

**Functional characterization of antigen repertoires in
HLA-associated complex diseases to investigate antagonistic
selection on HLA genes**



Dissertation
in fulfilment of the requirements for the degree
“Doctor rerum naturalium”
of the Faculty of Mathematics and Natural Sciences
at the Christian Albrechts University of Kiel

Submitted by
Jatin Arora
Emmy Noether group for Evolutionary Immunogenomics
Max Planck Institute for Evolutionary Biology

Plön, December 2018

First referee: Dr. Tobias Lenz

Second referee: Prof. Tal Dagan

Date of oral examination: 22.02.2019

Table of content

<i>Summary</i>	5
<i>Zusammenfassung</i>	7
<i>Introduction</i>	10
<i>Thesis outline</i>	18
<i>Chapter 1</i>	21
Introduction	21
Results	23
Discussion	31
Material and methods	33
<i>Chapter 2</i>	40
<i>Chapter 3</i>	70
Introduction	70
Results	72
Discussion	78
Material and methods	81
<i>Chapter 4</i>	85
Introduction	85
Results	86
Discussion	92
Material and methods	94
<i>Conclusion</i>	113
<i>Perspectives</i>	117
<i>References</i>	118
<i>Annex</i>	126
<i>Acknowledgements</i>	143
<i>Curriculum vitae</i>	144
<i>Declaration</i>	145

Summary

The interaction between hosts and pathogens represents a major driver of their evolution, and a core interest in the evolutionary biology. In vertebrate hosts, the classical genes of Major Histocompatibility Complex (MHC) make a central component of their adaptive immune system. The cell-surface molecules encoded by the MHC genes present peptide fragments (derived from both self-antigens and pathogens) to T-cells, which upon recognizing them as foreign, initiate a specific immune response. To explore the general MHC evolution, I used humans as a study system where MHC is designated as Human Leukocyte Antigen (HLA). Although a vast allelic diversity of HLA genes exists at the population level, potentially maintained by pathogen-mediated balancing selection, only a fraction of that is seen at the individual level in the form of a limited number of HLA genes and their alleles. This setting has been proposed as an optimum between HLA-conferred resistance to pathogens and the risk of autoimmunity, which represent major antagonistic selection forces on HLA genes. The fine-mapping of various HLA's association with different infectious and autoimmune diseases has suggested a major role of the peptide-repertoire of HLA alleles in determining their specific effect on diseases. However, the exact mechanisms that underlie HLA's association with the diseases, and by that modulate the antagonistic selection on HLA genes remain elusive.

In order to elucidate them, I started by investigating the functional basis of the previously known protective effect of HLA heterozygosity at the HLA class-I genes on HIV-1 progression. I used a dataset of 6,311 HIV-1 infected individuals and predicted the HLA-bound peptides derived from HIV-1 proteome for each HLA allele represented in the dataset. The individual-specific repertoire of HLA-bound peptides suggested that HLA heterozygote advantage against HIV-1 could be mediated by both a broader array of HLA-bound peptides and a higher likelihood of carrying specific protective alleles in heterozygotes compared to homozygotes. The comparison of the peptide-repertoire of risk and protective alleles suggested that individual alleles could confer disease control by binding either a large number of peptides or specific immunodominant peptides. The separate analysis of the individual HLA genes indicated that different mechanism for the heterozygote advantage might work at different genes, such as either T-cell or NK-cell-mediated immune attack on the virus, possibly resulting in different evolutionary constraints on different HLA genes. Overall, the findings suggested that the pathogen-mediated selection might favor both HLA heterozygosity and individual alleles.

Further, hypothesizing that not all HLA-bound peptides would be relevant for disease control, we developed a new approach, named Peptidome-wide association study (PepWAS), that can predict HLA-bound disease-associated epitopes from a given peptidome. The PepWAS-predicted HIV-1-associated epitopes accounted for as much variation (12%) in HIV-1 viral load as by the genetic variants in HLA class-I genes, providing a functional basis for the association between HLA and HIV-1 control.

I then focused on the association between HLA class-II genes and Type 1 Diabetes (T1D). Using a case-control dataset of 16,029 individuals, I first showed that heterozygosity at the HLA class-II genes conferred T1D risk. To investigate functional basis of this HLA heterozygote disadvantage, I predicted individual-specific repertoires of HLA-bound peptides from 17 T1D-relevant human proteins. The comparison of individual-specific peptide-repertoires between HLA heterozygous and homozygous individuals suggested that both a broader array of HLA-bound self-peptides and higher odds of carrying risk alleles might contribute to HLA heterozygote disadvantage. The characterization of the allele-specific peptide-repertoire suggested that an allele might confer T1D risk due to its low peptide-binding affinity possibly contributing to inefficient removal of autoreactive T-cells in thymus and (or) by binding specific disease-causing peptides, e.g. post-translationally deamidated peptides. The PepWAS-predicted T1D-associated epitopes accounted for even more deviance in T1D status than HLA class-II haplotypes (33.1 vs. 29.6%). Moreover, the sequence homology between predicted T1D-associated epitopes and pathogenic peptides suggested pathogens as the potential trigger of autoimmunity. Overall, the insights from this thesis shed light on different mechanisms that possibly underlie the differential association of HLA genes with infectious and autoimmune diseases, which, in turn, potentially shape the antagonistic selection on the classical HLA genes.

Zusammenfassung

This summary was kindly translated from English to German by Marc Ritter and Sina Schirmer.

Die Wechselwirkungen zwischen Wirten und Pathogenen sind ein wesentlicher Treiber ihrer Evolution und von besonderem Interesse in der Evolutionsbiologie. In Wirbeltieren bilden die klassischen Gene des Haupthistokompatibilitätskomplex (MHC von engl. „major histocompatibility complex“) eine zentrale Komponente ihres adaptiven Immunsystems. Die Moleküle der Zelloberfläche, die von den MHC-Genen kodiert werden, präsentieren T-Zellen Peptidfragmente (sowohl die eigenen Antigene als auch von Krankheitserregern), die wenn sie als fremd erkannt werden, eine spezifische Immunantwort auslösen. Um die allgemeine MHC-Evolution zu untersuchen, habe ich den Menschen als Studiensystem verwendet, in dem MHC als Human Leukocyte Antigen (HLA) bezeichnet wird. Obwohl auf Populationsebene eine große Allel-Diversität von HLA-Genen vorhanden ist, die möglicherweise durch pathogenvermittelte balancierte Polymorphismen erhalten wird, ist nur ein Bruchteil davon auf der individuellen Ebene in Form einer begrenzten Anzahl von HLA-Genen und Allelen zu sehen. Dieser Zustand wurde als ein Optimum zwischen der durch HLA vermittelten Resistenz gegen Krankheitserreger und dem Risiko der Autoimmunität vorgeschlagen, welche die stärksten gegensätzlichen Selektionskräfte für HLA-Gene darstellen. Die genauere Untersuchung der Verbindung von HLA mit verschiedenen Infektions- und Autoimmunkrankheiten legt nahe, dass das Peptid-Repertoire von HLA-Allelen eine wichtige Rolle bei der Bestimmung ihrer spezifischen Wirkung auf Krankheiten spielt. Die genauen Mechanismen, die der Verbindung von HLA mit den Krankheiten zugrunde liegen und so die antagonistische Selektion auf HLA-Gene beeinflussen bleiben dabei schwer fassbar.

Daher untersuchte ich zunächst die funktionale Basis der schon bekannten schützenden Wirkung von HLA-Klasse-I Heterozygotie auf HIV-1-Progression. Ich verwendete dazu einen Datensatz von 6.311 HIV-1-infizierten Individuen und prognostizierte die HLA-gebundenen Peptide, welche ich vom HIV-1-Proteom abgeleitet hatte, für jedes im Datensatz dargestellte HLA-Allel voraus. Das individuelle spezifische Repertoire an HLA-gebundenen Peptiden legt nahe, dass der Vorteil von HLA-Heterozygoten gegenüber HIV-1 sowohl durch ein breiteres Spektrum an HLA-gebundenen Peptiden als auch durch eine

höhere Wahrscheinlichkeit in Heterozygoten im Vergleich zu Homozygoten, Träger eines spezifische Schutzallele zu sein. Der Vergleich des Peptid-Repertoires von Risiko und schützenden Allelen legt nahe, dass einzelne Allele die Krankheit durch das Binden entweder einer großen Anzahl von Peptiden oder aber von spezifischen immundominanten Peptiden kontrollieren. Die getrennte Analyse der einzelnen HLA-Gene zeigte, dass unterschiedliche Mechanismen beim Vorteil von heterozygoten bei verschiedenen Genen wirken können, beispielsweise bei T-Zellen- oder NK-Zell-vermittelten Immunangriffen auf das Virus, was möglicherweise zu unterschiedlichen evolutionären Einschränkungen bei verschiedenen HLA-Genen führt. Insgesamt deuteten die Ergebnisse darauf hin, dass die durch Pathogene vermittelte Selektion sowohl HLA-Heterozygoten als auch einzelne Allele begünstigen könnte.

Mit der Hypothese, dass nicht alle HLA-gebundenen Peptide für die Krankheitsbekämpfung relevant sind, entwickelten wir einen neuen Ansatz, Peptidome-wide Association Study (PepWAS), welcher HLA-gebundene krankheitsassoziierte Epitope aus einem gegebenen Peptidom vorhersagen kann. Die von PepWAS-vorhergesagten HIV-1-assoziierten Epitope machten soviel der Variation (12 %) in der HIV-1 Viruslast aus, wie die genetischen Varianten in HLA-Klasse-I-Genen, und bildeten eine funktionelle Grundlage für die Verbindung zwischen HLA und HIV-1 Kontrolle.

Im Anschluss konzentrierte ich mich auf die Assoziation zwischen HLA-Klasse-II-Genen und Typ-1-Diabetes (T1D). An einem Fall-Kontroll-Datensatz von 16.029 Individuen zeigte ich zuerst, dass Heterozygoten bei den HLA-Klasse-II-Genen, T1D-Risiko mit sich brachte. Um die funktionelle Basis dieses HLA-Heterozygoten-Nachteils zu untersuchen, habe ich aus 17 T1D-relevanten humanen Proteinen individuelle spezifische Repertoires an HLA-gebundenen Peptiden vorhergesagt. Der Vergleich von individualspezifischen Peptid-Repertoires zwischen HLA-heterozygoten und homozygoten Individuen legt nahe, dass sowohl ein breiteres Spektrum an HLA-gebundenen Selbstpeptiden als auch höhere Chancen für das Tragen von Risiko-Allelen zum Nachteil für HLA-Heterozygote beitragen können. Die Charakterisierung des Allel-spezifischen Peptid-Repertoires legt nahe, dass ein Allel aufgrund seiner geringen Peptidbindungsaffinität ein T1D-Risiko verursachen kann, das möglicherweise zu einer ineffizienten Entfernung autoreaktiver T-Zellen im Thymus und (oder) durch Bindung spezifischer krankheitsverursachender Peptide, z. posttranslational desamidierte Peptide. Die mit PepWAS vorhergesagten T1D-assoziierten Epitope erklärten im T1D-Status sogar noch mehr Abweichungen als die HLA-Klasse-II-Haplotypen (33,1 vs. 29,6%). Darüber hinaus deutete die

Sequenzhomologie zwischen vorhergesagten T1D-assoziierten Epitopen und pathogenen Peptiden auf Krankheitserreger als möglichen Auslöser der Autoimmunität hin.

Insgesamt werfen die Erkenntnisse aus dieser Arbeit Licht auf verschiedene Mechanismen, die möglicherweise der unterschiedlichen Assoziation von HLA-Genen mit Infektions- und Autoimmunkrankheiten zugrunde liegen, welche wiederum die antagonistische Selektion von klassischen HLA-Genen prägen können.