Plots showing unique strong-binding epitopes of the VL-causing species (*L. donovani* and *L. infantum*) that bind across the protective-associated alleles known for these species. The set size denotes the total number of strong-binding epitopes predicted for each species, while the intersection size denotes the number of strong-binding epitopes shared across the species.

Lastly, Table 3 lists the strong-binding epitopes conserved across species that overlap with those that bind across different alleles. A total of 14 strong-binding epitopes are conserved across all species and bound across at least two different HLA allotypes (range 2-4).

Table 3. Prioritized in silico predicted strong-binding epitopes that are A) conserved across the included species and B) bind across different HLA alleles. Ranked by the number of HLA alleles and the protein these epitopes are derived from.

HLA alleles	Peptide	Binding Core	Protein
DRB1*15:01;DRB1*15:02;DRB1*16:01;DRB1*16:02	QTFTLFKSLRAHMLP	FTLFKSLRA	Eukaryotic translation initiation factor eIF2A
DRB1*16:01;DRB1*01:01;DRB1*16:02	TFTLFKSLRAHMLPL	FKSLRAHML	Eukaryotic translation initiation factor eIF2A
DRB1*16:01;DRB1*01:01;DRB1*16:02	FTLFKSLRAHMLPLT	FKSLRAHML	Eukaryotic translation initiation factor eIF2A
DRB1*16:01;DRB1*01:01;DRB1*16:02	HFTSYRHLPALRLLS	YRHLPALRL	hypothetical protein, conserved
DRB1*16:01;DRB1*01:01;DRB1*16:02	FTSYRHLPALRLLSA	YRHLPALRL	hypothetical protein, conserved
DRB1*16:01;DRB1*01:01;DRB1*16:02	NDEFHMLRSASIKII	FHMLRSASI	Carbamoyl-phosphate synthase small chain
DRB1*16:01;DRB1*01:01;DRB1*16:02	HRLFILLHGQPIAQS	FILLHGQPI	Ankyrin repeats
DRB1*16:01;DRB1*01:01;DRB1*16:02	RLFILLHGQPIAQSP	FILLHGQPI	Ankyrin repeats
DRB1*15:02;DRB1*01:01;DRB1*16:02	TNAFRLQLSNPIIFS	FRLQLSNPI	Ring finger domain containing protein
DRB1*15:02;DRB1*16:02	NPTNAFRLQLSNPII	FRLQLSNPI	Ring finger domain containing protein
DRB1*15:02;DRB1*16:02	PTNAFRLQLSNPIIF	FRLQLSNPI	Ring finger domain containing protein
DRB1*15:02;DRB1*16:02	NPGYISLFSTPIVKI	ISLFSTPIV	topoisomerase IV, subunit A
DRB1*15:02;DRB1*16:02	FIILPFIFIPSNTIS	FIFIPSNTI	ABC-2 family transporter protein
DRB1*15:02;DRB1*16:02	DRIFRSFRTNNVKMT	FRSFRTNNV	hypothetical protein, conserved

4. Discussion

We carried out the first multi-species and proteome-wide T cell epitope binding predictions for leishmaniasis-associated HLA alleles (6 protective-associated and 11 risk-associated) and further studied the underlying properties of HLA-DRB1 alleles associated with susceptibility or resistance to infection. We demonstrated that several functionally important amino acid variants may differentiate protective- from risk-associated HLA-DRB1 alleles. This further translated in common epitope binding motifs across the protective-associated alleles, altering the epitope binding repertoire, in sharp contrast to the risk-associated HLA profile which displayed a

promiscuous nature. Finally, this knowledge was used to identify species-conserved and HLA-unrestricted binding epitopes across the vast amount of possible epitopes. This prioritized list of epitopes can serve as guidance for the study of *Leishmania*-specific immunity and future epitope discovery, and may drive the development of novel subunit, multi-epitope vaccines.

To explore the mechanisms underpinning the HLA associations in leishmaniasis, we identified several amino acid variants that were completely shared across either the risk- or the protective-associated HLA-DRB1 alleles included in this study. These amino acid variants are thought to confer distinct (physicochemical and structural) properties on functionally important positions in the HLA-DRB1 structure, and this may ultimately impact the disease association status by altered antigen presentation. In a similar manner, although related to auto-immunity, three amino acid variants located in the binding groove of the HLA-DRB1 locus are known to mediate disease association in rheumatoid arthritis [51]. For leishmaniasis, the identified amino acid variants include residues 9, 37 and 140, which fully differentiate between the leishmaniasis protective- and risk-associated HLA alleles, but also residues 10-13, 96, 133, 142, 149, and 233 which were shared across the majority of the alleles within the association groups but not shared in one allele. In general, DRB1*04:07 and DRB1*01:01 were the alleles for which these exceptions were noted. Some degree of influence has been observed between the epitope binding pocket positions, where the binding of a specific amino acid at a certain binding pocket position alters the range of amino acids able to bind at another position [49]. DRB1*01:01 and DRB1*04:07 both exhibit several distinct mutations in binding pocket positions that may be permissive for variants at other positions. Residues 9 and 37 of the HLA-DRB1 sequence dictate the strength and stability of binding of the ninth amino acid of an epitope to the epitope binding pocket of the MHC molecule [23]. Residue 37 is the main contributor in shaping the electrostatic properties of position 9 of the epitope binding pocket [52]. At this residue, asparagine induces a positive charge while tyrosine induces a negative charge, ultimately restricting the range of distinct epitopes that are able to bind this pocket [52]. Residue 140 interacts with the CD4 co-receptor, and a hydrophobic alanine is shared across the protective-associated alleles while the risk-associated alleles share a hydrophilic threonine. This indicates that these structural and physicochemical variations may alter the epitope binding repertoire and explain the disease association status [53].

Earlier work of our lab on the Varicella-Zoster virus (VZV) demonstrated that HLA class I alleles associated with risk for postherpetic neuralgia demonstrated a significantly lower affinity for VZV-derived epitopes than those associated with lower risk [17]. Thus, we postulated that the structural and physicochemical variations identified in leishmaniasis-associated HLA alleles would alter their binding affinity. Yet, the epitope binding affinity distributions were not significantly different between protective- and risk-associated HLA alleles in leishmaniasis. However, we employed leishmaniasis-wide relative affinity comparisons, treating all *Leishmania*-derived epitopes as equals, while studies for other pathogens have shown that disease susceptibility is more directly associated with epitope-affinity within specific timely-expressed proteins rather than the full proteome [54]. Consequentially, the effect of the HLA disease-association on the binding affinity of (not yet known) leishmaniasis epitopes, that play a key role in the immune response, may be masked by predicted epitopes that are not functionally important. This can, however, not yet be verified as a consequence of a lack of well-characterized immune response proteins and derived epitopes.

We confirmed a common combination of binding motifs in the binding cores of predicted epitopes that bind to the known protective-associated HLA alleles. These common binding motifs at position 4, 6 and 9 of the binding core map to several of the identified variants in the amino acid sequence of the HLA-DRB1 alleles. In a prior attempt to identify the effect of leishmaniasis-associated HLA variants on antigen presentation, Singh et al. characterized the binding pocket preferences of one protection-associated allele and one susceptibility-associated allele for *L. donovani*-derived epitopes of a subset of immunogenic proteins [40]. Similar to our findings, they demonstrate that the binding pocket positions 4 and 6 of the protective-associated HLA-DRB1 allele exhibit preference for hydrophobic and polar amino acids, while the binding pocket positions of the risk-associated HLA-DRB1 allele preferred basic amino acids and were generally more promiscuous. To minimize the confounding effect of natural variation in the highly polymorphic HLA locus, we successfully expanded on their findings in only 2 disease-associated HLA alleles, replicating it across 10 different disease-associated HLA-DRB1 alleles and whole proteomes of multiple *Leishmania* species.
Additionally, Singh et al demonstrated that *in vitro* stimulation of whole blood derived from cured patients, homozygous for the disease-associated HLA alleles, with peptides captured from the protective-associated HLA-DRB1*15:01 allele resulted in a higher IFN-γ to IL-10 ratio response than those captured from the risk-associated HLA-DRB1*13:01 allele. IFN-γ is linked to increased parasite clearance through increased parasitotoxic nitrate production [6, 55]. This finding could indicate that preferential binding of an epitope to a protective- or risk-associated allele can elicit a protective- or risk-associated T cell response. This means dis-

ease-associated epitopes may be directly wheeled for generating insight in functional *Leishmania*-specific T cell immune responses, and to guide the immunological evaluation of these T cell responses in future vaccine trials.

Our unbiased epitope prediction approach used stringent filter criteria to identify only those epitopes that are likely to get presented *in vivo.* This approach still yielded several epitopes from antigens known to be immunogenic (atleast in animal models), including for example amastin, kinesin, 60S ribosomal subunit L31, GP63 and LeIF (see supplementary spreadsheet) [56-61]. Our epitope candidate prioritization method, which placed preference on conservancy across species and promiscuous binding across alleles, resulted in several epitopes of antigens we speculate to be immunogenic in humans. More specifically, several peptides derived from LeIF2A were predicted to be both conserved across all species and to bind across different protective-associated HLA alleles. Of note, the LeIF variant commonly thought to be immunogenic in animals is typically LeIF4A instead of LeIF2A. Moreover, an epitope of Ribosomal protein S3 was predicted to be conserved and promiscuous as well. This antigen was shown to be involved in the Th1/Th2 skewing of the immune response in a BALB/c mice model [62]. Thus, we postulate that with the *in silico* predicted and prioritized epitopes, generated by this study, we provide an important resource for future applications in rational vaccine discovery. Before these epitopes are included for further vaccine candidate prioritization, broader promiscuous binding than to protective-associated HLA alleles needs to be tested to acquire a wide population coverage. It is not known whether any possible protective effect elicited by a protective-associated epitope is fully restricted to protective-associated HLA alleles, or if these epitopes elicit an immunodominant response in protective-associated HLA alleles and a subdominant response in non-disease associated allotypes. In addition, the subset of the population that requires protection by means of induced immunity through vaccination may differ from those that carry protective-associated HLA allotypes. In this case, a multi-epitope vaccine including multiple subdominant epitopes may be enough to elicit protective effects in a broader population.

Although we demonstrate a feasible approach for generating insight in the mechanisms that contribute to HLA disease-associations, this approach is heavily dependent on the quality of prior HLA association studies. The HLA association studies included in this work are limited in number, and vary considerably in species studied, study age, statistical power and methods of HLA typing. The methods of HLA typing used in the earlier included studies are often restricted to either class I or class II, and are low resolution enabling only typing up to the level of the allele group [63]. It is also worth mentioning that the majority of these HLA association studies are restricted to the South-American continent, and little data has been generated for other endemic regions such as the African continent. Future endeavors that cover this research gap are warranted. Nevertheless, despite the small number of HLA association studies, of variable quality, conducted to date, we were still able to generate valuable knowledge on the antigen presentation in leishmaniasis.

5. Conclusions

The data presented in this study strongly suggests that polymorphism in the HLA-DRB1 region, and the resulting variation in the epitope binding pocket, mediates the association between HLA-DRB1 and leishmaniasis disease outcome. Protective-associated HLA alleles display common epitope binding motifs, and these motifs map to amino acid substitutions largely or fully shared across these alleles. This knowledge can be used to support future evaluations of epitope-specific immune responses in *Leishmania*. Moreover, our in silico prioritization of potential T cell epitopes can be directly wheeled as a resource for future reverse-vaccinology approaches, guiding the development of novel subunit multi-epitope vaccines.

Author Contributions: Conceptualization: B.C., N.D.V., P.M and S.G.; Methodology: B.C., N.D.V., P.M and S.G.; Formal analysis and coding: N.D.V.; Investigation: B.C., N.D.V., P.M.; Data curation: N.D.V.; Writing – original draft preparation: B.C., N.D.V. and P.M.; Writing – review and editing: B.C., K.L., N.D.V., P.M., S.G. and W.A.; Visualization: N.D.V.; Supervision: B.C., K.L., P.M., S.G. and W.A.; Project administration: K.L.; Funding acquisition: B.C., K.L., P.M. and W.A.; All authors have read and agreed to the published version of the manuscript.

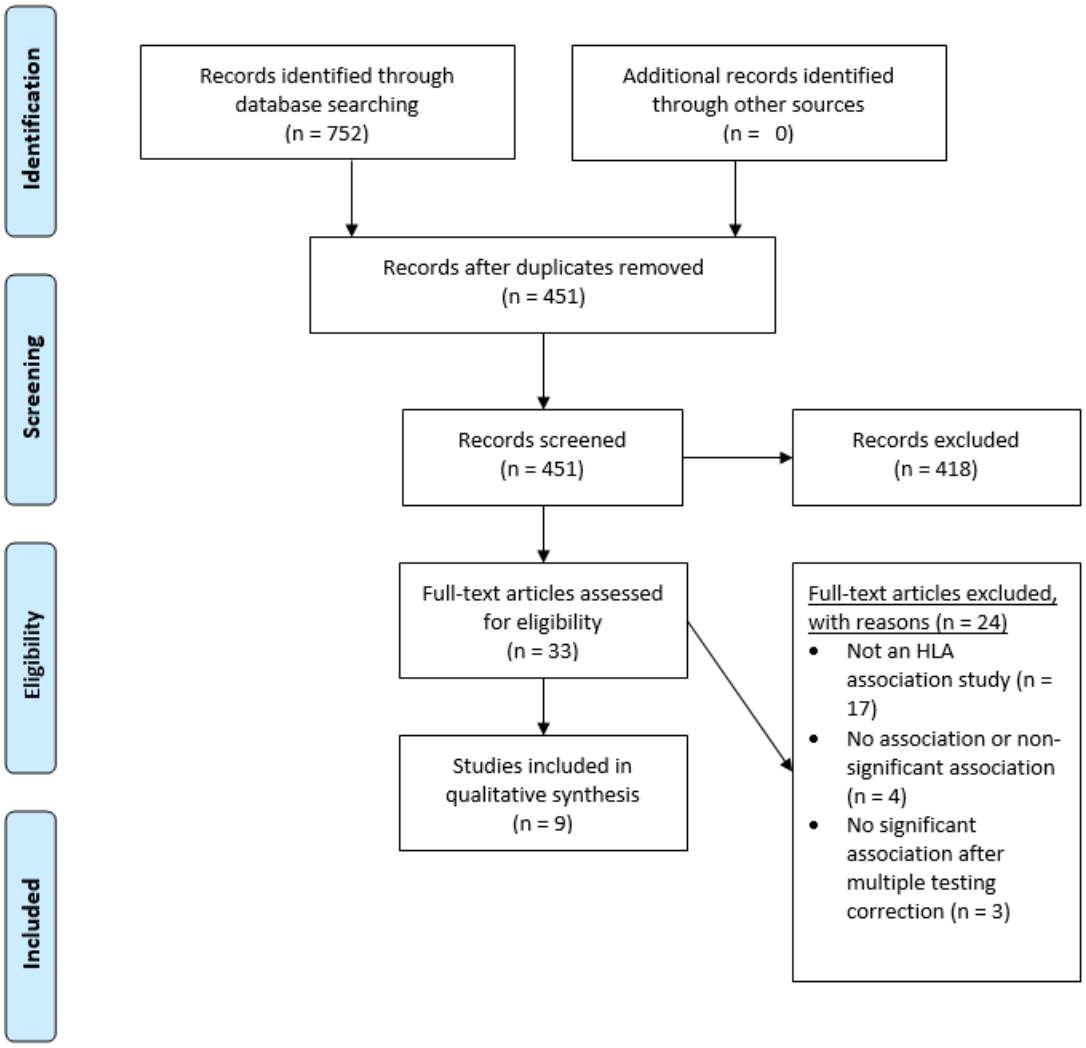
Funding: This work was supported by the Research Foundation Flanders [1S71721N to N.D.V., 1S48819N to S.G. and 11O1614N to B.C.]

Data Availability Statement: Data and code are provided at a Github repository for this project: https://github.com/BioinformaNicks/LeisHLA.

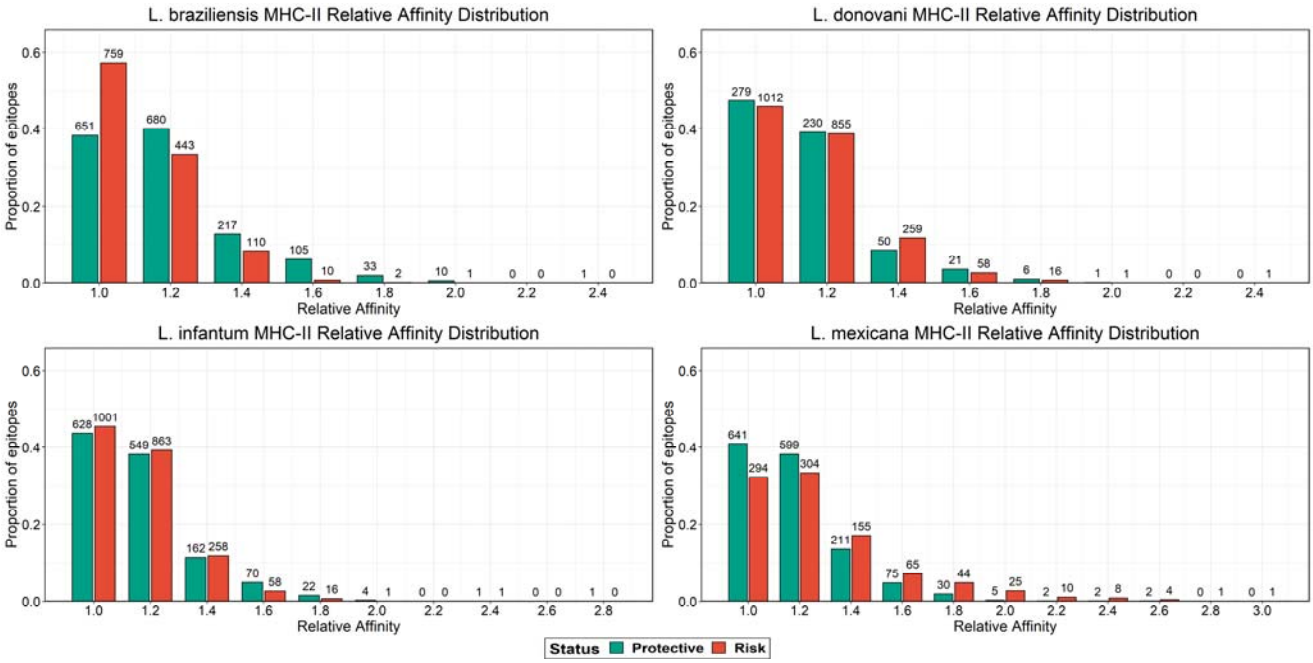
Acknowledgments: The computational resources used for this work were provided by the VSC (Flemish Supercomputer Center) at the University of Antwerp.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results

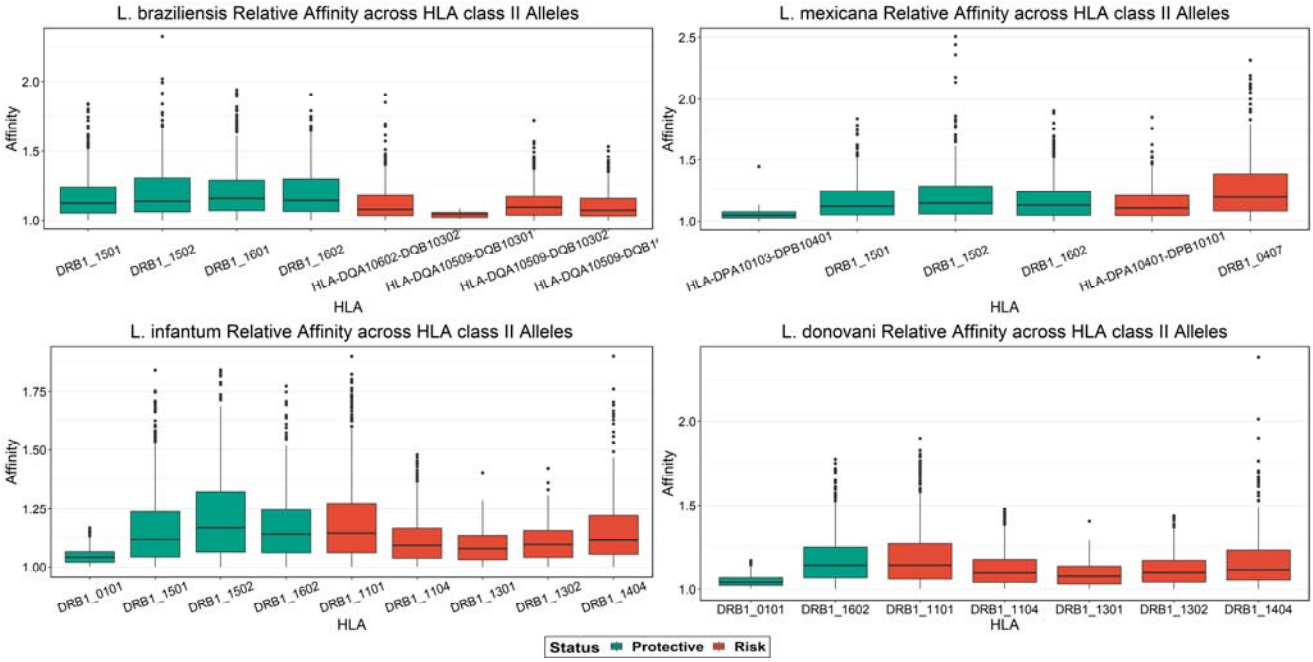
Supplementary figures



Supplementary Figure 1. The PRISMA Flow Diagram used in this study to describe the literature review workflow [22].



Supplementary Figure 2. Relative affinity distribution plots for the HLA class II alleles associated with L. braziliensis, L. donovani, L. infantum and L. mexicana. This plot shows the number of epitopes, and the proportion of epitopes relative to the total number of epitopes, with a specific relative affinity.



Supplementary Figure 3. Relative affinity distribution across HLA class II alleles associated with protection and risk for L. braziliensis, L. donovani, L. infantum and L. mexicana.

References

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. PLoS One. 2012;7(5):e35671. Epub 2012/06/14. doi: 10.1371/journal.pone.0035671.

2. WHO. Surveillance of leishmaniasis in the WHO European Region, 2016 and Global leishmaniasis surveillance update, 1998–2016. WHO. 2018;40: 521–540.

3. Escobar LE, Romero-Alvarez D, Leon R, Lepe-Lopez MA, Craft ME, Borbor-Cordova MJ, et al. Declining Prevalence of Disease Vectors Under Climate Change. Scientific Reports. 2016;6(1):39150. doi: 10.1038/srep39150.

4. Kholoud K, Denis S, Lahouari B, El Hidan MA, Souad B. Management of Leishmaniases in the Era of Climate Change in Morocco. International journal of environmental research and public health. 2018;15(7). Epub 2018/07/25. doi: 10.3390/ijerph15071542.

5. Semenza JC, Suk JE. Vector-borne diseases and climate change: a European perspective. FEMS Microbiol Lett. 2018;365(2). Epub 2017/11/18. doi: 10.1093/femsle/fnx244.

6. Kaye PM, Cruz I, Picado A, Van Bocxlaer K, Croft SL. Leishmaniasis immunopathology—impact on design and use of vaccines, diagnostics and drugs. Seminars in Immunopathology. 2020;42(3):247-264. doi: 10.1007/s00281-020-00788-y.

7. Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. Lancet. 2005;366(9496):1561-1577. Epub 2005/11/01. doi: 10.1016/S0140-6736(05)67629-5.

8. Aliaga L, Ceballos J, Sampedro A, Cobo F, Lopez-Nevot MA, Merino-Espinosa G, et al. Asymptomatic Leishmania infection in blood donors from the Southern of Spain. Infection. 2019;47(5):739-747. Epub 2019/03/20. doi: 10.1007/s15010-019-01297-3.

9. Fakhar M, Motazedian MH, Hatam GR, Asgari Q, Kalantari M, Mohebali M. Asymptomatic human carriers of Leishmania infantum: possible reservoirs for Mediterranean visceral leishmaniasis in southern Iran. Ann Trop Med Parasitol. 2008;102(7):577-583. Epub 2008/09/27. doi: 10.1179/136485908X337526.

10. Torres-Guerrero E, Quintanilla-Cedillo MR, Ruiz-Esmenjaud J, Arenas R. Leishmaniasis: a review. F1000Res. 2017;6:750. Epub 2017/06/27. doi: 10.12688/f1000research.11120.1.

11. Blackwell JM, Fakiola M, Castellucci LC. Human genetics of leishmania infections. Hum Genet. 2020;139(6-7):813-819. Epub 2020/02/15. doi: 10.1007/s00439-020-02130-w.

12. Samaranayake N, Fernando SD, Neththikumara NF, Rodrigo C, Karunaweera ND, Dissanayake VH. Association of HLA class I and II genes with cutaneous leishmaniasis: a case control study from Sri Lanka and a systematic review. BMC Infect Dis. 2016;16:292. Epub 2016/06/16. doi: 10.1186/s12879-016-1626-8.

13. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Res. 2020;48(D1):D948-D955. Epub 2019/11/02. doi: 10.1093/nar/gkz950.

14. Tripathi P, Singh V, Naik S. Immune response to leishmania: paradox rather than paradigm. FEMS Immunology & Medical Microbiology. 2007;51(2):229-242. doi: 10.1111/j.1574-695X.2007.00311.x.

15. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Research. 2018;47(D1):D339-D343. doi: 10.1093/nar/gky1006.

16. Nielsen M, Andreatta M, Peters B, Buus S. Immunoinformatics: Predicting Peptide–MHC Binding. Annual Review of Biomedical Data Science. 2020;3(1):191-215. doi: 10.1146/annurev-biodatasci-021920-100259.

17. Meysman P, Ogunjimi B, Naulaerts S, Beutels P, Van Tendeloo V, Laukens K. Varicella-zoster virus-derived major histocompatibility complex class I-restricted peptide affinity is a determining factor in the HLA risk profile for the development of postherpetic neuralgia. J Virol. 2015;89(2):962-969. Epub 2014/10/31. doi: 10.1128/jvi.02500-14.

18. Aebischer T. Leishmania spp. Proteome Data Sets: A Comprehensive Resource for Vaccine Development to Target Visceral Leishmaniasis. Front Immunol. 2014;5:260. Epub 2014/06/25. doi: 10.3389/fimmu.2014.00260.

19. Akya A, Farasat A, Ghadiri K, Rostamian M. Identification of HLA-I restricted epitopes in six vaccine candidates of Leishmania tropica using immunoinformatics and molecular dynamics simulation approaches. Infect Genet Evol. 2019;75:103953. Epub 2019/07/10. doi: 10.1016/j.meegid.2019.103953.

20. Hamrouni S, Bras-Goncalves R, Kidar A, Aoun K, Chamakh-Ayari R, Petitdidier E, et al. Design of multi-epitope peptides containing HLA class-I and class-II-restricted epitopes derived from immunogenic Leishmania proteins, and evaluation of CD4+and CD8+T cell responses induced in cured cutaneous leishmaniasis subjects. Plos Neglected Tropical Diseases. 2020;14(3). doi: 10.1371/journal.pntd.0008093.

21. Singh G, Pritam M, Banerjee M, Singh AK, Singh SP. Designing of precise vaccine construct against visceral leishmaniasis through predicted epitope ensemble: A contemporary approach. Comput Biol Chem. 2020;86:107259. Epub 2020/04/28. doi: 10.1016/j.compbiolchem.2020.107259.

22. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine. 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097.

23. Bondinas GP, Moustakas AK, Papadopoulos GK. The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics. 2007;59(7):539-553. Epub 2007/05/15. doi: 10.1007/s00251-007-0224-8.

24. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol [Internet]. 2011 2011/10//; 7:[539 p.]. Available from: http://europepmc.org/abstract/MED/21988835
https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/21988835/?tool=EBI
https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/21988835/pdf/?tool=EBI
https://doi.org/10.1038/msb.2011.75
https://europepmc.org/articles/PMC3261699
https://europepmc.org/articles/PMC3261699?pdf=render.

25. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-1191. doi: 10.1093/bioinformatics/btp033.

26. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Research. 2009;38(suppl_1):D457-D462. doi: 10.1093/nar/gkp851.

27. Dumetz F, Imamura H, Sanders M, Seblova V, Myskova J, Pescher P, et al. Modulation of Aneuploidy in Leishmania donovani during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. mBio. 2017;8(3):e00599-00517. doi: 10.1128/mBio.00599-17.

28. González-Galarza Faviel F, Takeshita Louise YC, Santos Eduardo JM, Kempson F, Maia Maria Helena T, Silva Andrea Luciana Soares d, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Research. 2014;43(D1):D784-D788. doi: 10.1093/nar/gku1166.

29. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 2020;48(W1):W449-W454. Epub 2020/05/15. doi: 10.1093/nar/gkaa379.

30. Trolle T, McMurtrey CP, Sidney J, Bardet W, Osborn SC, Kaever T, et al. The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. J Immunol. 2016;196(4):1480-1487. Epub 2016/01/20. doi: 10.4049/jimmunol.1501721.

31. Paul S, Karosiene E, Dhanda SK, Jurtz V, Edwards L, Nielsen M, et al. Determination of a Predictive Cleavage Motif for Eluted Major Histocompatibility Complex Class II Ligands. Front Immunol. 2018;9:1795. Epub 2018/08/22. doi: 10.3389/fimmu.2018.01795.

32. Cuypers B, Berg M, Imamura H, Dumetz F, De Muylder G, Domagalska MA, et al. Integrated genomic and metabolomic profiling of ISC1, an emerging Leishmania donovani population in the Indian subcontinent. Infect Genet Evol. 2018;62:170-178. Epub 2018/04/22. doi: 10.1016/j.meegid.2018.04.021.

33. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. 2017. 2017;82(13):26. Epub 2017-11-29. doi: 10.18637/jss.v082.i13.

34. Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics. 2017;33(22):3645-3647. doi: 10.1093/bioinformatics/btx469.

35. Sidney J, Peters B, Sette A. Epitope prediction and identification- adaptive T cell responses in humans. Semin Immunol. 2020;50:101418. Epub 2020/11/03. doi: 10.1016/j.smim.2020.101418.

36. Fakiola M, Strange A, Cordell HJ, Miller EN, Pirinen M, Su Z, et al. Common variants in the HLA-DRB1-HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis. Nat Genet. 2013;45(2):208-213. Epub 2013/01/08. doi: 10.1038/ng.2518.

37. Lara ML, Layrisse Z, Scorza JV, Garcia E, Stoikow Z, Granados J, et al. Immunogenetics of human American cutaneous leishmaniasis. Study of HLA haplotypes in 24 families from Venezuela. Hum Immunol. 1991;30(2):129-135. Epub 1991/02/01. doi: 10.1016/0198-8859(91)90081-j.

38. Hollenbach JA, Mack SJ, Thomson G, Gourraud PA. Analytical methods for disease association studies with immunogenetic data. Methods Mol Biol. 2012;882:245-266. Epub 2012/06/06. doi: 10.1007/978-1-61779-842-9_14.

39. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, et al. Genetic analysis of Leishmania donovani tropism using a naturally attenuated cutaneous strain. PLoS Pathog. 2014;10(7):e1004244. Epub 2014/07/06. doi: 10.1371/journal.ppat.1004244.

40. Singh T, Fakiola M, Oommen J, Singh AP, Singh AK, Smith N, et al. Epitope-Binding Characteristics for Risk versus Protective DRB1 Alleles for Visceral Leishmaniasis. J Immunol. 2018;200(8):2727-2737. Epub 2018/03/07. doi: 10.4049/jimmunol.1701764.

41. Faghiri Z, Tabei SZ, Taheri F. Study of the association of HLA class I antigens with kala-azar. Hum Hered. 1995;45(5):258-261. Epub 1995/09/01. doi: 10.1159/000154309.

42. Petzl-Erler ML, Belich MP, Queiroz-Telles F. Association of Mucosal Leishmaniasis with HLA. Human Immunology. 1991;32(4):254-260. doi: https://doi.org/10.1016/0198-8859(91)90088-Q.

43. Ribas-Silva RC, Ribas AD, Ferreira EC, Silveira TG, Borelli SD. Association between HLA-C*04 and American cutaneous leishmaniasis in endemic region of southern Brazil. Genet Mol Res. 2015;14(4):14929-14935. Epub 2015/11/26. doi: 10.4238/2015.November.18.58.

44. Olivo-Díaz A, Debaz H, Alaez C, Islas VJ, Pérez-Pérez H, Hobart O, et al. Role of HLA class II alleles in susceptibility to and protection from localized cutaneous leishmaniasis. Human Immunology. 2004;65(3):255-261. doi: https://doi.org/10.1016/j.humimm.2003.12.008.

45. El-Mogy MH, Abdel-Hamid IA, Abdel-Razic MM, Rizk RA, Romia SA. Histocompatibility antigens in Egyptians with cutaneous leishmaniasis: a preliminary study. Journal of Dermatological Science. 1993;5(2):89-91. doi: https://doi.org/10.1016/0923-1811(93)90075-Z.

46. Lara ML, Layrisse Z, Scorza JV, Garcia E, Stoikow Z, Granados J, et al. Immunogenetics of human american Cutaneous Leishmaniasis: Study of HLA haplotypes in 24 families from Venezuela. Human Immunology. 1991;30(2):129-135. doi: https://doi.org/10.1016/0198-8859(91)90081-J.

47. Hussain SK, Makgoeng SB, Everly MJ, Goodman MT, Martinez-Maza O, Morton LM, et al. HLA and Risk of Diffuse Large B cell Lymphoma After Solid Organ Transplantation. Transplantation. 2016;100(11):2453-2460. Epub 2016/10/25. doi: 10.1097/TP.0000000000001025.

48. Ferrante A, Gorski J. Cooperativity of Hydrophobic Anchor Interactions: Evidence for Epitope Selection by MHC Class II as a Folding Process. The Journal of Immunology. 2007;178(11):7181-7189. doi: 10.4049/jimmunol.178.11.7181.

49. James EA, Moustakas AK, Bui J, Nouv R, Papadopoulos GK, Kwok WW. The Binding of Antigenic Peptides to HLA-DR Is Influenced by Interactions between Pocket 6 and Pocket 9. The Journal of Immunology. 2009;183(5):3249-3258. doi: 10.4049/jimmunol.0802228.

50. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature. 1993;364(6432):33-39. doi: 10.1038/364033a0.

51. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee H-S, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature Genetics. 2012;44(3):291-296. doi: 10.1038/ng.1076.

52. Hov JR, Kosmoliaptsis V, Traherne JA, Olsson M, Boberg KM, Bergquist A, et al. Electrostatic modifications of the human leukocyte antigen-DR P9 peptide-binding pocket and susceptibility to primary sclerosing cholangitis. Hepatology. 2011;53(6):1967-1976. Epub 2011/03/18. doi: 10.1002/hep.24299.

53. Sollid LM, Pos W, Wucherpfennig KW. Molecular mechanisms for contribution of MHC molecules to autoimmune diseases. Curr Opin Immunol. 2014;31:24-30. Epub 2014/09/13. doi: 10.1016/j.coi.2014.08.005.

54. Meysman P, De Neuter N, Bartholomeus E, Elias G, Van den Bergh J, Emonds M-P, et al. Increased herpes zoster risk associated with poor HLA-A immediate early 62 protein (IE62) affinity. Immunogenetics. 2018;70(6):363-372. doi: 10.1007/s00251-017-1047-x.

55. Rossi M, Fasel N. How to master the host immune system? Leishmania parasites have the solutions! International Immunology. 2017;30(3):103-111. doi: 10.1093/intimm/dxx075.

56. Dey A, Sharma P, Redhu NS, Singh S. Kinesin motor domain of Leishmania donovani as a future vaccine candidate. Clin Vaccine Immunol. 2008;15(5):836-842. Epub 2008/03/21. doi: 10.1128/cvi.00433-07.

57. Dikhit MR, Das S, Mahantesh V, Kumar A, Singh AK, Dehury B, et al. The potential HLA Class I-restricted epitopes derived from LelF and TSA of Leishmania donovani evoke anti-leishmania CD8+ T lymphocyte response. Sci Rep. 2018;8(1):14175. Epub 2018/09/23. doi: 10.1038/s41598-018-32040-x.

58. ALI SA, REZVAN H, MCARDLE SE, KHODADADI A, ASTEAL FA, REES RC. CTL responses to Leishmania mexicana gp63-cDNA vaccine in a murine model. Parasite Immunology. 2009;31(7):373-383. doi: https://doi.org/10.1111/j.1365-3024.2009.01111.x.

59. Rezvan H. Immunogenicity of HLA-DR1 Restricted Peptides Derived from Leishmania major gp63 Using FVB/N-DR1 Transgenic Mouse Model. Iran J Parasitol. 2013;8(2):273-279. Epub 2013/08/06.

60. Silva LP, Paciello MO, Aviz Teixeira WP, Rivas AV, Agular RWS, Cangussu ASR, et al. Immunogenicity of HLA-DR1 and HLA-A2 peptides derived from Leishmania major Gp63 in golden hamsters. Parasite Immunol. 2020;42(12):e12780. Epub 2020/08/02. doi: 10.1111/pim.12780.

61. Stober CB, Lange UG, Roberts MT, Gilmartin B, Francis R, Almeida R, et al. From genome to vaccines for leishmaniasis: screening 100 novel vaccine candidates against murine Leishmania major infection. Vaccine. 2006;24(14):2602-2616. Epub 2006/01/13. doi: 10.1016/j.vaccine.2005.12.012.

62. Cordeiro-Da-Silva A, Borges MC, Guilvard E, Ouaissi A. Dual role of the Leishmania major ribosomal protein S3a homologue in regulation of T- and B-cell activation. Infect Immun. 2001;69(11):6588-6596. Epub 2001/10/13. doi: 10.1128/iai.69.11.6588-6596.2001.

63. Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens. 2012;80(1):1-11. doi: 10.1111/j.1399-0039.2012.01881.x.