Université de Montréal

Systems biology of the human MHC class I immunopeptidome

par
Diana Paola Granados

Programme de Biologie Moléculaire
Faculté de Médecine

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de
Docteur en biologie moléculaire
option biologie de systèmes

Octobre, 2013

Université de Montréal
Faculté de Médecine

Cette thèse intitulée :

Systems biology of the human MHC class I immunopeptidome

présentée par :
Diana Paola Granados

A été évaluée par un jury composé des personnes suivantes:

François Major, Président-rapporteur
Claude Perreault, Directeur de recherche
John David Rioux, Membre du jury
Silvia Vidal, Examinateur externe
Michel Desjardins, Représentant du Doyen

RÉSUMÉ

Le système de différenciation entre le « soi » et le « non-soi » des vertébrés permet la détection et le rejet de pathogènes et de cellules allogéniques. Il requiert la surveillance de petits peptides présentés à la surface cellulaire par les molécules du complexe majeur d'histocompatibilité de classe I (CMH I). Les molécules du CMH I sont des hétérodimères composés par une chaîne lourde encodée par des gènes du CMH et une chaîne légère encodée par le gène β_2-microglobuline. L'ensemble des peptides est appelé l'immunopeptidome du CMH I. Nous avons utilisé des approches en biologie de systèmes pour définir la composition et l'origine cellulaire de l'immunopeptidome du CMH I présenté par des cellules B lymphoblastoïdes dérivés de deux pairs de fratries avec un CMH I identique. Nous avons découvert que l'immunopeptidome du CMH I est spécifique à l'individu et au type cellulaire, qu'il dérive préférentiellement de transcrits abondants, est enrichi en transcrits possédant d'éléments de reconnaissance par les petits ARNs, mais qu'il ne montre aucun biais ni vers les régions génétiques invariables ni vers les régions polymorphiques. Nous avons également développé une nouvelle méthode qui combine la spectrométrie de masse, le séquençage de nouvelle génération et la bioinformatique pour l'identification à grand échelle de peptides du CMH I, dont ceux résultants de polymorphismes nucléotidiques simples non-synonymes (PNS-ns), appelés antigènes mineurs d'histocompatibilité (AMHs), qui sont les cibles de réponses allo-immunitaires. La comparaison de l'origine génomique de l'immunopeptidome de sœurs avec un CMH I identique a révélé que 0,5% des PNS-ns étaient représentés dans l'immunopeptidome et que 0,3% des peptides du CMH I seraient immunogéniques envers une des deux sœurs. En résumé, nous avons découvert des nouveaux facteurs qui modèlent l'immunopeptidome du CMH I et nous présentons une nouvelle stratégie pour l'indentification de ces peptides, laquelle pourrait accélérer énormément le développement d'immunothérapies ciblant les AMHs.

MOTS CLÉS : complexe majeur d'histocompatibilité, antigène de leucocytes humains, immunopeptidome, antigène mineur d'histocompatibilité, spectrométrie de masse, lignées de cellules B lymphoblastoïdes, séquençage de nouvelle génération

ABSTRACT

The self/nonself discrimination system of vertebrates allows detection and rejection of pathogens and allogeneic cells. It requires the surveillance of short peptides presented by major histocompatibility class I (MHC I) molecules on the cell surface. MHC I molecules are heterodimers that consist of a heavy chain produced by *MHC* genes and a light chain encoded by the *β_2-microglobulin* gene. The peptides presented by MHC I molecules are collectively referred to as the MHC I immunopeptidome. We employed systems biology approaches to define the composition and cellular origin of the self MHC I immunopeptidome presented by B lymphoblastoid cells derived from two pairs of MHC-identical siblings. We found that the MHC I immunopeptidome is subject- and cell-specific, derives preferentially from abundant transcripts, is enriched in transcripts bearing microRNA response elements and shows no bias toward invariant vs. polymorphic genomic sequences. We also developed a novel personalized approach combining mass-spectrometry, next-generation sequencing and bioinformatics for high-throughput identification of MHC I peptides including those caused by nonsynonymous single nucleotide polymorphisms (ns-SNPs), termed minor histocompatibility antigens (MiHAs), which are the targets of allo-immune responses. Comparison of the genomic landscape of the immunopeptidome of MHC-identical siblings revealed that 0.5% of ns-SNPs were represented in the immunopeptidome and that 0.3% of the MHC I-peptide repertoire would be immunogenic for one of the siblings. We discovered new factors that shape the self MHC I immunopeptidome and present a novel strategy for the identification of MHC I-associated peptides that could greatly accelerate the development of MiHA-targeted immunotherapy.

KEYWORDS : major histocompatibility complex, human leukocyte antigen, immunopeptidome, minor histocompatibility antigens, mass-spectrometry, B lymphoblastoid cell lines, next-generation sequencing

TABLE OF CONTENTS

RÉSUMÉ ii
ABSTRACT iii
LIST OF FIGURES ix
LIST OF TABLES xii
LIST OF ABBREVIATIONS xiii
ACKNOWLEDGEMENTS xviii
OVERVIEW xx

CHAPTER 1

1. Introduction 2
1.1 The innate and adaptive immune system 2
1.2 T lymphocytes 4
1.3 The human major histocompatibility complex (MHC) 5
1.4 The classical MHC class I and II molecules 6
1.4.1 The MHC II molecules and associated peptides 7
1.4.2 The MHC I molecules and associated peptides 7
1.5 Antigen processing and presentation 10
1.5.1 MHC II antigen processing and presentation 10
1.5.2 MHC I antigen processing and presentation 12
1.5.2.1 Peptide processing in the cytoplasm 13
1.5.2.2 Generation of peptide-receptive MHC I molecules 15
1.5.2.3 Peptide processing in the ER 16
1.5.2.4 Peptide loading and presentation 17
1.5.2.5 The end of MHC I life 18
1.5.3 Alternative MHC I antigen presentation pathways 19
1.5.3.1 Cross-presentation 21
1.6 The origin of MHC I-associated peptides 23
1.6.1 Endogenous proteins as source of peptides 24
1.6.1.1 Rapidly versus slowly degraded polypeptides 24
1.6.1.2 The DRiPs hypothesis 25
1.6.1.3 Stable proteins as source of peptides 28
1.6.1.4 Cryptic translation as a source of naturally processed peptides 28
1.6.1.5 Peptides derived from peptide splicing 29

1.6.1.6 Peptides derived from proteins destined to the secretory pathway 30
1.7 The MHC I immunopeptidome: exposing the inside of the cell to the immune system 30
1.7.1 Biological roles of the MHC I immunopeptidome 31
1.7.2 Characterization of the immunopeptidome by mass spectrometry 33
1.7.3 Immunoinformatics and prediction of the immunopeptidome 36
1.8 Objectives 37
1.8.1 Research questions and hypothesis 37
1.8.2 General objective 38
1.8.3 Specific objectives 38
1.9 Cellular model 39
1.8 References 40

CHAPTER 2

2 Origin and plasticity of MHC I-associated self peptides 60
2.1 Abstract 61
2.2 Authors' contributions 62
2.3 Background 63
2.4 The nature and role of the immune self recognized by CD8 T cells 66
2.5. A synopsis of MHC I processing – making the most out of misbegotten polypeptides 67
2.6 Different types of proteasomes generate different MIP repertoires 68
2.7 The SMII is complex and is not a representative excerpt from the proteome 70
2.8 The SMII conceals a tissue-specific signature 71
2.9. Neoplastic transformation has a broad impact on the SMII 72
2.10 Viral infection causes presentation of cryptic self MIPs 73
2.11 The SMII conveys to the cell surface an integrative view of cellular regulation 74
2.12 The immunogenicity of neo-MIPs 75
2.13 The complexity and plasticity of the SMII – A challenge for self tolerance 76
2.14 The MHC class II immunopeptidome 78

2.13 Perspective –toward a more comprehensive definition of the immune self 79
2.15 Take-home messages 80
2.16 Acknowledgements 81
2.17 References 82

CHAPTER 3

3. MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements 95
3.1 Authors' contributions 96
3.2 Abstract 97
3.3 Introduction 98
3.4 Material and Methods 99
3.5 Results 103
3.6 Discussion 118
3.7 Authorship 120
3.8 Acknowledgements 121
3.9 Disclosure of Conflict of Interests 121
3.10 References 122
3.11 Supplemental methods 126
3.12 Supplemental Figures 129
3.13 Supplemental Tables 132

CHAPTER 4

4 Minor histocompatibility antigens 134
4.1 Origin of MiHAs 134
4.2 MiHAs and allo-recognition 137
4.3 MiHA-based immunotherapy 138
4.4 The arduous identification of MiHA 141
4.6 References 150

CHAPTER 5

5 Impact of Genomic Polymorphisms on the Repertoire of Human MHC Class I-Associated Peptides 159
5.1 Authors' contributions 160

5.2 Graphical abstract 161
5.3 Abstract 162
5.4 Introduction 163
5.5 Results 165
5.6 Discussion 180
5.7 Methods 183
5.8 References 192
5.9 Acknowledgments 196
5.10 Author contributions 196
5.11 Additional information 196
5.13 Competing Financial Interests 197
5.14 Supplementary Information 198
5.15 Supplementary Data 205

CHAPTER 6

6 Discussion 207
6.1 The influence of specific HLA allelic products on the MHC I immunopeptidome 207
6.2 The MHC I immunopeptidome is cell-specific and subject-specific 208
6.3 Influence of the transcriptome on the MHC I immunopeptidome 210
6.4 The immunopeptidome preferentially derives from transcripts bearing microRNA response elements (MREs) 211
6.5 Which factors determine the number of MHC I peptides generated by a transcript? 213
6.6 A novel personalized approach for the identification of MHC I-associated peptides including MiHAs 214
6.7 Proposed improvements to our personalized approach: the database 216
6.8 Proposed improvements to our personalized approach: peptide elution and MS analyses 217
6.9 Impact of genomic polymorphisms on the MHC I immunopeptidome 219
6.10 MHC I peptides encoded by conserved and polymorphic genomic regions 221
6.11 Identification of potential MiHAs in the general population 222

6.12 How many human MiHA exist? 223
Perspectives 225
Study of the MHC I immunopeptidome from a systems biology perspective 225
Model to unravel the factors that control which MHC I peptides are displayed and in which quantity 226
Conclusion 228
References 230

APPENDIXES
APPENDIX 1 xxiii
ER stress affects processing of MHC class I-associated peptides xxiii
Abstract xxiv
Background xxv
Results xxvi
Discussion xli
Conclusions xlv
Methods xlvi
Abbreviations l
Authors' contributions l
Acknowledgements l
References lii
APPENDIX 2 lviii
Deletion of Immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules lviii

LIST OF FIGURES

CHAPTER 1

Figure 1. Location and organization of the human MHC complex on chromosome 6 6
Figure 2. Three-dimensional structures of peptide-bound MHC I and MHC II molecules 8
Figure 3. Examples of peptide-binding motifs of some HLA molecules..... 9
Figure 4. Simplified overview of the MHC II antigen processing and presentation pathway 11
Figure 5. General overview of the MHC I antigen processing and presentation pathway. 12
Figure 6. The MHC I antigen processing and presentation pathway..... 14
Figure 7. Basic and some alternative antigen processing and presentation pathways..... 20
Figure 8. Endogenous and exogenous sources of MHC I-associated peptides 24
Figure 9. The DRiPs hypothesis 26
Figure 10. Simplified flowchart of analysis of MHC I-associated peptides by MS 34

CHAPTER 2

Figure 1. Plasticity of the SMII..... 77

CHAPTER 3

Figure 1. General features of MIPs eluted from B-LCLs from 2 HLA-disparate sibships 104
Figure 2. MIP source proteins are very different in the two sibships, but are implicated in similar biological pathways and are functionally interconnected 106
Figure 3. MIPs derive preferentially from abundant transcripts 109
Figure 4. The MIP repertoire is functionally connected to the transcriptome 111
Figure 5. The MIP-coding transcriptome is enriched in transcripts containing miRNA-binding sites..... 113
Figure 6. MIPs from HLA-disparate sibships derive from different sets of

transcripts regulated by similar miRNomes 116
Figure S1. Comparison of the predicted MHC I binding affinity (IC50) of MIPs 129
Figure S2. Random changes of mRNA expression thresholds set to define expression categories do not affect the differences in expression between transcripts encoding MIPs and all the transcripts expressed in B-LCLs 129
Figure S3. The MIP-transcriptome connectivity correlates with the level of expression of the transcripts in the transcriptome 130
Figure S4. MIPs coded by miRNA targets are overrepresented in the mouse and human immunopetidomes 131

CHAPTER 4
Figure 1. Mechanisms of generation of MiHA disparities 135

CHAPTER 5
Figure 1. High-throughput genoproteomic strategy used for the identification of polymorphic MIPs on B-LCLs from 2 HLA-identical siblings 166
Figure 2. Integrative view of the genomic landscape of the MIP repertoire of HLA-identical siblings 168
Figure 3. HLA-identical siblings present similar but not identical MIP repertoires 171
Figure 4. Overview of MiHAs identified following analysis of genomic and peptidomic data from our two subjects 172
Figure 5. Unshared MIPs encoded by polymorphic loci are immunogenic 175
Figure 6. Frequency of ns-SNPs in the MIP coding exome 179
Supplementary Figure S1. Global false discovery rate (FDR) and predicted binding affinity allow discrimination between MIPs and contaminant (non-MIP) peptides 200
Supplementary Figure S2. The global false discovery rate (FDR) allows enrichment of MIPs and affects the proportion of small (8-9mers) and long peptides (10-11mers) identified 201
Supplementary Figure S3. Comparison of MIPs identified using UniProt vs. personalized databases built with next generation sequencing data 202
Supplementary Figure S4. Quantification of surface HLA-ABC before and after peptide elution 203

Supplementary Figure S6. Differential expression of non-polymorphic MIPs does not correlate with differences in MIP-coding genes or exons between subjects 205

CHAPTER 6

APPENDIXES

Figure 1. EL4 stable transfectants express the SIINFEKL peptide derived from HEL targeted to the ER or to the cytosol xxvii
Figure 2. Induction of ER stress in EL4 cells xxix
Figure 3. ER stress impairs MHC I surface expression xxxi
Figure 4. ER stress impairs cell surface MHC I expression through posttranscriptional mechanism(s) xxxiii
Figure 5. Differential effects of ER stress on surface expression of various glycosylated proteins xxxv
Figure 6. ER stress inhibits protein synthesis through phosphorylation of eIF2α in EL4 stable cell lines xxxvi
Figure 7. Increased presentation of SIINFEKL peptide derived from ER-localized relative to cytosolic HEL protein during ER stress xxxviii
Figure 8. Stability of cytosolic HEL and ER-retained HEL during ER stress .. xli

LIST OF TABLES

CHAPTER 3

Table 1. SNP frequency (SNPs/bp) in MIPs and in the human exome 107
Table 2. Representative set of enriched miRNA binding sites present in transcripts that are source of MIPs in B-LCLs from both sibships..... 114

CHAPTER 4

Table 1. Features of known MiHAs associated to HLA I and HLA II molecules 146

CHAPTER 5

Table 1. MiHAs resulting from ns-SNVs in the MIP-coding region and detected in (A) one of the two subjects or in (B) both subjects 173

LIST OF ABBREVIATIONS

A	
ACE	Angiotensin-converting enzyme
AE	Average expression
Allo-HSCT	Allogeneic hematopoietic stem cell transplantation
APC	Antigen-presenting cell
ARF	Alternative open-reading frame
B	
β_2-m	β_2-microglobulin
BCR	B cell receptor
BiP	Binding immunoglobulin protein
B-LCL	B lymphoblastoid cell line
bp	Base pairs
C	
C	Constant
CDS	Coding sequence
CEPH	Centre d'Etude Polymorphism Humain
CFSE	Carboxyfluorescein succinimidyl ester
CID	Collision induced dissociation
CLIP	Class II-associated Ii peptide
CNV	Copy number variants
CP	Constitutive proteasome
TECs	Thymic epithelial cells
CTLs	Cytotoxic T cells
D	
D	Divesity
DC	Dendritic cell
DLI	Donor lymphocyte infusion
DRiPs	Defective ribosomal products
E	
EBV	Epstein-Barr virus
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated degradation

	ERAAP	Endoplasmic reticulum aminopeptidase associated with antigen processing
	ERAP1	Endoplasmic reticulum aminopeptidase associated with antigen processing 1
	ERAP2	Endoplasmic reticulum aminopeptidase associated with antigen processing 2
F		
	FACS	Fluorescence-activated cell sorting
	FDR	False discovery rate
	FPKM	Fragments per kilobase of exon model per million mapped reads
G		
	GvHD	Graft-versus-host disease
	GvL	Graft-versus-leukemia
	GvT	Graft-versus-tumor
H		
	HCD	High collision induced dissociation
	HSV	Herpex simplex virus
	HLA	Human leucocyte antigen
	HPLC	High-performance liquid chromatography
	HSP	Heat shock protein
I		
	IDE	Insulin-degrading enzyme
	IEDB	Immune epitope database
	IFN-γ	Interferon gamma
	INDEL	Insertion/deletion
	IP	Immunoproteasomes
	IARP	Insulin-regulated aminopeptidase
	IRES	Internal ribosomal entry sites
	ITAM	Immunoreceptor tyrosine-based activating motifs
J, K		
	J	Joining
	KIR	Killer cell immunoglobulin-like receptor
M		
	m/z	Mass-charge ratio

MFI	Mean fluorescence intensity
MHC	Major histocompatibility complex
MHC I	Major histocompatibility complex class I
MHC II	Major histocompatibility complex class II
MHC III	Major histocompatibility complex class III
MIIC	MHC class II compartment
MiHA	Minor histocompatibility antigen
MIPs	MHC I-associated peptides
MiRNA	MicroRNA
MRE	microRNA response element
MS	Mass spectrometry
MSigDB	Molecular Signature Database
mTEC	Medullary thymic epithelial cells

N, O

NK	Natural killer
ns-SNP	Non-synonymous single nucleotide polymorphism
ns-SNV	Non-synonymous single nucleotide variation
ORF	Open reading frame

P

PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PCT	Protein coding transcriptome
PDI	Protein disulfide isomerase
PLC	Peptide-loading complex

R

RDP	Rapidly degraded polypeptides
RT-qPCR	Quantitative real-time reverse transcriptase polymerase chain reaction

S

SCX	Strong cation exchange
SOM	Self-organizing maps
s-SNP	Synonymous single nucleotide polymorphism
SDP	Slowly degraded polypeptides
siRNA	Small interfering RNA
SMII	Self MHC I immunopeptidome

	SNP	Single nucleotide polymorphism
T		
	TAP	Transporter associated with antigen processing
	TAPBPR	tapasin-related protein
	TCR	T cell receptor
	TOP	Thimet oligopeptidase
	TP	Thymoproteasome
	TPPII	Tripeptidyl peptidase II
U		
	UGT1	UDP-glucose-glycoprotein glucosyltransferase 1
	UTR	Untranslated region
V, W		
	V	Variable
	WGAs	Whole-genome association scanning

A mis padres

ACKNOWLEDGEMENTS

This work would have not been possible without the generous advice and support of many people. First, I wish to express my deepest thank to “le plus grand amateur de l’immunopeptidome”, Claude Perreault, for giving me the chance to be part of his team, for his wise guidance during my Masters and Ph.D. studies, his contagious passion for research and for being an excellent mentor.

I would like to thank all past and present members of the laboratory for sharing with me their expertise, for their support and for providing a pleasant working ambience. I am especially grateful to my first “lab mentor” Étienne, for teaching me different techniques and for all the valuable scientific discussions. Thanks to all the people whom I had the fortune to work with: Danielle, Marie-Pierre, Caro, Tara, Wafaa, Olivier, Antoine, Vydia and Céline. I specially thank Caro for her constant support and technical help, and Tariq and Dev for all the fruitful discussions and experiences we gained in building our multidisciplinary team. I learned a lot from all of you! I am specially grateful to Benjamin, Marie-Pierre and Dev for their comments on some sections of this thesis.

I would also like to thank the members of the jury and of my thesis committees for their valuable time and their contribution to my education and to the advancement of our work: Sébastien Lemieux, John Rioux, Silvia Vidal and François Major. I am also grateful to Pierre Thibault and Sébastien for thoughtful comments and discussions, as well as many people from IRIC’s core facilities for their excellent assistance and/or significant contribution to my work: Raphaëlle Lambert, Pierre Chagnon, Simon Drouin, Jean-Philippe Laverdure, Geneviève Boucher, Patrick Gendron, Danièle Gagné, Gaël Dulude and Christian Charbonneau.

I would like to express my deepest gratitude to Benjamin for his encouragement and his support in the good and difficult moments. I thank him for correcting my texts in French, listening to my presentation practices, but specially for giving me so many valuable advices during my graduate studies. I thank him for being “Mum and Dad” so many times. Without you, I wouldn’t have

been able to finish this long road. I thank “mis pequenos tesoritos”, Maïté and Salomé, for shining my life and providing me with a work-family balance. I am also very grateful to my parents, my sister and my family-in-law who even in the distance where always supporting me and encouraging me to make my dreams come true.

Finally, I would like to thank our blood donors as well as the *Programmes de biologie moléculaire de l’Université de Montréal,* IRIC and the following financial agencies for their contribution to my education: Foreign Affairs Canada, the Cole Foundation and the Canadian Institutes for Health Research (CIHR).

OVERVIEW

The ability to discriminate between self and non-self is a fundamental requirement for life. Multicellular organisms use self/non-self discrimination primarily in immune defense. The adaptive immune system of jawed vertebrates has taken advantage of the protein fragments (peptides) generated by the ubiquitin-proteasome degradation machinery to use them as flags in self/non-self discrimination. These protein fragments, collectively known as the MHC I immunopeptidome, need to be processed and presented on MHC class I molecules on the surface of the cell. Under steady state conditions, the MHC I immunopeptidome is composed solely of self peptides. The immune system keeps track of intracellular protein content by sampling the universe of self peptides produced in search of: i) altered (transformed) peptides or neo-self peptides that may reveal dysfunction (e.g. cancer, stress, inflammation) or ii) non-self peptides resulting from alteration of steady state conditions (e.g. infection, pregnancy, transplantation). While the adaptive immune system is vital for combating pathogens and neoplastic transformation, it represents a significant barrier for foreign (allogeneic) transplantation in a clinical context. Moreover, alteration of the self/nonself discrimination system may lead to autoimmune diseases.

The MHC I immunopeptidome is the end result of the antigen processing and presentation pathway, which behaves as a complex system involving input, processing and output of data. Complex systems can be studied from a holistic perspective. Systems biology is an inter-disciplinary field that combines high-content multiplexed measurements with computational methods to better understand and model biological function at various scales. Accordingly, recent large-scale (-omic) studies have yielded unprecedented insights into the genesis, molecular composition and plasticity of the MHC I immunopeptidome. The aim of the present multidisciplinary work was to unravel the biogenesis and composition of the MHC I immunopeptidome of human B lymphoblastoid cell lines by applying data-driven systems biology approaches.

The results from this work will be presented in six chapters and two appendixes. The first chapter includes the general introduction and objectives. The

second chapter corresponds to a published review article about the origin and plasticity of MHC-I associated self peptides. The third chapter includes a published article in which we report the relationship between the MHC I immunopeptidome, the transcriptome and the microRNAome. The forth chapter corresponds to a review article to be submitted about identification methods and molecular mechanisms responsible for the generation of MHC I peptides that cause immune reactions in allogeneic transplantation (i.e. minor histocompatibility antigens). The fifth chapter includes a submitted manuscript showing the impact of single nucleotide polymorphisms on the MHC I immunopeptidome. The results from this work are collectively discussed in the sixth chapter. The first appendix is a published work that I started during my Masters studies and completed during my Ph.D. studies. It shows the effect of ER stress on the processing of MHC class I-associated peptides. The second appendix is the summary of my contribution to a published article showing the impact of immunproteasome deletion on the repertoire of MHC I-associated peptides.

Our studies reveal how various factors in different functional genomic levels such as genomic polymorphisms, transcript abundance and the presence of microRNA response elements, influence the human self MHC I immunopeptidome. Moreover, this work significantly expand the number of sequenced and characterized human MHC I peptides and constitute a precious resource of exome, transcriptome and miRNAome sequencing and analyses of human B lymphoblasts. Lastly, we provide a novel approach relying on next-generation sequencing, bioinformatics and mass spectrometry for high-throughput discovery of minor histocompatibility antigens. Therefore, our work provides major insights on the biogenesis of the MHC I immunopeptidome and contributes to the advancement of cancer immunotherapy.

CHAPTER 1

1. Introduction

1.1 The innate and adaptive immune system

In vertebrates, the effective recognition and elimination or containment of infectious microorganisms is achieved by the synergic and complementary action of the **innate and adaptive immune systems** [1]. The immune system is not only equipped with potent effector mechanisms to clear pathogen, toxins and allergens, but it also has the ability to distinguish self from non-self and avoid damaging self-tissues that can lead to auto-immune diseases [2]. This immunosurveillance mechanism is known as **self tolerance** and it is manifested both in the innate and the adaptive immune responses [2].

The **innate immune (or non-specific) system** allows initial host defense against microbial pathogens and comprises *innate* mechanisms that are encoded in their mature functional forms by germline genes [2]. It includes i) physical barriers such as respiratory, gastrointestinal and genitourinary epithelia, ii) small molecules including complement proteins and defensins that are constitutively present in fluids or are released from activated cells, iii) soluble proteins such as cytokines, chemokines and enzymes, and iv) membrane-bound receptors and cytoplasmic proteins of phagocytic immune cells that bind molecular patterns of microbial origin [3]. The innate immune system is non-specific as the distinction of microbial pathogens from host cells is made through recognition of conserved molecules shared by many microbes [2]. These molecules are recognized by a limited repertoire of receptors, such as the Toll family, expressed on phagocytic immune cells [4]. Because the recognition molecules are expressed broadly on a large number of cells, the cells are rapidly activated (within hours of contact) and this constitutes the initial host response [2]. Of note, the efficacy of the innate response is not increased by previous exposure to the same pathogen [3].

By contrast, the **adaptive (or specific) immune system** is composed of small numbers of cells with specificity for any individual pathogen. These specialized cells are **T and B lymphocytes** that express antigen-specific receptors on their cell surface [2]. The T cell receptors (TCR) and the B-cell receptors (BCR)

are *acquired* during the lifetime of the organism as a result of somatic rearrangement of gene segments and allow the formation of millions of different antigen receptors, each with a unique specificity for a different antigen [2]. In this way, the adaptive immune system provides flexibility to respond to numerous and highly variable targets. This response is less rapid (within days of contact) than the innate response since the responding cells must proliferate after encountering the antigen to attain sufficient numbers to mount an effective response [2,3]. Hence, the adaptive response generally arrives temporarily after the innate response in host defense [2]. Following recognition of specific antigens of a given microorganism, the lymphocytes and the antibodies they produce persist as **immunological memory** and are rapidly protective on re-exposure to the same pathogen, albeit in an antigen-dependent manner [1]. The phenomenon of immunological memory is exploited in vaccination, in which antigens are inoculated to stimulate an individual's adaptive immunity to a pathogen [1].

Both B and T lymphocytes provide defense against extracellular pathogens via recognition of antigens, although by different mechanisms. **B lymphocytes (or B cells)** mature in the bone marrow and trigger what is typically known as the **humoral immune response (or antibody-mediated system)**. This response is characterized by recognition of intact antigens in the extracellular milieu through secreted immunoglobulins (antibodies) produced by B cells and surface immunoglobulins that compose the BCRs on the surface of B cells [3]. By contrast, **T lymphocytes (or T cells)** develop in the thymus from common lymphoid progenitors and trigger the **cellular immune response (or cell-mediated system)**. In this response T cells are activated by **antigen-presenting cells (APCs)** (e.g. dendritic cells (DCs), macrophages, B cells) and consequently eliminate infected cells and activate other cells of the immune system [3,5]. Activation of T lymphocytes is initiated upon recognition of peptide fragments of antigen (i.e. epitopes) presented by MHC molecules encoded by **the major histocompatibility complex (MHC)** on the surface of APCs [3]. Hence, T cells recognize a molecular complex composed of a self-component (the MHC) and a non-self structure (the epitope) [2].

1.2 T lymphocytes

T lymphocytes are classified into several T cell subsets that are distinguished based on the composition of their TCR (αβ or γδ), their antigenic specificity (determined by expression of the CD4+ or CD8+ coreceptor molecules) and their state (e.g. naive, effector, regulatory) [6]. The composition of the TCR determines its specificity for a given target [7]. While T cells that express the αβ TCR react to peptides presented by MHC molecules, those that express the γδ **TCR** are not MHC-restricted and play a role in the surveillance of tissue stress [6]. αβ **T cells** (referred as T cells hereinafter) constitute the majority of the T cell population in lymphoid organs [5]. The αβ TCR is an heterodimer made from two separate chains that arise from somatic gene rearrangement of variable (V), diversity (D), joining (J) and constant (C) gene fragments during T cell development [7]. In this process, nucleotides are inserted and deleted at V(D)J junctions in each chain, resulting in an extensive repertoire of TCRs, whose reactivity against self MHC-peptide complexes is screened thereafter during thymic selection [7].

In the thymus, **thymic selection** of T cells is achieved through positive and negative selection. During **positive selection**, immature double-positive (CD4+CD8+) thymocytes that are capable of forming a minimal interaction between the αβ TCR and self MHC-peptide complexes on cortical thymic epithelial cells, are rescued [8,9]. Positive selection is assumed to enrich the repertoire of self-MHC-restricted T cells capable of react against potential foreign antigens [10]. Hence, double-positive T cells bearing a TCR that do not bind MHC-self-peptide complexes die by neglect. During **negative selection (or clonal deletion)**, thymocytes are exposed to a variety of tissue-specific peptides ectopically expressed and presented by medullary epithelial cells and medullary APCs [6,8]. In this process, T thymocytes that bind with high avidity to self MHC-peptide complexes are eliminated [8]. The removal of these self-reactive T cells is essential for preventing autoimmunity and illustrates one mechanism of central tolerance [6]. Double-positive thymocytes passing both positive and negative selection develop into mature CD4+ or CD8+ T cells expressing large amounts of TCRs [5]. This pre-immune repertoire of **naive T lymphocytes** potentially reactive to foreign but not self antigens, exit the

thymus and migrate to secondary lymphoid organs, where they can encounter non-self peptides presented by MHC molecules on APCs and thereafter proliferate and differentiate into **effector T lymphocytes** [11]. Activation of **CD8+ (or cytotoxic) T cells** (CTLs) and **CD4+ (or helper) T cells** is triggered by recognition of peptides bound on MHC class I and class II molecules, respectively, encoded by the major histocompatibility complex (MHC) genes (see next section). Upon activation, CD4+ T cells produce cytokines that activate other cells including macrophages and B cells and thereby regulate cellular and humoral immune responses, while CD8+ T cells differentiate into effector T cells that directly contact infected or transformed cells and destroy them via release of perforin and granzymes [5].

1.3 The human major histocompatibility complex (MHC)

The **major histocompatibility complex (MHC)** is a large multigenic region found in most vertebrates containing at least 128 genes in humans, of which more than 20% encode proteins of the immune system [4,12]. The human MHC is the most gene-dense and polymorphic region and is part of the so-called extended MHC region, covering 7.6 Mb of the short arm of the chromosome 6 [4]. The MHC region contains, among others, highly polymorphic genes encoding the aforementioned **MHC molecules** that present protein-derived peptides (antigens) to T cells [1]. These type of antigen-presenting molecules are known as **classical MHC molecules** [3]. There are also structurally related molecules that are monomorphic or oligomorphic, known as **non-classical MHC molecules** [3,13]**.** Non-classical MHC molecules are involved in immune and non-immune processes and some can function in the presentation of peptide antigens [13]. Besides, the MHC region contains genes involved in the processing of antigens [4].

For historical reasons, the human MHC is also referred to as the **human leukocyte antigen (HLA)** complex, since the first MHC gene products were discovered on the surface of leukocytes [4]. Three distinct regions have been identified within the human MHC: the **MHC class I (MHC I)**, **MHC class II (MHC II)** and **MHC class III (MHC III)** (Figure 1). The MHC I and II regions encode both classical MHC and non-classical MHC molecules**.** The **MHC I region** contains

among other genes, the **classical MHC I genes (also known as MHC Ia genes)** *HLA-A*, *HLA-B* and *HLA-C* and the more ancient **non-classical MHC I genes (also known as MHC Ib genes)** *HLA-E*, *HLA-F* and *HLA-G* [4,13]. The **MHC II region** covers the **classical MHC II genes** *HLA-DP*, *HLA-DQ* and *HLA-DR*, as well as **non-classical MHC II genes** *HLA-DM* and *HLA-DO* [4]. Some non-classical MHC I molecules play a role in activating specialized classes of T cells, whereas non-classical MHC II molecules regulate peptide loading into classical MHC class II molecules. The **MHC III region** is located between the class I and class II regions and encloses miscellaneous non-HLA genes (e.g. *MICA, MICB*) encoding proteins with or without immune function [12].

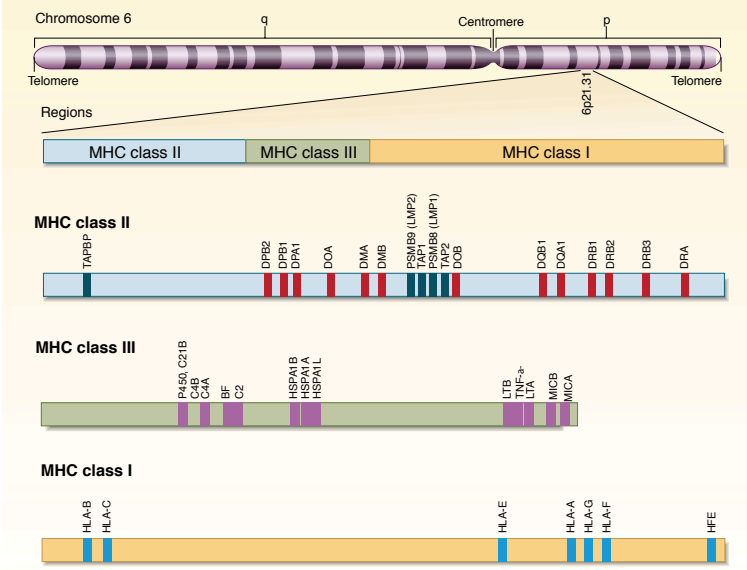


Figure 1. Location and organization of the human MHC complex on chromosome 6. Adapted with permission from [14]. Copyright Massachusetts Medical Society

1.4 The classical MHC class I and II molecules

MHC I and MHC II molecules are cell-surface glycoproteins with a similar three-dimensional structure and similar function in presenting peptide fragments or antigens to the immune system [15]. Nevertheless, these molecules differ in their tissue distribution and in the type of antigenic peptides they display that reflects different antigen processing pathways [16]. The 3 classical human MHC I genes (*HLA-A*, *HLA-B* and *HLA*-C) are co-dominantly expressed on the cell surface of all nucleated cells and play a major role in adaptive immunity.

On the contrary, surface expression of MHC II molecules is restricted mainly to APCs such as macrophages, B cells and DCs, although it can be induced by interferon-γ (IFN-γ) and other stimuli in non-APCs like mesenchymal stromal cells, fibroblasts, endothelial cells and activated human T cells [5,16].

1.4.1 The MHC II molecules and associated peptides

MHC II molecules are transmembrane glycoproteins with short cytoplasmic domains, composed of one α and one β chain. The membrane-proximal region consists of one conserved domain that is part of the α subunit and another conserved domain of the β subunit (Figure 2c). MHC II molecules exhibit enormous amino acid sequence variation in the region that interacts with the peptide named the **peptide-binding groove** [17] (Figure 2d). Consequently, different alleles bind different sets of peptides. The groove is formed by the juxtaposition of the N-terminal regions of the α and β chains. The peptide-binding groove is open at both ends and hence, long peptides, generally between 15 and 24 amino acids long, can bind and overhang the binding groove resulting in more or less restrictive binding motifs [17] (Figure 2d).

1.4.2 The MHC I molecules and associated peptides

The molecular target of the TCR of CD8+ T cells is the complex formed by the MHC I and the associated peptide [2]. MHC I molecules are cell-surface heterodimers composed of a polymorphic transmembrane 44-kd heavy (α) chain and a 12-kd invariant light chain, known as **β2-microglobulin (β2m)** [2,14,16] (Figure 2a). The heavy chain consists of 3 extracellular domains (α1, α2 and α3), a transmembrane domain and a short intracellular domain that anchors the protein in the cell membrane [18]. The α1 and α2 domains form a platform with a groove, in which antigenic peptides can bind noncovalently through the N and C termini [17] (Figure 2a). The main anchor residues of the peptide to the peptide-binding groove are frequently the second and the last C-terminal residue [15] (Figure 3b). In some exceptional cases, the main anchors are located in positions P3 (as in HLA-A1-associated peptides) and P5 (as in HLA-B8-associated peptides). The peptide-binding groove is blocked at both ends by bulky aromatic amino acids that typically limit the length of the bound peptide

to 8-10 residues [15] (Figure 2b). Less frequently, MHC I molecules can bind longer peptides [19-21], in which case the central part of the peptide protrudes and the binding is presumably less stable [15,22].

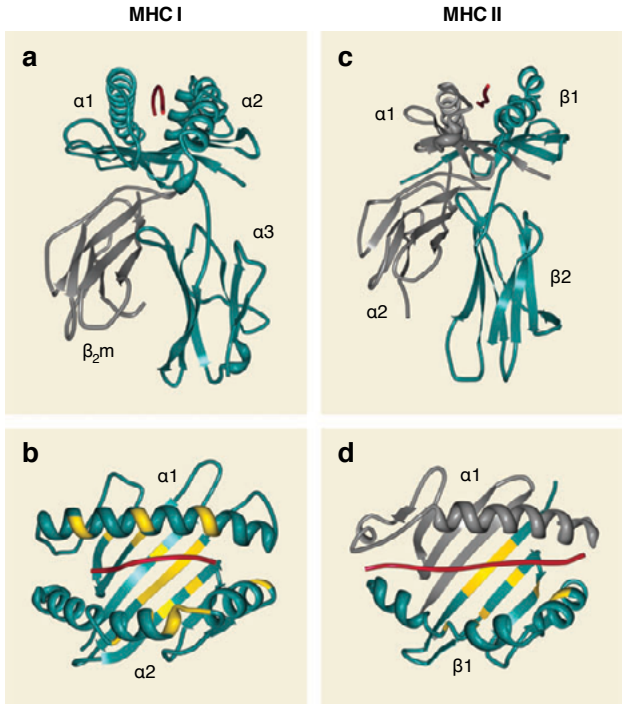


Figure 2. Three-dimensional structures of peptide-bound MHC I and MHC II molecules
(a) and (b) HLA-A2 molecule complexed with influenza derived peptide. (c) and (d) HLA-DR1 molecule complexed with influenza derived peptide. Highly polymorphic residues proximal to the peptide-binding groove are highlighted in yellow; the peptide is shown in red. Adapted with permission from [17].

A prominent characteristic of the human MHC I molecules is their high degree of polymorphism [1,23]. The IMGT/HLA database of the international ImMunoGeneTics project (http://www.ebi.ac.uk/ipd/imgt/hla) registers thousands of MHC I alleles in their last release (2013-07-25): 2,365 *HLA-A*, 3,005 *HLA-B* and 1,848 *HLA-C* alleles. Besides, *HLA-B* is the most polymorphic gene in the human genome [24]. Most polymorphisms in exons 2 and 3 of the *HLA-A*, *HLA-B* and *HLA-C* genes lead to amino acid substitutions in the floor and sides of the peptide-binding groove of the corresponding MHC I protein [17,25]. Thereby, different MHC I allelic products display distinct amino acid preferences at key

positions in the peptide sequence. These MHC-specific amino acid patterns are known as **peptide-binding motifs (or binding specificities)** [26] (Figure 3a)**.** Consequently, different MHC I allelic products bind different sets of peptides [27]. Moreover, MHC I molecules are promiscuous since each type of MHC I allelic product can bind a diversity of peptides that differ in their sequences but share 2-3 anchoring amino acid residues (i.e. the same binding motif) [14,28].

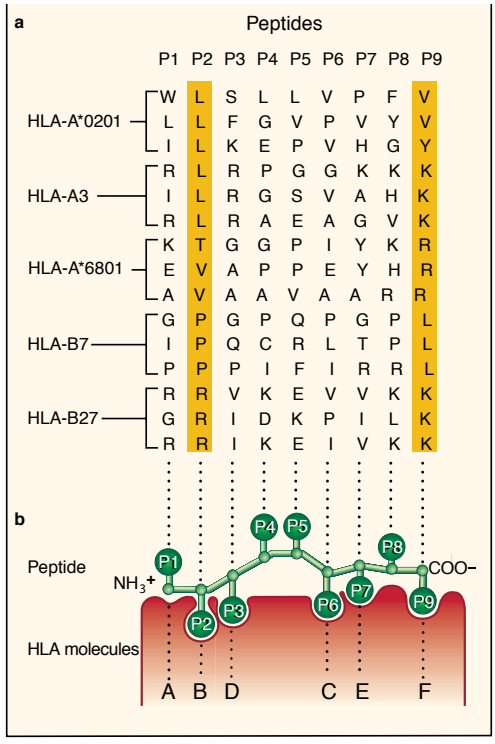


Figure 3. Examples of peptide-binding motifs of some HLA molecules
(a) Peptide-binding motifs. Anchor residues are highlighted in yellow. (b) Longitudinal section through the peptide-binding groove of an MHC I molecule. P1-P9 indicates amino acid positions. Adapted with permission from [14]. Copyright Massachusetts Medical Society

At the individual scale, this variability means that each subject who is heterozygous at the 3 HLA loci will have 6 distinct peptide-binding grooves. Consequently, on a population level, the diversity of peptide-binding motifs is colossal. This characteristic supports a hypothesis whereby variations in the MHC I binding groove allow the various MHC I molecules in the population to bind and present different portions of viral or intracellular pathogens for immune recognition [29]. Hence, this huge diversity protects the population against microbes that evade recognition through mutations in antigens that

impair MHC I binding [2,30]. It is believed that this overwhelming variability of antigenic structures and the ability of pathogens to mutate and avoid host detection have driven the evolution of the adaptive immune system [31]. The enormous degree of MHC polymorphism, although highly desirable in healthy individuals for dealing with infections, poses great problems in a transplant setting, as the search for a suitable bone marrow or organ donor may be extremely difficult.

1.5 Antigen processing and presentation

Typically, MHC I molecules bind peptides that derive from proteins translated within the cell (endogenous antigens), and present them to CD8+ T cells. By contrast, MHC II molecules sample the extracellular milieu through binding of peptides derived from exogenous proteins that have been ingested by APCs and degraded in the endocytic pathway, and expose them to CD4+ T cells [2,16]. In the next sections, the MHC II antigen processing pathway will be briefly described followed by a detailed review of the MHC I pathway.

1.5.1 MHC II antigen processing and presentation

The transmembrane α and β chains of MHC II are assembled in the **endoplasmic reticulum (ER)**, where they associate with the invariant chain (Ii) and form the Ii-MHC II complex (Figure 4). The complex is transported to endosomal compartments through the recognition of sorting motifs in the cytoplasmic tail of Ii [16]. Once the complexes reach late endosomal compartment named MHC class II compartment (MIIC), Ii is digested leaving a residual Ii peptide named class II-associated Ii peptide (CLIP). CLIP lies in the peptide-binding groove where it is exchanged for a specific peptide through the action of HLA-DM. In B cells, HLA-DO associates with HLA-DM and restricts HLA-DM activity to more acidic compartments [32]. MHC II-peptide complexes are exported from the MIIC via the trans-Golgi network to the plasma membrane.

MHC II-associated peptides arise from the degradation of exogenous proteins in the endosomal pathway (Figure 2b). Antigens are internalized through endocytosis and targeted into lysosomes for processing by lysosomal proteases

and loading on MHC II molecules [16]. Cytosolic (endogenous) proteins degraded through autophagy, as well as membrane proteins degraded in the lysosome, can be presented by MHC II molecules [33-36]. Information about the MHC II-peptide repertoire can be found in the section "*The MHC class II immunopeptidome*" of the review article included in chapter 2.

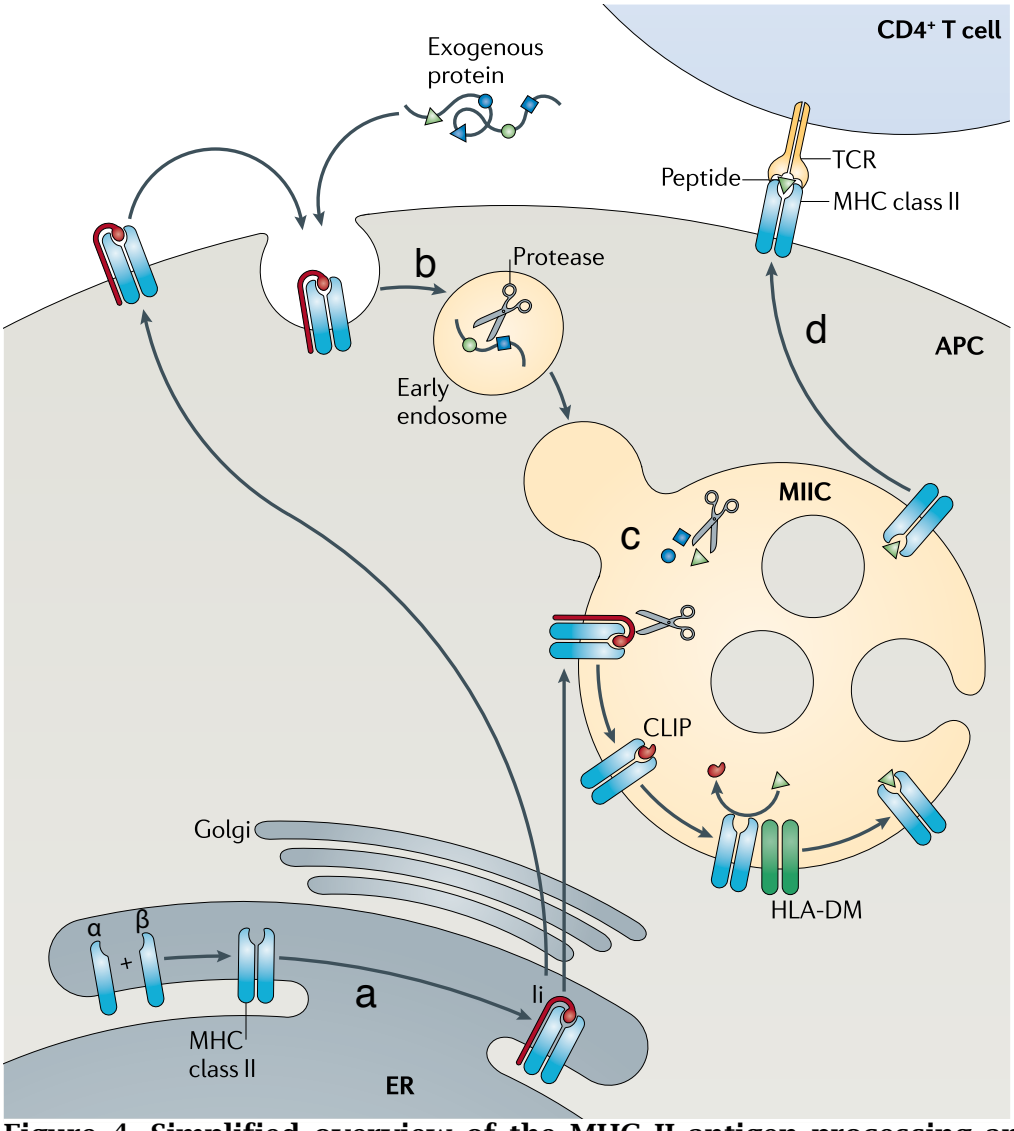


Figure 4. Simplified overview of the MHC II antigen processing and presentation pathway

(a) MHC II-Ii assembly. (b) Internalization of antigens. (c) Ii digestion and peptide binding in the MIIC (MHC class II compartment). (d) Export of MHC II-peptide complexes. Adapted with permission from [17].

1.5.2 MHC I antigen processing and presentation

Peptide-bound MHC I molecules are the end product of the MHC I-antigen processing and presentation pathway. This pathway begins with the degradation of proteins source of peptides inside the cell. In a stepwise manner, proteolytic intermediate fragments are generated and protected until they yield the final peptides that can fit MHC I molecules in the ER and be exported as MHC I-peptide complexes to the cell surface [37]. Accordingly, the antigen presentation pathway is composed of two merging cellular processes (Figure 5). In the first one, peptides suitable for loading the MHC I molecules are generated, whereas in the second pathway, peptide-receptive MHC I molecules are assembled [38]. This highly specialized system, operating with essentially conserved components, is nonetheless capable of generating highly diverse sets of peptides that satisfy a large number of different MHC I molecules

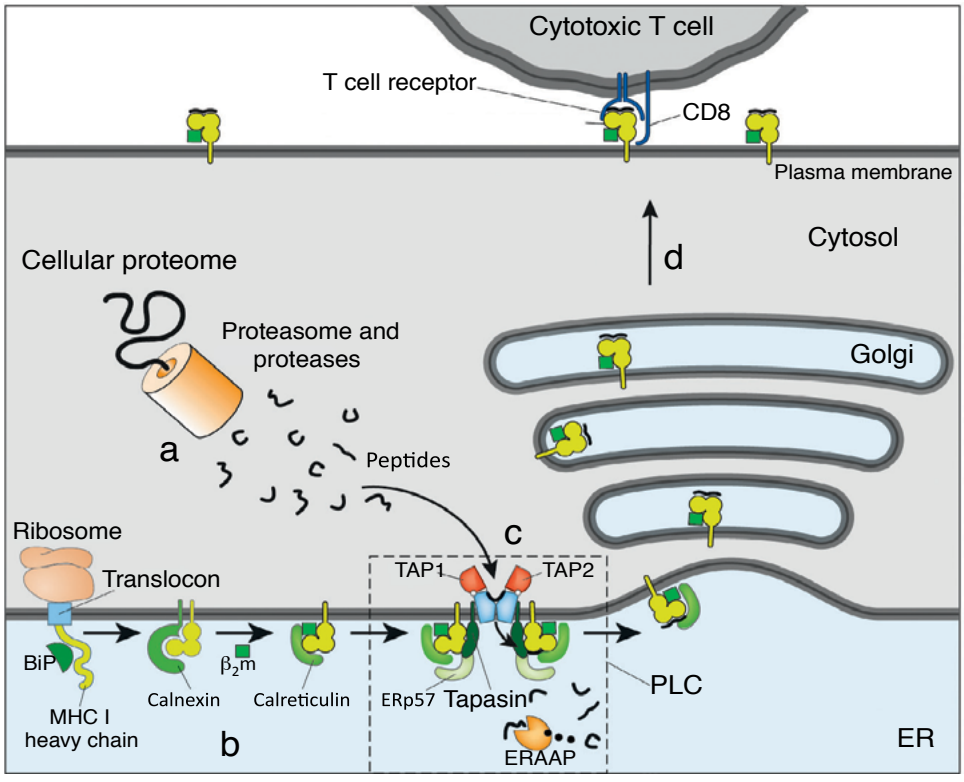


Figure 5. General overview of the MHC I antigen processing and presentation pathway.
(a) Production of peptides. (b) Assembly of MHC I heterodimers (c) Loading of MHC I molecules with peptide. (d) Export of MHC I-peptide complexes. PLC: peptide-loading complex. Adapted with permission from [39].

1.5.2.1 Peptide processing in the cytoplasm

Generation of peptides suitable for loading MHC I molecules starts in the cytoplasm. This is the major site of protein turnover, because even ER proteins are retrieved into the cytosol for degradation [37]. In the cytoplasm, peptide precursors of variable lengths (2-25 amino acids) are generated from degradation of endogenous proteins through the action of the **proteasome** and other **proteases** [22,40,41] (Figure 6). Cytosolic proteins first associate with the chaperone Hsp90α, are then ubiquitinated by the ubiquitin ligase CHIP and consequently degraded by the multicatalytic proteasome [17,42,43]. Of note, different proteasome variants such as the **immunoproteasome** (induced by IFN-γ and IFN-α) [44], **mixed proteasomes** [45] and the **thymoproteasome** (restricted to cortical thymic epithelial cells) [46] are suppliers of peptide precursors (3-22 residues in length) [22,47,48]. Treatment of cells with proteasome inhibitors has revealed that most but not all MHC I peptides require the proteasome for their generation [49,50], suggesting the contribution of other proteases to the peptide pool [22]. Additionally, proteasomes are thought to generate the final C-terminal residues of MHC I-binding peptides [43,51], although this notion have been questioned by recent studies [52-57].

Proteasomal degradation yield truncated protein fragments that can associate with the chaperone protein **TRiC** [58,59]. In this way, protein fragments are protected from cleavage by cytosolic amino- and endopeptidases that recycle amino acids and prevent their accumulation [60]. Before transport into the ER, most peptide epitopes require additional trimming at the amino terminus [61]. Various cytosolic aminopeptidases, such as leucine aminopeptidases, puromycin-sensitive aminopeptidase and bleomycin hydrolase have been shown to cleave the amino terminus [22,39,50,62] (Figure 6). Of particular interest is the cytoplasmic **tripeptidyl peptidase II (TPPII)**, involved in trimming of proteasomal products over 15 amino acids in length, in contrast to most other aminopeptidases that only cleave peptides smaller than 14 residues [22,63]. Studies using TPPII inhibitors have shown that TPPII plays a role in the presentation of selected peptides [22,53]. TPPII has also been shown to participate in proteasome-independent pathway for epitope generation [53]. Recently, the cytosolic peptidases **insulin-degrading enzyme (IDE)**, **thimet oligopeptidase**

(TOP), **nardilysin** and **DPP9** have been implicated in the generation of some epitopes in a proteasome-dependent or independent manner [54,55,64]. Nevertheless, mice deficient for different cytosolic peptidases have shown that despite being essential for the generation of particular MHC I peptides, they are not required for production of the bulk peptide pool [22,50,65-67]. Furthermore, a caspase-mediated epitope production has been discovered in apoptotic cells upon infection with vaccinia virus, and this mechanism has been proposed to explain the increased immunogenicity of apoptotic cells [68].

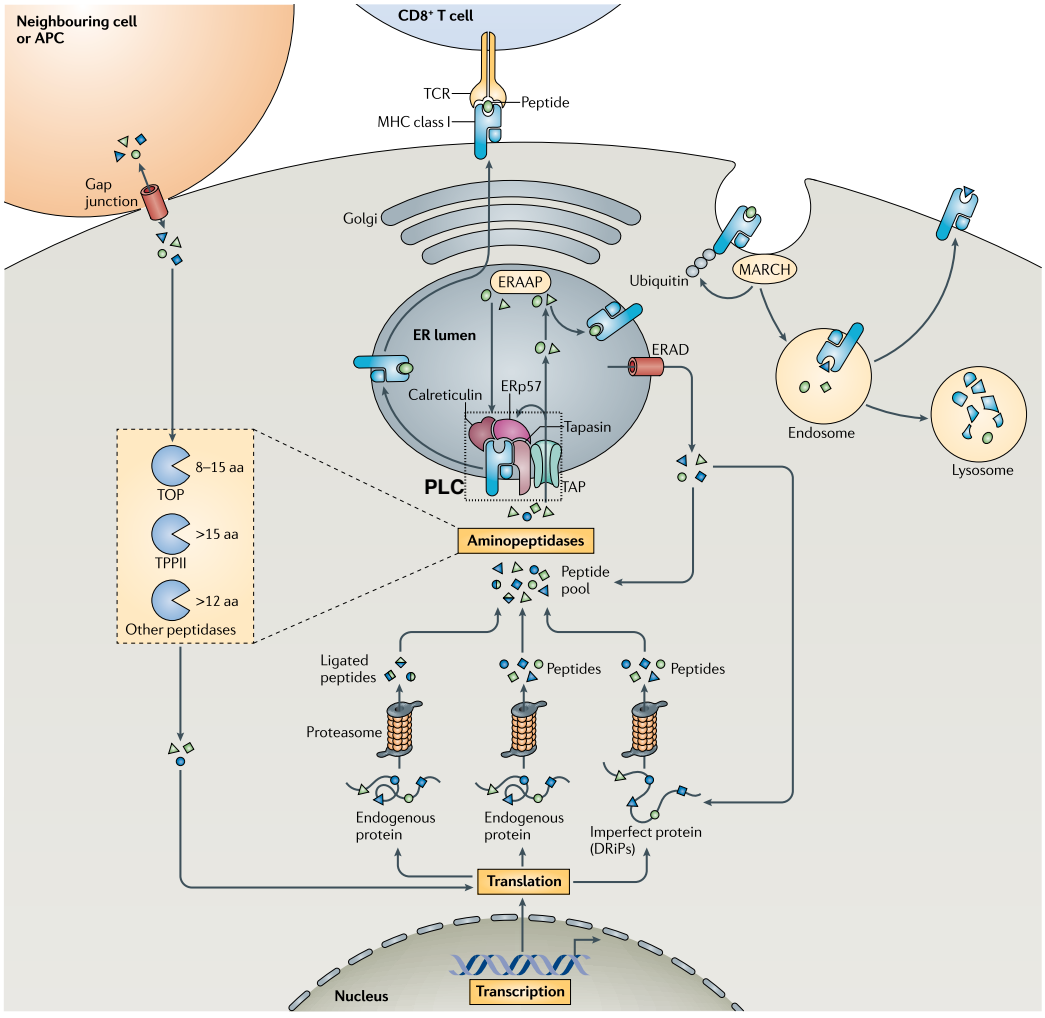


Figure 6. The MHC I antigen processing and presentation pathway
PLC: peptide-loading complex. Adapted with permission from [16].

Paradoxically, the cytoplasm constitutes not only the place where antigenic peptides are born, but also the site where most peptides are rapidly destroyed owed to excess trimming by cytosolic peptidases [60]. Indeed, peptides have a half-life of 6-10 seconds in the cytosol [60] and more than 99% of peptides are degraded by cytosolic peptidases before they reach the ER [69]. Cytosolic chaperones, such as Hsp70 and Hsp90, are thought to protect peptides from exhaustive degradation [42,58,70] This rapid degradation by the proteasome and cytoplasmic enzymes limits the availability of peptides and accounts for the inefficiency of the peptide presentation pathway [37,71]. For instance, cytosolic efficiency has been measured to produce one MHC I-peptide complex for each 500-5,000 viral translation products degraded [72].

At this point, the resulting protein fragments constitute precursors of the MHC I-peptide repertoire and are ready to be translocated into the ER. Nonetheless, these precursors need further trimming in the ER in order to be suitable for loading receptive and correctly assembled MHC I molecules [73].

1.5.2.2 Generation of peptide-receptive MHC I molecules

Production of MHC I peptides requires a concomitant pathway that generates peptide-receptive MHC I molecules in the ER. Similar to what occurs with the peptide precursors, both the MHC I heavy chain (α chain) and the β2m polypeptides are cotranslationally translocated into the ER [74]. During translocation, the nascent MHC I heavy chains are bound by the chaperone **binding immunoglobulin protein (BiP)** and modified into a monoglucosylated form by **glucosidases I and II** [75] (Figure 5). Subsequently, early folding and oxidation of MHC I heavy chain is mediated by the chaperone **calnexin** [39,76]. These events are followed by 2m association and replacement of calnexin by another lectin-like chaperone, **calreticulin**, which is part of the **peptide-loading complex (PLC)** [39] (Figures 5 and 6). The multisubunit PLC includes the MHC I heavy chain, 2m, carltericulin, the transmembrane protein **tapasin**, the oxidoreductase **ERp57** and the **transporter associated with antigen processing (TAP)** [39]**.** The luminal domain of TAP acts as a binding platform for calreticulin and ERp57, which provide quality control and mediate the formation of disulphide bonds, supporting the correct folding of MHC I in the PLC [76-78]. In this way,

this multisubunit structure keeps the MHC I molecules in a peptide-receptive state [39]. Binding to a high-affinity peptide is needed for the stabilization of MHC I heterodimers, which otherwise are destroyed by the ERAD system [76].

1.5.2.3 Peptide processing in the ER

Besides retaining empty MHC I molecules, **TAP** is responsible for the active transport of peptide precursors (8-16 amino acids length) from the cytosol into the lumen of the ER [79], where they can access empty MHC I heterodimers [78]. The TAP complex is a heterodimer composed of the **TAP1** and **TAP2** molecules and shows substrate specificity when selecting peptides for translocation [80,81] (Figure 5). Most MHC I peptides are TAP-dependent, as evidenced by impaired surface expression of MHC I-peptide complexes in cells lacking TAP1 or TAP2 [81].

Peptide precursors are translocated into the ER where they can follow one or more different fates. They can bind MHC I molecules, be protected by ER chaperones, be trimmed by ER aminopeptidases, be degraded or be retrotranslocated back into the cytosol [82]. Together, these processes keep a low concentration of peptides in the ER such that only the most recent peptides are available for MHC I binding and do not have to compete with those that arrived earlier [69].

Peptides entering the ER via TAP bind to newly synthesized MHC I molecules immediately, as long as the appropriate motif is present [62]. At this point, the length is not yet crucial [62]. This initial peptide binding is followed by peptide exchange and editing in the ER [83]. Generation of correct MHC I peptides sometimes requires trimming by ER aminopeptidases. The key enzyme responsible for generation of quality peptides and final N-terminal trimming are the ubiquitously expressed **ER amino peptidases associated with antigen processing, ERAP1 and ERAP2** in humans (ERAAP in mice) [84] (Figure 6). ERAP1 recognizes the peptide carboxyl terminus and trims the amino terminus to generate peptides 8-10 residues in length [85]. Of note, ERAP1 is induced by IFN-γ [84,86] and it serves a unique function in modifying the quantity and the quality of the MHC I-peptide repertoire and influencing CD8+ T cell responses

[87,88]. ERAP1 deficiency (through small interfering RNA or in ERAAP knockout mice) leaves some peptides unaffected, whereas others are either absent or dramatically upregulated [22,65,85]. Consistent with these findings, mass spectrometry analysis revealed a drastic increase in MCH I peptide length in mice lacking ERAAP [52,89]. ERAP2 also trims residues from the N terminus of the peptide, but in contrast to ERAP1, the former is more active in cleaving N-terminal basic residues, does not stop trimming peptides smaller than 8-9 residues in length and its silencing only reduced overall MHC I expression by ~10% [56,65]. Additionally, the C-terminal editing of proteasome-produced peptide precursors can be performed by the carboxypeptidase **angiotensin-converting enzyme (ACE)** in the ER [57].

Peptide precursors can also bind ER chaperones that protect them from degradation [82,90]. The **protein disulfide isomerase (PDI)** appears to be the most efficient peptide-binding ER chaperone, as it binds to peptides of different length and sequence [91]. Peptides not bound either to ER chaperons nor to MHC I molecules, can be trimmed and destroyed by ERAP1/ERAP2 [92], or retrotranslocated back into the cytosol for **ER-associated degradation (ERAD)** [82]. In this way, they no longer compete for space in the local compartment.

1.5.2.4 Peptide loading and presentation

The PLC orchestrates the final assembly of MHC I molecules with peptides (now 8-11 amino acids), delivered into the ER by TAP, for generation of stable MHC class I-peptide complexes [61]. The transfer of translocated peptides to peptide-receptive MHC I molecules is facilitated by 2 **tapasin** molecules that act as bridge. Tapasin interacts not only with TAP and ERp57, but also recruits MHC I-β2m heterodimers and calreticulin to the PLC [17,93] (Figure 5). Moreover, tapasin plays a peptide-editing role, mediating the binding of high-affinity peptides at the expense of low-affinity peptides, as shown in tapasin-negative cells in which MHC I surface complexes are less stable [94]. A recent study using -deficient mice has revealed that this enzyme plays a role in defining the C-terminus of MHC I peptides [52], a function previously attributed to the proteasome [43]. Tapasin function is well complemented with that of the **UDP-glucose-glycoprotein glucosyltransferase 1 (UGT1)** enzyme, which

reglucosylates the heavy chain of MHC I molecules with suboptimal peptides, allowing reentry of the MHC I into the PLC and exchange for high affinity peptides [17,75,95]. Thus, the chief function of the PLC is to provide 'quality control' by selectively retaining MHC I molecules loaded with suboptimal peptides for replacement by higher affinity peptides [74]. High-quality peptides that confer stability to the MHC I molecules share two important properties: the precise length and amino acid sequence required for a given MHC class I-binding motif [74]. When any of the PLC constituents are missing or inhibited by viruses, intracellular MHC I molecules can suffer unfolding, degradation and indiscriminate peptide loading, all of which can compromise the stability, expression and function of MHC I-peptide complexes at the surface (see some exceptions in the next section) [74,88].

Successful peptide loading induces dissociation of the MHC I molecule from tapasin, removal of the glucose residue by glucosidase II and export of the MHC I-peptide complexes escorted by the **tapasin-related protein (TAPBPR)** [96] through the Golgi cisternae [61],. In the Golgi, suboptimally loaded MHC I molecules can be retrieved to the ER or be directed through the constitutive secretory pathway to the cell surface [39]. MHC I-peptide complexes segregate into peptide-specific clusters that have been associated with increased T cell sensitivity [97].

1.5.2.5 The end of MHC I life

Surface MHC I-peptide trimolecular complexes can be released through metaloprotease cleavage [98], or be recycled via the endocytic pathway [16] (Figure 6). A fraction of endocytosed MHC I molecules can be recycled after peptide exchange with high affinity endosomal peptides [16,99] (see next section). The proteins **MARCH4** and **MARCH9** have been shown to control the half-life of MHC I molecules through ubiquitination that promotes MHC I internalization and lysosomal degradation [100].

Peptide-free MHC I dimers are also recycled via the endocytic pathway [99]. Peptide-free MHC I heavy chains without or with very weakly bound β2m are known as **open MHC I conformers** and can result from dissociation of the

heterodimer caused by low-affinity peptides or acidic pH [99,101]. Empty MHC I heavy chains have a shorter half-life on the cell surface than trimeric MHC I-peptide complexes [99]. Empty MHC I heavy chains have been shown to interact in *cis* and *trans* with various receptor molecules including those for insulin, interleukin 2 and growth factors [102].

1.5.3 Alternative MHC I antigen presentation pathways

The proteasome-TAP pathway is considered as the conventional processing route. However, MHC I peptide presentation can take place (albeit frequently modified) in the absence of one or more molecules of the pathway [103-105]. Viral evasion strategies targeting different stages of the antigen processing pathway have been very useful not only to understand the classical pathway, but also to elucidate non-traditional alternative pathways [106-108]. Moreover, polymorphisms in MHC I molecules not only result in different peptide-binding grooves, but can also affect aspects of the antigen processing such as MHC I assembly and the rate of transport to the cell surface [16,30,109,110]. Accordingly, studies using proteasome inhibitors have shown a variable effect on MHC surface expression depending on the HLA class I allomorph, suggesting the existence proteasome-independent mechanisms for the generation of peptides that bind specific HLA I variants such as HLA-B27 [111,112]. Interestingly, characterization of proteasome-independent ligands associated to HLA-B27 has revealed that these peptides originate mostly from small and basic proteins and suggest the contribution of additional proteases [113].

Allelic variation in MHC I molecules also influences their dependence on tapasin for peptide loading [103] and their association with TAP, calreticulin and the PLC [30,103,109,114-119]. Neisig and coworkers have tested the *in vitro* association of HLA-A, -B and –C heavy chains with TAP and demonstrated that most HLA-A and HLA-C alleles efficiently interacted with TAP, whereas the majority of HLA-B alleles showed inefficient association [114]. Additionally, HLA-C molecules are more selective in their peptide binding than HLA-A and HLA-B, resulting in prolonged association with TAP [120]. This prolonged association has been proposed to explain the lower surface levels of HLA-C molecules in comparison to HLA-A and HLA-B molecules [120,121].

The import of some peptides appears to be TAP-independent [122]. TAP-independent peptides include peptides derived from signal sequences (or leader sequences) of certain ER-targeted proteins that are cleaved by signal peptide peptidases, can be further processed in the ER and bind particularly HLA-A2 and HLA-B51 molecules [123-125]. However, it is unclear how the loading of these peptides occurs [105]. Interestingly, mutations in the TAP gene are responsible for HLA I deficiency in human subjects who manifest a reduced functional CD8+ population but nevertheless may appear asymptomatic for long periods of their lives [126,127]. Also, it has been demonstrated that the proteases **furin** and **PC7**, located in the trans-Golgi and/or endocytic vesicles network, can process some peptide epitopes from secretory proteins [127-130].

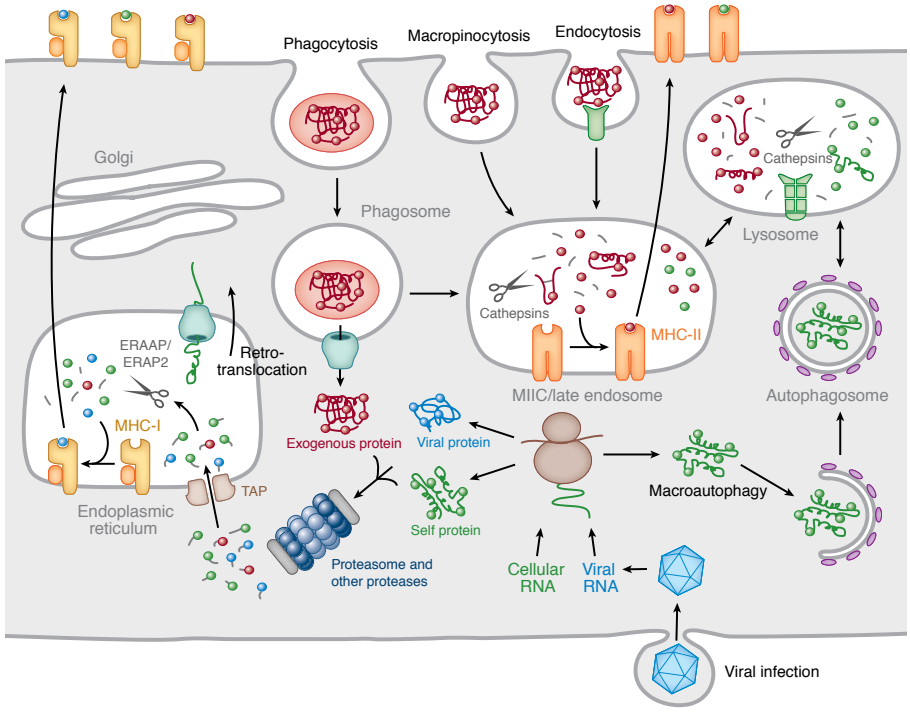


Figure 7. Basic and some alternative antigen processing and presentation pathways
Adapted from [17]

There is a growing amount of evidence suggesting that there is a variety of alternative pathways for generation of MHC I peptides (Figure 7). For instance, **autophagy**, which is responsible for clearance of old or damaged cellular com-

ponents, might contribute to the peptide repertoire [17,34,35]. During autophagy, acidic double-membrane vacuoles (autophagosomes) target ubiquitylated cytoplasmic protein aggregates for degradation through fusion to lysosomes without involving the endocytic or vacuolar protein sorting pathways [35]. Autophagy has been shown to contribute to MHC I presentation of endogenous viral peptides in macrophages infected with Herpes simplex virus (HSV) [131,132].

1.5.3.1 Cross-presentation

One of the most remarkable alternative MHC I presentation routes is the **cross-presentation (or cross-priming) pathway**. Cross-presentation occurs routinely in APCs such as macrophages and DCs, in which exogenous proteins are internalized to generate antigens that are presented by MHC I molecules [35,133]. The display of exogenous peptides on MHC I molecules links the MHC I and II processing pathways and is important for effective immune responses to tumors and viral infections [134]. The molecular mechanisms of cross-presentation are not completely understood yet and many pathways have been described, including TAP-dependent and TAP-independent mechanisms, even though the former seem to dominate [17,35]. APCs can internalize antigens by endocytosis, phagocytosis, receptor-mediated endocytosis, pinocytosis and even gap-junctions [50]. Moreover, different intracellular compartments have been proposed for the loading of cross-presented peptides with MHC I molecules such as endosomes or phagosomes, the ER and lysosomes/vacuoles [135]. In either case, peptides in the phagosome destined to cross-presentation have to enter MHC I presentation before lysosomal degradation and before loading onto MHC II molecules [134]. Besides, cross-presentation routes seem to vary according to the maturation state and subset of APCs [136].

The two studied popular cross-presentation pathways are the **phagosome-to-cytosol pathway** and the **vacuolar pathway**. In the phagosome-to-cytosol pathway, antigens need to be exported from the endosome or phagosome to the cytosol for proteasomal degradation and further processing via the ER and the classical pathway in a TAP-dependent manner [137,138]. It remains unclear yet how does the transport from the endocytic pathway to the cytosol

occur. Proposed mechanisms involve transient physical rupture or leaking of the endosomal membrane and the action of a specific channel or translocator, such as SEC61 or the ER translocon of the ERAD machinery resulting from ER-phagosome fusion [35,129,139,140]. Alternatively, antigens might make use of an established retrograde pathway leading from endosomes to the ER via the Golgi. Form the ER, they may reach the cytosol using the translocation channel involved in retrotranslocation during protein degradation [140]. It has also been suggested that an alternative transport pathway (such as lipid droplets) could connect the phagosome and the ER without any direct ER-phagosome fusion [35].

In the vacuolar pathway, antigens are bound onto MHC I molecules within phagosomal compartments and thus outside of the ER. Indeed, internalized surface MHC I molecules can be present in phagolysosomal compartments [34,35,99]. The peptides are either processed in the cytosol (phagosome to cytosol to phagosome model) or in the phagosome itself [34,50]. In either case, the phagosome needs to be equipped with molecules necessary for peptide loading onto MHC I molecules.

Internalized antigens can take many different forms, ranging from cell debris from apoptotic or necrotic cells to proteins or chaperone-associated peptides [140]. For instance, DCs can ingest infected non-immune cells, cancer cells or cell-derived fragments and generate antigens from these sources [141]. Also, exogenous proteins introduced directly into the cytosol of a cell are recognized by CD8+ T cells [142]. This is the case of listeriolysin, a protein secreted by the intracellular pathogen *Listeria monocytogenes* after its internalization by macrophages [143]. The mannose receptor has been implicated in the internalization and routing of some antigens into a distinct endosome subpopulation where they are protected from lysosomal degradation and can be further processed for cross-presentation [144]. This early endosome subpopulation do not mature to late endosomes, in which peptides are loaded into MHC II molecules. Besides, it has been shown that MHC I-peptides can be transferred from an infected cells to APCs through gap-junctions [145].

Internalized antigens can be processed (or destructed) by endosomal and ly-

sosomal proteases. Mechanisms that reduce the activity of endosomal hydrolases in DCs have been shown to enhance the efficiency of cross-presentation [61]. **Cathepsins** have been involved in the generation of MHC I peptides in endolysosomal compartments in a proteasome- and TAP-independent manner [146]. Additionally, the **insulin-regulated aminopeptidase (IRAP)** exclusively trims peptides in endosomes for cross-presentation [56,135]. Proteases acting in the endogenous pathway are presumably also involved in the processing of ligands for cross-presentation [50].

Despite being less common, cross-presentation is essential to trigger immune responses to viruses that do not infect APCs such as Epstein-Barr virus (EBV) and hepatitis B virus [147]. Also, tumor antigens released from tumors can be cross-presented in draining lymph nodes and this property is boosted by chemotherapy and can lead to stimulation of antitumor immunity [134]. Moreover, cross-presentation allows CD8+ T cell recognition of antigens coming from tissues that do not express MHC I molecules, such as placental trophoblasts [148]. In this way, antigens of fetal origin can be presented by maternal APCs thus eliciting a maternal immune response to the fetus [148].

1.6 The origin of MHC I-associated peptides

Peptides displayed by MHC I molecules derive from degradation of proteins acquired from an exogenous source (cross-presentation) or from proteins endogenously synthesized (direct presentation) [37] (Figure 8). The complexity of the MHC I-immunopeptidome reflects the equally complex milieu of intracellular proteins [37]. When and how intracellular proteins are chosen for entry into the antigen processing pathway? This is an interesting yet not completely solved question. What is now clear based on various studies including ours, however, is that the MHC I-peptide repertoire is not a random sample of the proteome. More details about this bias can be found in the section "*The SMII is complex and is not a representative excerpt from the proteome*" of our published review article included in chapter 2 and in our published article "*MHC I-associated peptides preferentially derive from transcripts bearing microRNA response elements*" included in chapter 3. In the following section, the mechanisms by which endogenous proteins can give rise to peptides are described.

1.6.1 Endogenous proteins as source of peptides

Most of the peptides destined for presentation on MHC I molecules are generated by proteasome-mediated cleavage of endogenous polypeptides [149]. However, the source of proteasomal substrate is quite varied. Peptide ligands for MHC I molecules can derive from the degradation of correctly folded "stable" proteins, **defective ribosomal products (DRiPs)**, non-classical transcription or translation products or from proteins destined to the secretory pathway that are retrotranslocated from the ER to the cytosol (Figure 8). Identification of the primary source of MHC I-associated peptides has remained a very controversial topic for years. The DRiPs hypothesis [150], which points rapidly degraded and defective proteins as the main source of MHC I-associated peptides, is currently the most popular yet controversial hypothesis.

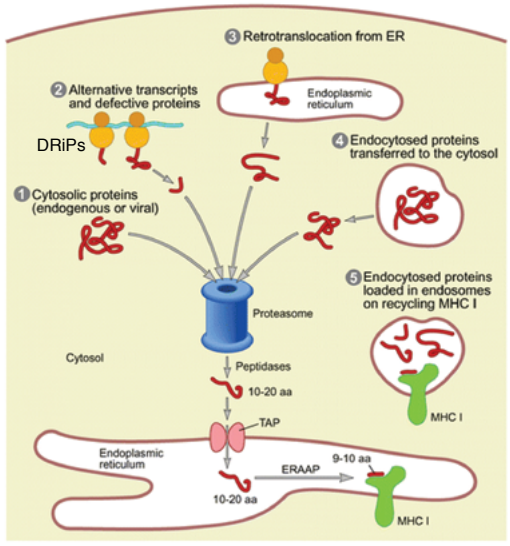


Figure 8. Endogenous and exogenous sources of MHC I-associated peptides Sources include (i) cytosolic proteins, (2) alternative translation products and DRiPs, (3) proteins retrotranslocated from the ER to the cytosol, (4) and (5) endocytosed proteins. Adapted with permission from [140].

1.6.1.1 Rapidly versus slowly degraded polypeptides

Proteins exhibit a wide range of degradation rates: from minutes to weeks, with an overall half-life of 1-2 days [151]. Polypeptides have been segregated

into two general pools: (i) those degraded with an average half-life of ~30 minutes, named rapidly degraded polypeptides (RDPs), and (ii) those degraded with an average half-life from hours to weeks (in average ~2000 minutes), referred to as slowly degraded polypeptides (SDPs) [151]. In mammalian cells, approximately 30% of total proteins correspond to RDPs [149].

1.6.1.2 The DRiPs hypothesis

If MHC I peptides derive exclusively from the degradation of proteins that are at the end of their functional lives, it would be expected that the time between the synthesis of a protein and the presentation of its peptides should reflect the half-life of the protein. Yewdell and coworkers pointed out that this rate is inconsistent with *in vitro* assays showing that cells become recognizable by CD8+ T cells soon after they are infected and thereby that peptide production must begin very shortly after protein synthesis [150]. In fact, one remarkable aspect of the antigen presentation pathway is the speed with which peptides can be generated following infection.

Early experiments showed that expression in cells of unstable truncated proteins generated MHC I-peptide complexes as effectively as full-length proteins [17,152]. Consequently, Yewdell *et al.* proposed that immediate peptide supply is driven not by degradation of mature proteins but by newly synthesized proteins that are defective, termed DRiPs [150] (Figure 9). DRiPs include polypeptides that fail to achieve its native structure, owing to imperfections in transcription, splicing, translation, alternative reading frame usage, post-translational modifications or protein folding, and are flagged by the quality-control machinery and rapidly degraded [72,150,153-156]. Accordingly, ~30% of all proteins made are immediately degraded after synthesis before forming functional proteins [149,157].

A significant proportion of peptides appear to result from the degradation of newly synthesized, but rapidly degraded polypeptides as opposed to slowly degraded polypeptides [149,157,159]. This is consistent with MS-based studies reveling the absence of correlation between the proteome and the MHC I-peptide repertoire of the same cell [160]. Accordingly, some MHC I peptides

derive from proteins that are undetectable in the cell [161,162]. It is still not clear what proportion of RDPs represents truly short-lived proteins and how much represents DRiPs. Strong evidence has been published supporting the idea that translation and protein folding must be error prone [61] and that newly generated polypeptides, including DRiPs, represent the main source of antigens entering the MHC I processing pathway [151,163]. Thus, it has been suggested that MHC I molecules preferentially sample what is being translated as opposed to what has been translated [157,161,164]. Therefore, the concentration of peptides, which is related to the rate of protein synthesis, is the rate-limiting step for MHC I-peptide presentation [16].

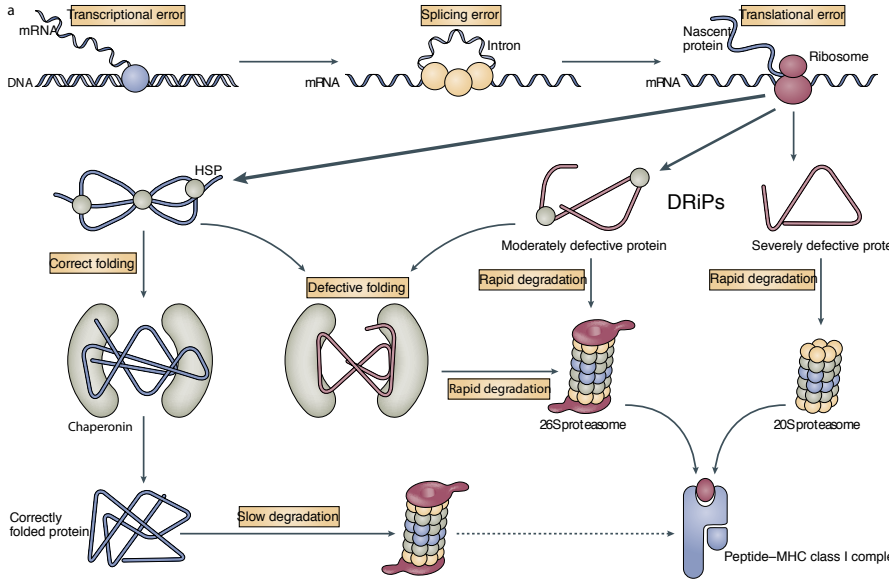


Figure 9. The DRiPs hypothesis
Accumulated errors can lead to defective proteins that are targeted for rapid degradation with some products ultimately becoming MHC I-associated peptides. Adapted with permision from [158].

Some indirect evidence initially supported the DRiPs hypothesis. First, early studies showed that mutant and misfolded proteins induced in the presence of certain compounds are immediately degraded after their synthesis [158,165,166]. Second, a considerable fraction of newly synthesized proteins is rapidly turned over. This fraction has been estimated to correspond to 30-40% per hour [149]. Third, increasing the degradation rate of an antigen considerably augments peptide production [167]. The DRiPs hypothesis was sub-

sequently supported by studies showing that peptide production ceased 30 minutes after inhibition of protein synthesis, suggesting that mature protein turnover is too slow to contribute significantly to the peptide pool [157,163].

It is not entirely clear what exact mechanisms drive DRiPs formation. Recent work from Fahraeus and coworkers points toward normal translational processes such as the nonsense-mediated decay [168]. They demonstrated that an mRNA with a premature stop codon and degraded after a single round of translation, can generate an epitope efficiently recognized by T cells [168]. Notably, this process is distinct from typical translation as it is eIF4G-dependent and eIF4E-independent. Also, a recent study points toward hydrophobicity as a key signal for immediate degradation and consequently peptide presentation [169]. Presence of longer hydrophobic sequences such as transmembrane domains enhanced generation of MHC I-peptide complexes [169]. It has also been proposed that DRiPs are produced predominantly by ribosomes lacking associated chaperones that directly transfer DRiPs for proteasomal degradation [156]. This model extends the concept of the "**immunoribosome**", a ribosome subset that exclusively and efficiently generates proteins targeted to antigen processing and that is distinct from ribosomes responsible for conventional translation [151]. Accordingly, Apcher and coworkers demonstrated that the production of antigenic peptides and full length proteins do not occur at the same time and suggested that they could be produced by alternative ribosomal complexes [168]. In line with this, Eisenlohr and coworkers have suggested that most nascent polypeptides are not inherently defective as the linear sequence is correct, and that defectiveness might be considered to lie more with the ribosome and its associated chaperones [169].

Some aspects of the DRiPs hypothesis have been questioned [158]. First, a DRiP has not yet been conclusively identified nor produced [163,170]. Second, a window of 30 minutes for peptide production implies a very short half-life (15 minutes or less) for the substrates from which the peptides are derived and a fast disposal of defective proteins [72]. This seems not to be in line with current concepts of protein production and quality control arguing against the DRiPs model [158]. There is an ever growing list of 'natively' or 'intrinsically' unfolded proteins that bypass the quality control machinery and are not

degraded [158,171]. Additionally, many misfolded proteins can be rescued by prolonged interaction with heat shock proteins [172]. Accordingly, increased degradation is not always the fate of misfolded proteins as they can enter aggregates that resolve very slowly or become candidates for ubiquitin-mediated autophagy and not proteasomal degradation [173]. The DRiPs model has also been called into question by evidence showing that newly synthesized polypeptides are mostly protected from proteasomal degradation during and immediately after translation and that preexisting proteins represent the main proteasome substrates [159].

An alternative model not excluding the DRiPs hypothesis, proposes that a subset of nascent polypeptides is neglected by the folding machinery and stochastically delivered to the 20S proteasome independent of quality control decisions [158]. For a given antigen, the basal level of peptide presentation from immediately degraded substrate by the 20S proteasome will be complemented by the subset of newly synthesized proteins that is successfully captured by the folding machinery [158]. According to this hypothesis, the more defective the protein is, the sooner and greater the presentation of the peptide will be, which is due to more rapid rejection by the quality control machinery [158].

1.6.1.3 Stable proteins as source of peptides

Long-lived intact proteins can also contribute to the peptide pool although possibly in a lesser extent [151]. MHC I-peptide ligands can be obtained from folded and functional proteins that have passed the quality control, as evidenced by presentation of several species of posttranslationally modified peptides [174]. Post-translational modifications evidenced in MHC I peptides include N-glycosylation [175,176], cytosolic O-GlcNAc glycosylation [177], deamidation [175,178-181], methylation [182], acetylation [183], phosphorylation on serine and threonine residues [184-186] and cysteinylation [187,188].

1.6.1.4 Cryptic translation as a source of naturally processed peptides

In addition to conventional translation products, cells can also generate peptide ligands for MHC I molecules from cryptic translation products. Cryptic trans-

lation refers to polypeptides that are synthesized by unconventional translational mechanisms [189]. These include peptides encoded by open reading frames (ORF) contained within 5' and 3' "untranslated" regions (UTR), alternative open-reading frames (ARF), introns, intron-exon junctions or from non-AUG start codon initiation on both endogenous and viral mRNAs [189-191]. The list of MHC I-peptides derived from cryptic translation has been steadily growing and many examples of peptides of viral origin or in tumor cells have been described [189,192-194]. Some nonclassical peptides were found to arise from aberrant transcription of intronic sequences [195-197] or reverse strand sequences [198]. Others resulted from translation of UTRs [199] or of alternative open reading frames [20,200,201], even involving leucine-tRNA-mediated initiation of translation at a CUG codon, instead of classical AUG codon [202-204]. Cryptic peptide also result from translational errors including ribosomal frameshifting whereby ribosomes may initiate at the primary ORF start codon but slip either forward or backwards and continue translation in an ARF [189,205]. Importantly, cryptic peptides can induce tolerance in transgenic mice that generate cryptic peptides, and elicit CD8+ T-cell responses in normal mice [206]. Overall, studies on cryptic peptides have shown that they are immunologically significant and can provide a protective role in viral infections [189].

1.6.1.5 Peptides derived from peptide splicing

It has been reported that MHC I ligands can contain sequences not encoded in the genome. Reported examples include peptides derived from non-contiguous sequences in the original protein and resulting from the splicing of neighboring or non-contiguous peptides by the proteasome either in the initial or reverse order [179,207-211]. For instance, Vigneron and coworkers described an HLA-A32-associated nonamer that was derived from a 13 amino acid precursor by excision of 4 internal residues followed by splicing by the proteasome [208]. Another example includes one peptide resulting from the splicing of two non-contiguous peptides from the same protein in the reverse order by the proteasome [210].

1.6.1.6 Peptides derived from proteins destined to the secretory pathway

Secretory and membrane proteins are a known source of MHC I-peptides. Many peptides derived from transmembrane or secretory proteins correspond to sequences in the transmembrane regions or signal sequences [37]. In order to be degraded, these proteins have to exit the ER because there are no proteasomes present in the ER lumen [212]. Proteins destined to the secretory pathway can gain access to the cytosol after being retrotranslocated from the ER, in a process that typically results in ubiquitination and proteasomal degradation via the ERAD pathway [176]. Generation of peptides via ERAD of glycoproteins was evidenced by the discovery of epitopes in which the asparagine residue (normally present in the source glycoprotein) was replaced by an aspartic acid residue. This deamidation of asparagine residues reflected the conventional deglycosylation step of glycoproteins in the cytosol that follow the ERAD pathway [178,179].

1.7 The MHC I immunopeptidome: exposing the inside of the cell to the immune system

The repertoire of peptides presented by MHC I molecules, is collectively referred to as the MHC I immunopeptidome (for the sake of simplicity hereafter referred to as immunopeptidome) [213,214]. Different estimations have been made regarding the number of surface MHC I molecules per cell: 50,000-100,000 [215], 30,000-120,000 [216], 10,000-500,000 [217] or at least 50,000 [218], although these numbers may vary depending on the cell type and cell condition [219]. The immunopeptidome is estimated to be composed of more than 10,000 different complexes reflecting its high complexity [219]. Many peptides are present at one or less copies per cell [220,221] and most peptides are represented at 10-400 copies per cell [216]. Notably, a single viral peptide can be presented at 100,000 copies per cell [222]. Recent semi-quantitative MS-based studies have estimated that the number of peptide copies associated to intracellular HLA molecules varies from 1 to 16,500 copies per cell [218]. Notably, even a single MHC I-peptide complex may be sufficient to trigger a T cell [223,224].

Here, a brief introduction of the immunopeptidome is presented and more details can be found in the published review article "*Origin and plasticity of MHC I-associated self peptides*" included in chapter 2.

1.7.1 Biological roles of the MHC I immunopeptidome

The primary function of the immunopeptidome is to present antigens to T-cell receptors and thus to provide protection against pathogens and neoplastic transformation [4]. In the absence of infection, cell surface MHC I molecules are associated solely with self-peptides that play vital and diverse roles [213].

The immunopeptidome shape the repertoire of developing thymocytes through regulation of positive and negative selection [225-227]. In addition, self MHC I-peptide complexes transmit survival signals to mature CD8+ T cells [228]. The immunopeptidome also allows immunosurveillance of neoplastic cells [229,230]. It reflects the state of the cell as mutations, or genes overexpressed in tumors modify and shape the self immunopeptidome [153,231,232].

Self MHC I-peptide complexes can also contribute to the recognition of nonself MHC I-peptide complexes by the TCR and in this way amplify T cell responses against intracellular pathogens [233]. Upon infection, the MHC I-peptide repertoire reflects not only the intracellular protein milieu but also additional proteins such as those derived from intracellular pathogens [153,232,234,235]. Notably, viral infection can not only generate viral antigens, but it can modify the presentation of self MHC I-peptides and/or lead to the presentation of neo-self MHC I peptides [236,237], reflecting metabolic changes in the cell [238]. Altogether, these modifications render the otherwise invisible internal proteome available for surveillance by cytotoxic CD8+ T cells, which have the ability to detect and eliminate cells expressing viral proteins or tumor antigens [191].

MHC molecules also provide a link between the innate and the adaptive immune systems. Changes in MHC I surface expression can be a sign of infection, cancer or invading cells from another person (e.g. pregnancy) [239]. Natural

killer (NK) cells monitor other cells for the quality and quantity of their MHC I expression through **killer cell immunoglobulin-like receptors (KIRs)** on their cell surface [239]. Interaction of KIRS with MHC I molecules at a site distant to the peptide binding groove provides positive or negative signals to NK cells through **immunoreceptor tyrosine-based activating motifs (ITAM)** or **immunoreceptor tyrosine-based inhibitory motifs (ITIM)** localized within KIR molecules [1,240,241]. In this way, immune evasion of a pathogen or a tumor cell that has downregulated MHC I antigen presentation, can be avoided [4,240]. Interestingly, inhibition of NK cells resulting from KIR recognition of MHC I molecules can be impaired by certain incompatible amino acids at positions P7 and P8 of nonamer peptides bound to the MHC I, leading to NK cell activation [241].

The MHC I immunopeptidome is also involved in immunopathology. It can be targeted by autoreactive T cells that initiate autoimmune diseases such as diabetes [242-245]. In addition, tumor-associated antigens can elicit paraneoplastic autoimmune disorders [246]. Moreover, MHC I-peptide complexes can be targeted by alloreactive T cells that cause graft rejection and graft-versus-host disease following transplantation [247] (more details are given in chapters 4 and 5). It has been recently shown that the drug abacavir used for the treatment of HIV, can affect the MHC I peptide repertoire through binding to one of the HLA-B*5701 pockets [248]. Also, inhibitors of the carboxipeptidase ACE are used to treat patients with hypertension and congestive heart failure and this enzyme has been shown to increase, decrease or have not effect on the surface expression of individual epitopes [57]. Thus, ACE inhibitors presumably change the peptide repertoire with potential consequences difficult to predict with our current knowledge.

The aforementioned roles of the immunopeptidome highlight its immunotherapeutic potential. In line with this, peptides overexpressed and/or specific to neoplastic cells can be used as targets in cancer immunotherapy [153,232,234,249,250], whereas viral antigens could be used in vaccination strategies against infections [251,252]. On the other hand, knowledge about the origin of viral or tumor-associated antigens could be exploited to induce immunity [253].

Apart from its role in immune function, the MHC I-peptide repertoire has been shown to influence mating preferences and other behaviors in mice [254] and humans [255]. According to these studies males and females favor mates expressing dissimilar MHC molecules and females prefer males exhibiting MHC heterozygosity [255-257]. These observations could be related with a seminal work demonstrating that dissociated MHC I peptides (and not MHC I-peptide complexes) can activate sensory neurons in the vomeronasal organ, which is specialized in initiation of behavioral or endocrine responses in mice [254,258]. These studies have revealed that MHC genes play an essential role in determining individual differences and preferences in body smell [256]. Furthermore, MHC I molecules have been shown to participate in neuronal development and function and regulate synaptic plasticity in the hippocampus through recognition by non-TCR receptors [259].

All these examples illustrate the global role of the MHC I immunopeptidome in communicating the intracellular milieu to the surrounding environment [191,260].

1.7.2 Characterization of the immunopeptidome by mass spectrometry

The first attempts to characterize MHC I peptides were done by the group of Rammensee in 1990, who purified MHC I molecules from whole cells, extracted the bound antigens and fractionated them by **high-performance liquid chromatography (HPLC)**. After analyzing the peptide mixture by Edman degradation they discovered the presence of dominant amino acids that were dependent on the MHC I molecule from which the peptides were isolated, i.e. the anchor residues of peptide-binding motifs [26,261,262]. These seminal studies were followed by the first characterization of MHC I peptides by HPLC and tandem **mass spectrometry (MS)** in 1992 performed by the groups of D. Hunt and V. Engelhard, who identified endogenous peptides bound to HLA-A*0201 [220]. These studies significantly contributed to the understanding of the factors that control the binding of peptides to MHC I molecules and in conjunction with increasingly sensitivity of mass spectrometers, have prompted the large-scale identification of MHC I-peptides [16].

MS has also enable to estimate the representation of individual peptides and the global complexity of the immunopeptidome [216]. Besides, MS-based studies have allow the identification of different types of peptides: tumor-associated antigens [178,216,263], transplantation antigens (see chapters 4 and 5), antigens derived from intracellular bacteria [264], antigens with posttranslational modifications [174] and antigens associated to autoimmune disease [265]. In addition, studies of the immunopeptidome by MS under particular conditions such as absence of tapasin or TAP or variations of the proteasome, have contributed to the understanding of the antigen presentation pathway [19,48,108,123-125,216]. A review of large-scale studies aiming to characterize the immunopeptidome in different human and mouse cell types or tissues is shown in tables 1 and 2 of the published review article "*Origin and plasticity of MHC I-associated self peptides*" included in chapter 2.

Numerous strategies for qualitative and quantitative analysis of the immunopeptidome have been developed. These strategies require first the isolation of MHC I peptides, followed by their identification and their quantification [219]

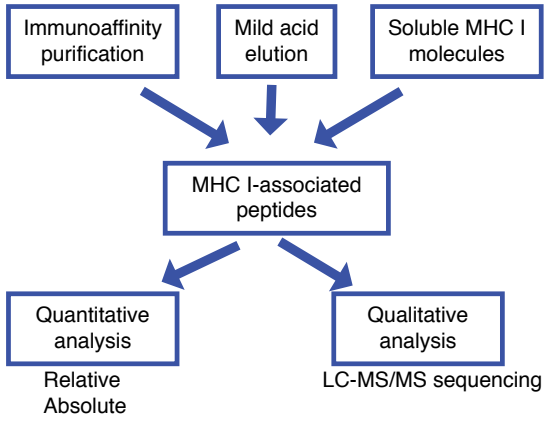


Figure 10. Simplified flowchart of analysis of MHC I-associated peptides by MS

(Figure 10).
Isolation of MHC I peptides can be performed by mild acid elution, immunoaffinity purification or by using soluble MHC I molecules. Mild acid elution induces the dissociation of β2m, denaturation of the MHC and peptide dissociation [101]. Mild acid elution of peptides from viable cells has the advantage

of capturing peptides that are indeed exposed on the cell surface and that bind to all MHC I molecules with all ranges of affinities, but its main caveat is the elution of non-MHC contaminant peptides [101,219,266]. Immunoaffinity purification is a more targeted method in which the MHC I is immunoprecipitated by affinity column from cell lysates and peptides are eluted with strong acid [267]. This is currently the most used method and although the isolated peptides are MHC-specific, the lysis detergent constitutes another source of contamination [219]. Moreover, the immunopurified peptides correspond to intracellular peptides and the isolated repertoire is influenced by the antibody's affinity and availability for any given MHC allotype [219]. Consequently, peptides with strong binding affinity are favored [266]. The third isolation method is based on the transfection of cells with MHC molecules lacking the transmembrane domain [268-270]. These molecules are secreted and further purified by immunoprecipitation. This technique has the advantage of providing high amounts of peptides but its main disadvantage is the induction of non-physiological conditions since MHC overexpression could affect the quality and quantify of presented peptides [219]. Furthermore, it is only applicable to "transfectable" cell lines impeding its use in primary tissues [219]. A very recent approach by Admon and coworkers has combined immunoprecipitation technique to analyze soluble MHCs. They have found that large amounts of soluble peptide-bound HLA molecules are present in the plasma of cancer patients and thus they have isolated the immunopeptidome from soluble plasma MHC I molecules [271]. Notably, the peptide repertoire of soluble HLA was similar to that of membranal HLA, highlighting the potential of this approach in a clinical context [272].

Following peptide isolation, the peptide mixture is fractionated and peptides in each fraction are separated, ionized and analyzed by MS [273].
MS is a way to accurately measure the weight of a molecule, or more precisely its mass-to-charge ratio (m/z), using electromagnetic fields in a vacuum [273]. To do so, molecules must first be electrically charged and transferred into the gas phase by electrospray ionization, where the m/z ratio of molecules can be determined with different types of instruments (e.g. TOF, Orbitrap, etc). In addition to the exact mass of a peptide, the peptide sequence is determined by fragmentation of each peptide by collision with an inert gas (CID, collision

induced dissociation or HCD, high collision dissociation) [273]. This results in a list of m/z ratios for different fragments of each peptide, called an MS/MS (or tandem MS) spectrum. Peptides are very difficult to detect in very complex mixtures such as mixtures of MHC I-peptides and hence, peptides are separated and ionized by HPLC coupled directly via electrospray to the mass spectrometer (LC-MS/MS) [214]. Because of the complexity of the eluted peptide sample mixture, an additional separation step such as strong cation exchange (SCX) is performed before the analysis to enrich a specific peptide population [273,274]. The measured fragment spectra and peptide masses are matched against protein databases using search engines (e.g. SEQUEST, MASCOT, etc.) [219].

Several MS techniques can be used to do an absolute or relative quantification of MHC I peptides [216]. For instance, a relative quantification can be made by measuring the intensity of detection of a given peptide, which is an estimate of its abundance. Peptide intensities can be compared for each peptide in two different samples analyzed in the same run and in this way detect changes in peptide abundance [48,232,260]. Absolute quantitation (i.e. number of peptide copies per cell) is made by comparing the abundance of isolated peptides to known amounts of their corresponding isotopically-labelled versions. More details on MS-based approaches to quantify and analyze MHC I peptides can be found in [274].

1.7.3 Immunoinformatics and prediction of the immunopeptidome

Advances in mass spectrometry have allowed the identification of increasingly number of peptides and hence a better characterization of peptide-binding motifs specific for most common HLA molecules [219,220,275,276]. Information on peptide-binding motifs as well as the precise amino acid sequence for MHC I-peptides characterized so far, is readily available in public databases such as the immune epitope database (IEDB) [277] and the SYFPEITHI database [28]. The progress in MHC I peptide discovery and in unraveling peptide processing has prompted the development of prediction algorithms for antigen processing [278-280], peptide binding to MHC molecules [281-285] and interaction between MHC I-peptide complexes and the TCR [286,287]. In particular,

MHC I binding prediction algorithms have been shown to be very accurate and have facilitated the characterization of the immunopeptidome [276,283,288].

1.8 Objectives

The ensemble of MHC I peptides or MHC I immunopeptidome presented on the surface of a cell establishes its immunologic identity (chapter 2). The self MHC I immunopeptidome regulates all key events during the lifetime of classic adaptive CD8 T cells: positive and negative selection in the thymus and survival in the periphery. Eventually, CD8T cells detect and/or react to changes in the self immunopeptidome and to non-self or transformed self MHC I peptides (chapter 2). Moreover, MHC I peptides are the targets of several immune reactions including autoimmunity, graft rejection, graft-vs-host disease and antitumor activity (chapter 4). Despite the important role of the MHC I immunopeptidome in defining the self-nonself boundary in health and disease, we know little about its biogenesis.

Recent progress in high-throughput mass spectrometry analyses of the immunopeptidome has remarkably contributed to defining the molecular definition of self for CD8 T cells (chapter 2). Nevertheless, current MS approaches are still blind to MHC I peptides resulting from non-synonymous genomic polymorphisms, known as MiHAs. MiHAs are the targets of nonself-driven immune reactions of great importance in clinical transplantation. The identification of MiHAs is technically challenging and no more than 30 MiHAs were known at the beginning of this work despite the vast number of genetic polymorphisms present in the human population. Therefore, we were particularly interested in implementing high-throughput mass spectrometry combined with other "omic" approaches to study the biogenesis and composition of the human MHC I immunopeptidome and to discover MiHAs.

1.8.1 Research questions and hypothesis

A variety of molecular players and numerous steps at all functional genomic levels (e.g. genome, transcriptome, proteome, degradome) underlie the generation of MHC I peptides and thus have the potential to shape the MHC I im-

munopeptidome. Moreover, a major proportion of MHC I-peptides is believed to derive from the rapid degradation of defective ribosomal products (DRiPs), whose physical nature remains unclear. This complexity and recent studies showing a minimal correlation between the proteome and the immunopeptidome of the same cell, support the hypothesis that the MHC I immunopeptidome is not a stochastic representation of the cellular proteome and that some intrinsic factors may favor the presentation of some self peptides with a particular origin over others.

Of note, we performed data-driven (as opposed to purely hypothesis-driven) studies to address two main questions:

- Which cell intrinsic factors play a role in defining the MHC I immunopeptidome?
- What is the impact of genomic polymorphisms on the human MHC I immunopeptidome?

1.8.2 General objective

The main goal of this work was to unravel the biogenesis and composition of the MHC I immunopeptidome of human B lymphoblastoid cell lines by means of systems biology approaches.

1.8.3 Specific objectives

Aim 1 : To characterize the global landscape of the MHC I-peptide repertoire of two pairs of MHC-identical siblings and to gain insights into the MHC I immunopeptidome biogenesis (chapter 3).

Aim 2 : To discover the impact of non-MHC polymorphisms on the immunopeptidome of two MHC-identical siblings. In other words, we wished to elucidate the relation between the genomic self and the immune self (chapter 5).

1.9 Cellular model

We studied the MHC I immunopeptidome presented by human EBV-infected B lymphocytes (B-LCLs). B-LCLs were derived from peripheral blood mononuclear cells (PBMCs) purified from the blood of four subjects. EBV infection of normal human B cells generally results in the establishment of autonomously proliferating lymphoblastoid cell lines. We used B-LCLs because they can be obtained from practically any subject, they proliferate extensively *in vitro*, express high levels of MHC I molecules at the surface and have been shown to be a reliable tool for high-throughput genomic studies.

1.8 References

1. Blackwell, J.M. et al. (2009) HLA and Infectious Diseases. Clinical Microbiology Reviews 22, 370–385
2. Chaplin, D.D. (2010) Overview of the immune response. Journal of Allergy and Clinical Immunology 125, S3–S23
3. DeFranco, A.L. et al. (2006) Immunity The immune response in infectious and inflammatory disease, (2007 edn) New Science Press.
4. Horton, R. et al. (2004) Gene map of the extended human MHC. Nature Reviews Genetics 5, 889–899
5. Francisco A Bonilla MD, P. and Hans C Oettgen MD, P. (2010) Adaptive immunity. Journal of Allergy and Clinical Immunology 125, S33–S40
6. Carpenter, A.C. and Bosselut, R. (2010) Decision checkpoints in the thymus. Nature Immunology 11, 666–673
7. Sewell, A.K. (2012) Why must T cells be cross-reactive? Nature Reviews Immunology 12, 669–677
8. Klein, L. et al. (2009) Antigen presentation in the thymus for positive selection and central tolerance induction. Nature Reviews Immunology 9, 833–844
9. D'Orsogna, L.J. et al. (2013) Endogenous-peptide-dependent alloreactivity: new scientific insights and clinical implications. Tissue Antigens 81, 399–407
10. Nitta, T. et al. (2010) Thymoproteasome Shapes Immunocompetent Repertoire of CD8+ T Cells. Immunity 32, 29–40
11. Jenkins, M.K. et al. (2010) On the Composition of the Preimmune Repertoire of T Cells Specific for Peptide–Major Histocompatibility Complex Ligands. The Annual Review of Immunology 28, 275–294
12. Robinson, J. et al. (2012) The IMGT/HLA database. Nucleic Acids Research 41, D1222–D1227
13. Rodgers, J.R. and Cook, R.G. (2005) MHC class Ib molecules bridge innate and acquired immunity. Nature Reviews Immunology 5, 459–471
14. Klein, J. and Sato, A. (2000) The HLA system. First of two parts. The New England Journal of Medicine 343, 702–709
15. Yaneva, R. et al. (2010) Peptide binding to MHC class I and II proteins: New avenues from new methods. Molecular Immunology 47, 649–657
16. Neefjes, J. et al. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nature Reviews Immunology 11, 823–836
17. Blum, J.S. et al. (2013) Pathways of Antigen Processing. The Annual Review of Immunology 31, 443–473
18. Flutter, B. and Gao, B. (2004) MHC Class I Antigen Presentation- Recently Trimmed and Well Presented. Cellular and Molecular Immunology 1, 22–30

19. Chen, Y. et al. (1994) Naturally processed peptides longer than nine amino acid residues bind to the class I MHC molecule HLA-A2.1 with high affinity and in different conformations. The Journal of Immunology 152, 2874–2881
20. Probst-Kepper, M. et al. (2001) An alternative open reading frame of the human macrophage colony-stimulating factor gene is independently translated and codes for an antigenic peptide of 14 amino acids recognized by tumor-infiltrating CD8 T lymphocytes. The Journal of Experimental Medicine 193, 1189–1198
21. Escobar, H. et al. (2008) Large Scale Mass Spectrometric Profiling of Peptides Eluted from HLA Molecules Reveals N-Terminal-Extended Peptide Motifs. The Journal of Immunology 181, 4874–4882
22. Rock, K.L. et al. (2010) Proteases in MHC class I presentation and cross-presentation. The Journal of Immunology 184, 9–15
23. Consortium, T.1.G.P. et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073
24. J, M.A. et al. (2003) The DNA sequence and analysis of human chromosome 6. Nature 425, 805–811
25. Warren, E.H. et al. (2012) Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. Blood 120, 2796–2806
26. Falk, K. et al. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351, 290–296
27. Rammensee, H.G. et al. (1995) MHC ligands and peptide motifs: first listing. Immunogenetics 41, 178–228
28. Rammensee, H. et al. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50, 213–219
29. Prugnolle, F. et al. (2005) Pathogen-Driven Selection and Worldwide HLA Class I Diversity. Current Biology 15, 1022–1027
30. H, H.W. et al. (2002) HLA Class I Polymorphism Has a Dual Impact on Ligand Binding and Chaperone Interaction. Human Immunology 63, 248–255
31. Cooper, M.D. and Alder, M.N. (2006) The Evolution of Adaptive Immune Systems. Cell 124, 815–822
32. Costantino, C.M. et al. (2012) Class II MHC Self-Antigen Presentation in Human B and T Lymphocytes. PLoS ONE 7, e29805
33. Crotzer, V.L. and Blum, J.S. (2010) Autophagy and adaptive immunity. Immunology 131, 9-17
34. Chemali, M. et al. (2011) Alternative pathways for MHC class I presentation: a new function for autophagy. Cellular and Molecular Life Sciences 68, 1533–1541
35. Vyas, J.M. et al. (2008) The known unknowns of antigen processing and presentation. Nature Reviews Immunology 8, 607–618
36. Dissanayake, S.K. et al. (2005) Presentation of Endogenously Synthesized MHC

Class II-Restricted Epitopes by MHC Class II Cancer Vaccines Is Independent of Transporter Associated with Ag Processing and the Proteasome. The Journal of Immunology 174, 1811–1819
37. Shastri, N. et al. (2005) All the peptides that fit: the beginning, the middle, and the end of the MHC class I antigen-processing pathway. Immunological Reviews 207, 31–41
38. Hammer, G.E. et al. (2007) The Final Touches Make Perfect the Peptide-MHC Class I Repertoire. Immunity 26, 397–406
39. Hulpke, S. and Tampé, R. (2013) The MHC I loading complex: a multitasking machinery in adaptive immunity. Trends in Biochemical Sciences 38, 412–420
40. Stoltze, L. et al. (2000) The function of the proteasome system in MHC class I antigen processing. Immunology Today 21, 317–319
41. Stoltze, L. et al. (2000) Two new proteases in the MHC class I processing pathway. Nature Immunology 1, 413–418
42. Kunisawa, J. and Shastri, N. (2006) Hsp90α Chaperones Large C-Terminally Extended Proteolytic Intermediates in the MHC Class I Antigen Processing Pathway. Immunity 24, 523–534
43. Cascio, P. et al. (2001) 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. The EMBO Journal 20, 2357–2366
44. Ortiz-Navarrete, V. et al. (1991) Subunit of the “20S” proteasome (multicatalytic proteinase) encoded by the major histocompatibility complex. Nature 353, 662–664
45. Guillaume, B. et al. (2010) Two abundant proteasome subtypes that uniquely process some antigens presented by HLA class I molecules. Proc.Natl.Acad.Sci.U.S.A 107, 18599–18604
46. Murata, S. et al. (2007) Regulation of CD8+ T Cell Development by Thymus-Specific Proteasomes. Science 316, 1349–1353
47. Sijts, E.J.A.M. and Kloetzel, P.M. (2011) The role of the proteasome in the generation of MHC class I ligands and immune responses. Cellular and Molecular Life Sciences 68, 1491–1502
48. de Verteuil, D. et al. (2010) Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. Molecular and Cellular Proteomics 9, 2034–2047
49. Luckey, C.J. et al. (1998) Proteasomes can either generate or destroy MHC class I epitopes: evidence for nonproteasomal epitope generation in the cytosol. The Journal of Immunology 161, 112–121
50. van Endert, P. (2011) Post-proteasomal and proteasome-independent generation of MHC class I ligands. Cellular and Molecular Life Sciences 68, 1553–1567

51. Craiu, A. et al. (1997) Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. Proc.Natl.Acad.Sci.U.S.A 94, 10850–10855
52. Kanaseki, T. et al. (2013) ERAAP and Tapasin Independently Edit the Amino and Carboxyl Termini of MHC Class I Peptides. The Journal of Immunology 191, 1547-1555
53. Guil, S. et al. (2006) Need for tripeptidyl-peptidase II in major histocompatibility complex class I viral antigen processing when proteasomes are detrimental. The Journal of Biological Chemistry 281, 39925–39934
54. Parmentier, N. et al. (2010) Production of an antigenic peptide by insulin-degrading enzyme. Nature Immunology 11, 449–454
55. Kessler, J.H. et al. (2010) Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. Nature Immunology 12, 45–53
56. Weimershaus, M. et al. (2013) Peptidases trimming MHC class I ligands. Current Opinion in Immunology 25, 90–96
57. Shen, X.Z. et al. (2011) The carboxypeptidase ACE shapes the MHC class I peptide repertoire. Nature Immunology 12, 1078–1085
58. Kunisawa, J. and Shastri, N. (2003) The group II chaperonin TRiC protects proteolytic intermediates from degradation in the MHC class I antigen processing pathway. Molecular Cell 12, 565–576
59. Norbury, C.C. and Tewalt, E.F. (2006) Upstream toward the “DRiP-”ing Source of the MHC Class I Pathway. Immunity 24, 503–506
60. Reits, E. et al. (2003) Peptide Diffusion, Protection, and Degradation in Nuclear and Cytoplasmic Compartments before Antigen Presentation by MHC Class I. Immunity 18, 97–108
61. Jensen, P.E. (2007) Recent advances in antigen processing and presentation. Nature Immunology 8, 1041–1048
62. Rammensee, H.-G. (2006) Peptides Made to Order. Immunity 25, 693–695
63. York, I.A. et al. (2006) Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides in vivo and plays an important role in immunodominance. Proc.Natl.Acad.Sci.U.S.A 103, 9202–9207
64. Geiss-Friedlander, R. et al. (2009) The cytoplasmic peptidase DPP9 is rate-limiting for degradation of proline-containing peptides. The Journal of Biological Chemistry 284, 27211–27219
65. Saveanu, L. et al. (2005) Complexity, contradictions, and conundrums: studying post-proteasomal proteolysis in HLA class I antigen presentation. Immunological Reviews 207, 42–59
66. Kawahara, M. et al. (2009) Analysis of the role of tripeptidyl peptidase II in MHC class I antigen presentation in vivo. The Journal of Immunology 183, 6069–6077

67. Firat, E. et al. (2007) Analysis of direct and cross-presentation of antigens in TPPII knockout mice. The Journal of Immunology 179, 8137–8145
68. Lopez, D. et al. (2010) Caspases in virus-infected cells contribute to recognition by CD8+ T lymphocytes. The Journal of Immunology 184, 5193–5199
69. Yewdell, J.W. et al. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nature Reviews Immunology 3, 952–961
70. Callahan, M.K. and Garg, M. (2008) Heat-shock protein 90 associates with N-terminal extended peptides and is required for direct and indirect antigen presentation. Proc.Natl.Acad.Sci.U.S.A 105, 1662–1667
71. Chapiro, J. et al. (2005) Destructive Cleavage of Antigenic Peptides Either by the Immunoproteasome or by the Standard Proteasome Results in Differential Antigen Presentation. The Journal of Immunology 276, 1053–1061
72. Princiotta, M.F. et al. (2003) Quantitating protein synthesis, degradation and endogenous antigen processing. Immunity 18, 343–354
73. Neefjes, J.J. et al. (1993) Folding and assembly of major histocompatibility complex class I heterodimers in the endoplasmic reticulum of intact cells precedes the binding of peptide. The Journal of Experimental Medicine 178, 1971–1980
74. Hammer, G.E. and Shastri, N. (2007) Construction and destruction of MHC class I in the peptide-loading complex. Nature Immunology 8, 793-794
75. Zhang, W. et al. (2011) A role for UDP-glucose glycoprotein glucosyltransferase in expression and quality control of MHC class I molecules. Proc.Natl.Acad.Sci.U.S.A 108, 4956–4961
76. Peaper, D.R. and Cresswell, P. (2008) Regulation of MHC Class I Assembly and Peptide Binding. Annual Review of Cell and Developmental Biology 24, 343–368
77. Cresswell, P. et al. (2005) Mechanisms of MHC class I-restricted antigen processing and cross- presentation. Immunological Reviews 207, 145–157
78. Cresswell, P. (2005) Antigen processing and presentation. Immunological Reviews 207, 5–7
79. Neefjes, J.J. et al. (1993) Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. Science 261, 769–771
80. Shepherd, J.C. et al. (1993) TAP1-dependent peptide translocation in vitro is ATP dependent and peptide selective. Cell 74, 577–584
81. Androlewicz, M.J. et al. (1993) Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. Proc.Natl.Acad. Sci.U.S.A 90, 9130–9134
82. Elliott, T. and Neefjes, J. (2006) The Complex Route to MHC Class I-Peptide Complexes. Cell 127, 249–251
83. Elliott, T. and Williams, A. (2005) The optimization of peptide cargo bound to MHC

class I molecules by the peptide-loading complex. Immunological Reviews 207, 89–99
84. Serwold, T. et al. (2002) ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. Nature 419, 480–483
85. York, I.A. et al. (2002) The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. Nature Immunology 3, 1177–1184
86. Saric, T. et al. (2002) An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. Nature Immunology 3, 1169–1176
87. Hammer, G.E. et al. (2006) The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules. Nature Immunology 7, 103–112
88. Hammer, G.E. et al. (2007) In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. Nature Immunology 8, 101–108
89. Blanchard, N. et al. (2010) Endoplasmic reticulum aminopeptidase associated with antigen processing defines the composition and structure of MHC class I peptide repertoire in normal and virus-infected cells. The Journal of Immunology 184, 3033–3042
90. Spee, P. et al. (1999) Identification of novel peptide binding proteins in the endoplasmic reticulum: ERp72, calnexin, and grp170. Biochemistry 38, 10559–10566
91. Park, B. et al. (2006) Redox Regulation Facilitates Optimal Peptide Selection by MHC Class I during Antigen Processing. Cell 127, 369–382
92. Kanaseki, T. et al. (2006) ERAAP Synergizes with MHC Class I Molecules to Make the Final Cut in the Antigenic Peptide Precursors in the Endoplasmic Reticulum. Immunity 25, 795–806
93. Kienast, A. et al. (2007) Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. Nature Immunology 8, 864–872
94. Wearsch, P.A. and Cresswell, P. (2007) Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. Nature Immunology 8, 873–881
95. Zhang, Y. and Williams, D.B. (2007) Assembly of MHC class I molecules within the endoplasmic reticulum. Immunologic Research 34, 151–162
96. Boyle, L.H. et al. (2013) Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway. Proc.Natl.Acad.Sci.U.S.A 110, 3465–3470
97. Lu, X. et al. (2012) Endogenous viral antigen processing generates peptide-specific MHC class I cell-surface clusters. Proc.Natl.Acad.Sci.U.S.A 109, 15407–15412

98. Demaria, S. and Bushkin, Y. (2000) Soluble HLA proteins with bound peptides are released from the cell surface by the membrane metalloproteinase. Human Immunology 61, 1332–1338
99. Mahmutefendic, H. et al. (2013) Endosomal trafficking of open Major Histocompatibility Class I conformers--implications for presentation of endocytosed antigens. Molecular Immunology 55, 149–152
100. Bartee, E. et al. (2004) Downregulation of major histocompatibility complex class I by human ubiquitin ligases related to viral immune evasion proteins. Journal of Virology 78, 1109–1120
101. Storkus, W.J. et al. (1993) Identification of T-cell epitopes: rapid isolation of class I-presented peptides from viable cells by mild acid elution. Journal of Immunotherapy with Emphasis on Tumor Immunology 14, 94–103
102. Arosa, F.A. et al. (2007) Open conformers: the hidden face of MHC-I molecules. Trends in Immunology 28, 115–123
103. Peh, C.A. et al. (1998) HLA-B27–Restricted Antigen Presentation in the Absence of Tapasin Reveals Polymorphism in Mechanisms of HLA Class I Peptide Loading. Immunity 8, 531–542
104. Garbi, N. et al. (2005) Accessory molecules in the assembly of major histocompatibility complex class I/peptide complexes: how essential are they for CD8+ T-cell immune responses? Immunological Reviews 207, 1–12
105. Oliveira, C.C. and van Hall, T. (2013) Importance of TAP-independent processing pathways. Molecular Immunology 55, 113–116
106. Lilley, B.N. and Ploegh, H.L. (2005) Viral modulation of antigen presentation: manipulation of cellular targets in the ER and beyond. Immunological Reviews 207, 126–144
107. Hansen, T.H. and Bouvier, M. (2009) MHC class I antigen presentation: learning from viral evasion strategies. Nature Reviews Immunology 9, 503–513
108. Lorente, E. et al. (2013) Diversity of Natural Self-Derived Ligands Presented by Different HLA Class I Molecules in Transporter Antigen Processing-Deficient Cells. PLoS ONE 8, e59118
109. Zernich, D. (2004) Natural HLA Class I Polymorphism Controls the Pathway of Antigen Presentation and Susceptibility to Viral Evasion. The Journal of Experimental Medicine 200, 13–24
110. Sieker, F. et al. (2008) Differential tapasin dependence of MHC class I molecules correlates with conformational changes upon peptide dissociation: A molecular dynamics simulation study. Molecular Immunology 45, 3714–3722
111. Benham, A.M. et al. (1998) Allelic differences in the relationship between proteasome activity and MHC class I peptide loading. The Journal of Immunology 161, 83–89

112. Luckey, C.J. et al. (2001) Differences in the expression of human class I MHC alleles and their associated peptides in the presence of proteasome inhibitors. The Journal of Immunology 167, 1212–1221
113. Marcilla, M. et al. (2007) Proteasome-independent HLA-B27 ligands arise mainly from small basic proteins. Molecular and cellular proteomics 6, 923–938
114. Neisig, A. et al. (1996) Allele-Specific Differences in the Interaction of M H C Class I Molecules with Transporters Associated with Antigen Processing. The Journal of Immunology 156, 3196–3206
115. Turnquist, H.R. et al. (2002) Disparate binding of chaperone proteins by HLA-A subtypes. Immunogenetics 53, 830–834
116. Park, B. et al. (2003) A Single Polymorphic Residue Within the Peptide-Binding Cleft of MHC Class I Molecules Determines Spectrum of Tapasin Dependence. The Journal of Immunology 170, 961–968
117. Groothuis, T.A.M. et al. (2005) MHC class I alleles and their exploration of the antigen-processing machinery. Immunological Reviews 207, 60–76
118. Greenwood, R. et al. (1994) Novel allele-specific, post-translational reduction in HLA class I surface expression in a mutant human B cell line. The Journal of Immunology 153, 5525–5536
119. Turnquist, H.R. et al. (2000) HLA-B polymorphism affects interactions with multiple endoplasmic reticulum proteins. European Journal of Immunology. 30, 3021–3028
120. Neisig, A. et al. (1998) Reduced Cell Surface Expression of HLA-C Molecules Correlates with Restricted Peptide Binding and Stable TAP Interaction. The Journal of Immunology 160, 171–179
121. Snary, D. et al. (2005) Molecular structure of human histocompatibility antigens the HLA-C series. European Journal of Immunology. 8, 58–585
122. Lautscham, G. et al. (2003) TAP-independent antigen presentation on MHC class I molecules: lessons from Epstein-Barr virus. Microbes and Infection 5, 291–299
123. Weinzierl, A.O. et al. (2008) Features of TAP-independent MHC class I ligands revealed by quantitative mass spectrometry. European Journal of Immunology. 38, 1503–1510
124. Henderson, R.A. et al. (1992) HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. Science 255, 1264–1266
125. Wei, M.L. and Cresswell, P. (1992) HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. Nature 356, 443–446
126. Lorente, E. et al. (2011) Multiple Viral Ligands Naturally Presented by Different Class I Molecules in Transporter Antigen Processing-Deficient Vaccinia Virus-Infected Cells. Journal of VIrology 86, 527–541
127. Del Val, M. et al. (2011) Generation of MHC class I ligands in the secretory and

vesicular pathways. Cellular and Molecular Life Sciences 68, 1543–1552
128. Gil-Torregrosa, B.C. et al. (2013) Generation of MHC class I Peptide antigens by protein processing in the Secretory Route by furin. Traffic 1, 641–651
129. Desjardins, M. et al. (2005) Phagocytosis: the convoluted way from nutrition to adaptive immunity. Immunological Reviews 207, 158–165
130. Tiwari, N. et al. (2007) A Transporter Associated with Antigen-Processing Independent Vacuolar Pathway for the MHC Class I-Mediated Presentation of Endogenous Transmembrane Proteins. The Journal of Immunology 178, 7932–7942
131. English, L. et al. (2009) Autophagy enhances the presentation of endogenous viral antigens on MHC class I molecules during HSV-1 infection. Nature Immunology 10, 480–487
132. English, L. et al. (2009) Nuclear membrane-derived autophagy, a novel process that participates in the presentation of endogenous viral antigens during HSV-1 infection. Autophagy 5, 1026–1029
133. Bevan, M.J. (1976) Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. The Journal of Experimental Medicine. 143, 1283–1288
134. Kurts, C. et al. (2010) Cross-priming in health and disease. Nature Reviews Immunology 10, 403–414
135. Saveanu, L. et al. (2009) IRAP identifies an endosomal compartment required for MHC class I cross-presentation. Science 325, 213–217
136. Segura, E. et al. (2009) Different cross-presentation pathways in steady-state and inflammatory dendritic cells. Proc.Natl.Acad.Sci.U.S.A 106, 20377–20381
137. Ackerman, A.L. et al. (2006) A Role for the Endoplasmic Reticulum Protein Retrotranslocation Machinery during Crosspresentation by Dendritic Cells. Immunity 25, 607–617
138. Kovacsovics-Bankowski, M. and Rock, K.L. (1995) A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. Science 267, 243–246
139. Monu, N. and Trombetta, E.S. (2007) Cross-talk between the endocytic pathway and the endoplasmic reticulum in cross-presentation by MHC class I molecules. Current Opinion in Immunology 19, 66–72
140. Trombetta, E.S. and Mellman, I. (2005) CELL BIOLOGY OF ANTIGEN PROCESSING IN VITRO AND IN VIVO. Annual Review of Immunology. 23, 975–1028
141. Albert, M.L. et al. (1998) Immature dendritic cells phagocytose apoptotic cells via alphavbeta5 and CD36, and cross-present antigens to cytotoxic T lymphocytes. The Journal of Experimental Medicine. 188, 1359–1368
142. Pamer, E.G. et al. (1997) MHC class I antigen processing of Listeria monocytogenes proteins: implications for dominant and subdominant CTL responses. Im-

munological Reviews 158, 129–136
143. Villanueva, M.S. et al. (1995) Listeriolysin is processed efficiently into an MHC class I-associated epitope in Listeria monocytogenes-infected cells. The Journal of Immunology 155, 5227–5233
144. Burgdorf, S. et al. (2008) Spatial and mechanistic separation of cross-presentation and endogenous antigen presentation. Nature Immunology 9, 558–566
145. Neijssen, J. et al. (2005) Cross-presentation by intercellular peptide transfer through gap junctions. Nature 434, 83–88
146. Shen, L. et al. (2004) Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo. Immunity 21, 155–165
147. Rock, K.L. and Shen, L. (2005) Cross-presentation: underlying mechanisms and role in immune surveillance. Immunological Reviews 207, 166–183
148. Linscheid, C. and Petroff, M.G. (2013) Minor histocompatibility antigens and the maternal immune response to the fetus during pregnancy. American Journal of Reproductive Immunology 69, 304–314
149. Schubert, U. et al. (2007) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 404, 770–774
150. Yewdell, J.W. et al. (1996) Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? The Journal of Immunology 157, 1823–1826
151. Yewdell, J.W. and Nicchitta, C.V. (2006) The DRiP hypothesis decennial: support, controversy, refinement and extension. Trends in Immunology 27, 368–373
152. Townsend, A.R. et al. (1985) Cytotoxic T cells recognize fragments of the influenza nucleoprotein. Cell 42, 457–467
153. Bruggen, P.V.D. and Eynde, B.J.V.D. (2006) Processing and presentation of tumor antigens and vaccination strategies. Current Opinion in Immunology 18, 98–104
154. Dolan, B.P. et al. (2011) Distinct pathways generate peptides from defective ribosomal products for CD8+ T cell immunosurveillance. The Journal of Immunology 186, 2065–2072
155. Yewdell, J.W. and Hickman, H.D. (2007) New lane in the information highway: alternative reading frame peptides elicit T cells with potent antiretrovirus activity. The Journal of Experimental Medicine. 204, 2501–2504
156. Yewdell, J.W. (2011) DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends in Immunology 32, 548–558
157. Reits, E.A. et al. (2000) The major substrates for TAP in vivo are derived from newly synthesized proteins. Nature 404, 774–778
158. Eisenlohr, L.C. et al. (2007) Rethinking peptide supply to MHC class I molecules. Nature Reviews Immunology 7, 403–410
159. Vabulas, R.M. (2005) Protein Synthesis upon Acute Nutrient Restriction Relies on

Proteasome Function. Science 310, 1960–1963
160. Milner, E. et al. (2006) The Turnover Kinetics of Major Histocompatibility Complex Peptides of Human Cancer Cells. Molecular & Cellular Proteomics 5, 357–365
161. Mackay, L.K. et al. (2009) T cell detection of a B-cell tropic virus infection: newly-synthesised versus mature viral proteins as antigen sources for CD4 and CD8 epitope display. PLoS Pathog 5, e1000699
162. Weinzierl, A.O. et al. (2006) Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface. Molecular & Cellular Proteomics 6, 102–112
163. Yewdell, J.W. (2005) The seven dirty little secrets of major histocompatibility complex class I antigen processing. Immunological Reviews 207, 8–18
164. Qian, S.-B. et al. (2006) Tight Linkage between Translation and MHC Class I Peptide Ligand Generation Implies Specialized Antigen Processing for Defective Ribosomal Products. The Journal of Immunology 177, 1–7
165. Knowles, S.E. et al. (1975) Increased degradation rates of protein synthesized in hepatoma cells in the presence of amino acid analogues. Biochemical Journal 146, 595–600
166. Rieder, R.F. et al. (1975) Rapid postsynthetic destruction of unstable haemoglobin Bushwick. Nature 254, 725–727
167. Townsend, A. et al. (1988) Defective presentation to class I-restricted cytotoxic T lymphocytes in vaccinia-infected cells is overcome by enhanced degradation of antigen. The Journal of Experimental Medicine. 168, 1211–1224
168. Apcher, S. et al. (2011) Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc.Natl. Acad.Sci.U.S.A108, 11572–11577
169. Huang, L. et al. (2011) Hydrophobicity as a driver of MHC class I antigen processing. The EMBO Journal 30, 1634–1644
170. Yewdell, J.W. (2012) Amsterdamming DRiPs. Molecular Immunology 55, 110-112
171. Golovina, T.N. et al. (2005) The impact of misfolding versus targeted degradation on the efficiency of the MHC class I-restricted antigen processing. The Journal of Immunology 174, 2763–2769
172. Markossian, K.A. and Kurganov, B.I. (2004) Protein folding, misfolding, and aggregation. Formation of inclusion bodies and aggresomes. Biochemistry (Moscow) 69, 971–984
173. Bukau, B. et al. (2006) Molecular chaperones and protein quality control. Cell 125, 443–451
174. Engelhard, V.H. et al. (2006) Post-translational modifications of naturally processed MHC-binding epitopes. Current Opinion in Immunology 18, 92–97
175. Ferris, R.L. et al. (1999) Processing of HIV-1 envelope glycoprotein for class I-

restricted recognition: dependence on TAP1/2 and mechanisms for cytosolic localization. The Journal of Immunology 162, 1324–1332
176. Wang, Y. et al. (1999) The role of endoplasmic reticulum-associated protein degradation in MHC class I antigen processing. Immunological Reviews 172, 67–72
177. Haurum, J.S. et al. (1999) Presentation of cytosolic glycosylated peptides by human class I major histocompatibility complex molecules in vivo. The Journal of Experimental Medicine. 190, 145–150
178. Skipper, J.C. et al. (1996) An HLA-A2-restricted tyrosinase antigen on melanoma cells results from posttranslational modification and suggests a novel pathway for processing of membrane proteins. The Journal of Experimental Medicine. 183, 527–534
179. Dalet, A. et al. (2011) An antigenic peptide produced by reverse splicing and double asparagine deamidation. Proc.Natl.Acad.Sci.U.S.A 108, 323–331
180. Selby, M. et al. (1999) Hepatitis C virus envelope glycoprotein E1 originates in the endoplasmic reticulum and requires cytoplasmic processing for presentation by class I MHC molecules. The Journal of Immunology 162, 669–676
181. Mosse, C.A. et al. (1998) The class I antigen-processing pathway for the membrane protein tyrosinase involves translation in the endoplasmic reticulum and processing in the cytosol. The Journal of Experimental Medicine. 187, 37–48
182. Yague, J. et al. (2000) A post-translational modification of nuclear proteins, N(G),N(G)-dimethyl-Arg, found in a natural HLA class I peptide ligand. Protein Science 9, 2210–2217
183. Zamvil, S.S. et al. (1986) T-cell epitope of the autoantigen myelin basic protein that induces encephalomyelitis. Nature 324, 258–260
184. Zarling, A.L. et al. (2000) Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. The Journal of Experimental Medicine. 192, 1755–1762
185. Zarling, A.L. et al. (2006) From the Cover: Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc.Natl.Acad. Sci.U.S.A103, 14889–14894
186. Meyer, V.S. et al. (2009) Identification of Natural MHC Class II Presented Phosphopeptides and Tumor-Derived MHC Class I Phospholigands. Journal of Proteome Research 8, 3666–3674
187. Meadows, L. et al. (1997) The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. Immunity 6, 273–281
188. Chen, W. et al. (1999) Modification of cysteine residues in vitro and in vivo affects the immunogenicity and antigenicity of major histocompatibility complex class I-restricted viral determinants. The Journal of Experimental Medicine. 189,

1757–1764
189. Starck, S.R. and Shastri, N. (2011) Non-conventional sources of peptides presented by MHC class I. Cellular and Molecular Life Sciences 68, 1471–1479
190. Cardinaud, S. (2004) Identification of Cryptic MHC I-restricted Epitopes Encoded by HIV-1 Alternative Reading Frames. The Journal of Experimental Medicine 199, 1053–1063
191. Shastri, N. et al. (2002) Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules. Annual Review of Immunology. 20, 463–493
192. Ho, O. and Green, W.R. (2006) Alternative Translational Products and Cryptic T Cell Epitopes: Expecting the Unexpected. The Journal of Immunology 177, 8283–8289
193. Vigneron, N. and Van den Eynde, B.J. (2011) Insights into the processing of MHC class I ligands gained from the study of human tumor epitopes. Cellular and Molecular Life Sciences 68, 1503–1520
194. Ho, O. and Green, W.R. (2006) Cytolytic CD8+ T cells directed against a cryptic epitope derived from a retroviral alternative reading frame confer disease protection. The Journal of Immunology 176, 2470–2475
195. Coulie, P.G. et al. (1995) A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma. Proc.Natl.Acad. Sci.U.S.A 92, 7976–7980
196. Guilloux, Y. et al. (1996) A peptide recognized by human cytolytic T lymphocytes on HLA-A2 melanomas is encoded by an intron sequence of the N-acetylglucosaminyltransferase V gene. The Journal of Experimental Medicine. 183, 1173–1183
197. Robbins, P.F. et al. (1997) The intronic region of an incompletely spliced gp100 gene transcript encodes an epitope recognized by melanoma-reactive tumor-infiltrating lymphocytes. The Journal of Immunology 159, 303–308
198. Spaapen, R.M. et al. (2009) Rapid identification of clinical relevant minor histocompatibility antigens via genome-wide zygosity-genotype correlation analysis. Clinical Cancer Research 15, 7137–7143
199. Weinzierl, A.O. et al. (2008) A cryptic vascular endothelial growth factor T-cell epitope: identification and characterization by mass spectrometry and T-cell assays. Cancer Research 68, 2447–2454
200. Wang, R.F. et al. (1996) Identification of TRP-2 as a human tumor antigen recognized by cytotoxic T lymphocytes. The Journal of Experimental Medicine. 184, 2207–2216
201. Bullock, T.N. and Eisenlohr, L.C. (1996) Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. The Journal of Experimental Medicine. 184, 1319–1329
202. Malarkannan, S. et al. (1999) Presentation of out-of-frame peptide/MHC class I

complexes by a novel translation initiation mechanism. Immunity 10, 681–690
203. Dolstra, H. et al. (1999) A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. The Journal of Experimental Medicine. 189, 301–308
204. Ronsin, C. et al. (1999) A non-AUG-defined alternative open reading frame of the intestinal carboxyl esterase mRNA generates an epitope recognized by renal cell carcinoma-reactive tumor-infiltrating lymphocytes in situ. The Journal of Immunology 163, 483–490
205. Zook, M.B. et al. (2006) Epitopes Derived by Incidental Translational Frameshifting Give Rise to a Protective CTL Response. The Journal of Immunology 176, 6928–6934
206. Schwab, S.R. (2003) Constitutive Display of Cryptic Translation Products by MHC Class I Molecules. Science 301, 1367–1371
207. Hanada, K.-I. et al. (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. Nature 427, 252–256
208. Vigneron, N. (2004) An Antigenic Peptide Produced by Peptide Splicing in the Proteasome. Science 304, 587–590
209. Hanada, K.-I. and Yang, J.C. (2005) Novel biochemistry: post-translational protein splicing and other lessons from the school of antigen processing. Journal of Molecular Medicine 83, 420–428
210. Warren, E.H. et al. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. Science 313, 1444–1447
211. Dalet, A. et al. (2010) Splicing of distant peptide fragments occurs in the proteasome by transpeptidation and produces the spliced antigenic peptide derived from fibroblast growth factor-5. The Journal of Immunology 184, 3016–3024
212. Wojcik, C. and DeMartino, G.N. (2003) Intracellular localization of proteasomes. The International Journal of Biochemestry and Cell Biology 35, 579–589
213. Istrail, S. et al. (2004) Comparative immunopeptidomics of humans and their pathogens. Proc.Natl.Acad.Sci.U.S.A 101, 13268–13272
214. Purcell, A.W. and Gorman, J.J. (2004) Immunoproteomics: Mass spectrometry-based methods to study the targets of the immune response. Molecular and cellular proteomics 3, 193–208
215. Gakamsky, D.M. et al. (2000) Assembly and dissociation of human leukocyte antigen (HLA)-A2 studied by real-time fluorescence resonance energy transfer. Biochemistry 39, 11163–11169
216. Engelhard, V.H. (2012) The contributions of mass spectrometry to understanding of immune recognition by T lymphocytes. International Journal of Mass Spectrometry 259, 32–39
217. Stevanovic, S. and Schild, H. (1999) Quantitative aspects of T cell activation--pep-

tide generation and editing by MHC class I molecules. Seminars in Immunology 11, 375–384
218. Hassan, C. et al. (2013) The Human Leukocyte Antigen-presented Ligandome of B Lymphocytes. Molecular and Cellular Proteomics 12, 1829–1843
219. Mester, G. et al. (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cellular and Molecular Life Sciences 68, 1521–1532
220. Hunt, D.F. et al. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 255, 1261–1263
221. Crotzer, V.L. et al. (2000) Immunodominance among EBV-derived epitopes restricted by HLA-B27 does not correlate with epitope abundance in EBV-transformed B-lymphoblastoid cell lines. The Journal of Immunology 164, 6120–6129
222. van Els, C.A. et al. (2000) A single naturally processed measles virus peptide fully dominates the HLA-A*0201-associated peptide display and is mutated at its anchor position in persistent viral strains. European Journal of Immunology. 30, 1172–1181
223. Sykulev, Y. et al. (1996) Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. Immunity 4, 565–571
224. Purbhoo, M.A. et al. (2004) T cell killing does not require the formation of a stable mature immunological synapse. Nature Immunology 5, 524–530
225. Crites, T.J. and Varma, R. (2010) On the issue of peptide recognition in T cell development. Self Nonself 1, 55–61
226. Huseby, E.S. et al. (2003) Negative selection imparts peptide specificity to the mature T cell repertoire. Proc.Natl.Acad.Sci.U.S.A 100, 11565–11570
227. Wang, B. et al. (2009) A single peptide-MHC complex positively selects a diverse and specific CD8 T cell repertoire. Science 326, 871–874
228. Marrack, P. and Kappler, J. (2004) Control of T Cell Viability. Annual Review of Immunology. 22, 765–787
229. Dunn, G.P. et al. (2004) The Immunobiology of Cancer Immunosurveillance and Immunoediting. Immunity 21, 137–148
230. Zitvogel, L. et al. (2006) Cancer despite immunosurveillance: immunoselection and immunosubversion. Nature Reviews Immunology 6, 715–727
231. Van Der Bruggen, P. et al. (2002) Tumor-specific shared antigenic peptides recognized by human T cells. Immunological Reviews 188, 51–64
232. Fortier, M.H. et al. (2008) The MHC class I peptide repertoire is molded by the transcriptome. Journal of Experimental Medicine 205, 595–610
233. Anikeeva, N. et al. (2006) Quantum dot/peptide-MHC biosensors reveal strong CD8-dependent cooperation between self and viral antigens that augment the T cell response. Proc.Natl.Acad.Sci.U.S.A103, 16846–16851

234. Purcell, A.W. et al. (2007) More than one reason to rethink the use of peptides in vaccine design. Nature Reviews Drug Discovery 6, 404–414
235. Wahl, A. et al. (2009) HLA class I molecules consistently present internal influenza epitopes. Proc.Natl.Acad.Sci.U.S.A 106, 540–545
236. Hickman, H.D. et al. (2003) Cutting Edge: Class I Presentation of Host Peptides Following HIV Infection. The Journal of Immunology 171, 22–26
237. Wahl, A. et al. (2010) HLA class I molecules reflect an altered host proteome after influenza virus infection. Human Immunology 71, 14–22
238. Hickman-Miller, H.D. and H, H.W. (2004) The immune response under stress: the role of HSP-derived peptides. Trends in Immunology 25, 427–433
239. Parham, P. and Moffett, A. (2013) Parham- Variable NK cell receptors and their MHC class I ligands. Nature Reviews Immunology 13, 133–144
240. Waldhauer, I. and Steinle, A. (2008) NK cells and cancer immunosurveillance. Oncogene 27, 5932–5943
241. Long, E.O. et al. (2013) Controlling natural killer cell responses: integration of signals for activation and inhibition. Annual Review of Immunology. 31, 227–258
242. Liblau, R.S. et al. (2002) Autoreactive CD8 T cells in organ-specific autoimmunity: emerging targets for therapeutic intervention. Immunity 17, 1–6
243. Chaparro, R.J. et al. (2008) Rapid identification of MHC class I-restricted antigens relevant to autoimmune diabetes using retrogenic T cells. Journal of Immunological Methods 335, 106–115
244. Pinkse, G.G.M. et al. (2005) Autoreactive CD8 T cells associated with beta cell destruction in type 1 diabetes. Proc.Natl.Acad.Sci.U.S.A 102, 18425–18430
245. Forrester, J.V. and Cornall, R.J. (2003) Tolerance and autoimmunity in the eye: a role for CD8 T cells in organ-specific autoimmunity in the retina. Immunology 110, 293–295
246. Maverakis, E. et al. (2012) The etiology of paraneoplastic autoimmunity. Clinical Reviews in Allergy and Immunology 42, 135–144
247. Perreault, C. et al. (1990) Minor histocompatibility antigens. Blood 76, 1269–1280
248. Illing, P.T. et al. (2012) Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. Nature 486, 554–558
249. Klug, F. et al. (2009) Characterization of MHC ligands for peptide based tumor vaccination. Current Pharmaceutical Design 15, 3221–3236
250. Weinschenk, T. et al. (2002) Integrated functional genomics approach for the design of patient-individual antitumor vaccines. Cancer Research 62, 5818–5827
251. Uchida, T. (2011) Development of a cytotoxic T-lymphocyte-based, broadly protective influenza vaccine. Microbiology and Immunology 55, 19–27
252. Azizi, A. and Diaz-Mitoma, F. (2007) Viral peptide immunogens: current challenges and opportunities. Joural of Peptide Science 13, 776–786

253. Pastor, F. et al. (2010) Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature 465, 227–230
254. Slev, P.R. et al. (2006) Sensory neurons with MHC-like peptide binding properties: disease consequences. Current Opinion in Immunology 18, 608–616
255. Chaix, R. et al. (2008) Is Mate Choice in Humans MHC-Dependent? PLoS Genetics 4, e1000184
256. Milinski, M. et al. (2005) Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proc.Natl.Acad.Sci.U.S.A 102, 4414–4418
257. Ziegler, A. et al. (2010) Self/nonself perception, reproduction and the extended MHC. Self Nonself 1, 176–191
258. Leinders-Zufall, T. et al. (2004) MHC class I peptides as chemosensory signals in the vomeronasal organ. Science 306, 1033–1037
259. Shatz, C.J. (2009) MHC class I: an unexpected role in neuronal plasticity. Neuron 64, 40–45
260. Caron, E. et al. (2011) The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Molecular Systems Biology 7, 533
261. Rotzschke, O. et al. (1990) Characterization of naturally occurring minor histocompatibility peptides including H-4 and H-Y. Science 249, 283–287
262. Rotzschke, O. et al. (1990) Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. Nature 348, 252–254
263. Cox, A.L. et al. (1994) Identification of a peptide recognized by five melanoma-specific human cytotoxic T cell lines. Science 264, 716–719
264. Flyer, D.C. et al. (2002) Identification by mass spectrometry of CD8(+)-T-cell Mycobacterium tuberculosis epitopes within the Rv0341 gene product. Infection and Immunity 70, 2926–2932
265. Lieberman, S.M. et al. (2003) Identification of the beta cell antigen targeted by a prevalent population of pathogenic CD8+ T cells in autoimmune diabetes. Proc. Natl.Acad.Sci.U.S.A 100, 8384–8388
266. Torabi-Pour, N. et al. (2002) Comparative study between direct mild acid extraction and immunobead purification technique for isolation of HLA class I-associated peptides. Urologia Internationalis 68, 38–43
267. Van Bleek, G.M. and Nathenson, S.G. (1990) Isolation of an endogenously processed immunodominant viral peptide from the class I H-2Kb molecule. Nature 348, 213–216
268. Prillimanm, K.R. et al. (1998) Complexity among constituents of the HLA-B*1501 peptide motif. Immunogenetics 48, 89–97
269. Hinrichs, J. et al. (2010) The nature of peptides presented by an HLA class I low expression allele. Haematologica 95, 1373–1380
270. Scull, K.E. et al. (2012) Secreted HLA recapitulates the immunopeptidome and

allows in-depth coverage of HLA A*02:01 ligands. Molecular Immunology 51, 136–142
271. Bassani-Strenberg, M. et al. (2010) Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc.Natl.Acad.Sci.U.S.A DOI: 10.1073/pnas.1008501107/-/DCSupplemental/pnas.1008501107_SI.pdf
272. Hickman, H.D. and Yewdell, J.W. (2010) Mining the plasma immunopeptidome for cancer peptides as biomarkers and beyond. Proc.Natl.Acad.Sci.U.S.A 107, 18747–18748
273. Walther, T.C. and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology. The Journal of Cell Biology 190, 491–500
274. Fortier, M.-H. (2009), Développement de méthodes analytiques pour la protéomique et l'identification de peptides MHC I issus de cellules leucémiques. Université de Montréal
275. Huczko, E.L. et al. (1993) Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. The Journal of Immunology 151, 2572–2587
276. Hoppes, R. et al. (2010) Technologies for MHC class I immunoproteomics. Journal of Proteomics 73, 1945–1953
277. Kim, Y. et al. (2012) Immune epitope database analysis resource. Nucleic Acids Res 40, W525–30
278. Tenzer, S. et al. (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage,TAP transport and MHC class I binding. Cellular and Molecular Life Sciences 62, 1025–1037
279. Stevanovic, S. (2005) Antigen processing is predictable: From genes to T cell epitopes. Transplant Immunology 14, 171–174
280. Stranzl, T. et al. (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 62, 357–368
281. Karosiene, E. et al. (2011) NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. Immunogenetics 64, 177–186
282. Luo, F. et al. (2013) Luo- Integrating peptides' sequence and energy of contact residues information improves prediction of peptide. BMC Bioinformatics 14, S1
283. Lundegaard, C. et al. (2010) State of the art and challenges in sequence based T-cell epitope prediction. Immunome Research 6, S3
284. Hu, X. et al. (2010) MetaMHC: a meta approach to predict peptides binding to MHC molecules. Nucleic Acids Research 38, W474–W479
285. Lin, H. et al. (2008) Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. BMC Immunology 9, 8
286. Roomp, K. and Domingues, F.S. (2011) Predicting interactions between T cell receptors and MHC-peptide complexes. Molecular Immunology 48, 553–562

287. Tung, C.W. and Ho, S.Y. (2007) POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties. Bioinformatics 23, 942–949
288. DeLuca, D.S. and Blasczyk, R. (2007) The immunoinformatics of cancer immunotherapy. Tissue Antigens 70, 265–271

CHAPTER 2

2 Origin and plasticity of MHC I-associated self peptides

Danielle de Verteuil[a,b,1], Diana Paola Granados[a,b,1], Pierre Thibault[a,c] and Claude Perreault[a,b,*]

[a] Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montreal, Canada
[b] Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada
[c] Department of Chemistry, Université de Montréal, Montreal, Quebec, Canada

[1]Co-first authors with equal contribution

Keywords: Major histocompatibility complex; peptide; CD8 T cells; mass spectrometry; antigen presentation; autoimmunity; tolerance.

* Correspondence to: Claude Perreault

This article was submitted for publication to **Autoimmunity Reviews** on October 21, 2011, accepted on November 2, 2011 and published online on November 12, 2011.

Autoimmunity Reviews, Volume 11(9), p. 627-635 (2012)

2.1 Abstract

Endogenous peptides presented by MHC I molecules represent the essence of self for CD8 T lymphocytes. These MHC I peptides (MIPs) regulate all key events that occur during the lifetime of CD8 T cells. CD8 T cells are selected on self-MIPs, sustained by self-MIPs, and activated in the presence of self-MIPs. Recently, large-scale mass spectrometry studies have revealed that the self-MIP repertoire is more complex and plastic than previously anticipated. The composition of the self-MIP repertoire varies from one cell type to another and can be perturbed by cell–intrinsic and –extrinsic factors including dysregulation of cellular metabolism and infection. The complexity and plasticity of the self-MIP repertoire represent a major challenge for the maintenance of self tolerance and can have pervasive effects on the global functioning of the immune system.

2.2 Authors' contributions

DPG*: Review of large-scale studies reporting identification of MHC I-associated peptides in human and mice (tables 1 and 2), review of some examples of MHC II-associated autoantibodies implicated in autoimmune diseases (table 3), writing of some sections, general discussion and revision

DDV*: Summary figure (figure 1), writing of some sections, general discussion and revision

CP: Writing of first draft of the manuscript, general discussion and revision

PT: Writing of sections related to mass spectrometry, general discussion and revision

*Co-first authors

2.3 Background

Self/non-self discrimination is a fundamental requirement of life. All organisms rely on their capacity of self/nonself discrimination to detect and reject allogeneic cells and microbes. While unicellular eukaryotes primarily employ self/nonself discrimination to avoid self-mating and germline parasitism, multicellular organisms use self/nonself discrimination primarily in immune defense [1, 2]. In a remarkable example of convergent evolution, agnathans and jawed vertebrates have evolved adaptive immune systems based on somatically diversified and clonally expressed Ag receptors [3, 4]. By allowing generation of exceedingly diversified repertoires of Ag receptors, somatic diversification conveys a decisive advantage in recognition of nonself. However somatic diversification comes at a price: some Ag receptors on adaptive lymphocytes happen to be self-reactive [4]. Therefore, while failure to respond to nonself can lead to death from infection, untoward adaptive immune response to self paves the way to autoimmunity. Furthermore, recognition of self has a pervasive influence on the development and function of the immune system because the adaptive lymphocytes of jawed vertebrates are eminently self-referential: they are selected on self-molecules, sustained by self-molecules, and activated in the presence of self-molecules [5, 6]. This raises the fundamental question: what is the molecular definition of self for the adaptive immune system? We will focus herein on the self recognized by CD8 T cells because recent large-scale (-omic) studies have yielded unprecedented insights into its genesis, molecular composition and plasticity [7-30] (Tables 1 and 2).

Table 1. MS-based studies of human MIPs

Cell line/ Tissue	Origin	MHC molecules	Reference
Panc-1	Pancreas	HLA-A02/11, B38/38, Cw12/12	[19]
MCF-7	Breast cancer	HLA-A02/02, B18/44, Cw05/05	[19]
PC3	Prostate cancer	HLA-A2, B7	[7]
UCI-107, UCI-101	Ovarian cancer	HLA-A2, B7	[7]
MDA-231, MCF-7	Breast cancer	HLA-A2, B7	[7]
C1R	B cell leukemia	HLA-A2, B7	[7]
Fresh plasma	Healthy donors and MM, AML and ALL patients	Many	[23]
HeLa	Cervical cancer	HLA-B27	[24]
SW-1353, C-20/A4	Chondro- sarcoma and immortalized chondrocytes)	HLA-B27	[24]
UCI-107	Ovarian cancer	HLA-Cw4	[24]
MDA-231	Breast cancer	HLA-Cw4	[9]
K562	Myelogenous leukemia	HLA-B*3501, 3502, 3503, 3504, 3506, 3508	[16]
Central nervous system	Brain autopsy samples of patients with multiple sclerosis	HLA-A01/02/03/11/25/30/68, B07/08/14/15/18/35/44/51, Cw-03/04/05/07/08	[22]
MCF-7, MDA-MB-231, and BT-20	Breast cancer	HLA-A*0201	[17]
MCF10A	Breast mammary gland	HLA-A*0201	[17]
HIV-infected and uninfected Sup-T1 cells	Lymphoma, T cell	HLA-B*0702	[8]
721.221	EBV-transformed B lymphoblasts	HLA-B*1801	[10]
Colon carcinoma	Colon carcinoma	HLA-A01/68, B08/44	[11]
Colon tissue	Normal colon	HLA-A01/68, B08/44	[11]
Awellis	EBV-transformed B lymphoblasts	HLA-A*0201/0201, HLA-B*4402/4402	[11]
UCI-107	Ovarian cancer	HLA-A02	[12]

SKOV3-A2, OV-CAR3	Ovarian cancer	HLA-A2/3/29/68, B7/18/35/58	[15]
CRL-5865	Human lung adenocarcinoma	HLA-A02, B35/50	[26]
CRL-5944	Human lung adenocarcinoma	HLA-A2/32, B14/51	[26]
Normal and neoplastic renal tissue	Surgical biopsies	HLA-A01/02/03, B07/08/50, Cw06/07	[21]
Influenza A virus- infected and uninfected HeLa cells	Cervical cancer	HLA-A*0201	[25]
Normal and neoplastic renal tissue	Surgical biopsies	HLA-A02/03/68, B07/18/27/ B57	[14]
TAP1/2–deficient and –sufficient cell lines	EBV-transformed B lymphoblasts	HLA-A02	[18]
DM331 and SLM2	Melanoma	HLA-A*0201	[13]
COV413	Ovarian carcinoma	HLA-A*0201	[13]
B lymphoblasts	EBV-transformed B lymphoblasts	HLA-A*0201	[13]

MM: multiple myeloma; AML: acute myeloid leukemia; ALL: acute lymphoblastic leukemia; EBV: Epstein-Barr virus

Table 2. MS-based studies of mouse MIPs

Cell line/ Tissue	Origin	MHC molecules	Reference
Bone-marrow-derived dendritic cells	C57BL/6 mice, *Lmp7*$^{-/-}$ *Mecl1*$^{-/-}$ mice	H2K^{b}, H2D^{b}, Qa1, Qa2	[29]
Fresh splenocytes	C57BL/6 mice	H2K^{b}, H2D^{b}	[20]
Fresh thymocytes	C57BL/6 mice	H2K^{b}, H2D^{b}, Qa1, Qa2	[28]
EL4	Thymoma cell line	H2K^{b}, H2D^{b}, Qa1, Qa2	[28, 30]

2.4 The nature and role of the immune self recognized by CD8 T cells

The TCR of classic adaptive CD8 T cells recognizes MHC I-associated peptides (MIPs). MHC I genes are polygenic, extremely polymorphic and represent the most conserved MHC genes [31]. In most modern human populations, the majority of MHC I alleles have been acquired by introgression from archaic humans (Neanderthals and Denisovans) [32]. Under steady state conditions (in the absence of infection), all MIPs derive from endogenous self proteins: these MIPs are referred to as self MIPs or the self MHC I immunopeptidome (SMII) [33, 34]. Upon infection, pathogen derived nonself MIPs become the proverbial needle in the haystack (of self MIPs) against which the immune system must quickly respond. Self MIPs constitute the essence of self for classical adaptive CD8 T cells. The SMII regulates all key events that occur during the lifetime of CD8 T cells: positive and negative selection in the thymus and survival in the periphery [35-37]. The role of the SMII is not limited to orchestration of CD8 T-cell development and homeostasis. Constitutive expression of self MIPs allows CD8 T cells to monitor expression of neo-self MIPs on neoplastic cells and to behave as an extrinsic tumor suppressor system [38, 39]. Furthermore, evidence suggests that self MIPs excreted in body fluids act as chemosensory signals for neurons in the vomeronasal organs and may thereby influence mate selection and social behaviours in several vertebrates [40-42].

The MHC complex contains multiple MHC I loci that belong to two major classes: modern classical MHC Ia genes (e.g., HLA-A,-B,-C in humans) and more ancient MHC Ib genes (e.g., HLA-E and -G) [43]. MHC Ib are oligomorphic, some but not all present MIPs, they are involved in several immune and non-immune processes but are less important in adaptive immunity than MHC Ia genes [43]. In the present review, we will focus on MHC Ia genes which play a dominant role in adaptive immunity [44]. All MHC Ia allelic products bind MIPs and MHC Ia genes are the most polymorphic genes known [45]. In humans, HLA-B, with its 1,605 alleles, is the most polymorphic gene of the entire genome (http://www.ebi.ac.uk/imgt/hla/stats.html). MHC Ia molecules have a peptide binding groove containing in general 6 pockets; the size, shape and electrochemical properties of these pockets determine the peptide-binding motif of each MHC I allelic product [46]. The combination of polygenicity and polymorphism has

two important consequences: it ensures that each individual will be able to present a broad range of MIPs and that populations will consist of individuals presenting different MIP repertoires [27]. Hence, in outbred animals, individuals present different SMII and consequently different T-cell repertoires. As a corollary, susceptibility to a variety of infectious and autoimmune diseases is dictated by the MHC I genotype because it regulates the ability to present and respond to various self and nonself MIPs [47-52].

2.5. A synopsis of MHC I processing – making the most out of misbegotten polypeptides

The general mechanisms of MIP genesis and presentation have been elucidated by extensive studies on microbial Ags and model Ags such as ovalbumin [53-56]. The molecular composition of the SMII is intertwined with protein metabolism and is ultimately shaped by two processes: protein translation and degradation [54, 57]. Hence, generation of MHC I-associated peptides ceases abruptly in the presence of molecules that inhibit protein synthesis or proteasome function [58-61]. In brief, MIP genesis begins with polypeptide degradation by the proteasome in the cytosol and nucleus [53]. Peptides generated by proteasomal digestion are then exposed to several peptidases [62, 63]. These peptides have only a few seconds to bind to the transporter for antigen processing (TAP; an heterodimeric peptide transporter located in the endoplasmic reticulum and Golgi), and thereby escape total digestion [54]. TAP-bound peptides are translocated into the endoplasmic reticulum or the Golgi where they can be further trimmed by aminopeptidases [55, 64]. Peptides of appropriate length (8-11 mers) which are able to bind MHC I allelic products expressed by the cell can then be inserted into MHC I proteins by the MHC I loading complex and exported at the cell surface [65]. Large-scale MS studies have established the major impact of several components of the MHC I processing pathway on the MIP repertoire [18, 66].

Only a small proportion of peptides derived from proteasomal digestion become MIPs: a cell generates approximately 2 million peptides per second, but only 150 are presented by MHC I molecules at the cell surface [54]. Perhaps the most salient feature of MIPs is that they derive mainly from defective ribosom-

al products (DRiPs), that is, polypeptides that fail to achieve native structure owing to imperfections in transcription, translation, post-translational modifications or protein folding [54, 67-70]. Several MIPs have been shown to originate from mRNAs that do not yield native full-length proteins [71, 72]. While the mean half-life of native (well structured and folded) proteins is about 24 h, that of DRiPs is about 10 min. Using DRiPs as a source of MIPs provides a mechanism for monitoring the expression of 25% of cellular proteins that are targeted to the secretory pathway: cell surface or secreted proteins are degraded by extracellular proteases but their DRiPs are degraded by the proteasome and can therefore be inserted in the SMII. While compelling evidence suggests that DRiPs are the main source of MIPs, two basic questions have yet to be elucidated. First, why are DRiPs more successful than other polypeptides in generating MIPs? Perhaps because MIPs originate preferentially from a subset of ribosomes (immunoribosomes) that would possess specific features such as a high DRiP rate, lack of chaperones or tethering to TAP (which would couple translation to entry in the MHC I processing pathway) [73-75]. Second, what is the physical nature of DRiPs? In theory, DRiPs may include miscoded or misfolded proteins, premature translation-termination products, polypeptides produced by non-conventional translation mechanisms or intrinsically disordered proteins [74]. Studies on cells transfected with shRNA or mRNAs carrying premature stop codons have revealed that the nonsense-mediated decay pathway is a source of both DRiPs and MIPs [72, 76]. However, in the absence of large-scale study of “natural” MIPs, the physical nature of self MIPs remains elusive.

2.6 Different types of proteasomes generate different MIP repertoires

All eukaryotes possess constitutive proteasomes (CPs) whose 20S proteolytic core is hollow and provides an enclosed cavity open at both ends in which proteins are degraded [77]. The 20S particle is composed of 14 different subunits organized in a barrel-shaped complex with the stoichiometry α7β7β7α7. Three subunits of the two inner β-rings (β1, β2, and β5) participate directly in peptide bond cleavage. Jawed vertebrates also express immunoproteasomes (IPs). In IPs, the three catalytic β-subunits expressed in CPs are replaced by three IFN-γ–inducible homologues (immunosubunits): low molecular weight protein

(LMP)-2 (or β1i) for β1, multicatalytic endopeptidase complex–like (MECL)-1 (or β2i) for β2, and LMP7 (or β5i) for β5 [78]. Under steady-state conditions, IPs are expressed at high levels in some cell types (e.g., thymocytes and DCs) but in low amounts in other cell types [79]. IFN-γ secretion enhances IP biogenesis in all cell types [80]. The overall impact of IPs on the SMII was assessed by comparing MIPs present on primary DCs obtained from either wild type (WT) or *Mecl1-/-Lmp7-/-* (dKO) mice [29]. Out of 417 MIPs eluted from WT DCs, 212 were expressed at similar levels on dKO DCs. However, 199 peptides were overexpressed in WT relative to dKO DCs. Among those 199 peptides, 60 were detected exclusively in WT DCs. Only 6 peptides were slightly overexpressed (3 to 5-fold) in dKO relative to WT DCs and none were unique to dKO DCs. Consistent with these findings, expression of cell surface H2Db and H2Kb was decreased by 2-fold on dKO relative to WT DCs [29, 81]. Furthermore, following immunization with WT DCs, both dKO and *Lmp7-/-* mice generated WT-specific cytotoxic T cells whereas WT mice did not generate cytotoxic effectors against *Lmp7-/-* or dKO cells [29, 82] Therefore the presence of IPs has a major impact on the SMII, by increasing both the abundance and the diversity of MIPs. There are however rare exceptions to this rule since IPs have been shown to destruct several tumor-associated Ags that can be produced by CPs [83]. In particular, CPs are more effective than IPs for generation of spliced MIPs [84, 85]. It should be noted that generation of MIPs is not the sole function of IPs. Indeed, IPs have non-redundant effects on gene expression and protein homeostasis that are probably of considerable biological relevance [29, 86, 87].

The most recently discovered proteasome subunit is being at the origin of a fascinating story. Murata and colleagues found that cortical thymic epithelial cells (TECs) express a unique variety of proteasome, called the thymoproteasome (TP) [88]. The catalytic subunits of TPs are β1i, β2i and β5t. β5t (encoded by the *Psmb11* gene) is found exclusively in cortical TECs [88]. Cortical TECs from *Psmb11−/−* express normal amounts of MHC I at the cell surface and contain IPs rather than TPs. However, *Psmb11−/−* mice are unable to support positive selection of most polyclonal and TCR-transgenic CD8 T cells [89]. Furthermore, *Psmb11−/−* CD8 T cells are defective in mounting immune responses to allogeneic and viral antigens [89]. Together, these studies provide compelling, though indirect, evidence that cortical TECs express a unique set

of MIPs that are essential for positive selection of classic TCRαβ CD8 T cells. The global impact of TPs on the MIP repertoire of cortical TECs has yet to be evaluated. Nonetheless, *in vitro* activity-based profiling studies suggest that β5t has a substrate preference distinct from β5 and β5i [90].

2.7 The SMII is complex and is not a representative excerpt from the proteome

A typical cell expresses about 10^5 MHC I molecules. What is the breadth and nature of their peptide repertoire? Currently, this question can be answered only by mass spectrometry (MS) analyses of MIPs. Progress in this field has been heralded by the development of MS instruments whose sensitivity, dynamic range and mass accuracy are orders of magnitude superior to those of analyzers available a decade ago [91-93]. State-of-the-art MS instruments can sequence low femtomole amounts of peptides and have greatly facilitated high-throughput analyses of MIPs. MS can provide absolute peptide quantitation (i.e., number of peptide copies per cell) by comparing the abundance of endogenous MIPs to known amounts of their corresponding isotopically-labeled counterparts. However, given the considerable complexity of the SMII, this effort has been largely undermined by the sizable peptide synthesis challenges needed for absolute quantitation of SMII components. Most quantitative proteomics analyses have therefore focused on the relative quantification of MIPs in different samples [27].

Theoretical estimates suggest that the SMII accommodates only a small portion of the proteome. Thus, a typical cell (human leukocytes for instance) expresses about 18,000 different proteins [94] with a mean length of 466 amino acids [95] (containing 458 potential nonamers) and would therefore contain over 8×10^6 distinct nonamers (the mean length of MIPs). Furthermore, by adding peptides generated by translation of alternate reading frames or post-translational modifications [71, 96-103], the total number of nonamers is probably greater than 10^7. On the other hand, MS analyses suggest that the number of distinct MIPs present at the cell surface is about 10^4 [27, 104]. This would mean that the SMII comprises about 0.1% ($10^4/10^7$) of the polypeptide sequences found in the proteome. Does the SMII derive essentially from highly

abundant proteins? Milner *et al.* performed MS analyses of MIPs eluted from a tumor cell line transfected with an expression vector coding soluble secreted HLA-A2.1 (lacking a functional transmembrane domain) to evaluate the relationship between protein abundance and the SMII [12]. Remarkably, they only found a minimal correlation (about 6%) between the amounts of MIPs and the relative abundance of their source protein: a large proportion of MIPs derived from low abundance proteins, and many highly abundant proteins did not generate MIPs in detectable amounts [12]. In order to evaluate the identity and properties of proteins that generate MIPs, Hickman *et al.* sequenced by MS over 200 MIPs derived from a B cell line engineered to secrete soluble HLA-B*1801 This seminal study showed that MIPs derive from a wide variety of proteins coded by genes located on all chromosomes [10]. Consistent with subsequent reports, MIP source proteins were distributed in all cell compartments but showed a significant enrichment in intracellular relative to extracellular proteins [10, 28, 105]. These and other high-throughput studies provided direct evidence that the SMII is complex [10, 12, 14, 28-30]. The poor correlation between the proteome and the SMII can be rationalized by the DRiP model [12]. Almost all protein species have two half-lives: a longer one for the well-conformed native protein and a short one for DRiPs. While the proteome essentially contains native proteins, MIPs originate mostly from DRiPs [54, 67]. It is therefore assumed that proteins have different DRiP rates and that self MIPs originate mostly from proteins with high DRiP rates [106, 107].

2.8 The SMII conceals a tissue-specific signature

High-throughput MS-based analyses of self MIPs eluted from primary mouse thymocytes revealed that the SMII is molded by the transcriptome. In thymocytes, the percentages of transcripts expressed at high/intermediate/low levels were 9/29/62 for total mRNAs and 42/38/20 for those coding for MIPs [28]. Hence, even though some low abundance transcripts can generates MIPs, the SMII is biased toward MIPs translated from highly abundant mRNAs. This concept was confirmed in a bioinformatic study of the expression profile of transcripts coding for MIPs listed in the SYFPEITHI database [108]: out of the 2.5% most abundant mRNAs expressed in a pool of hematopoietic tissues, 41% encoded MIPs [105]. The correlation between mRNA abundance and MIP abun-

dance is consistent with evidence that the rate of translation initiation directly correlates with MIP presentation [58, 61]. As a corollary, since abundance of discrete mRNAs varies among different cell types, different tissues should display different SMIIs. This assertion was evaluated by comparing the SMII of primary thymocytes and myeloid dendritic cells (DCs) from C57BL/6 mice. Since MHC I molecules are more abundant on DCs than on thymocytes, more MIPs were eluted from DCs than from thymocytes. Nonetheless, the key finding was that over 40% of MIPs were cell type specific [29]. The MIP repertoire of myeloid DCs was enriched in peptides encoded by genes regulating proteasome function, myeloid differentiation and TLR signaling. In thymocytes, MIP source genes were biased toward cell cycle regulation, purine metabolism and tight junction formation. The fact that a substantial proportion of self MIPs are cell type-specific increases the complexity of the immune self at the organismal level.

2.9. Neoplastic transformation has a broad impact on the SMII

Neoplastic transformation perturbs the two key processes that mould the SMII: protein translation and proteasomal degradation [109, 110]. With hindsight, MIPs expressed exclusively or at increased levels on tumor cells (tumor-associated MIPs) were therefore bound to exist. Over the last two decades, numerous tumor-associated MIPs have been molecularly defined and many of them were shown to be immunogenic for CD8 T cells [111, 112]. Tumor-associated MIPs are encoded by a variety of mutated and non mutated genes. A well curated and up-to-date database on human tumor-associated MIPs can be found at http://www.cancerimmunity.org/peptidedatabase/Tcellepitopes.htm. The majority of tumor-associated MIPs were discovered using biased approaches targeting principally oncogenes and other genes overexpressed in cancer cells. Recently, two studies have used unbiased large-scale MS-based strategies to evaluate more globally how neoplastic transformation impinges on the SMII [14, 28]. Though the experimental design of these studies was different, they yielded strikingly concordant results. Weinzierl *et al.* evaluated the relative abundance of MIPs isolated from three human renal cell carcinomas and autologous normal kidney tissue. They sequenced a total of 273 MIPs, of which 45 (16%) were differentially expressed on neoplastic relative to normal kidney

tissue: 30 MIPs were more abundant on tumor tissue (including 10 MIPs found only in carcinoma cells) and 15 MIPs were more abundant on normal tissue [14]. Fortier *et al.* compared the SMII of normal vs. neoplastic C57BL/6 mouse thymocytes. Out of 196 MIPs, 43 (22%) were differentially expressed: 21 MIPs were more abundant on tumor cells (including 9 tumor-specific MIPs) while 22 were more abundant on normal thymocytes [28]. Tumor-associated MIPs are of prime interest, mainly for two reasons: they represent potential targets in cancer immunotherapy and they can elicit paraneoplastic autoimmunity [113, 114]. Notably, in both renal carcinoma and thymoma cells, changes in the SMII induced by neoplastic transformation did not correlate closely with changes in the transcriptome [14, 28]. Thus, for 74% of thymoma-associated MIPs, increased peptide abundance did not correlate with increased mRNA levels [28]. Two conclusions can be drawn from these two studies. First, neoplastic transformation has a broad impact on the SMII as it led to differential expression of about 20% of self MIPs. The breadth of this impact is commensurate with the profound perturbations in protein metabolism that characterize cancer cells [38]. Second, global changes in the SMII can be detected only by MS-based expression profiling approaches and cannot be inferred from mRNA expression levels. As a corollary, we infer that genesis of tumor-associated MIPs involves mainly posttranscriptional mechanisms that have yet to be characterized.

Cells release low amounts of soluble HLA molecules into the blood. Remarkably, Admon's team recently demonstrated that MS analyses of the SMII bound to plasma soluble HLA molecules allowed identification of tumor-associated MIPs in subjects with leukemia or multiple myeloma [23]. MS analyses of the plasmatic SMII may hold considerable potential for diagnosis and monitoring of various illnesses.

2.10 Viral infection causes presentation of cryptic self MIPs

Alike neoplastic transformation, intracellular pathogens drastically impinge on the synthesis and degradation of host cell proteins [115, 116]. This led Hildebrand's group to test whether viral infection might lead to presentation of neo-self MIPs. To this end, they analyzed HLA-B*0702-bound peptides extracted from a cell line infected or not with HIV-1 [8]. MS analyses led to the iden-

tification of 15 MIPs unique to or upregulated on HIV-1 infected cells. These MIPs derived from a variety of proteins: chaperones, proteins involved in RNA transcription or translation, and components of the ubiquitin-proteasome system. The abundance of transcripts coding for MIPs unique to or upregulated on infected cells was not significantly increased. Thus, differential MIP presentation was likely due to posttranscriptional mechanisms [8]. More recently, MS-based comparison of HLA-A*0201-transfected HeLa cells infected or not with influenza A virus identified 20 MIPs unique to infected cells [25]. One salient implication emerges from these studies: infected cells present two classes of neo-MIPs to the immune system, some derive from the pathogen others from the host. The appearance of neo-self MIPs on infected cells raises the possibility that the SMII might conceal a molecular signature of host proteins hijacked by the virus.

2.11 The SMII conveys to the cell surface an integrative view of cellular regulation

While neoplastic transformation and infection are relatively rare events, cells are constantly subjected to variations in nutrients and growth factors concentrations. Is the molecular composition of the SMII affected by cellular metabolic activity? To investigate this issue, our group used a quantitative high-throughput MS-based approach to study the MIP repertoire of EL4 mouse cells treated or not with the mTOR inhibitor rapamycin. mTOR integrates environmental cues (nutrients and growth factors), and mTOR signaling is regulated in numerous physiological and pathological conditions in all types of cells [117-119]. Treatment with rapamycin for up to 48h led to a progressive increase (≥ 2.5-fold) in the abundance of 98 MIPs, 6 of which were present only on treated cells [30]. Analyses of transcripts levels, translational activity (polysomal loading), protein abundance and stability revealed that variations in MIP abundance were regulated at multiple levels inside the cell. Upregulation of MIPs could be ascribed to enhanced transcriptional activity in some cases (≤ 36%) and to co- or post-translational events in most cases (≥ 64%). Importantly, analyses with the STITCH and STRING interaction databases showed that genes coding for differentially expressed MIPs segregated in 7 main functional modules and were tightly connected to the mTOR interactome. These data

show that the SMII is extremely plastic and that it projects at the cell surface a representation of biochemical networks and metabolic events regulated at multiple levels inside the cell. Since perturbation of a single signaling pathway can lead to significant changes in the composition of the SMII, cells can communicate their metabolic status to the adaptive immune system.

2.12 The immunogenicity of neo-MIPs

It is now clear that several MIPs that are absent on resting healthy cells are present at the surface of transformed, infected or metabolically perturbed cells. But are these neo-MIPs recognized by CD8 T cells? For neo-MIPs that appear on cancer cells, the answer is clearly yes: CD8 T cells specific for tumor-associated neo-MIPs spontaneously expand in tumor-bearing hosts and can expand further following immunization [112, 120, 121]. In a few cases, infections have been found to induce autoimmune CD8 T cell responses against self MIPs. Specific examples include autoreactivity to MIPs derived from vinculin in HIV-infected subjects and from HSP90 and IFI-6-16 following measles (rubeola) [122, 123]. Finally, when coated on dendritic cells, MIPs upregulated on rapamycin-treated cells were found to elicit CD8 T cell responses in mice, suggesting that metabolic perturbations lead to expression of immunogenic neo-MIPs [30].

The biologic relevance of neo-MIPs expressed on tumor cells is compelling. Though contradictory views exist [124], the dominant paradigm holds that recognition of tumor-associated MIPs by CD8 T cells prevents the occurrence or progression of a substantial proportion of cancers [38, 39, 125]. Furthermore, adoptive immunotherapy targeted to tumor-associated MIPs can mediate durable complete responses in patients with metastatic melanoma and represents a promising treatment for other types of cancer [114, 126]. However, we ignore whether self MIPs presented on infected or metabolically stressed cells can elicit meaningful immune responses. It has been speculated that CD8 T-cell responses against stress-induced self MIPs could be of considerable importance because under real life conditions, the cells that CD8 T lymphocytes must deal with (infected and transformed cells) are stressed [127, 128]. Indeed, infection and transformation increase protein synthesis and thereby cause endoplasmic

reticulum stress [129, 130]. This is particularly true for cells submitted to hypoxia, nutrient deprivation or low pH in poorly vascularized tumors and sites of inflammation [131, 132]. A correlation between infection and anti-tumor immunity is supported by epidemiologic evidence that childhood infections (e.g., mumps and measles) lower the risk for certain adult cancers including ovarian cancer and non-Hodgkin's lymphoma [133]. Also, in patients with acute myeloid leukemia undergoing allogeneic hematopoietic cell transplantation, cytomegalovirus infection reduces the leukemic relapse risk [134]. However, the potential protective value of immune response against neo-MIPs expressed on non-neoplastic stressed cells has yet to be addressed experimentally.

2.13 The complexity and plasticity of the SMII – A challenge for self tolerance

A key finding emerging from large-scale peptidomic studies is that the SMII is more complex and plastic than previously anticipated (Fig. 1). At the organismal level, the composition of SMII varies from one cell type to another. Furthermore, the SMII of a cell can be perturbed by dysregulation of cellular metabolism and infection, hence by both cell–intrinsic and –extrinsic factors. Therefore, events that occur frequently under real life conditions lead to presentation of neo-self MIPs to the immune system [30]. The complexity and plasticity of the SMII most likely represent a major challenge for the maintenance of self tolerance. They may justify the need for self tolerance to be maintained by multiple dominant and recessive mechanisms involving diverse hematopoietic and non-hematopoietic stromal cells in the thymus and secondary lymphoid organs [35, 135-137]. Despite these sophisticated mechanisms of tolerance, autoimmune diseases represent a major medical burden [52, 138].

2.13.1 Neo-self MIPs induced by infection

Environmental factors play a major role in the occurrence of autoimmune disease. In particular, a variety of microbial infections are associated with autoimmune diseases such as type 1 diabetes, multiple sclerosis, rheumatic fever, systemic sclerosis and spondyloarthritis [52, 139, 140]. Arguably the most popular hypothesis to explain the association between infection and au-

toimmunity is molecular mimicry between microbial peptides and self MIPs. Though evidence for a role of molecular mimicry has been gathered for selected cases [141], its overall importance in autoimmunity remains difficult to assess. Demonstration that infection leads to presentation of neo-self MIPs [8, 25] raises an alternative explanation: autoreactive T cells could be specific neo-self MIPs presented on infected cells. The latter hypothesis does not require any crossreactivity between microbial peptides and self MIPs. That neo-self MIPs elicit biologically relevant responses would be consistent with the fact that our immune system is not tolerant to all peptides that can be encoded by the genome, but only to self MIPs presented at physiological levels [142]. This principle was recently reiterated following the immune rejection of autologous induced pluripotent stem cells [143]. These cells had a perfectly normal genome but were recognized as non-self by autologous CD8 T cells that reacted to undefined neo-self MIPs.

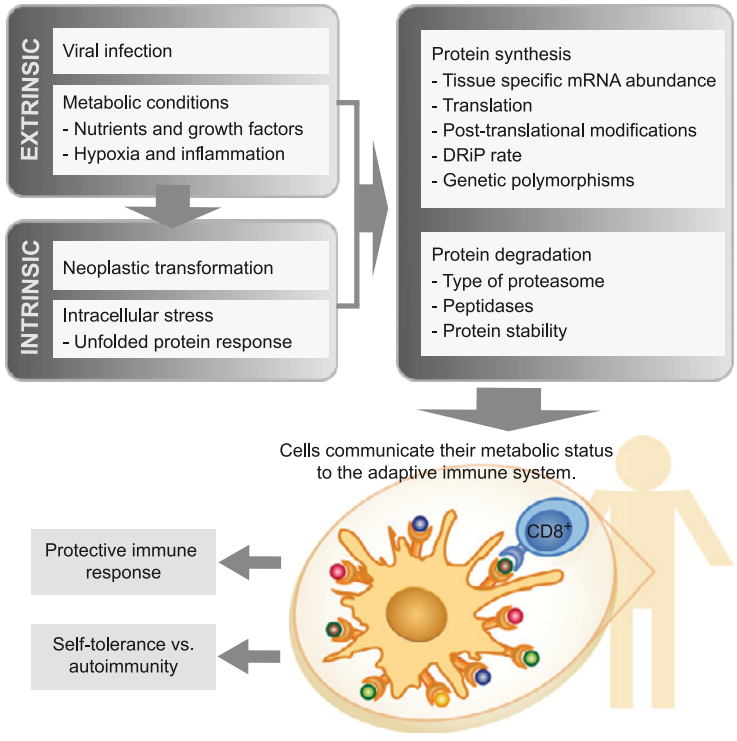


Figure 1. Plasticity of the SMII
The SMII is shaped by both intrinsic and extrinsic factors that affect the abundance and the diversity of MIPs.

2.13.2 Neo-self MIPs induced by metabolic disturbances

Growing interest in the field of immunometabolism is being fuelled by the recognition that obesity-induced inflammation promotes a variety of chronic conditions including atherosclerosis [144]. Evidence suggests that obesity-induced inflammation is initiated by CD8 T cells: the adipose tissue attracts and activates CD8 T cells that in turn initiate and maintain adipose tissue inflammation and systemic insulin resistance [145]. Excessive amounts of nutrients induce endoplasmic reticulum stress in adipocytes [146]. Given that metabolically stressed cells express neo-self MIPs [30], it would be interesting to evaluate whether neo-self MIPs on overloaded adipocytes might be specifically recognized by CD8 T cells that initiate obesity-induced inflammation.

2.14 The MHC class II immunopeptidome

For most autoimmune disorders, MHC genes have by far the strongest single genetic effect. Autoimmunity is frequently linked primarily to MHC II alleles that are strongly associated with the presence of specific circulating autoantibodies [147]. Table 3 shows some representative examples of autoantibodies associated with specific *HLA-DR* and *HLA-DQ* alleles. These autoantibodies can be both pathognomonic and pathogenic [148, 149]. Disease-associated MHC II alleles presumably present self peptides to CD4 T cells that "help" autoreactive B cells. *HLA-DR* and *HLA-DQ* alleles can also be associated with autoimmune disorders where no antibody marker has been identified [150]. These data beg the question: what is the nature of self MHC II-associated peptides recognized by autoreactive CD4 T cells. In comparison to the SMII, characterization of the self MHC II-immunopeptidome represents a greater challenge due to the less-restrictive binding motifs of MHC II alleles and the scarcity of peptide sequencing data [151]. High-throughput analysis of MHC II-associated peptides are absolutely required to help improve current prediction algorithms and understand the structure of the MHC II-restricted self [152]. Systems level analyses of the self MHC II immunopeptidome should also pave the way to a molecular definition of pathogenic MHC II-associated self peptides.

Table 3. Representative examples of autoantibodies associated with specific MHC II alleles

MHC II allele	Autoantibody	Disease	Reference
HLA-DRB1*13	F-actin	Autoimmune hepatitis 1, 2	[147]
HLA-DRB1*03	Antinuclear antibodies (ANA)	Autoimmune hepatitis 1, 2	[147]
HLA-DRB1*03	Liver cytosol type 1 (LC1)	Autoimmune hepatitis 2	[162]
HLA-DRB1*07	Liver kidney microsomes type 1 (LKM-1)	Autoimmune hepatitis 2	[147, 162]
HLA-DQβ1*0301	Anti-basement membrane zone (anti-BMZ)	Pemphigoid	[163]
HLA-DRB1*11-DQB1*0301	Anti-topoisomerase I (Topo)	Systemic sclerosis	[164]
HLA-DRB1*01-DQB1*0501	Anti-centromere (ACA)	Systemic sclerosis	[164]
HLA-DRB1*0404-DQ8, DR3-DQ2	Anti-21-hydroxylase (21-OH)	Type 1 diabetes	[165, 166]
HLA-DR4, HLA-DQ8	Anti-insulin (IAA)	Type 1 diabetes	[165, 167]
HLA-DRB1 SE [a], HLA-DRB1*04	Anti-citrullinated α-enolase peptide 1 (CEP-1)	Rheumatoid arthritis	[168] [169]

[a]HLA-DRB1 SE: HLA-DRB1 shared epitope

2.13 Perspective –toward a more comprehensive definition of the immune self

It is noteworthy that the complexity of the SMII is also spelled in words that go beyond a simple germline-encoded 20-amino acids alphabet. The complexity of the SMII is enhanced by three processes: non-conventional translation mechanisms, post-translational modifications and genetic polymorphisms. Thus, cryptic MIPs include peptides that arise from untranslated regions of the mRNA as well as peptides encoded in alternate translational reading frames and from non-AUG start codon initiation on mRNAs [71, 96-98]. Furthermore, peptides can undergo several post-translational modifications such deamidation, phosphorylation, transpeptidation and proteasomal splicing [13, 99-103].

Peptide splicing leads to the production of MIPs made of two noncontiguous fragments of a protein that are linked together after the excision of the intervening segment [85, 153]. Finally, genetic polymorphisms that hinder MIP generation (e.g., gene deletion) or the structure of a MIP (e.g., single nucleotide polymorphisms) can impinge on the SMII [154-156]. Accordingly, some MIPs are present in some individuals but absent in other MHC-matched subjects [156, 157]. These polymorphic MIPs are traditionally referred to as minor histocompatibility antigens (MiHAs). While MiHAs are of great biological relevance [114, 158], the global impact of genomic polymorphisms on the SMII (i.e. the number of MiHA differences between MHC-identical subjects) is totally unknown, with theoretical estimates ranging from ≈15 to thousands [154, 155, 159]. Importantly, peptide identification by tandem MS is generally based on search algorithms that evaluate matches between observed peptide fragments and a reference proteome. It follows that peptides with post-translational modifications and peptides generated by alternative reading frames, protein splicing or genomic variations are largely missed by current database search algorithms. Furthermore, current binding motif prediction algorithms rely on known peptide sequences, limiting our knowledge of peptides associated to rare alleles and to MHC class II molecules. Thus, unmined MS datasets harbor a whole immunopeptidome universe that awaits to be uncovered by novel analytical approaches including *de novo* MS/MS peptide sequencing and enrichment of peptide bearing discrete post-translational modifications [160, 161].

2.15 Take-home messages

- The self-MIP repertoire is complex and is not a representative excerpt from the proteome.
- Different types of proteasomes generate different MIP repertoires.
- The self-MIP repertoire conceals a tissue-specific signature.
- Neoplastic transformation, viral infection and metabolic changes have a broad impact on the MIP repertoire and lead to presentation of cryptic self-MIPs.
- The self-MIP repertoire conveys to the cell surface an integrative view of cellular regulation.
- The complexity and plasticity of the MHC I-restricted self represent a ma-

jor challenge for the maintenance of self tolerance.

2.16 Acknowledgements

We are grateful to Étienne Caron for insightful discussions on the genesis of the MIP repertoire. DdV and DPG are supported by studentships from the Fonds de la Recherche en Santé du Québec (FRSQ) and the Canadian Institutes for Health Research (CIHR), respectively. CP and PT are supported by the Canada Research Chairs Program. Work in the authors' lab has been supported by grants from the CIHR (MOP-42384), the FRSQ (#22049) and the Canadian Cancer Society (#019475).

2.17 References

[1] Khalturin K, Bosch TC. Self/nonself discrimination at the basis of chordate evolution: limits on molecular conservation. Curr Opin Immunol 2007;19:4-9.
[2] Boehm T. Quality control in self/nonself discrimination. Cell 2006;125:845-58.
[3] Guo P, Hirano M, Herrin BR, Li J, Yu C, Sadlonova A, et al. Dual nature of the adaptive immune system in lampreys. Nature 2009;459:796-801.
[4] Boehm T. Design principles of adaptive immune systems. NatRevImmunol 2011;11:307-17.
[5] Janeway CA, Jr. How the immune system works to protect the host from infection: a personal view. Proc Natl Acad Sci U S A 2001;98:7461-8.
[6] Davis MM, Krogsgaard M, Huse M, Huppa J, Lillemeier BF, Li QJ. T cells as a self-referential, sensory organ. Annu Rev Immunol 2007;25:681-95.
[7] Barnea E, Beer I, Patoka R, Ziv T, Kessler O, Tzehoval E, et al. Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens. Eur J Immunol 2002;32:213-22.
[8] Hickman HD, Luis AD, Bardet W, Buchli R, Battson CL, Shearer MH, et al. Cutting edge: class I presentation of host peptides following HIV infection. J Immunol 2003;171:22-6.
[9] Buchsbaum S, Barnea E, Dassau L, Beer I, Milner E, Admon A. Large-scale analysis of HLA peptides presented by HLA-Cw4. Immunogenetics 2003;55:172-6.
[10] Hickman HD, Luis AD, Buchli R, Few SR, Sathiamurthy M, VanGundy RS, et al. Toward a definition of self: proteomic evaluation of the class I peptide repertoire. J Immunol 2004;172:2944-52.
[11] Lemmel C, Weik S, Eberle U, Dengjel J, Kratt T, Becker HD, et al. Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling. Nat Biotechnol 2004;22:450-4.
[12] Milner E, Barnea E, Beer I, Admon A. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. Mol Cell Proteomics 2006;5:357-65.
[13] Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, et al. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A 2006;103:14889-94.
[14] Weinzierl AO, Lemmel C, Schoor O, Muller M, Kruger T, Wernet D, et al. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. Mol Cell Proteomics 2007;6:102-13.
[15] Philip R, Murthy S, Krakover J, Sinnathamby G, Zerfass J, Keller L, et al. Shared immunoproteome for ovarian cancer diagnostics and immunotherapy: potential theranostic approach to cancer. J Proteome Res 2007;6:2509-17.

[16] Escobar H, Crockett DK, Reyes-Vargas E, Baena A, Rockwood AL, Jensen PE, et al. Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs. J Immunol 2008;181:4874-82.
[17] Hawkins OE, Vangundy RS, Eckerd AM, Bardet W, Buchli R, Weidanz JA, et al. Identification of breast cancer peptide epitopes presented by HLA-A*0201. J Proteome Res 2008;7:1445-57.
[18] Weinzierl AO, Rudolf D, Hillen N, Tenzer S, van Endert P, Schild H, et al. Features of TAP-independent MHC class I ligands revealed by quantitative mass spectrometry. Eur J Immunol 2008;38:1503-10.
[19] Antwi K, Hanavan PD, Myers CE, Ruiz YW, Thompson EJ, Lake DF. Proteomic identification of an MHC-binding peptidome from pancreas and breast cancer cell lines. Mol Immunol 2009;46:2931-7.
[20] Delgado JC, Escobar H, Crockett DK, Reyes-Vargas E, Jensen PE. Identification of naturally processed ligands in the C57BL/6 mouse using large-scale mass spectrometric peptide sequencing and bioinformatics prediction. Immunogenetics 2009;61:241-6.
[21] Stickel JS, Weinzierl AO, Hillen N, Drews O, Schuler MM, Hennenlotter J, et al. HLA ligand profiles of primary renal cell carcinoma maintained in metastases. Cancer Immunol Immunother 2009;58:1407-17.
[22] Fissolo N, Haag S, de Graaf KL, Drews O, Stevanovic S, Rammensee HG, et al. Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients. Mol Cell Proteomics 2009;8:2090-101.
[23] Bassani-Sternberg M, Barnea E, Beer I, Avivi I, Katz T, Admon A. Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc Natl Acad Sci U S A 2010;107:18769-76.
[24] Ben Dror L, Barnea E, Beer I, Mann M, Admon A. The HLA-B*2705 peptidome. Arthritis Rheum 2010;62:420-9.
[25] Wahl A, Schafer F, Bardet W, Hildebrand WH. HLA class I molecules reflect an altered host proteome after influenza virus infection. Hum Immunol 2010;71:14-22.
[26] Shetty V, Sinnathamby G, Nickens Z, Shah P, Hafner J, Mariello L, et al. MHC class I-presented lung cancer-associated tumor antigens identified by immunoproteomics analysis are targets for cancer-specific T cell response. J Proteomics 2011;74:728-43.
[27] Mester G, Hoffmann V, Stevanovic S. Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell Mol Life Sci 2011;68:1521-32.
[28] Fortier MH, Caron E, Hardy MP, Voisin G, Lemieux S, Perreault C, et al. The MHC class I peptide repertoire is molded by the transcriptome. J Exp Med 2008;205:595-

610.
[29] de Verteuil D, Muratore-Schroeder TL, Granados DP, Fortier MH, Hardy MP, Bramoulle A, et al. Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. Mol Cell Proteomics 2010;9:2034-47.
[30] Caron E, Vincent K, Fortier MH, Laverdure JP, Bramoulle A, Hardy MP, et al. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Mol Syst Biol 2011;7:533.
[31] Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, et al. The genome sequence of Atlantic cod reveals a unique immune system. Nature 2011;477:207-10.
[32] Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science 2011;334:89-94.
[33] Rammensee HG, Falk K, Rotzschke O. Peptides naturally presented by MHC class I molecules. Annu Rev Immunol 1993;11:213-44.
[34] Istrail S, Florea L, Halldorsson BV, Kohlbacher O, Schwartz RS, Yap VB, et al. Comparative immunopeptidomics of humans and their pathogens. ProcNatlAcadSciUSA 2004;101:13268-72.
[35] Klein L, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. Nat Rev Immunol 2009;9:833-44.
[36] Sprent J, Surh CD. Normal T cell homeostasis: the conversion of naive cells into memory-phenotype cells. Nature Immunology 2011;131:478-84.
[37] Labrecque N, Baldwin T, Lesage S. Molecular and genetic parameters defining T-cell clonal selection. Immunol Cell Biol 2011;89:16-26.
[38] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
[39] Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science 2011;331:1565-70.
[40] Leinders-Zufall T, Brennan P, Widmayer P, Pc S, Maul-Pavicic A, Jager M, et al. MHC class I peptides as chemosensory signals in the vomeronasal organ. Science 2004;306:1033-7.
[41] Leinders-Zufall T, Ishii T, Mombaerts P, Zufall F, Boehm T. Structural requirements for the activation of vomeronasal sensory neurons by MHC peptides. Nat Neurosci 2009;12:1551-8.
[42] Ortega-Hernandez OD, Kivity S, Shoenfeld Y. Olfaction, psychiatric disorders and autoimmunity: is there a common genetic association? Autoimmunity 2009;42:80-8.

[43] Rodgers JR, Cook RG. MHC class Ib molecules bridge innate and acquired immunity. Nat RevImmunol 2005;5:459-71.
[44] Perreault C. The origin and role of MHC class I-associated self-peptides. Prog Mol Biol Transl Sci 2010;92:41-60.
[45] Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, et al. The DNA sequence and analysis of human chromosome 6. Nature 2003;425:805-11.
[46] Rammensee HG, Friede T, Stevanoviic S. MHC ligands and peptide motifs: first listing. Immunogenetics 1995;41:178-228.
[47] Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. Clin Microbiol Rev 2009;22:370-85, Table of Contents.
[48] Kosmrlj A, Read EL, Qi Y, Allen TM, Altfeld M, Deeks SG, et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. Nature 2010;465:350-4.
[49] Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 2010;330:1551-7.
[50] Noble JA, Valdes AM, Varney MD, Carlson JA, Moonsamy P, Fear AL, et al. HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes 2010;59:2972-9.
[51] Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet 2010;42:985-90.
[52] Shapira Y, Agmon-Levin N, Shoenfeld Y. Geoepidemiology of autoimmune rheumatic diseases. Nat Rev Rheumatol 2010;6:468-76.
[53] Rock KL, York IA, Saric T, Goldberg AL. Protein degradation and the generation of MHC class I-presented peptides. Adv Immunol 2002;80:1-70.
[54] Yewdell JW, Reits E, Neefjes J. Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol 2003;3:952-61.
[55] Hammer GE, Kanaseki T, Shastri N. The final touches make perfect the peptide-MHC class I repertoire. Immunity 2007;26:397-406.
[56] Wearsch PA, Cresswell P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. NatImmunol 2007;8:873-81.
[57] Granados DP, Tanguay PL, Hardy MP, Caron E, de Verteuil D, Meloche S, et al. ER stress affects processing of MHC class I-associated peptides. BMC Immunol 2009;10:10.
[58] Qian SB, Reits E, Neefjes J, Deslich JM, Bennink JR, Yewdell JW. Tight linkage between translation and MHC class I peptide ligand generation implies specialized antigen processing for defective ribosomal products. J Immunol 2006;177:227-33.

[59] Michalek MT, Grant EP, Gramm C, Goldberg AL, Rock KL. A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. Nature 1993;363:552-4.
[60] Wherry EJ, Golovina TN, Morrison SE, Sinnathamby G, McElhaugh MJ, Shockey DC, et al. Re-evaluating the generation of a “proteasome-independent” MHC class I-restricted CD8 T cell epitope. J Immunol 2006;176:2249-61.
[61] Apcher S, Daskalogianni C, Manoury B, Fahraeus R. Epstein Barr virus-encoded EBNA1 interference with MHC class I antigen presentation reveals a close correlation between mRNA translation initiation and antigen presentation. PLoS Pathog 2010;6:e1001151.
[62] Parmentier N, Stroobant V, Colau D, de Diesbach P, Morel S, Chapiro J, et al. Production of an antigenic peptide by insulin-degrading enzyme. Nature Immunology 2010;11:449-54.
[63] Shen XZ, Billet S, Lin C, Okwan-Duodu D, Chen X, Lukacher AE, et al. The carboxypeptidase ACE shapes the MHC class I peptide repertoire. Nature Immunology 2011;Epub October 2.
[64] Ghanem E, Fritzsche S, Al-Balushi M, Hashem J, Ghuneim L, Thomer L, et al. The transporter associated with antigen processing (TAP) is active in a post-ER compartment. J Cell Sci 2010;123:4271-9.
[65] Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. Immunity 2009;30:21-32.
[66] Blanchard N, Kanaseki T, Escobar H, Delebecque F, Nagarajan NA, Reyes-Vargas E, et al. Endoplasmic reticulum aminopeptidase associated with antigen processing defines the composition and structure of MHC class I peptide repertoire in normal and virus-infected cells. J Immunol 2010;184:3033-42.
[67] Yewdell JW, Nicchitta CV. The DRiP hypothesis decennial: support, controversy, refinement and extension. Trends Immunol 2006;27:368-73.
[68] Dolan BP, Bennink JR, Yewdell JW. Translating DRiPs: progress in understanding viral and cellular sources of MHC class I peptide ligands. Cell Mol Life Sci 2011;68:1481-9.
[69] Cardinaud S, Starck SR, Chandra P, Shastri N. The synthesis of truncated polypeptides for immune surveillance and viral evasion. PLoS One 2010;5:e8692.
[70] Dolan BP, Li L, Veltri CA, Ireland CM, Bennink JR, Yewdell JW. Distinct pathways generate peptides from defective ribosomal products for CD8+ T cell immunosurveillance. J Immunol 2011;186:2065-72.
[71] Boon T, Van Pel A. T cell-recognized antigenic peptides derived from the cellular genome are not protein degradation products but can be generated directly by transcription and translation of short subgenic regions. A hypothesis. Immunoge-

netics 1989;29:75-9.
[72] Apcher S, Daskalogianni C, Lejeune F, Manoury B, Imhoos G, Heslop L, et al. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc Natl Acad Sci U S A 2011;108:11572-7.
[73] Lev A, Princiotta MF, Zanker D, Takeda K, Gibbs JS, Kumagai C, et al. Compartmentalized MHC class I antigen processing enhances immunosurveillance by circumventing the law of mass action. Proc Natl Acad Sci U S A 2010;107:6964-9.
[74] Dolan BP, Knowlton JJ, David A, Bennink JR, Yewdell JW. RNA polymerase II inhibitors dissociate antigenic peptide generation from normal viral protein synthesis: a role for nuclear translation in defective ribosomal product synthesis? J Immunol 2010;185:6728-33.
[75] Yewdell JW. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends Immunol 2011;32:548-58.
[76] Gu W, Cochrane M, Leggatt GR, Payne E, Choyce A, Zhou F, et al. Both treated and untreated tumors are eliminated by short hairpin RNA-based induction of target-specific immune responses. Proc Natl Acad Sci U S A 2009;106:8314-9.
[77] Kloetzel PM. Antigen processing by the proteasome. Nat Rev Mol Cell Biol 2001;2:179-87.
[78] Murata S, Yashiroda H, Tanaka K. Molecular mechanisms of proteasome assembly. Nat Rev Mol Cell Biol 2009;10:104-15.
[79] Pelletier S, Schuurman KG, Berkers CR, Ovaa H, Heck AJ, Raijmakers R. Quantifying cross-tissue diversity in proteasome complexes by mass spectrometry. Mol Biosyst 2010;6:1450-3.
[80] Heink S, Ludwig D, Kloetzel PM, Kruger E. IFN-gamma-induced immune adaptation of the proteasome system is an accelerated and transient response. Proc Natl Acad Sci U S A 2005;102:9241-6.
[81] Zaiss DM, de Graaf N, Sijts AJ. The proteasome immunosubunit multicatalytic endopeptidase complex-like 1 is a T-cell-intrinsic factor influencing homeostatic expansion. Infect Immun 2008;76:1207-13.
[82] Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med 2001;194:1-12.
[83] Vigneron N, Van den Eynde BJ. Insights into the processing of MHC class I ligands gained from the study of human tumor epitopes. Cell Mol Life Sci 2011;68:1503-20.
[84] Dalet A, Stroobant V, Vigneron N, Van den Eynde BJ. Differences in the production of spliced antigenic peptides by the standard proteasome and the immunoproteasome. Eur J Immunol 2011;41:39-46.
[85] Dalet A, Robbins PF, Stroobant V, Vigneron N, Li YF, El-Gamil M, et al. An an-

tigenic peptide produced by reverse splicing and double asparagine deamidation. Proc Natl Acad Sci U S A 2011;108:E323-31.
[86] Seifert U, Bialy LP, Ebstein F, Bech-Otschir D, Voigt A, Schroter F, et al. Immunoproteasomes preserve protein homeostasis upon interferon-induced oxidative stress. Cell 2010;142:613-24.
[87] Muchamuel T, Basler M, Aujay MA, Suzuki E, Kalim KW, Lauer C, et al. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. Nat Med 2009;15:781-7.
[88] Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, et al. Regulation of CD8+ T cell development by thymus-specific proteasomes. Science 2007;316:1349-53.
[89] Nitta T, Murata S, Sasaki K, Fujii H, Ripen AM, Ishimaru N, et al. Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. Immunity 2010;32:29-40.
[90] Florea BI, Verdoes M, Li N, van der Linden WA, Geurink PP, van den Elst H, et al. Activity-based profiling reveals reactivity of the murine thymoproteasome-specific subunit beta5t. Chem Biol 2010;17:795-801.
[91] Nilsson T, Mann M, Aebersold R, Yates JR, 3rd, Bairoch A, Bergeron JJ. Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods 2010;7:681-5.
[92] Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 2009;11:49-79.
[93] Depontieu FR, Qian J, Zarling AL, McMiller TL, Salay TM, Norris A, et al. Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy. Proc Natl Acad Sci U S A 2009;106:12073-8.
[94] Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 2009;54:15-39.
[95] Princiotta MF, Finzi D, Qian SB, Gibbs J, Schuchmann S, Buttgereit F, et al. Quantitating protein synthesis, degradation, and endogenous antigen processing. Immunity 2003;18:343-54.
[96] Schwab SR, Li KC, Kang C, Shastri N. Constitutive display of cryptic translation products by MHC class I molecules. Science 2003;301:1367-71.
[97] Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N. Unanticipated antigens: translation initiation at CUG with leucine. PLoS Biol 2004;2:e366.
[98] Starck SR, Shastri N. Non-conventional sources of peptides presented by MHC class I. Cell Mol Life Sci 2011;68:1471-9.
[99] Skipper JC, Hendrickson RC, Gulden PH, Brichard V, Van Pel A, Chen Y, et al. An HLA-A2-restricted tyrosinase antigen on melanoma cells results from posttranslational modification and suggests a novel pathway for processing of membrane

proteins. J Exp Med 1996;183:527-34.
[100] Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. Nature 2004;427:252-6.
[101] Vigneron N, Stroobant V, Chapiro J, Ooms A, Degiovanni G, Morel S, et al. An antigenic peptide produced by peptide splicing in the proteasome. Science 2004;304:587-90.
[102] Warren EH, Vigneron NJ, Gavin MA, Coulie PG, Stroobant V, Dalet A, et al. An antigen produced by splicing of noncontiguous peptides in the reverse order. Science 2006;313:1444-7.
[103] Mohammed F, Cobbold M, Zarling AL, Salim M, Barrett-Wilt GA, Shabanowitz J, et al. Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nature Immunology 2008;9:1236-43.
[104] Engelhard VH, Brickner AG, Zarling AL. Insights into antigen processing gained by direct analysis of the naturally processed class I MHC associated peptide repertoire. Mol Immunol 2002;39:127-37.
[105] Juncker AS, Larsen MV, Weinhold N, Nielsen M, Brunak S, Lund O. Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands. PLoS One 2009;4:e7448.
[106] Caron E, Charbonneau R, Huppe G, Brochu S, Perreault C. The structure and location of SIMP/STT3B account for its prominent imprint on the MHC I immunopeptidome. Int Immunol 2005;17:1583-96.
[107] Yewdell JW. Plumbing the sources of endogenous MHC class I peptide ligands. Curr Opin Immunol 2007;19:79-86.
[108] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 1999;50:213-9.
[109] Richter JD, Sonenberg N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature 2005;433:477-80.
[110] Nakayama KI, Nakayama K. Ubiquitin ligases: cell-cycle control and cancer. Nat Rev Cancer 2006;6:369-81.
[111] Rosenberg SA. A new era for cancer immunotherapy based on the genes that encode cancer antigens. Immunity 1999;10:281-7.
[112] Boon T, Coulie PG, Van den Eynde BJ, van der Bruggen P. Human T cell responses against melanoma. Annu Rev Immunol 2006;24:175-208.
[113] Roberts WK, Deluca IJ, Thomas A, Fak J, Williams T, Buckley N, et al. Patients with lung cancer and paraneoplastic Hu syndrome harbor HuD-specific type 2 CD8+ T cells. J Clin Invest 2009;119:2042-51.
[114] Vincent K, Roy DC, Perreault C. Next-generation leukemia immunotherapy. Blood 2011;118:2951-9.

[115] Isaacson MK, Ploegh HL. Ubiquitination, ubiquitin-like modifiers, and deubiquitination in viral infection. Cell Host Microbe 2009;5:559-70.
[116] Ho BC, Yu SL, Chen JJ, Chang SY, Yan BS, Hong QS, et al. Enterovirus-induced miR-141 contributes to shutoff of host protein translation by targeting the translation initiation factor eIF4E. Cell Host Microbe 2011;9:58-69.
[117] Caron E, Ghosh S, Matsuoka Y, Ashton-Beaucage D, Therrien M, Lemieux S, et al. A comprehensive map of the mTOR signaling network. Mol Syst Biol 2010;6:453.
[118] Sengupta S, Peterson TR, Sabatini DM. Regulation of the mTOR complex 1 pathway by nutrients, growth factors, and stress. Mol Cell 2010;40:310-22.
[119] Araki K, Turner AP, Shaffer VO, Gangappa S, Keller SA, Bachmann MF, et al. mTOR regulates memory CD8 T-cell differentiation. Nature 2009;460:108-12.
[120] Lee PP, Yee C, Savage PA, Fong L, Brockstedt D, Weber JS, et al. Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. Nat Med 1999;5:677-85.
[121] Rosenberg SA, Restifo NP, Yang JC, Morgan RA, Dudley ME. Adoptive cell transfer: a clinical path to effective cancer immunotherapy. Nat Rev Cancer 2008;8:299-308.
[122] di Marzo Veronese F, Arnott D, Barnaba V, Loftus DJ, Sakaguchi K, Thompson CB, et al. Autoreactive cytotoxic T lymphocytes in human immunodeficiency virus type 1-infected subjects. J Exp Med 1996;183:2509-16.
[123] Herberts CA, van Gaans-van den Brink J, van der Heeft E, van Wijk M, Hoekman J, Jaye A, et al. Autoreactivity against induced or upregulated abundant self-peptides in HLA-A*0201 following measles virus infection. Hum Immunol 2003;64:44-55.
[124] Uhr JW, Pantel K. Controversies in clinical cancer dormancy. Proc Natl Acad Sci U S A 2011;108:12396-400.
[125] Nelson BH. The impact of T-cell immunity on ovarian cancer outcomes. Immunol Rev 2008;222:101-16.
[126] Rosenberg SA, Yang JC, Sherry RM, Kammula US, Hughes MS, Phan GQ, et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clin Cancer Res 2011;17:4550-7.
[127] Gleimer M, Parham P. Stress management: MHC class I and class I-like molecules as reporters of cellular stress. Immunity 2003;19:469-77.
[128] Hickman-Miller HD, Hildebrand WH. The immune response under stress: the role of HSP-derived peptides. Trends Immunol 2004;25:427-33.
[129] Holcik M, Sonenberg N. Translational control in stress and apoptosis. Nat Rev Mol Cell Biol 2005;6:318-27.
[130] Mamane Y, Petroulakis E, LeBacquer O, Sonenberg N. mTOR, translation initiation and cancer. Oncogene 2006;25:6416-22.
[131] Bi M, Naczki C, Koritzinsky M, Fels D, Blais J, Hu N, et al. ER stress-regulated

translation increases tolerance to extreme hypoxia and promotes tumor growth. EMBO J 2005;24:3470-81.
[132] Moenner M, Pluquet O, Bouchecareilh M, Chevet E. Integrated endoplasmic reticulum stress responses in cancer. Cancer Res 2007;67:10631-4.
[133] Cramer DW, Finn OJ. Epidemiologic perspective on immune-surveillance in cancer. Curr Opin Immunol 2011;23:265-71.
[134] Elmaagacli AH, Steckel NK, Koldehoff M, Hegerfeldt Y, Trenschel R, Ditschkowski M, et al. Early human cytomegalovirus replication after transplantation is associated with a decreased relapse risk: evidence for a putative virus-versus-leukemia effect in acute myeloid leukemia patients. Blood 2011;118:1402-12.
[135] Levings MK, Allan S, d'Hennezel E, Piccirillo CA. Functional dynamics of naturally occurring regulatory T cells in health and autoimmunity. Adv Immunol 2006;92:119-55.
[136] Cohen JN, Guidi CJ, Tewalt EF, Qiao H, Rouhani SJ, Ruddell A, et al. Lymph node-resident lymphatic endothelial cells mediate peripheral tolerance via Aire-independent direct antigen presentation. J Exp Med 2010;207:681-8.
[137] Fletcher AL, Lukacs-Kornek V, Reynoso ED, Pinner SE, Bellemare-Pelletier A, Curry MS, et al. Lymph node fibroblastic reticular cells directly present peripheral tissue antigen under steady-state and inflammatory conditions. J Exp Med 2010;207:689-97.
[138] Marrack P, Kappler J, Kotzin BL. Autoimmune disease: why and where it occurs. Nat Med 2001;7:899-905.
[139] Benoist C, Mathis D. Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? Nature Immunology 2001;2:797-801.
[140] Grossman C, Dovrish Z, Shoenfeld Y, Amital H. Do infections facilitate the emergence of systemic sclerosis? Autoimmun Rev 2011;10:244-7.
[141] Harkiolaki M, Holmes SL, Svendsen P, Gregersen JW, Jensen LT, McMahon R, et al. T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. Immunity 2009;30:348-57.
[142] Schild H, Rotzschke O, Kalbacher H, Rammensee HG. Limit of T cell tolerance to self proteins by peptide presentation. Science 1990;247:1587-9.
[143] Zhao T, Zhang ZN, Rong Z, Xu Y. Immunogenicity of induced pluripotent stem cells. Nature 2011;474:212-5.
[144] Mathis D, Shoelson SE. Immunometabolism: an emerging frontier. Nat Rev Immunol 2011;11:81.
[145] Nishimura S, Manabe I, Nagasaki M, Eto K, Yamashita H, Ohsugi M, et al. CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. Nat Med 2009;15:914-20.
[146] Zhang K, Kaufman RJ. From endoplasmic-reticulum stress to the inflammatory

response. Nature 2008;454:455-62.
[147] Oliveira LC, Porta G, Marin ML, Bittencourt PL, Kalil J, Goldberg AC. Autoimmune hepatitis, HLA and extended haplotypes. Autoimmun Rev 2011;10:189-93.
[148] Csorba TR, Lyon AW, Hollenberg MD. Autoimmunity and the pathogenesis of type 1 diabetes. Crit Rev Clin Lab Sci 2010;47:51-71.
[149] Iaccarino L, Ghirardello A, Canova M, Zen M, Bettio S, Nalotto L, et al. Anti-annexins autoantibodies: their role as biomarkers of autoimmune diseases. Autoimmun Rev 2011;10:553-8.
[150] Sadovnick AD. Genetic background of multiple sclerosis. Autoimmun Rev 2011. Epub May 18, 2011.
[151] Dengjel J, Nastke MD, Gouttefangeas C, Gitsioudis G, Schoor O, Altenberend F, et al. Unexpected abundance of HLA class II presented peptides in primary renal cell carcinomas. Clin Cancer Res 2006;12:4163-70.
[152] Liao WW, Arthur JW. Predicting peptide binding to Major Histocompatibility Complex molecules. Autoimmun Rev 2011;10:469-73.
[153] Liepe J, Mishto M, Textoris-Taube K, Janek K, Keller C, Henklein P, et al. The 20S proteasome splicing activity discovered by SpliceMet. PLoS Comput Biol 2010;6:e1000830.
[154] Perreault C, Decary F, Brochu S, Gyger M, Belanger R, Roy D. Minor histocompatibility antigens. Blood 1990;76:1269-80.
[155] Roopenian D, Choi EY, Brown A. The immunogenomics of minor histocompatibility antigens. Immunol Rev 2002;190:86-94.
[156] Spierings E, Hendriks M, Absi L, Canossi A, Chhaya S, Crowley J, et al. Phenotype frequencies of autosomal minor histocompatibility antigens display significant differences among populations. PLoS Genet 2007;3:e103.
[157] Rotzschke O, Falk K, Wallny HJ, Faath S, Rammensee HG. Characterization of naturally occurring minor histocompatibility peptides including H-4 and H-Y. Science 1990;249:283-7.
[158] Brochu S, Baron C, Hetu F, Roy DC, Perreault C. Oligoclonal expansion of CTLs directed against a restricted number of dominant minor histocompatibility antigens in hemopoietic chimeras. J Immunol 1995;155:5104-14.
[159] Lindahl KF. Minor histocompatibility antigens. Trends Genet 1991;7:219-24.
[160] Courcelles M, Lemieux S, Voisin L, Meloche S, Thibault P. ProteoConnections: a bioinformatics platform to facilitate proteome and phosphoproteome analyses. Proteomics 2011;11:2654-71.
[161] Admon A, Bassani-Sternberg M. The Human Immunopeptidome Project, a Suggestion for yet another Postgenome Next Big Thing. Mol Cell Proteomics 2011;10:O111 011833.
[162] Djilali-Saiah I, Fakhfakh A, Louafi H, Caillat-Zucman S, Debray D, Alvarez F.

HLA class II influences humoral autoimmunity in patients with type 2 autoimmune hepatitis. J Hepatol 2006;45:844-50.
[163] Zakka LR, Reche P, Ahmed AR. Role of MHC Class II Genes in the pathogenesis of pemphigoid. Autoimmun Rev 2011;11:40-47.
[164] Dieude P, Boileau C, Allanore Y. Immunogenetics of systemic sclerosis. Autoimmun Rev 2011;10:282-90.
[165] Taplin CE, Barker JM. Autoantibodies in type 1 diabetes. Autoimmunity 2008;41:11-8.
[166] Barker JM, Yu J, Yu L, Wang J, Miao D, Bao F, et al. Autoantibody "subspecificity" in type 1 diabetes: risk for organ-specific autoimmunity clusters in distinct groups. Diabetes Care 2005;28:850-5.
[167] Ziegler R, Alper CA, Awdeh ZL, Castano L, Brink SJ, Soeldner JS, et al. Specific association of HLA-DR4 with increased prevalence and level of insulin autoantibodies in first-degree relatives of patients with type I diabetes. Diabetes 1991;40:709-14.
[168] Mahdi H, Fisher BA, Kallberg H, Plant D, Malmstrom V, Ronnelid J, et al. Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. Nat Genet 2009;41:1319-24.
[169] Montes A, Dieguez-Gonzalez R, Perez-Pampin E, Calaza M, Mera-Varela A, Gomez-Reino JJ, et al. Particular association of clinical and genetic features with autoimmunity to citrullinated alpha-enolase in rheumatoid arthritis. Arthritis Rheum 2011;63:654-61.

CHAPTER 3

3. MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements

Running title : MHC and miRNAs
Scientific category : Immunobiology

Diana Paola Granados[1,2], Wafaa Yahyaoui[1,3], Céline M. Laumont[1,2], Tariq Daouda[1,4], Tara L. Muratore-Schroeder[1,3], Caroline Côté[1], Jean-Philippe Laverdure[1], Sébastien Lemieux[1,4], Pierre Thibault[1,3]* & Claude Perreault[1,2,5]*

Institute for Research in Immunology and Cancer[1]; Departments of Medicine[2], Chemistry[3] and Computer Science and Operations Research[4], Université de Montréal; Division of Hematology, Hôpital Maisonneuve-Rosemont[5]. Montréal, QC, Canada.

*Equal contribution and senior authors

Correspondence to: Claude Perreault or Pierre Thibault

This article was submitted for publication to **Blood** on February 20, 2012, accepted on March 21, 2012 and published online on March 21, 2012.

Blood, Volume 119(26), p. e181-191 (2012)

3.1 Authors' contributions

DPG: Study design, conducting all experiments except MS experiments, data analysis and discussion, preparation of all main and supplementary figures (except for figure 3) and tables (except for table 1), writing of first draft of the manuscript

WY: Conducting and analysis of MS experiments

CML: Contribution to analysis of relation between the MHC I immunopeptidome and the transcriptome, preparation of figure 3 and general discussion

TD: Analysis comparing the SNP frequency in the immunopeptidome and the human exome (table 1)

TLMS: Conducting and analysis of MS experiments

CC: technical assistance in cell culture

JPL: Calculation of a connectivity score and bootstrapping analysis (figures 2c and 4b)

SL: Study design, statistical analysis and general discussion

PT: Study design, writing of some sections related to mass spectrometry, general discussion and revision

CP: Study design, data analysis, general discussion and manuscript writing

3.2 Abstract

MHC I-peptides (MIPs) play an essential role in normal homeostasis and diverse pathological conditions. MIPs derive mainly from defective ribosomal products (DRiPs), a subset of nascent proteins that fail to achieve a proper conformation and whose physical nature remains elusive. We used high-throughput proteomic and transcriptomic methods to unravel the structure and biogenesis of MIPs presented by HLA-A and -B molecules on human EBV-infected B lymphocytes from four subjects. We found that although HLA-different subjects present distinctive MIPs derived from different proteins, these MIPs originate from proteins that are functionally interconnected and implicated in similar biological pathways. Secondly, the MIP repertoire of human B cells showed no bias toward conserved vs. polymorphic genomic sequences, derived preferentially from abundant transcripts, and conveyed to the cell surface a cell type-specific signature. Finally, we discovered that MIPs derive preferentially from transcripts bearing miRNA response elements. Furthermore, while MIPs of HLA-disparate subjects are coded by different sets of transcripts, these transcripts are regulated by mostly similar miRNAs. Our data support an emerging model in which i) the generation of MIPs by a transcript depends on its abundance and DRiP rate, and ii) the DRiP rate is regulated to a large extent by miRNAs.

3.3 Introduction

All nucleated cell types constitutively express MHC class I molecules that present self peptides to CD8 T cells. MHC I-associated peptides (MIPs) play crucial roles in many processes including development and homeostasis of CD8 T cells, self/non-self discrimination, tumor immunosurveillance, tissue rejection, graft-vs.-host disease and odorant-based mate selection.[1-6] Classical MHC I molecules are encoded by three genes in humans, HLA-A, -B, and –C, which have an astounding allelic diversity.[7] Genetic polymorphisms that distinguish HLA allomorphs affect mainly their peptide-binding groove. Accordingly, studies on presentation of various peptides suggest that a minority of peptides may bind to several MHC I molecules but that different HLA allomorphs present largely non-overlapping sets of peptides.[8, 9] Generation of MIPs is tightly linked to protein economy because it depends on protein synthesis and degradation.[10, 11] Nevertheless, the most striking (and intriguing) feature of MIPs is that they derive mainly from defective ribosomal products (DRiPs).[12, 13] DRiPs are a subset of nascent proteins that fail to achieve a proper conformation, and are therefore rapidly degraded by proteasomes. A key unresolved issue is the physical nature of DRiPs. In theory, DRiPs could originate from a variety of processes that impinge on transcription, translation, post-translational modifications and protein folding.[12, 13]

No algorithm can predict the amount of DRiPs produced by a gene.[14] Furthermore, the composition of the self MIP repertoire (or immunopeptidome) does not correlate with the abundance of MIP source proteins.[15] Finally, studies on the relation between the transcriptome and the immunopeptidome have yielded seemingly conflicting results.[9, 16] Therefore, large-scale mass spectrometry (MS) studies represent the sole approach that can yield a global appraisal of the MIP landscape. MS studies have highlighted the complexity of the MIP repertoire and shown that MIPs derive from all cell compartments.[5, 16-19] They also revealed that the immunopeptidome is plastic, conveys to the cell surface an integrative view of cellular regulation, and that it can be affected by cell-intrinsic and –extrinsic factors.[17, 18]

Given the quintessential importance of self MIPs, it is imperative to develop

and exploit systems-level quantitative methods that will yield insights into MIP biogenesis and enable modeling of the immunopeptidome. Here, we used high-throughput proteomic and transcriptomic methods to unravel the structure and biogenesis of MIPs presented by HLA-A and -B molecules on human EBV-infected B lymphocytes. Our results show that different HLA allomorphs present MIPs derived from distinct but functionally related source proteins. We further demonstrate that two features determine whether a transcript will generate MIPs: the transcript abundance and whether it contains microRNA (miRNA) response elements (MREs).

3.4 Material and Methods

3.4.1 Cell culture and HLA typing

This study was approved by the Comité Éthique Recherche de l'Hôpital Maisonneuve-Rosemont and all subjects provided written informed consent. Peripheral blood mononuclear cells (PBMCs) were isolated from blood samples of 4 subjects (2 pairs of HLA-identical siblings). Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell lines (B-LCLs) were derived from PBMCs with Ficoll-Paque Plus (Amersham) as described.[20] Established B-LCLs were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum, 25 mM of HEPES, 2mM L-glutamine and penicillin-streptomycin (all from Invitrogen). HLA genotyping was performed at the Maisonneuve-Rosemont Hospital. Siblings of sibship 1 are HLA-A*0101/*0205, HLA-B*1501/*5001, HLA-C*0602/*0701 and DRB1*0101/*1104. Siblings of sibship 2 are: HLA-A*0301/2902, HLA-B*0801/*4403, HLA-C*0701/*1601 and DRB1*0301/*0701. HeLa and HEK293 cell lines were maintained in DMEM medium supplemented with 10% fetal bovine serum, 2mM L-glutamine and penicillin-streptomycin (all from Invitrogen).

3.4.2 Mass spectrometry and peptide sequencing

Three to four biological replicates of $4x10^8$ exponentially growing B-LCLs were prepared from each subject. MIPs were released by mild acid treatment and separated by strong cation exchange chromatography (SCX) using an off-line 1100 series binary LC system (Agilent Technologies) as previously described.[16]

MIP fractions were analyzed by LC-MS/MS using an Eksigent LC system coupled to a LTQ-Orbitrap mass spectrometer (Thermo Electron).[16-18] Additional details are provided as supplemental methods. MS/MS of all peptide identifications are available through ProteoConnections[24] at http://www.thibault.iric.ca/proteoconnections.

3.4.3 MIPs selection, predicted binding affinity and identification of MIP source proteins

We filtered peptides by their length and kept those peptides with the canonical MIP length (8-11mers) and predicted binding affinity (IC50)<500 nM. Also, only MIPs detected in at least 3 replicates of both siblings per sibship (i.e. MIPs detected in 6-8 samples derived from 2 different HLA-identical siblings) were considered for further analyses. The predicted binding affinity of peptides to the allelic products was obtained using NetMHC version 3.2 (http://www.cbs.dtu.dk/services/NetMHC/)[21] for the frequent alleles HLA-A*0101/*0301/*2902 and HLA-B*0801/*1501/*4403) or NetMHCpan version 2.2 (http://www.cbs.dtu.dk/services/NetMHCpan-2.2/)[22] for the rare alleles HLA-A*0205 and HLA-B*5001. MIPs were further inspected for mass accuracy and MS/MS spectra were validated manually. The list of MIPs reported in the present work has been provided to The Immune Epitope Database and Analysis Resource (IEDB) (http://www.immuneepitope.org/)[23] under the reference ID 1022782. MIP source proteins were identified using the Sidekick resource (http://www.bioinfo.iric.ca/sidekick/) and only MIPs unambiguously assigned to one source gene were further analyzed.

3.4.4 Analysis of enriched pathways and functional categories

The Ingenuity Pathway Analysis software (Ingenuity Systems, http://www.ingenuity.com/) was used to identify overrepresented canonical pathways and functional categories for MIP source proteins. The right-tailed Fisher's exact test was used to calculate a p-value determining the probability that each overrepresented pathway or biological function was due to chance alone.

3.4.5 Functional connectivity score

An in-house developed all-pairs-shortest-path matrix[18] was used to calculate the functional association between two lists of nodes (transcripts or proteins), in this case: i) MIP source proteins of sibship 1 and MIPs source proteins of sibship 2 and ii) all MIP source proteins and all protein-coding transcripts expressed in B-LCLs. The functional association corresponds to a connectivity score. Details on the calculation of the score are provided as supplemental methods. A bootstrapping procedure was used as a statistical sampling method to calculate control connectivity scores from 10^5 sets of randomly selected human protein coding transcripts (1,165 transcripts in Figure 2C and 12,384 in Figure 4B). The *P* value corresponds to the number of times that the score of the bootstrap is smaller than the score of the sample/number of bootstrap iterations (105).

3.4.6 Analysis of transcripts containing miRNA-binding sites

Transcripts containing MREs were obtained from 2 sources: i) TargetScan v5.1 (http://www.targetscan.org/)[27], which includes both validated and predicted miRNA targets and ii) the Molecular Signature Database v3.0 (MSigDB) (http://www.broadinstitute.org/gsea/msigdb/index.jsp)[28], which includes gene sets with a known 3'-UTR miRNA binding motif. The MIP datasets used in the analyses were obtained from MS-based studies by our group[16, 17], by others[19, 29] or were available at IEDB (http://www.immuneepitope.org/, accessed on May 2011)[23]. The human and mouse proteome were downloaded from NCBI by excluding entries with a gene type for non-coding RNA, 'transposon' or 'pseudogene'. The proportion of transcripts containing MREs in the different datasets was compared with the two-tailed Fisher's exact test.

3.4.7 miRNA enrichment analysis

The miRNA target analysis module of the Web-based Gene Set Analysis Toolkit (WebGestalt) (http://bioinfo.vanderbilt.edu/webgestalt/)30 was used to identify transcripts containing MREs that are overrepresented among MIP source transcripts in B-LCLs. This tool also allowed us to predict miRNAs recognizing MREs found in enriched MIP source transcripts. A hypergeometric test and a

Benjamini and Hochberg correction were used to identify enriched miRNAs and target transcripts.

3.4.8 RNA extraction and cell sorting

5 x 10^6 PBMCs were labeled with FITC-conjugated anti CD19 (clone HIB19), APC-conjugated anti CD20 (clone 2H7) and propidium iodide to exclude cells in later apoptotic stages (all from BD Biosciences). $CD19^+CD20^+$ B cells were sorted on a FACSAria cell sorter (BD Biosciences) before RNA extraction. 0.5-1.0 x 10^6 cells were lysed in Qiazol (Qiagen) and total RNA was isolated of with the miRNeasy mini kit including DNase I treatment (Qiagen) according to the manufacturer's instructions. Total RNA was quantified using the NanoDrop 2000 (Thermo Scientific) and RNA quality was assessed with the 2100 Bioanalyzer (Agilent Technologies).

3.4.9 miRNA profiling

miRNA labeling, hybridization and washing steps were carried out with the miRNA complete labeling and hybridization kit (Agilent Technologies) according to the manufacturer's instructions. 100 ng of each RNA sample were hybridized to Agilent Human miRNA Microarray Release 16.0 (G4872A-031181 Agilent Technologies) containing 60K probes for 1,205 human and 144 human viral miRNAs (including 25 specific for EBV). Images were acquired with a GenePix 4000B scanner (Molecular Devices) at a 5 uM/pixel resolution and features were extracted with the GenePix 6.1 software. Analyses were performed using BRB-ArrayTools Version 4.2.0 Stable Release developed by Dr. Richard Simon and BRB-ArrayTools Development Team (http://en.bio-soft.net/chip/BRBArrayTools.html). The data were background-subtracted and quantile normalized. To estimate a single intensity measure for each miRNA we calculated the mean of its different probes. We applied the Self-Organizing Maps (SOM) method followed by hierarchical clustering analysis using the average linkage method and uncentered correlation as similarity metric. Clustering analyses were carried out with the Cluster v3.0 program (http://bonsai.hgc.jp/~mdehoon/software/cluster/)[31] and visualized with Java TreeView (http://jtreeview.sourceforge.net/).[32] The miRNA array data have been deposited into

the GEO database under accession number GSE35319.

3.5 Results

3.5.1 High-throughput MS-based identification of the MIP repertoire of B-LCLs

To characterize the MIP repertoire of B-LCLs, we performed mild acid elution of peptides from 4 B-LCLs derived from PBMCs of 4 different subjects. The 4 subjects belonged to 2 HLA-disparate sibships, each composed of one pair of HLA-identical siblings. Eluted peptides were fractionated by ion exchange chromatography and analyzed by nanoLC-MS/MS.[16-18] To identify MIPs consistently expressed by B-LCLs we selected those MIPs detected in at least 3 replicates of both siblings in each sibship (i.e. MIPs constantly detected in 6-8 samples derived from 2 different HLA-identical siblings). We further filtered peptides according to the canonical MIP length (8-11mers) and kept only those that were predicted to bind to their corresponding HLA-A and B allelic products for further analysis. In order to evaluate the HLA binding affinity of our peptides, we used NetMHC and NetMHCpan,[21, 22] which are the best-performing MHC-binding predictors.[33, 34] HLA-C-associated peptides were not included in this study because predictors for HLA-C allomorphs are not fully developed.

We identified a total of 2375 unique 8-11mers associated to 8 HLA-A and B allelic products: HLA-A*0101/*0205/*0301/*2902 and HLA-B*0801/*1501/*4403/*5001 (Table S1). Of the identified MIPs, 29 are listed in the IEDB and have been previously identified by various techniques including MHC ligand elution, MHC binding and T-cell response assays. Most of the peptides identified here are therefore novel and significantly enlarge the list of MIPs associated to MHC allomorphs, especially those for which less than 10 human peptides are currently known (HLA-A*0205 and HLA-B*0801/*5001). We found that a large proportion of MIPs were specific to one HLA (Figure 1A). Less than 10% of MIPs were predicted to bind to more than one HLA (Figure 1A), though a stronger binding preference was found for a single HLA allelic product (Table S1). Accordingly, the overlap between MIPs from HLA-disparate subjects was minimal (0.4%) (Figure 1C). Of note, we observed locus-dependent

differences in the predicted binding affinity of our MIPs. On average, HLA-A-associated peptides had a significantly stronger binding affinity than HLA-B-associated peptides (Figure 1B and Figure S1). We conclude that HLA allomorphs present essentially distinct MIP repertoires and that at the organismal level, the HLA genotype ultimately defines the MIP repertoire of an individual.

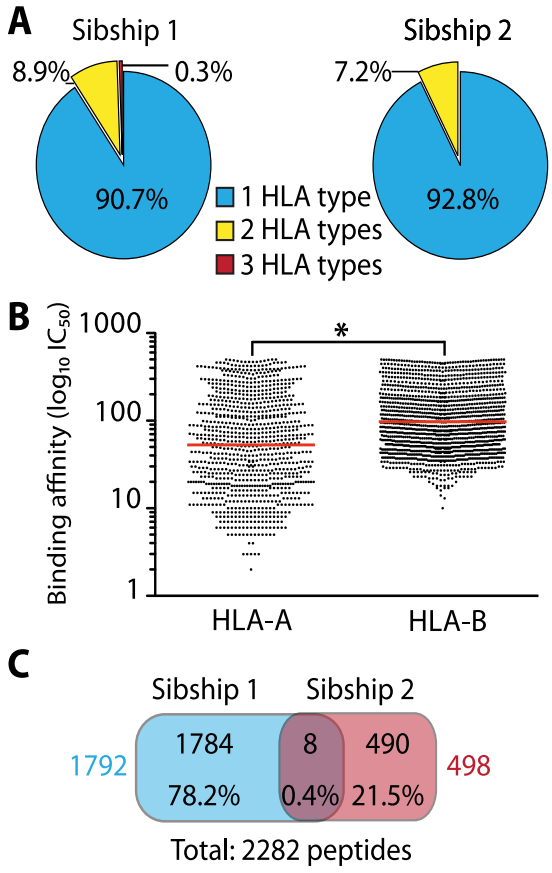


Figure 1. General features of MIPs eluted from B-LCLs from 2 HLA-disparate sibships

There were 2 HLA-identical siblings per sibship and each MIP included in the analysis was found in both siblings. A) Proportion of MIPs associated to 1, 2 or more different HLA allelic products, based on bioinformatic predictions with NetMHC/NetMHCpan. Most MIPs are associated to a single allelic product (HLA type). B) Predicted MHC I binding affinity (IC_{50}) of eluted peptides shows that HLA-B-associated peptides are weaker binders than HLA-A-associated peptides (*$P < 0.0001$; 2-tailed Mann Whitney test). Each dot represents a peptide and the red line corresponds to the mean binding affinity. C) Venn diagram showing minimal overlap between MIPs from HLA-disparate sibships. Peptide numbers and percentages are depicted for each category. Total number of MIPs for each sibship are shown in colors.

3.5.2 MIPs of HLA-disparate individuals derive from different sets of source proteins that are functionally interconnected and implicated in similar biological pathways

We unambiguously identified 1,750 proteins source of 2,290 peptides (Table S1), indicating that most MIP source proteins are represented by a single MIP. We asked whether distinct sets of MIPs associated to non-overlapping sets of HLA molecules originated from the same or from different sets of proteins. We found that the overlap between MIP source proteins of sibships 1 and 2 was minimal (7.5%) (Figure 2A). We then performed an enrichment analysis to identify pathways overrepresented in the proteins source of MIPs in each sibship. This analysis revealed that most of the overrepresented pathways were common to both sibships (Fisher's exact test, $P < 0.05$) (Figure 2B). Thus, despite the fact that MIPs of HLA-disparate subjects originated mostly from different proteins, a significant number of those proteins are implicated in the same biological pathways. We further confirmed this result by analyzing the functional connectivity between the sets of proteins giving rise to MIPs in sibships 1 and 2 using an in-house developed all-pairs-shortest-path matrix (Figure 2C). This analysis is based on the fact that proteins acting in the same biological pathway are closer in an interaction network, i.e. more functionally connected. Functional association was measured as a connectivity score, which corresponds to the mean of the shortest path distance between every pair of proteins (one protein from each sibship) in an interaction network. We found that the set of MIP source proteins of sibship 1 is highly connected to the set of proteins of sibship 2. A bootstrap procedure (100,000 iterations) failed to reveal a single randomly selected set of proteins in the human proteome that was so tightly connected to the MIP source proteins of our subjects ($P < 0.0001$). Collectively, these results show that B-LCLs from subjects with no shared HLA alleles present different immunopeptidomes that originate from different sets of proteins that are nevertheless functionally interconnected and implicated in the same biological pathways.

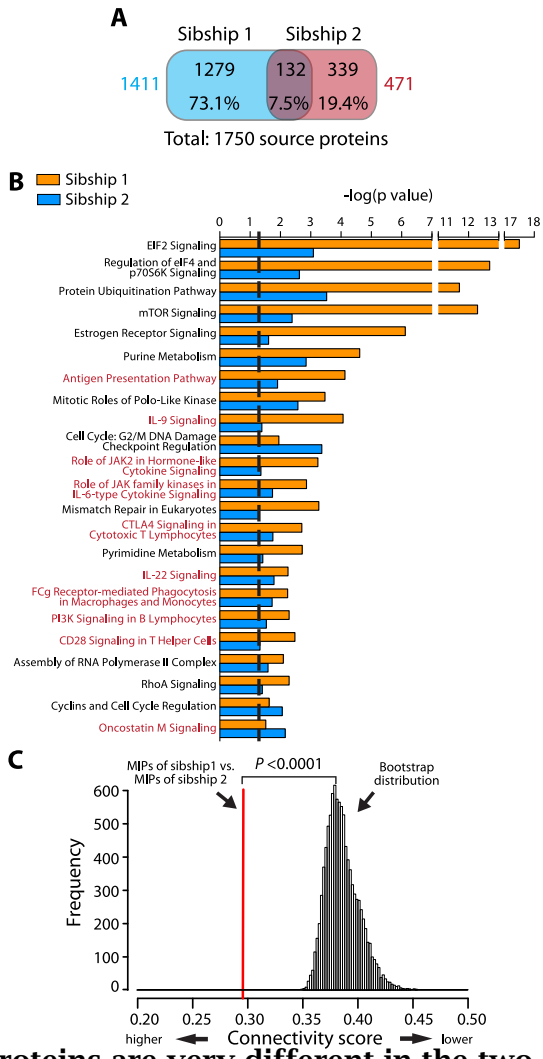


Figure 2. MIP source proteins are very different in the two sibships, but are implicated in similar biological pathways and are functionally interconnected

A) Venn diagram showing the minimal overlap between MIP source proteins from sibships 1 vs. 2. Protein numbers and percentages are depicted for each category. Total number of MIP source proteins for each sibship are shown in colors. B) The Ingenuity Pathway Analysis resource was used to identify overrepresented biological pathways for MIPs source proteins in sibships 1 and 2. Pathways are sorted by their statistical significance [-log(p value, calculated with the right-tailed Fisher's exact test)]. Higher scores indicate increased association between the MIP source proteins and a given pathway. The dotted line represents the threshold for statistical significance ($P < 0.05$). Immune-associated pathways are highlighted in red. C) MIP source proteins from HLA-disparate sibships are functionally interconnected. An all-pairs-shortest-path matrix was used to calculate the functional association between MIP source proteins of sibship 1 and MIPs source proteins of sibship 2 (see material and methods for details). The functional association was measured as a connectivity score, which

corresponds to the mean of the shortest path distance between every pair of proteins (one protein from each sibship) in an interaction network. Lower scores indicate increased connectivity. The red line represents the connectivity score between the MIP source proteins of sibship 1 and the MIP source proteins of sibship 2. A bootstrap procedure, represented by the Gaussian distribution, was used to calculate control connectivity scores between MIPs source proteins (of sibships 1 or 2) and random sets of proteins from the human proteome (*$P < 0.0001$, calculated as the number of times that the score of the bootstrap is smaller than the score of the sample/ number of bootstraps).

3.5.3 MIPs encoded by conserved and polymorphic genomic sequences

One interesting, yet unexplored question is whether MIPs preferentially derive from conserved vs. polymorphic genomic regions. Polymorphic MIPs, commonly referred to as minor histocompatibility antigens, are important in hematology because they elicit graft-vs.-host disease and graft-vs.-leukemia effect following allogeneic hematopoietic cell transplantation.[35-38] The most common form of polymorphism is single nucleotide polymorphism (SNP). We used an in-house developed software (pyGeno) to estimate the frequency of SNPs (SNPs/base pairs) in the MIP coding DNA sequences and compared it to the SNP frequency in the human exome (Table 1). Our analysis took into account all validated synonymous and non-synonymous SNPs reported in dbSNP. We found no significant differences between the frequency of synonymous or non-synonymous SNPs within MIP-coding sequences vs. the rest of the human exome. This result suggests that the MIP repertoire shows no bias toward polymorphic or invariant genomic sequences.

Table 1. SNP frequency (SNPs/bp) in MIPs and in the human exome
MIPs correspond to 2,282 unique peptides eluted from B-LCLs. The human exome corresponds to 26,4401 CDS extracted from Ensembl (GRCh37.65). dbSNP (Build 135) was used to calculate SNP frequency (SNPs/bp). SNP: single-nucleotide polymorphism, bp: base pair.

Dataset	Length (bp)	Non-synonymous SNPs		Synonymous SNPs		Total SNPs	
		Number	SNPs/ bp[a]	Number	SNPs/ bp[b]	Number	SNPs/bp[c]
MIPs	59994	230	0.0038	216	0.0036	446	0.0074

Exome	35201356	150158	0.0043	115376	0.0033	265534	0.0075

[a]$P = 0.12$, [b]$P = 0.16$, [c]$P = 0.50$, two-tailed Fisher's exact test when comparing MIP coding sequences to the exome not encoding MIPs.

3.5.4 MIPs derive preferentially from abundant transcripts

We previously reported that in primary mouse thymocytes, MIPs derived preferentially from highly abundant mRNAs.[16] However, studies on human cells have casted some doubts on the correlation between the immunopeptidome and the transcriptome.[9] To directly evaluate this relationship, we compared the expression level of 15,737 protein-coding transcripts expressed by B-LCLs[26] with the expression level of 1,707 MIP coding transcripts (Figure 3A-B). We set 4 expression categories based on Toung *et al.*[26]: very low, low, medium and high, and compared the proportion of transcripts belonging to each category in the 2 sets of transcripts. We found that MIP coding transcripts were skewed toward higher expression categories (Figure 3B). While 54% of all transcripts were expressed at a medium to high levels, 87% of those coding for MIPs did so. Furthermore, the proportion of highly abundant transcripts was enriched by 3.5 times in MIP-coding transcripts relative to the whole protein coding transcriptome. Consistent with this, while 46% of transcripts are expressed at low or very low levels, only 13% of MIP coding transcripts fell into these categories (Figure 3B). We further demonstrated that changing the thresholds that defined the various categories did not affect the differential expression of MIP coding transcripts relative to all transcripts expressed in B-LCLs (Figure S2).

Since MIPs preferentially derived from abundant transcripts, we next asked whether the proportion of MIP-coding transcripts was increased in a region with high transcriptional activity. The 6p21 chromosomal region containing the extended human MHC is recognized as a canonical transcriptional hotspot.[7] We first confirmed that this region generates increased proportion of highly abundant transcripts and further calculated the average expression of transcripts derived from 6p21 vs. the whole transcriptome of B-LCLs (Figure 3C). We found that the average gene expression level measured in FPKM (fragments per kilobase of exon per million fragments mapped) was 2 times greater for 6p21 than for all expressed protein-coding transcripts. We then compared the proportion of MIP-coding transcripts located in 6p21 vs. in the rest of the

genome (Figure 3D). We found that the proportion of MIP-coding transcripts was significantly higher for transcripts located in 6p21 than elsewhere in the

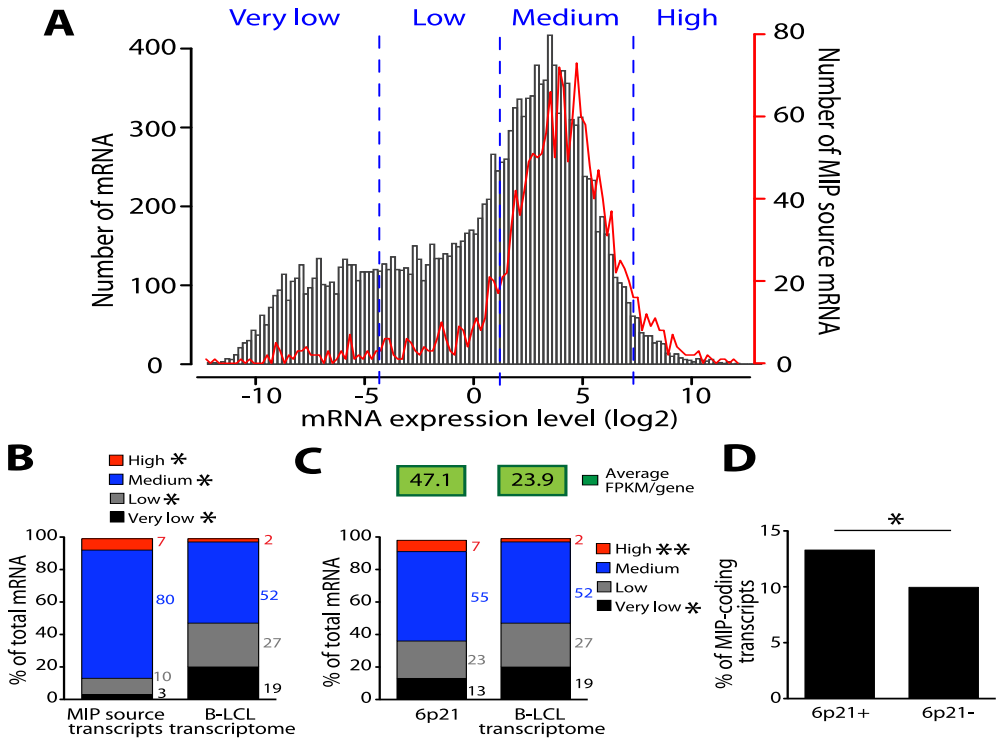


Figure 3. MIPs derive preferentially from abundant transcripts

A) MIP source transcripts are expressed at a wide range of levels. Expression levels were obtained from *Toung et al.*[26] and measured in FPKM by RNA-sequencing. Frequency distribution of expression levels of 15,737 protein-coding transcripts expressed by B-LCLs (black bars) and 1,707 (out of 1,750) transcripts source of MIPs (red line) for which an entry was found in the RNA-seq data. The frequencies of mRNAs and of MIPs source mRNAs are displayed on the left and right y-axis, respectively. Frequencies were calculated for 0.2 FPKM-bin increments. Expression categories (very low, low, medium and high) were set based on Toung *et al.*[26] and confirmed to be unbiased (Figure S2). B) Proportion of transcripts belonging to the expression categories shown in A among the two sets of transcripts indicate that MIPS derive preferentially from transcripts expressed at moderate to high levels as opposed to transcripts expressed at very low to low levels (*$P < 2.2 \times 10^{-6}$, Fisher's exact test). C) The 6p21 chromosomal region is a transcriptional hotspot in B-LCLs. Graph shows the proportion of transcripts belonging to the expression categories shown in A among 316 protein-coding transcripts located in the 6p21 chromosomal region and the protein-coding transcriptome of B-LCLs (*$P < 0.008$, **$P < 2.05 \times 10^{-7}$, Fisher's exact test). The average expression level (FPKM per gene) is shown on the top for both datasets. The 6p21 region analyzed was comprised between positions 30400001 and 46200000 (UCSC genome browser and NCBI Map viewer). D) The 6p21 transcriptional hotspot is a preferential source of MIPs. Comparison of the proportion of MIP-coding transcripts located in 6p21 (6p21+) and in the rest of the genome (6p21-) (*$P = 0.013 \times 10^{-6}$, Fisher's exact test).

genome. This results show that the 6p21 transcriptional hotspot is a preferential source of MIPs.

3.5.5 The MIP repertoire is functionally connected to the transcriptome

In mouse thymocytes and dendritic cells, we found that the MIP repertoire conceals a cell type-specific signature.[16, 17] To test whether this is also true in human cells, we first used the Ingenuity Pathway Analysis resource to identify major functional categories that were overrepresented in the set of proteins encoding MIPs in B-LCLs (Figure 4A). Some overrepresented categories were associated to basic cellular biological processes such as cell cycle, protein synthesis and gene expression, reflecting intracellular events occurring in any proliferating cell. The salient finding is that several immune-specific functional categories were also overrepresented, including hematopoiesis, lymphoid tissue structure and development, and humoral and cell-mediated immune responses (Figure 4A, categories in red). Furthermore, these major immune categories included B cell-specific functions such as proliferation, development and differentiation of B lymphocytes, class switching and development of pre-B and pro-B lymphocytes (Table S2). Moreover, the MIP repertoire reflected immune-associated, as well as intracellular pathways important for B cell biology (Figure 2B, pathways in red). Among the overrepresented pathways, we found antigen presentation, IL-6 signaling (required for B cell maturation), PI3K signaling in B lymphocytes (crucial in B-cell development) and JAK2 in cytokine signaling (very active in stimulated B cells).

If the MIP repertoire is molded by the cell type-specific transcriptome, then the set of MIP-encoding proteins in B-LCLs should be functionally connected to the B-LCLs' transcriptome. To test this hypothesis, we used our shortest path matrix method[18] to evaluate the functional connectivity between the set of MIP source proteins and the protein-coding transcriptome of B-LCLs and compare this connectivity to the connectivity scores of 10^5 random human transcriptome sets (Figure 4B). The functional connectivity found between MIP coding transcripts and the transcriptome of B-LCLs was significantly greater than the MIP connectivity to control transcriptomes (bootstrapping, $P < 0.0001$). Interestingly, the MIP-transcriptome connectivity increased as a function of the ex-

pression level of MIP coding transcripts (Figure S3). Thus, the connectivity of MIP-encoding transcripts to the transcriptome reached its maximum value when only highly abundant transcripts were considered. We conclude that the MIP repertoire of human cells is functionally connected to the cells' transcriptome in a transcript level-dependent manner.

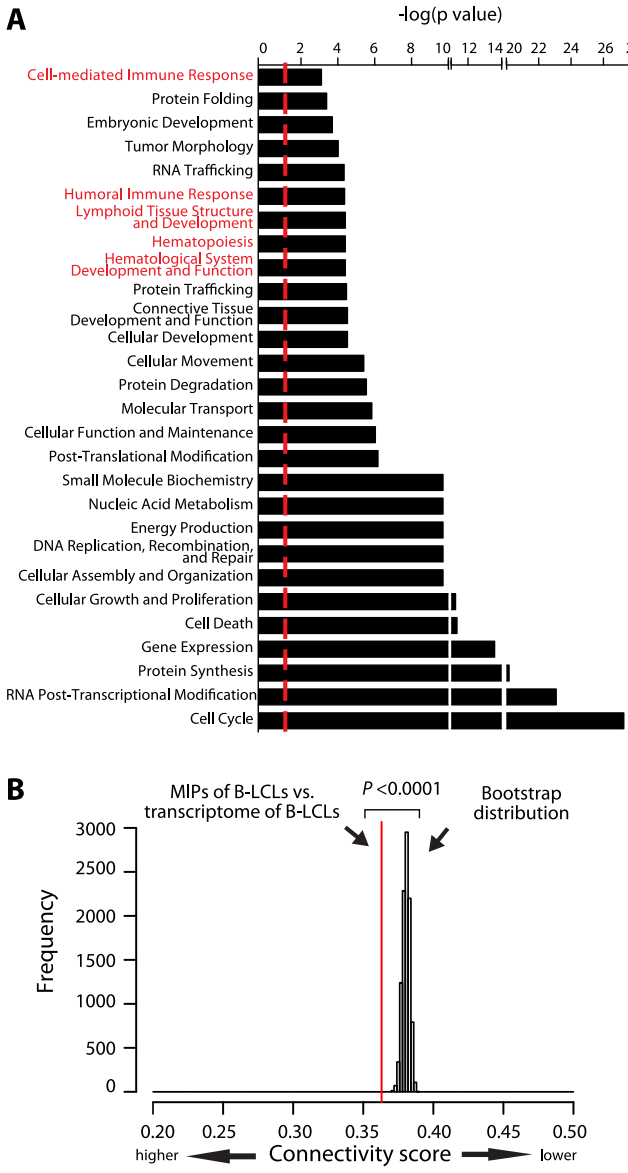


Figure 4. The MIP repertoire is functionally connected to the transcriptome
A) Graph showing the major functional categories that are overrepresented in the MIP repertoire of B-LCLs. Overrepresented functional categories were identified with the Ingenuity Pathway Analysis resource and are sorted by their statistical significance [-log(p value), calculated with the right-tailed Fisher's exact test]. Higher val-

ues indicate increased association between the MIP source proteins and a given category. The dotted line represents the threshold for statistical significance ($P < 0.05$). Immune-associated categories are highlighted in red. More details on functional annotations represented in immune-associated categories are provided in Table S2. B) The immunopeptidome of B-LCLs is functionally connected to the transcriptome of B-LCLs. An all-pairs-shortest-path matrix was used to calculate the functional association between MIP source proteins and the protein-coding transcriptome of B-LCLs (see material and methods for details). The functional association was measured as a connectivity score, which corresponds to the mean of the shortest path distance between every pair of proteins in an interaction network, i.e. one MIP-source protein and one expressed transcript of B-LCLs. The red line represents the connectivity score between MIP source proteins and transcripts expressed by B-LCLs. A bootstrap procedure, represented by the Gaussian distribution, was used to calculate control connectivity scores between MIPs source proteins and random sets of the human transcripts.

3.5.6 MIPs preferentially derive from transcripts containing miRNA-binding sites

According to the dominant paradigm, MIPs derive primarily from DRiPs generated by yet elusive biochemical processes.[10, 13] Theoretically, the amount of MIPs generated by a transcript should therefore be dictated by two factors: the transcript abundance and the transcript DRiP rate. In accordance with this assumption, medium and highly abundant transcripts were a preferential source of MIPs (Figure 3A), However, some MIPs derived from transcripts expressed at low or very low levels (Figure 3A). We therefore speculated that low abundance transcripts/proteins that generate MIPs, do so because they have a high DRiP rate. miRNAs are a class of non-protein-coding RNAs that bind to MREs on target transcripts causing destabilization of the target transcript or inhibition of its translation.[39] Destabilization of target mRNAs by miRNAs could constitute a possible mechanism by which DRiPs are generated. To test whether miRNA targets are a preferential source of MIPs, we first calculated the proportion of transcripts containing MREs in 3 datasets: transcripts source of MIPs in our B-LCLs, all human protein-coding transcripts (Figures 5A and S4A), and protein-coding transcripts expressed in our B-LCLs (Figures S4C-D). We used 2 different databases to identify validated and predicted transcripts recognized by miRNAs, TargetScan v5.1[27] and the Molecular Signature Database v3.0 (MSigDB)[28], and performed independent analyses with both databases (Figures

5A and S4). These analyses revealed that when compared to the B-LCL transcriptome (Figures S4C and D) or to the human protein-coding transcriptome (Figure 5A), the set of transcripts encoding MIPs identified in B-LCLs from each sibship was significantly enriched in transcripts containing MREs.

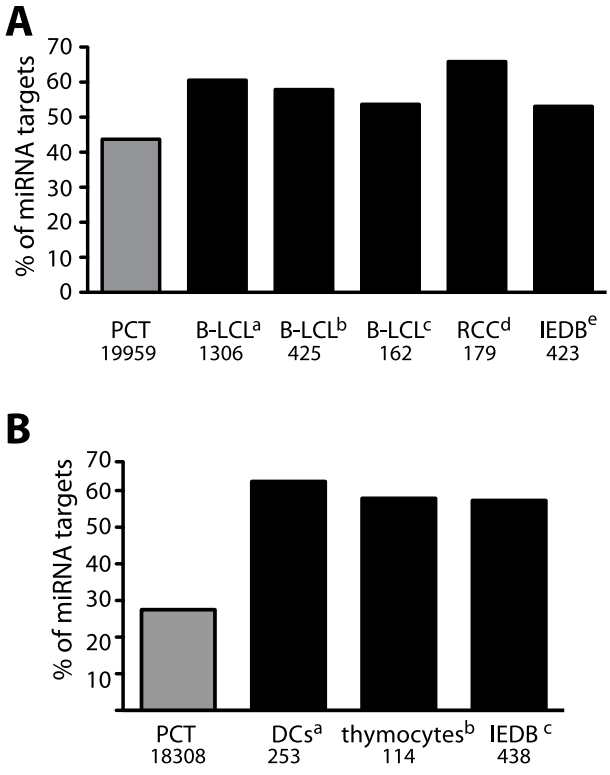


Figure 5. The MIP-coding transcriptome is enriched in transcripts containing miRNA-binding sites

A) Human transcripts containing miRNA-binding sites are a preferential source of MIPs. The proportion of MIPs that derive from miRNA targets was calculated for various datasets: B-LCLs from sibships 1 and 2 (this study), B-LCLs[19], renal cell carcinomas[29] and all human peptides listed in the IEDB[23]. A list of 8,725 human protein-coding miRNA targets was extracted from TargetScan.[27] Numbers on the bottom indicate the number of entries for each dataset. The proportion of MIPs that derive from miRNA targets for all datasets was significantly higher than the proportion of miRNA targets in the human protein coding transcriptome (PCT) a) $P < 2.2 \times 10^{-16}$, b) $P < 3.3 \times 10^{-9}$, c) $P < 0.01$, d) $P < 2.12 \times 10^{-9}$, e) $P < 1.38 \times 10^{-4}$, Fisher's exact test). B) Mouse transcripts containing MREs are preferential sources of MIPs. The proportion of MIPs that derive from miRNA targets was calculated for various datasets: dendritic cells (DCs)[17], thymocytes16 and all mouse peptides listed in the IEDB[23]. A list of 504 mouse protein-coding miRNA targets was extracted from TargetScan.[27] Numbers on

the bottom indicate the number of entries for each dataset. The proportion of MIPs that derive from miRNA targets for all MIP datasets was significantly higher than the proportion of miRNA targets in the mouse protein coding transcriptome (PCT) (a) $P < 1.0 \times 10^{-11}$, b,c) $P < 2.2 \times 10^{-16}$,Fisher's exact test).

To determine whether this feature of the MIP-coding transcriptome was cell type-specific, we carried out the same analysis on publicly available datasets of human MIPs: B-LCLs expressing HLA-B*1801[19], renal cell carcinomas[29] and a list of all human peptides identified in at least 20 different cell types and extracted from the IEDB.[23] In all cases, the proportion of MIP source transcripts bearing MREs was higher than expected (Figures 5A and S4A). We then tested whether this was also true for mouse MIPs (Figures 5B and S4B). Analysis of mouse MIPs previously identified by our group in DCs[17] and thymocytes[16], as well as MIPs from different cell lines listed in the IEDB[23], showed that the mouse MIP-encoding transcriptome was also significantly enriched in miRNA targets. In summary, these results show that irrespective of cell type, both mouse and human MIP-coding transcripts are enriched in miRNA targets.

3.5.7 MIPs from HLA-disparate sibships derive from different sets of transcripts regulated by similar miRNomes

We used the WebGestalt analysis program[30] to identify the specific MREs enriched in the 3'-UTR of MIP coding transcripts identified in B-LCLs from our 2 HLA-disparate sibships (Figure 6A). A representative set of enriched MREs is shown in Table 2 and the complete list can be found in Table S3.

Table 2. Representative set of enriched miRNA binding sites present in transcripts that are source of MIPs in B-LCLs from both sibships

The WebGestalt analysis tool was used to calculate the enrichment. N: number of transcripts source of MIPs that contain the 3'-UTR binding site recognized by the specified miRNAs. RE: Ratio of enrichment (observed/expected number of transcripts). *P* value: calculated from the hypergeometric test.

3'-UTR binding site	miRNA	sibship 1			sibship 2		
		N	RE	P value	N	RE	P value
TGCACTT	MIR-519C MIR-519B MIR-519A	51	4.21	5.93E-18	9	2.23	2.15E-02

ACATTCC	MIR-1 MIR-206	40	4.94	8.63E-17	11	4.07	9.95E-05
TGCCTTA	MIR-124A	53	3.53	2.87E-15	11	5.02	1.31E-02
CTTGTAT	MIR-381	30	5.64	1.66E-14	5	2.81	3.36E-02
ACCAAAG	MIR-9	48	3.48	9.38E-14	16	3.48	2.01E-05
TAGCTTT	MIR-9	48	3.48	9.38E-14	16	3.48	1.95E-02
TGAATGT	MIR-181A MIR-181B MIR-181C MIR-181D	47	3.51	1.23E-13	9	2.01	3.72E-02
GTGCCTT	MIR-506	57	2.93	7.65E-13	20	3.08	1.16E-05
ACTTTAT	MIR-142-5P	34	4.28	1.18E-12	11	4.15	8.38E-05
TTTGCAC	MIR-19A MIR-19B	46	3.28	2.41E-12	14	2.99	3.00E-04
TTTGTAG	MIR-520D	35	3.93	6.75E-12	7	2.36	3.08E-02
GTATTAT	MIR-369-3P	25	4.51	3.62E-10	6	3.24	1.11E-02
GTTTGTT	MIR-495	27	4.05	8.41E-10	8	3.6	1.90E-03
TTTGCAG	MIR-518A-2	24	4.33	1.88E-09	7	3.79	2.70E-03
CTTTGTA	MIR-524	36	3.01	5.82E-09	13	3.26	2.00E-04
CAGTGTT	MIR-141 MIR-200A	28	3.33	3.32E-08	10	3.57	6.00E-04
TATTATA	MIR-374	26	3.44	5.41E-08	8	3.17	4.10E-03
TTTTGAG	MIR-373	23	3.77	5.75E-08	8	3.93	1.10E-03
CTCAGGG	MIR-125B MIR-125A	28	3.20	7.76E-08	7	2.4	2.84E-02
TGTTTAC	MIR-30A-5P MIR-30C MIR-30D MIR-30B MIR-30E-5P	40	2.54	9.80E-08	11	2.09	1.78E-02
TGGTGCT	MIR-29A MIR-29B MIR-29C	37	2.65	9.83E-08	12	2.58	2.80E-03
GCAAAAA	MIR-129	20	4.09	1.07E-07	6	3.68	6.20E-03
ATTCTTT	MIR-186	25	3.35	1.63E-07	6	2.41	4.03E-02
ATACTGT	MIR-144	21	3.81	1.77E-07	10	5.44	1.83E-05

We then asked whether the MIP repertoire would reflect the miRNome. In accordance to the observation that MIPs of HLA-disparate sibships derive from different sets of proteins (Figure 2A), the sets of MIP source transcripts bearing enriched MREs in the two sibships displayed minimal overlap (7%; Figure 6A). Having compared miRNA targets, we next compared the miRNAs recog-

nizing MREs in MIP source transcripts. Enriched miRNA targets were classified according to their specific 3'-UTR MREs. We then compared the miRNAs that were predicted to regulate MIP source transcripts in the 2 sibships. Notably, we found a 57% overlap among enriched miRNAs regulating MIP source transcripts in the two sibships (Figure 6B). Collectively, these results suggest that MIPs from HLA-disparate subjects derive from different sets of transcripts regulated by mostly similar miRNAs.

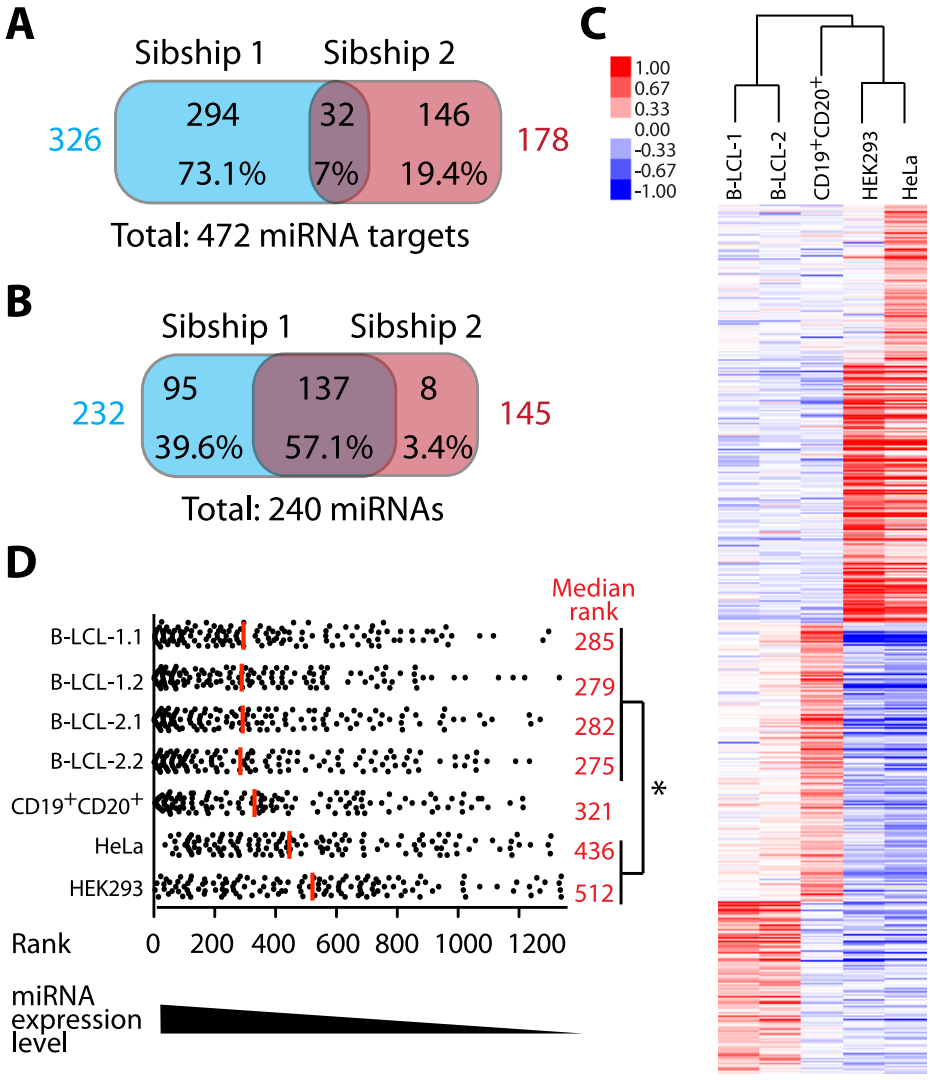


Figure 6. MIPs from HLA-disparate sibships derive from different sets of transcripts regulated by similar miRNomes

A) Venn diagram shows that MIPs of HLA-disparate sibships derive from different sets of miRNA targets. The WebGestalt analysis program30 was used to identify transcripts significantly enriched in 3'-UTR binding sites of specific miRNAs among all

MIP source transcripts identified in sibships 1 and 2 ($P < 0.05$, hypergeometric test). Total number of MIP source transcripts that are miRNA targets for each sibship are shown in colors. B) Different sets of MIP source transcripts are predicted to be regulated by mostly the same miRNAs. The WebGestalt analysis program[30] was used to identify miRNAs that recognize the 3'-UTR binding site present in enriched MIP source transcripts. Numbers and percentages of miRNAs are depicted for each category. Total number of miRNAs that target MIP source transcripts in each sibship are shown in colors. C) miRNA profiling shows that B-LCLs from HLA-disparate sibships have similar miRNA expression profiles. Human miRNA microarrays were used to assess the expression of 1,205 human and 144 viral miRNAs, in B-LCLs (2 subjects from each sibship), $CD19^+CD20^+$ cells isolated from PBMCs of one member of sibship 2, and two non-lymphoid cell lines (HEK293 and HeLa). Hierarchical clustering was used to calculate the correlation between various profiles (represented by a dendrogram). D) miRNAs that target MIP source transcripts in B-LCLs are expressed at higher levels in B-LCLs than in other cell types. All miRNAs studied were ranked from high to low according to their expression level in each cell line. Among them, the ranking of each of the 133 miRNAs predicted to regulate MIP source transcripts in B-LCLs was determined in B-LCLs, primary $CD19^+CD20^+$ cells, HEK293 cells and HeLa cells. Each dot corresponds to one miRNA. The median rank is shown for each cell population (red line and values). In average, miRNAs predicted to recognize MIP source transcripts are ranked significantly higher (i.e. more expressed) in B-LCLs than in non-lymphoid cell lines ($^*P < 0.05$, one-way ANOVA and Bonferroni's multiple comparison test). B-LCL-1.1: from subject 1 of sibship 1, B-LCL-1.2: from subject 2 of sibship 1, B-LCL-2.1: from subject 1 of sibship 2, B-LCL-2.2: from subject 2 of sibship 2.

To confirm that the miRNome of our 2 sibships was indeed similar, we performed genome-wide miRNA profiling in 8 different cell lines: B-LCLs from the 4 subjects (2 siblings from each sibship), primary $CD19^+CD20^+$ B cells from one of the subjects and 2 non-lymphoid human cell lines, HEK293 and HeLa (Figure 6C). Hierarchical clustering analysis revealed that the highest similarity was found between B-LCLs from the 2 sibships, which formed a separate cluster. Hence, as expected, B-LCLs express a cell type-specific miRNome that is independent of the HLA genotype or other interindividual genetic differences. We then asked whether enriched miRNAs that were predicted to regulate MIP source transcripts in the 2 sibships (Tables 2 and S3) were indeed expressed in the B-LCLs of our four subjects and if so, at what level. All 1,349 profiled miRNAs were ranked from high to low according to their expression level in each

cell line. Among them, the ranking of each of the 133 miRNAs predicted to regulate MIP source transcripts in B-LCLs was determined in B-LCLs, primary CD19+CD20+ cells, HEK293 cells and HeLa cells (Figure 6D). Importantly, enriched MREs can be recognized by more than one miRNA (Table 2), and we cannot assign which of the predicted miRNAs was indeed acting on the transcript source of MIPs. Despite this source of uncertainty, we found that in average, miRNAs predicted to recognize MIP source transcripts ranked significantly higher in B-LCLs than in control cell lines. Thus, miRNAs predicted to recognize MIP source transcripts in B-LCLs were expressed at higher levels in B-LCLs than in other cell types. This supports the concept that the immunopeptidome of a cell is molded by its miRNome.

3.6 Discussion

Self MIPs regulate all key events in the life of CD8 T cells. Indeed, CD8 T cells are selected on self MIPs, sustained by self MIPs, and activated in the presence of self MIPs.[40] Moreover, MIPs are the targets of several immune processes including autoimmunity, graft rejection, graft-.vs.-host disease and the graft-vs. leukemia effect.[5, 6, 37] Estimates suggest that the immunopeptidome comprises about 0.1% of the nonapeptide sequences (the typical length of MIPs) found in the proteome.[5] Despite the important role of MIPs in health and disease, we know little about their biogenesis except for the fact that they derive from DRiPs whose physical nature remains ill-defined.[13] In this study, we aimed to characterize the global landscape of MIPs on human cells and to gain insights in the biogenesis of the human immunopeptidome. We used B-LCLs as they can be obtained from practically any subject, they proliferate extensively *in vitro*, express high levels of MHC I molecules at the surface and have been shown to be a reliable tool for high-throughput genomic studies.[41] We found that most MIPs bind one unique HLA among all available MHC allelic products, and they do so with different predicted binding affinities. In general, HLA-A-associated peptides had a significantly higher binding affinity than HLA-B-associated peptides. This probably results from differences in the permissiveness of the HLA binding motifs. We also observed that subjects with no shared HLA-A and -B alleles present different MIP repertoires, showing that the immunopeptidome of an individual is ultimately determined by the combination of its HLA allelic products. This result was expected based on the specificities of the various

HLA binding motifs[9] and is in accordance with a recent MS-based study that showed minimal overlap in the identity of soluble plasma MIPs found in two HLA-different individuals.

Three salient findings emerged from the present work. Firstly, even though HLA-different subjects present different MIPs derived from different proteins, these MIPs originate from proteins that are functionally interconnected and implicated in the same biological pathways. In our B-LCLs, many MIP source proteins were associated to functional categories or involved in signaling pathways characteristic of the immune system or specific to B cell biology. Accordingly, we found that the MIP repertoire of B-LCLs was intimately linked to the transcriptome of B-LCLs. This further supports our previous results on mouse MIPs identified on thymocytes and DCs showing that the MIP repertoire conceals a tissue/cell-specific signature.[16, 17] This means that the immunopeptidome is not monopolized by MIPs derived from ubiquitous housekeeping genes. The notion that the MIP repertoire is cell type-specific has to be taken into account in immunotherapeutic interventions such as leukemia immunotherapy.[42]

Secondly, our results show that under basal conditions, the MIP repertoire of human B-LCLs derived preferentially from abundant transcripts. This was true when taking into consideration the whole exome (Figure 3 A-B) or a single transcriptional hotspot (the 6p21 chromosomal region; Figure 3 C-D). Notably, we used as a reference the average mRNA expression estimated by RNA-seq in B-LCLs from 20 unrelated individuals[26], which provides an accurate estimation of transcript abundance in this cell type. A correlation between transcript abundance and MIP presentation was also observed in MS studies on primary mouse thymocytes.[16] However, this correlation was absent in studies analyzing changes in MIP abundance induced by neoplastic transformation or metabolic stress.[16, 18, 29] All of these reports can be reconciled with a parsimonious explanation: i) in healthy normal cells, MIPs preferentially derive from the most abundant transcripts, and ii) in stressed cells, changes in MIP abundance result to a large extent from co- or post-translational processes that regulate the DRiP rate.

Finally, perhaps the most exciting finding reported herein is that MIPs derive preferentially from transcripts bearing MREs. miRNA profiling experiments revealed that the miRNAs which recognize MIP source transcripts in our B-LCLs were expressed at higher levels in B-LCLs than in other cell types. Furthermore, even though MIPs from HLA-disparate subjects derived from different transcripts, miRNA profiling experiments confirmed that these sets of transcripts are regulated by similar miRNomes as predicted by bioinformatic analysis. Notably, we found that transcripts containing MREs were a preferential source of MIPs not only in our B-LCLs but also in various mouse and human cell types analyzed by our group and others. The relation between MREs and MIPs is therefore very robust as it holds true across species and cell types. Though we still have to investigate how miRNAs could enhance the generation of MIPs, our working hypothesis is that destabilization of mRNAs by miRNAs generates DRiPs. Consistent with this idea, studies on cells transfected with shRNA or mRNAs carrying premature stop codons have revealed that the nonsense-mediated decay pathway can generate MIPs, presumably as a result of DRiP formation.[43, 44] Indeed, miRNAs can decrease the levels of the targeted protein via various mechanisms including slowing of elongation, ribosome drop-off and nascent polypeptides degradation.[45, 46] We therefore postulate that miRNAs are major regulators of the DRiP rate. Given the pervasive influence of miRNAs on all cellular processes including normal and neoplastic lymphohematopoiesis,[47-50] further studies are warranted to understand how miRNA may mold the immunopeptidome of normal and neoplastic cells. In addition we propose that the engineering of the MIP repertoire via the miRNome might be relevant in immunotherapy.

3.7 Authorship

DPG designed the study, carried out experiments, analyzed the data, prepared the figures and wrote the first draft of the manuscript. WY and TLMS performed MS experiments and analysis. CML participated in data analysis and prepared the figures. TD and JPL performed bioinformatics analyses. SL contributed to the study design, statistical analysis and discussion of the results. CP and PT designed the study, discussed the results and wrote the manuscript. All authors edited and approved the final manuscript.

3.8 Acknowledgements

We are grateful to Simon Drouin, Raphaëlle Lambert and Pierre Chagnon at the Genomics facility, Olivier Caron Lizotte and Eric Bonneil at the proteomics facility, and Danièle Gagné and Gaël Dulude at the Flow Cytometry falicility of the Institute for Research in Immunology and Cancer (IRIC). We also thank Véronique Lisi, Moutih Rafei, Chantal Baron and Marie-Christine Meunier for advice and thoughtful comments. This work was supported by a grant from the Fonds d'Innovation Pfizer-Fonds de la Recherche en Santé du Québec (FRSQ). DPG is supported by a studentship from the Canadian Institutes for Health Research. CL is supported by the Perseverance Program of IRIC. CP and PT are supported by the Canada Research Chairs Program. IRIC is supported in part by the Canadian Center of Excellence in Commercialization and Research, the Canada Foundation for Innovation, and the FRSQ.

3.9 Disclosure of Conflict of Interests

The authors declare that they have no conflict of interest.

3.10 References

(1) Perreault C, Decary F, Brochu S, Gyger M, Belanger R, Roy D. Minor histocompatibility antigens. *Blood* 1990;76(7):1269-1280.
(2) Boehm T. Quality control in self/nonself discrimination. *Cell* 2006;125(5):845-858.
(3) Klein L, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol* 2009;9(12):833-844.
(4) Perreault C. The origin and role of MHC class I-associated self-peptides. *Prog Mol Biol Transl Sci* 2010;92:41-60.
(5) de Verteuil D, Granados DP, Thibault P, Perreault C. Origin and plasticity of MHC I-associated self peptides. *Autoimmun Rev* 2012;Epub November 2011.
(6) McPhee CG, Sproule TJ, Shin DM et al. MHC class I family proteins retard systemic lupus erythematosus autoimmunity and B cell lymphomagenesis. *J Immunol* 2011;187(9):4695-4704.
(7) Horton R, Wilming L, Rand V et al. Gene map of the extended human MHC. *Nat Rev Genet* 2004;5(12):889-899.
(8) Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol* 2008;9:1.
(9) Mester G, Hoffmann V, Stevanovic S. Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. *Cell Mol Life Sci* 2011;68:1521-1532.
(10) Yewdell JW, Reits E, Neefjes J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature Rev Immunol* 2003;3(12):952-961.
(11) Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 2011;11(12):823-836.
(12) Dolan BP, Bennink JR, Yewdell JW. Translating DRiPs: progress in understanding viral and cellular sources of MHC class I peptide ligands. *Cell Mol Life Sci* 2011;68(9):1481-1489.
(13) Yewdell JW. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. *Trends Immunol* 2011;32(11):548-558.
(14) Yewdell JW, Lacsina JR, Rechsteiner MC, Nicchitta CV. Out with the old, in with the new? Comparing methods for measuring protein degradation. *Cell Biol Int* 2011;35(5):457-462.
(15) Milner E, Barnea E, Beer I, Admon A. The turnover kinetics of MHC peptides of human cancer cells. *Mol Cell Proteomics* 2006;5(2):357-365.
(16) Fortier MH, Caron E, Hardy MP et al. The MHC class I peptide repertoire is

molded by the transcriptome. *J Exp Med* 2008;205(3):595-610.
(17) de Verteuil D, Muratore-Schroeder TL, Granados DP et al. Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. *Mol Cell Proteomics* 2010;9(9):2034-2047.
(18) Caron E, Vincent K, Fortier MH et al. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Mol Syst Biol* 2011;7:533.
(19) Hickman HD, Luis AD, Buchli R et al. Toward a definition of self: proteomic evaluation of the class I peptide repertoire. *J Immunol* 2004;172(5):2944-2952.
(20) Tosato G, Cohen JI. Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines. *Curr Protoc Immunol* 2007;Chapter 7:Unit.
(21) Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* 2008;36(Web Server issue):W509-W512.
(22) Nielsen M, Lundegaard C, Blicher T et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* 2007;2(8):e796.
(23) Vita R, Zarebski L, Greenbaum JA et al. The immune epitope database 2.0. *Nucleic Acids Res* 2010;38(Database issue):D854-D862.
(24) Courcelles M, Lemieux S, Voisin L, Meloche S, Thibault P. ProteoConnections: a bioinformatics platform to facilitate proteome and phosphoproteome analyses. *Proteomics* 2011;11(13):2654-2671.
(25) Szklarczyk D, Franceschini A, Kuhn M et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;39(Database issue):D561-D568.
(26) Toung JM, Morley M, Li M, Cheung VG. RNA-sequence analysis of human B-cells. *Genome Res* 2011;21(6):991-998.
(27) Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120(1):15-20.
(28) Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545-15550.
(29) Weinzierl AO, Lemmel C, Schoor O et al. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol Cell Proteomics* 2007;6(1):102-113.
(30) Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005;33(Web Server

issue):W741-W748.
(31) de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics* 2004;20(9):1453-1454.
(32) Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics* 2004;20(17):3246-3248.
(33) Peters B, Bui HH, Frankild S et al. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol* 2006;2(6):e65.
(34) Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol* 2008;9:8.
(35) Choi EY, Christianson GJ, Yoshimura Y et al. Real-time T-cell profiling identifies H60 as a major minor histocompatibility antigen in murine graft-versus-host disease. *Blood* 2002;100(13):4259-4264.
(36) Mullally A, Ritz J. Beyond HLA: the significance of genomic variation for allogeneic hematopoietic stem cell transplantation. *Blood* 2007;109(4):1355-1362.
(37) Vincent K, Roy DC, Perreault C. Next-generation leukemia immunotherapy. *Blood* 2011;118(11):2951-2959.
(38) Li N, Matte-Martone C, Zheng H et al. Memory T cells from minor histocompatibility antigen-vaccinated and virus-immune donors improves GVL and immune reconstitution. *Blood* 2011;118(22):5965-5976.
(39) Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;136(2):215-233.
(40) Davis MM, Krogsgaard M, Huse M, Huppa J, Lillemeier BF, Li QJ. T cells as a self-referential, sensory organ. *Annu Rev Immunol* 2007;25:681-695.
(41) Herbeck JT, Gottlieb GS, Wong K et al. Fidelity of SNP array genotyping using Epstein Barr virus-transformed B-lymphocyte cell lines: implications for genome-wide association studies. *PLoS ONE* 2009;4(9):e6915.
(42) Bleakley M, Otterud BE, Richardt JL et al. Leukemia-associated minor histocompatibility antigen discovery using T-cell clones isolated by in vitro stimulation of naive CD8+ T cells. *Blood* 2010;115(23):4923-4933.
(43) Gu W, Cochrane M, Leggatt GR et al. Both treated and untreated tumors are eliminated by short hairpin RNA-based induction of target-specific immune responses. *Proc Natl Acad Sci U S A* 2009;106(20):8314-8319.
(44) Apcher S, Daskalogianni C, Lejeune F et al. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. *Proc Natl Acad Sci U S A* 2011;108(28):11572-11577.
(45) Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 2008;9(2):102-114.

(46) Hendrickson DG, Hogan DJ, McCullough HL et al. Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol* 2009;7(11):e1000238.
(47) Guo S, Lu J, Schlanger R et al. MicroRNA miR-125a controls hematopoietic stem cell number. *Proc Natl Acad Sci U S A* 2010;107(32):14229-14234.
(48) Visone R, Veronese A, Rassenti LZ et al. miR-181b is a biomarker of disease progression in chronic lymphocytic leukemia. *Blood* 2011;118(11):3072-3079.
(49) Fabbri M, Croce CM. Role of microRNAs in lymphoid biology and disease. *Curr Opin Hematol* 2011;18(4):266-272.
(50) O'Connell RM, Zhao JL, Rao DS. MicroRNA function in myeloid biology. *Blood* 2011;118(11):2960-2969.

3.11 Supplemental methods

3.11.1 Mass spectrometry and peptide sequencing

MIPs were released by mild acid treatment and separated by ion exchange chromatography using an off-line 1100 series binary LC system (Agilent Technologies) as previously described.[16] Peptides were loaded at 8 uL/min on a homemade strong cation exchange (SCX) column (0.3 mm internal diameter × 50 mm length) packed with SCX bulk material (Polysulfoethyl ATM, PolyLC). Peptides were separated into five fractions using a linear gradient of 0–25% B in 25 min (solvent A: 5 mM ammonium formate, 15% acetonitrile, ACN, pH 3.0; solvent B: 2 M ammonium formate, 15% ACN, pH 3.0) and brought to dryness using a speedvac. MIP fractions were resuspended in 2% aqueous ACN (0.2% formic acid) and analyzed by LC-MS/MS using an Eksigent LC system coupled to a LTQ-Orbitrap mass spectrometer (Thermo Electron).[16-18] Peptides were separated in a custom C12 reversed phase column (150 μm i.d. X 100 mm, Jupiter Proteo 4 μm, Phenomenex) at a flow rate of 600 nL/min using a linear gradient of 3-60% aqueous ACN (0.2% formic acid) in 69 mins. Full mass spectra were acquired with the Orbitrap analyzer operated at a resolving power of 60 000 (at m/z 400) and collision-activated dissociation tandem mass spectra were acquired in data-dependent mode with the linear ion trap analyzer. Mass calibration used an internal lock mass (protonated (Si(CH3)2O))6; *m/z* 445.120029) and mass accuracy of peptide measurements was within 5 ppm.

3.11.2 MS/MS sequencing and peptide clustering

Mass spectra were analyzed using Xcalibur software and peak lists were generated using Mascot distiller version 2.1.1 (http://www.matrixscience.com). Database searches were performed against a non-redundant IPI human database containing 150,858 sequences (version 3.54, released January 2009) using Mascot (version 2.2, Matrix Science). A Mascot search against a concatenated target/decoy database consisting of combined forward and reverse versions of the IPI human database was performed to establish a cutoff score threshold of typically 25. Non-redundant peptide sequences with a Mascot score higher than 25 were selected. The tolerance for precursor and fragment mass values

were set to 0.02 and 0.5 Da, respectively. Searches were performed without enzyme specificity and a variable modification for oxidation (Met) and deamidation (Asn, Gln). Raw data files were converted to peptide maps comprising *m/z* values, charge state, retention time and intensity for all detected ions above a threshold of 15000 counts using in-house software (Mass Sense).[16-18] Peptide maps were aligned and clustered together to profile the abundance of Mascot identified peptides using hierarchical clustering with criteria based on m/z and time tolerance (±0.02 m/z and ±1 min). This resulted in a list of non-redundant peptide clusters for all replicates of all samples. MS/MS of all peptide identifications are available through ProteoConnections[24] at http://www.thibault.iric.ca/proteoconnections under the project 111.

3.11.3 Functional connectivity score

To calculate the score, we first computed a distance (D) between each pair of neighbor nodes in an interaction network, D = –log2(P), where P is the probability that an interaction exists between 2 nodes given by the STRING database 9.0 (http://string-db.org/)[25]. STRING is a database of both known and predicted protein-protein interactions. It includes direct (physical) and indirect (functional) associations, which are derived from different sources: genomic context, high-throughput experiments, coexpression, and prior knowledge. Procedures to compute experimental scores are described in http://string-stitch.blogspot.com/2008_06_01_archive.html. We then summed the distances between nodes composing the shortest path between each pair of proteins in the matrix (one protein from each list). The connectivity score was then obtained by calculating the mean of the shortest path distance between every pair of nodes in a given matrix.

3.11.4 Analysis of SNP frequency

We have used the in-house developed software pyGeno to estimate the frequency of single nucleotide polymorphisms (SNPs) in the MIP coding genomic sequences and in the human exome. PyGeno is a Python module that integrates Ensembl GTF files (GRCh37.65) for gene annotations, the reference human genome (GRCh37.p2) and the reference database of SNPs dbSNP (Build

135). We first located exomic sequences corresponding to each MIP and then searched for validated SNPs in the peptide-coding region. Similarly, validated SNPs were located in the whole human exome corresponding to 264,401 coding sequences (CDS) extracted from Ensembl. We then classified SNPs as synonymous (nucleotide variation not leading to an amino acid change) or non-synonynous (nucleotide variation leading to an amino acid change). Synonymous SNPs correspond to the dbSNP function class 'synonymous-codon' and non-synonymous SNPs comprise the dbSNP function classes 'stop-gained', 'missense', 'stop-lost', 'cds-indel' and 'frameshift-variant'. The number of SNPs inside all MIP-coding regions where summed and divided by the total length in base pairs (bp) of all MIP-coding regions to determine the SNP frequency (SNP/bp). The same approach was applied to estimate the SNP frequency in the exome, considering all CDS instead of MIP coding regions. The SNP frequency in MIP coding sequences and in the exome not encoding MIPs was compared with a two-tailed Fisher's exact test.

3.11.5 Transcriptome of B-LCLs

We used reported gene expression profile data of B-LCLs from 20 unrelated individuals performed by RNA-seq[26] and available at the GEO repository (http://www.ncbi.nlm.nih.gov/geo/, GSE29158). RNA expression levels are given in fragments per kilobase of exon model per million mapped reads (FPKM) and the expression level for a gene is the sum of the FPKM values of its isoforms. For each coding transcript we calculated the average expression (AE) of the 20 B-LCLs. Non-protein coding and not-expressed transcripts were excluded. A total of 15,737 protein-coding transcripts expressed in B-LCLs were classified into four expression categories based on Toung *et al.*[26]: very low (AE>0), low (0.05<AE≤ 2.3), medium (2.3<AE≤163) and high (163<AE). We used the UCSC genome browser to map the 6p21 chromosomal region between positions 30400001 and 46200000 bp and the NCBI Map viewer to obtain the list of genes comprised in that region. Of those, we considered only genes classified as 'protein-coding' and with a RefSeq status 'validated', 'reviewed' or 'model'. The average expression level (FPKM per gene) was calculated for all protein-coding genes in 6p21 and all protein-coding genes expressed in B-LCLs.

3.12 Supplemental Figures

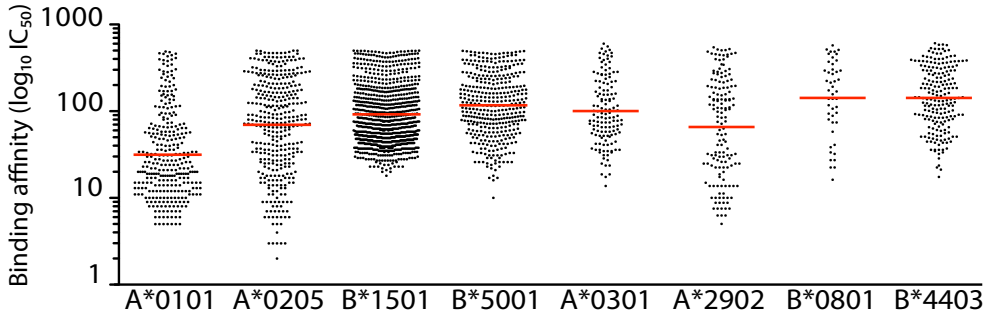


Figure S1. Comparison of the predicted MHC I binding affinity (IC50) of MIPs
Predicted MHC I binding affinity (IC_{50}) of eluted peptides to different HLA allelic products, based on bioinformatic predictions with NetMHC/NetMHCpan.

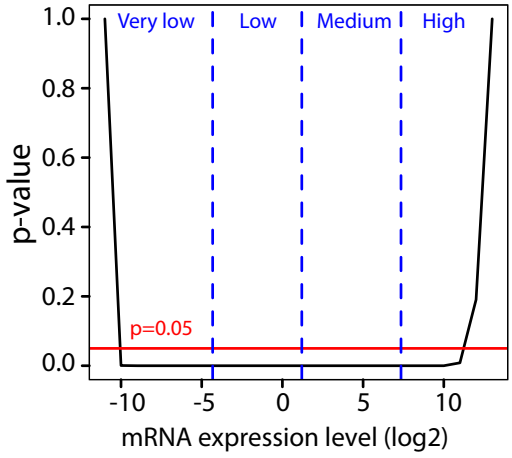


Figure S2. Random changes of mRNA expression thresholds set to define expression categories do not affect the differences in expression between transcripts encoding MIPs and all the transcripts expressed in B-LCLs
Random thresholds of mRNA expression level were set. For each threshold, a Fisher's exact T test was performed to compare the frequency of mRNA expression levels above and below a given threshold in MIP source transcripts vs. the transcriptome. The P value obtained by this test is shown for each random threshold of mRNA expression level analyzed (black line). The dotted blue lines represent the thresholds used to define expression categories in Figure 3. The red line indicates the limit P value for significance. Each significant P-value indicates that the proportions on both sides of the threshold in the 2 datasets are significantly different. Thus, this analysis shows that after setting the expression categories thresholds anywhere between -10

and 10, the comparison of the proportion of transcripts in each category is significantly different between MIP source transcripts and the transcriptome of B-LCLs.

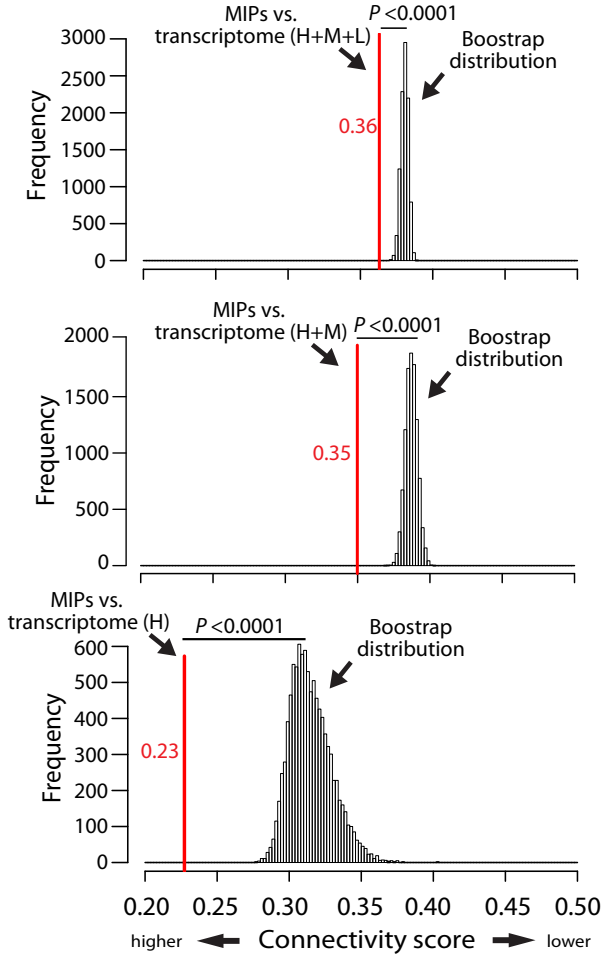


Figure S3. The MIP-transcriptome connectivity correlates with the level of expression of the transcripts in the transcriptome

The connectivity scores between MIP source proteins and 3 different sets of the protein-coding transcriptome of B-LCLs were calculated with the all-pairs-shortest-path matrix as described (Figure 4B). The 3 sets of transcripts analyzed were as follows: i) high, medium and low (H+M+L) (top), ii) high and medium (H+M) (middle) and iii) low (L) (bottom). The scores are indicated in red and represented by red lines. A bootstrap procedure, represented by the Gaussian distribution, was used to calculate control connectivity scores between MIPs source proteins and random sets of the human protein coding transcriptome. (*P < 0.0001, calculated as the number of times that the score of the bootstrap is smaller than the score of the sample/ number of bootstraps).

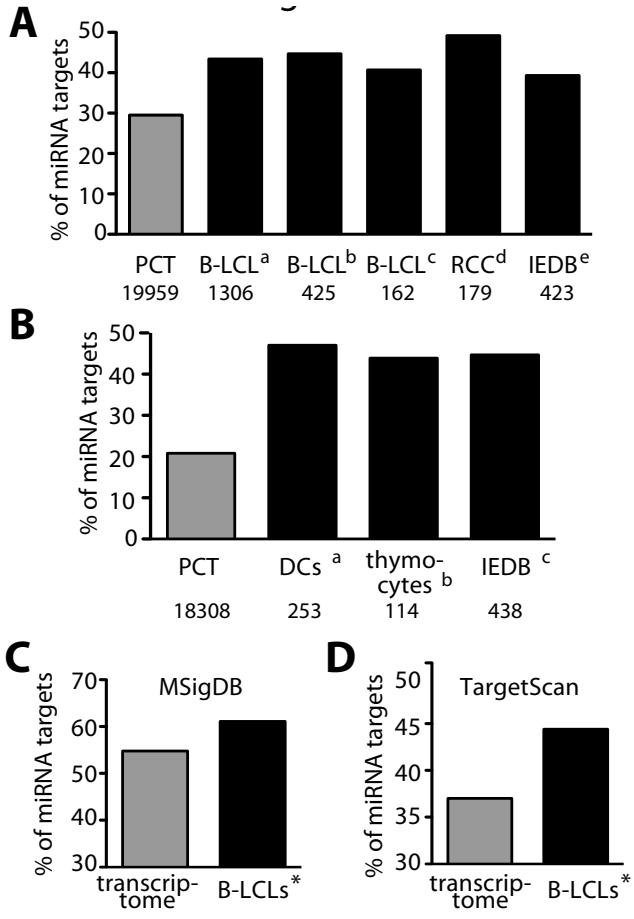


Figure S4. MIPs coded by miRNA targets are overrepresented in the mouse and human immunopetidomes

The same approach as the one described in Figure 5 A and B was used. In this case, the list of A) 5,887 human and B) 3,803 mouse protein-coding miRNA targets was extracted from MSigDB.[28] Numbers on the bottom indicate the number of entries for each dataset. A) The proportion of MIPs that derive from miRNA targets for all human datasets was significantly higher than the proportion of miRNA targets in the human protein coding transcriptome (PCT) (a) $P < 2.2 \times 10^{-16}$, b) $P < 2.43 \times 10^{-11}$, c) $P < 2.35 \times 10^{-3}$, d) $P < 3.3 \times 10^{-8}$, e) $P < 1.38 \times 10^{-6}$, Fisher's exact test). B) The proportion of MIPs that derive from miRNA targets for all mouse datasets was significantly higher than the proportion of miRNA targets in the mouse protein coding transcriptome (PCT) (a,c) P < 2.2 x 10-16, b) P < 3.52 x 10-8, Fisher's exact test). C) and D) Proportion of miRNA targets in MIP source transcripts vs. the protein coding transcriptome of B-LCLs. A list of C) 5,813 and D) 8,621 human protein-coding miRNA targets extracted from MSigDb[28] and TargetScan[27], respectively, was used to calculate the proportion of targets in 15,730 protein-coding transcripts expressed by B-LCLs[26] and 1,577 transcripts source of peptides eluted from B-LCLs of both sibships. The proportion of

miRNA targets is significantly higher in MIP coding transcripts than in the proteincoding transcriptome of B-LCLs (a) $P < 1.96 \times 10^{-10}$, bP $< 1.25 \times 10^{-7}$, Fisher's exact test).

3.13 Supplemental Tables

Supplemental tables S1-S3 are available on http://bloodjournal.hematologylibrary.org/content/119/26/e181/suppl/DC1.

CHAPTER 4

4 Minor histocompatibility antigens

Minor histocompatibility antigens (MiHAs) are a particular type of MHC I or MHC II-associated peptides that were originally discovered in a transplantation context, where they play a role in modulating immune reactions between donor and recipient cells [1,2]. In a transplantation setting between two perfectly HLA-matched (i.e. MHC-identical) individuals, recipient cells present a repertoire of MHC I-peptides that is not perfectly identical to the repertoire of peptides presented on donor cells [3]. Those MHC peptides that differ between patient and donor and that result from donor/recipient genetic disparity outside the MHC are known as MiHAs [4]. Hence, in the context of two HLA-identical subjects, MiHAs represent non-self peptides and therefore are not subject to tolerance mechanisms [5]. Consequently, MiHAs can induce and modulate T cell allo-responses in the recipient-to-donor or donor-to-recipient direction [5].

4.1 Origin of MiHAs

MiHAs result from any type of genetic variation in genes outside the MHC altering the amino acid sequence of the peptide or the expression of the protein source of peptide [6]. Most MiHAs discovered to date are caused by one or more nonsynonymous single nucleotide polymorphisms (ns-SNPs) in protein coding sequences that give rise to single or several amino acid substitutions in the polymorphic peptide [3]. Moreover, MiHAs can result from insertions, deletions, copy number variations (CNVs), frame-shift mutations and nonsense mutations [3,7-9]. The type and location of the genomic polymorphism can either affect the generation [9,10], the processing [11,12], the presentation [13-16] or the recognition of the peptides by the TCR [13,14,17,18] (Figure 1). Since in most human cases the alternative potential MiHA variant never reach the cell surface or is not immunogenic, most MiHAs possess only one immunogenic allele and only unidirectional recognition is seen [5].

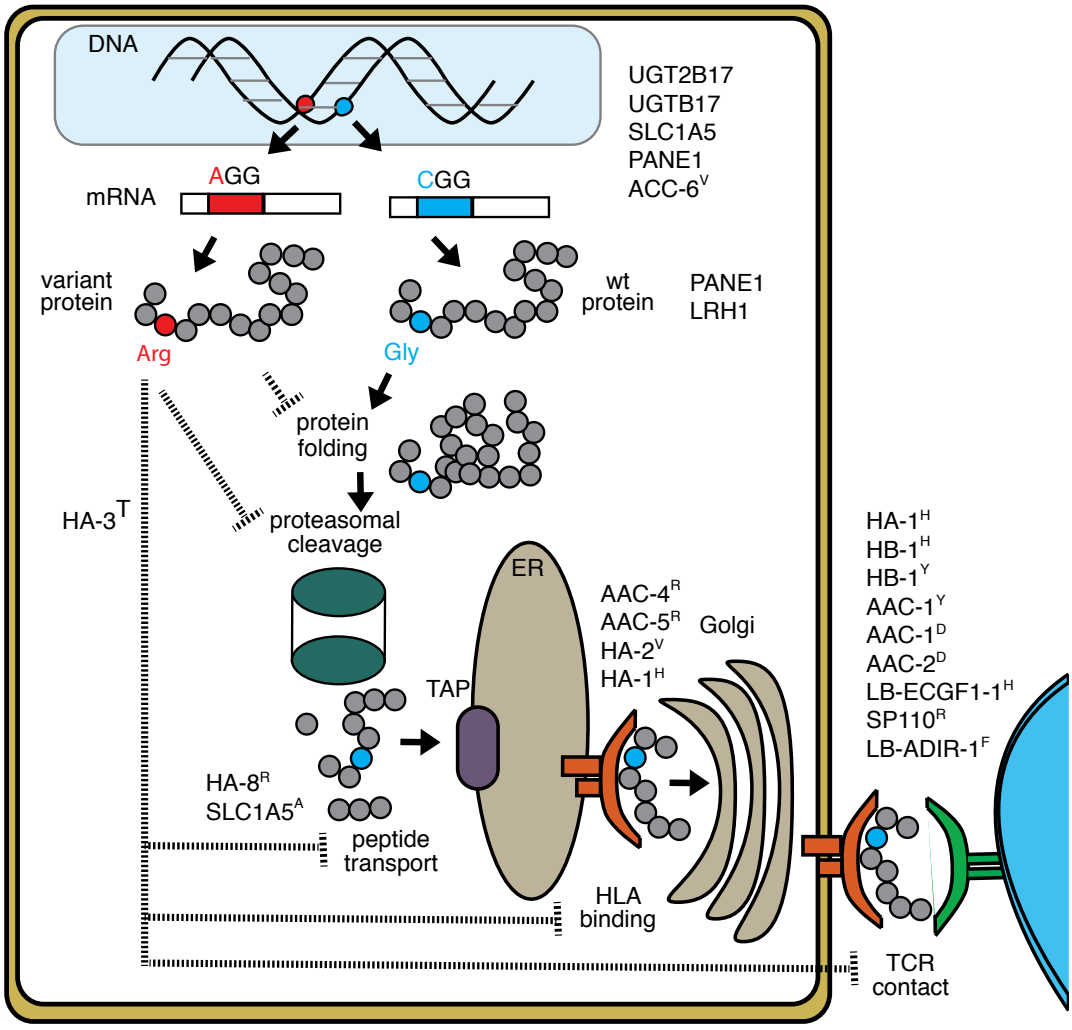


Figure 1. Mechanisms of generation of MiHA disparities
Genomic polymorphisms can affect multiple steps in the MHC I antigen processing and presentation pathway. Examples of human MiHAs are given in each step.

In various instances genetic variation impedes the generation of the protein source of peptides leading to MiHAs. The first MiHA was discovered following rejection of transplanted HLA-matched bone marrow cells from a male donor by his female recipient sibling [1]. The cause of rejection was the recognition of a male antigen derived from a gene on the Y- chromosome by the female T cells. Hence, the absence of Y chromosome-encoded proteins in female cells due to a “sex-mismatch” can lead to a particular type of MiHAs, referred to as HY MiHAs. The genes *KDM5D* and *UTY* located in the Y chromosome encode more than 6 distinct HY MiHAs [19]. Another example is the non-deleterious

homozygous deletion of the *UGT2B17* gene that is responsible for the generation of at least 2 known MiHAs [10,20,21]. Additionally, the PANE1 MiHA results from an nonsense SNP that generates a premature termination codon in the MiHA negative (variant) allele [9]. Similarly, the P2X5 MiHA results from a single nucleotide insertion/deletion (indel) that causes a frameshift and leads to a different protein sequence and thus the absence of peptide [8].

MiHAs can also result from SNPs in non-coding regions, such as 3' or 5' UTRs of protein-coding genes [22-24]. The HY-A33 MiHA results from a SNP present in the 5' UTR of the *TMSB4Y* gene leading to the expression of a protein variant in an alternative reading frame that is absent in the *TMSB4Y* homologue on chromosome X [23]. MiHAs can also be created by intronic SNPs in splicing sites leading to different splicing variants and thus the presence of a MiHA in some subjects and its absence in others [25,26]. This is the case of the ACC-6 MiHA, which derives from an alternative splice variant of the *HSMD* gene [25]. Likewise, MiHAs can be generated by SNPs in cryptic, noncanonical open reading frames in "genes" that are otherwise not known to generate a protein product. This is the case of the HB-1 MiHA encoded by a short 41 amino acids sequence whose function is unknown and that results from translation starting at a CUG codon [27].

Ns-SNPs have also been shown to affect the processing of the peptide [11,12,28-31]. For example, the methionine residue in the sequence of the HA-3T MiHA allelic variant (named HA-3M) causes the destruction of the precursor peptide by proteasomal cleavage C-terminal to the methionine residue [12]. Also, the ns-SNP in the coding region of HA-8 leads to inefficient translocation of peptide precursors into the ER by TAP and thus the corresponding variant peptide is absent on the cell surface [11]. A more peculiar example is the SP110 MiHA, produced by an exonic ns-SNP that causes the natural splicing of two non-contiguous peptide fragments in the reverse order by the proteasome resulting in a novel MiHA whose sequence is not encoded by the genome [30].

Finally, MiHAs can result from amino acid substitutions that affect the binding to MHC molecules or the recognition by the TCR [14,27]. In the first case, ns-SNPs that lead to changes in residues at anchor positions of the peptide

frequently affect the stability and binding of the peptide to MHC molecules. In the second scenario, the MiHA and its allelic counterpart are highly discriminated by T cells due to the fine specificity of the TCR in distinguishing non-self from self [32]. This situation is exemplified by the HA-1H MiHA variant (HA-1R) that do bind the HLA molecule but is not recognized by the TCR [14]. Since both MiHA variants are presented at the cell surface, these type of MiHAs can exhibit reciprocal antigenicity, such that both allelic peptide variants can elicit a T cell response [32,33]. Notably, posttranslational modifications can also lead to discrimination by the TCR [34-36]. For instance, the murine MiHA H4b contains a phosphorylated threonine residue that distinguishes it from the not immunogenic variant that contains an isoleucine residue instead [34]. Analogously, the cysteinylation of the cysteine residue contained in the HY-A2 MiHA can augment the recognition by a T cell clone, despite the fact that it slightly reduces the binding affinity for HLA-A*0201 [35]

4.2 MiHAs and allo-recognition

In an HLA-matched transplantation setting, MiHAs encoded by biallelic genes can differ between patient and donor due to genomic polymorphisms and trigger 3 types of bidirectional immune responses between donor and recipient cells: graft rejection, graft-versus-host disease (GvHD) and graft-versus-tumor (GvT) or graft-versus-leukemia (GvL) reactions. Graft rejection occurs when recipient T cells respond and target immunogenic peptides on the graft itself [37]. Inversely, donor T-cell responses directed against MiHAs of the recipient that are not present on donor cells, can cause the detrimental GvHD, but also elicit the curative GvT or GvL reactions [38]. The cell and tissue expression of the MiHAs determine in some extent the direction and type of reaction [39]. For instance, an HLA-matched donor who is homozygous for the MiHA ‘negative’ allele may have T cells that recognize recipient cells that are heterozygous or homozygous for the MiHA ‘positive’ allele [32,40]. In this donor-to-recipient scenario, MiHAs restrictively expressed in hematopoietic cells or tissue (including leukemic cells) induce specific allo-immune responses of donor T cells against recipient hematopoietic/leukemic cells that can lead to destruction of cancer cells (GvT/GvL reactions). On the contrary, MiHAs ubiquitously expressed in the recipient participate not only in GvL but also in GvHD [41,42], in

which donor T cells damage recipient target organs including the skin, gastrointestinal tract, liver and lung [43].

HY MiHA-specific T cells can be readily isolated from male recipients of MHC-matched female grafts [44]. Accordingly, the rate of relapse after transplantation is decreased in male recipients of MHC-matched female grafts owing to alloreactivity against male-specific HY MiHAs that contribute to GvL activity [45-48]. Paradoxically, male recipients of female donors do not necessarily experience increased survival due to increased morbidity and mortality caused by a higher risk of acute GVHD in these patients [41,45,46,49,50].

Apart from their central role in mediating transplantation outcomes, MiHAs are also important in "natural" allo-recognition conditions such as pregnancy. The role of MiHAs in pregnancy was first considered following the observation that parous female donors were more likely to elicit GvHD in transplant recipient than non-parous female or male donors [45]. Several studies have shown that the maternal immune system can recognize and respond to fetal MiHAs (reviewed in [51]). For instance, T cells specific for the MiHAs HA-1, HA-2 and HY have been found in maternal blood following pregnancy of mothers lacking the corresponding immunogenic alleles [51-53]. Indirect evidence suggests that fetal MiHAs might play a role in pregnancy complications. Increased expression of the MiHA HA-1 has been measured in preeclamptic placentas. Also, epidemiologic evidence suggests that recognition of fetal MiHAs might be implicated in secondary recurrent miscarriage, specially in women who have previously given birth to a male baby [51].

4.3 MiHA-based immunotherapy

Allogeneic hematopoietic stem cell transplantation (allo-HSCT) is an established potentially immunotherapeutic approach for several hematological malignancies [54]. Originally, the role of allo-HSCT was to reconstitute the patient hematopoietic compartment after sublethal chemoradiotherapy with the engrafted immune system from the donor. Nevertheless, it shortly became evident that donor lymphocytes could mediate the curative GvT/GvL effect after allo-HSCT, which could be exploited in cancer immunotherapy [54,55].

To reduce the development of GvHD following allo-HSCT, donor T cells can be depleted from the stem cell graft and re-administered after allo-HSCT in case of relapse in a process termed donor lymphocyte infusion (DLI). Compelling direct and indirect evidence suggest that the curative capacity of allo-HSCT results from a potent GvL/GvT effect mediated by donor lymphocytes [56]. First, it has been shown that following allo-HSCT, both CD8+ and CD4+ T cells specific for MiHAs on recipient cells are activated *in vivo* and can be isolated *in vitro* [2,57]. Second, human MiHA-specific CD8+ T cells have been shown to lyse primary leukemic cells *in vitro* [57-60]. Third, an increased incidence of leukemic relapse has been observed after autologous, syngeneic (identical twin) or T-cell depleted allo-HSCT [54]. Lastly, donor lymphocyte infusion is very effective to treat patients who relapsed after allo-HSCT [54]. Unfortunately, the successful application of the GvL/GvT effect in allo-HSCT is masked by the considerable morbidity and mortality of GvHD [56]. Currently, one of the main focus of research in the field is strategies to augment the GvL effect to prevent and treat relapse, while diminishing or avoiding GvHD [5].

The association of transplantation outcome and donor/recipient mismatching at loci encoding MiHAs has been evaluated by numerous retrospective studies yielding conflicting results even when analyzing the same MiHA [3,49,50,61-63]. Some studies have associated MiHA mismatch with GvL effect and long term survival in patients who received donor lymphocyte infusion to treat post-transplant relapse [14,60,64,65]. For instance a retrospective study has shown that one or more MiHA mismatches between HLA-matched donor and recipients in a large cohort of patients, who underwent T cell-depleted allo-SCT, resulted in improved relapse-free survival [49]. Other studies have not found a significant correlation between specific MiHA disparities and allo-HSCT outcome [61-63,66]. In opposition, some studies have shown that certain (but not all) MiHA mismatches between donor and recipient can increase the incidence and severity of GvHD [39,67,68], suggesting that some MiHAs might be preferred in allo-SCT. Despite the need of more systematic and comparable studies to establish the relation of allo-HSCT outcome and each MiHA mismatch, collectively these results suggest that early identification of MiHA mismatches between recipient and potential donors may help improving donor selection, improve allo-HSCT outcome, while minimizing GvHD [67].

The concept of targeting hematopoietic tissue-restricted MiHAs for immunotherapy of hematological malignancies was introduced in the mid-nineties. Various studies have shown that MiHA-based immunotherapy can very effectively destroy tumor cells in mouse models [69-72]. Moreover, it has been shown that adoptive transfer of CD8+ cytotoxic T cells directed at a single MiHA can effectively eradicate leukemia and melanoma without GvHD in mouse models [73,74]. Several strategies have been developed for using MiHAs to elicit GvL and GvT effects in humans [75]. Examples include *in vitro* isolation and expansion MiHA-specific T cells, vaccination of patients and donors with dendritic cells pulsed with MiHA or modified to present MiHA source proteins, mRNA or DNA and redirection of T cell specificity by gene transfer of MiHA-specific TCR [32,54]. An alternative approach is to isolate MiHA-specific T cells directly from the donor before allo-HSCT and administer them as part of the graft or shortly after transplant (reviewed in [5]).

Currently, the hematopoietic-restricted HA-1 [28] and HA-2 MiHAs are the most extensively explored MiHAs in immunotherapy. HA-1 is probably the most therapeutically-relevant MiHA, that could be applied in 6-12% of patients, depending on the transplantation setting (HLA-matched sibling vs. unrelated donor transplantation) [76]. Recently, Warren and coworkers have proposed the modulation of T cells specific for MiHAs to augment GvL and GvT effects, while reducing the need of immunosuppressive drugs [77]. In this approach, donor T cells specific for a particular MiHA expressed exclusively in leukemic/tumor cells are isolated from the recipient's blood following transplantation and expanded *ex vivo* [77]. The expanded T cells are re-infused into the recipient following a disease relapse. A preliminary clinical trial of adoptive immunotherapy using MiHA-specific T cells (although poorly characterized), showed some success in treating relapse but also pulmonary toxicity at high T-cell doses which correlated with MiHA expression in pulmonary endothelial cells [77].

The limited number of identified human MiHAs has been one major barrier for the broad clinical application of MiHAs in immunotherapy [5]. Ideally, a large panel of MiHAs would facilitate the selection of the best target as the tissue ex-

pression of the MiHA, the HLA-restricting molecule and the correct directional disparity between donor and recipient, reduce the real number of usable MiHA for each particular transplantation case. Preferably, MiHAs need to fulfill various characteristics for broad application in immunotherapy of leukemia. First, a restricted expression of MiHAs to hematopoietic cells would maximize the GvT effect without causing GvHD [55]. Second, MiHAs presented by common HLA molecules and that display an equally balance phenotype population frequency (to increase the chance of MiHA-mismatch) would have a broader application [78]. Additionally, the genetic nature of the MiHA could increase its potency. It has been suggested that MiHAs resulting from gene deletions (such as UGT2B17) could have more impact than MiHAs resulting from SNPs, as this will increase the potential of multiple antigenic epitopes associated to different HLA molecules to trigger an immune response [67,68].

4.4 The arduous identification of MiHA

Since the discovery of the first MiHA almost 4 decades ago [1], new MiHAs have been identified at a very slow rate. A summary of all currently known MiHAs associated to MHC I and MHC II is presented in table 1. There are currently 59 MiHAs reported in the literature, of which 12 are associated to MHC II and 47 to MHC I molecules. They derive from 44 autosomal and 6 Y-chromosome-encoded genes of broad or restricted expression. Notably, 33 of them (56 %) have been identified in the last 5 years, illustrating how recent advances in human genomics and bioinformatics have accelerated the pace of MiHA discovery.

Several techniques including both **forward and reverse** immunology approaches have been developed to identify MiHAs. Classically, the identification of MiHAs starts with a MiHA-reactive T cell clone obtained from a patient showing a clinical response to DLI after allo-HSCT, followed by elucidation of the antigens that are recognized. These T-cell based approaches are referred to as **forward immunology approaches**. The T cell clone is used to investigate the MiHA specificity, the HLA allele that presents it and the basis of its immunogenicity, using techniques such as cDNA-expression cloning [10,17,22,25,27,79-82], high performance liquid chromatography (HPLC) and mass spectrometry (MS) [13,18,28,35,36,83-85], genetic linkage analysis [8,11-13,28,29,86,87] or

whole genome association scanning [21,22,81,88,89].

The first MiHAs were discovered using T-cell based immunology approaches combined with HPLC and MS. In this strategy, peptides are extracted directly from the recipient cell and fractionated by HPLC to identify fractions that contain the epitope recognized by cytotoxic T cells. Positive fractions are then analyzed by MS [13,28,35,36,84]. Alternatively, T cell-based approaches have been combined with cDNA expression cloning. Here, MiHA-specific cytotoxic T cells are screened against COS7 or 293T cells cotransfected with pools of plasmid cDNA library prepared from MiHA-positive cells and with a plasmid encoding the HLA-restricting allele [17,24,25,27,30,33,90]. Positive cells are subcloned to identify those that express the cDNA encoding the MiHA, which is then localized by transfecting truncated versions of the gene or by prediction algorithms for HLA binding [6]. This technique has been more recently adapted for the identification of MHC II-associated MiHAs by expressing the cDNA library in bacteria and then loading the modified bacteria into MHC II-positive B-LCLs [82,91].

Alternatively, MiHAs have been discovered by testing the reactive T cell clones against a panel of B-LCLs from large pedigrees such as the Centre d'Etude Polymorphism Humain (CEPH) families that have been mapped for genetic markers, followed by genetic linkage analysis to identify the chromosomal region underlying the MiHA [8,81,92].

Important development in the field are genome-based strategies for MiHA identification such as genome-wide-association studies, HapMap screening, whole genome association scanning and genome-wide zygosity-genotype correlation analysis [21,22,88,89,92,93]. Basically, these approaches analyze the relationship between patterns of expression of the MiHA phenotype of B-cell lymphoblastoid cell lines (B-LCLs), measured by *in vitro* cytotoxicity or cytokine secretion assays, and their pattern of SNPs genotype obtained from publicly available B-LCLs of known genotypes or from databases such as HapMap [94]. In contrast to genetic linkage analysis, these approaches are more rapid and powerful because the genome-wide analysis is performed simultaneously for all SNPs across the whole genome.

The genome-wide zygosity-genotype correlation analysis allows to precisely map the genomic locus of MiHAs and identify the MiHA amino acid sequence [89,92]. This method starts with the isolation of CTLs from allo-SCT recipients with a strong GvT response. The CTLs are tested *in vitro* for recognition of recipient's but not donor's APCs. Selected CTLs are then screened using a panel of B-LCLs with different combinations of HLA molecules to identify the specific HLA molecule presenting the MiHA. Then, B-LCLs from multiple individuals carrying the specific HLA and from different father-mother-child trios are tested against the CTL. Using a Mendelian segregation pattern of the father-mother-child trios the MiHA zygosities are deduced for various individuals and a genome-wide zygosity-genotype correlation analysis allows the identification of candidate SNPs, of which only non-synonymous coding SNPs are retained. Lastly the SNP disparity in donor and recipient cells is confirmed and all possible nonameric sequences including the SNP are synthesized and tested in vitro with the CTL (IFN-γ response). This approach allowed the recent identification of the UTA2-1 MiHA, among others [76].

Alternative approaches not requiring father-mother-child trios are those based on whole-genome association scanning (WGAs) [21,22,81,88]. In this strategy, a panel of SNP-genotyped B-LCLs (either expressing the patient's HLA molecules or retrovirally transduced with the HLA genes of the patients) are tested for recognition by the CD8+ T cells isolated from patients with leukemia who responded to DLI with minor GvHD. The genetic origin of MiHAs is then identified based on an analysis of association between T cell recognition of each B-LCL and individual SNP genotypes. The precise identity of the MiHA is determined using synthetic peptides. Recently, a similar approach combining microarray-based SNP genotyping followed by peptide prediction and *in vitro* peptide binding assays has been used for the identification of MiHAs associated to MHC II [93]. Genome-based approaches represent one of the most promising methods currently available for high-throughput identification of MiHAs.

Forward approaches based on cytotoxic T cell clones have the advantage of identifying clinically relevant MiHAs, as frequently MiHA specific- T cells correlate with clinical responses [60]. However, one of the major constraints of

T cell-based approaches is the relative scarcity of CD8+ T cells in peripheral blood and target sites. Furthermore, MiHA-specific T-cell responses may go undetected, especially when analyzed directly *ex vivo* in clinical specimens that are not obtained during the peak GVL response. In addition, T cells being extremely cross-reactive, T-cell based antigen discovery (in the absence of MS) is associated with a high rate of false positives [95,96]. Hence, it is often necessary to fish a target population with MHC multimers and expand the T cells *ex vivo* using cytokines and antigen, potentially altering their functionality [97,98]. Novel alternative approaches have been proposed to facilitate the cumbersome task of isolating T cell clones from blood of allo-HSCT recipients. Bleakley and coworkers have successfully applied the *in vitro* stimulation of naive CD8+ T cells from unprimed donors with monocyte-derived dendritic cells from the HLA-identical sibling for the identification of novel MiHAs [40]. Hombrink and coworkers also developed an highthroughput approach to simultaneously isolate different reactive T cells from peripheral blood mononuclear cells (PBMCs) by using a large collection peptide-MHC tetramers and then expanding the cells *in vitro* prior to analysis by flow cytometry [99].

Because T cell are not always available, **reverse immunology approaches** based on bioinformatic predictions of MiHA have been developed [19,93,99-101]. These approaches start with prediction of candidate peptides and subsequently *in vitro* screening using proteasome digestion, HLA-binding, and testing for their ability to activate T cells. An example of this strategy is the recent work by Mortensen and coworkers, who predicted all possible 8-11mer peptides derived from the *UTY* gene located on the Y-chromosome that could bind the 15 most common HLA-A and HLA-B molecules [102]. The predicted peptides were ranked, synthesized and tested for induction of cytokine response upon contact with PBMCs from a lymphoma patient who was in remission after allo-HSCT. Subsequently, the candidate MiHA was confirmed with HLA-peptide binding assays and tetramer staining. Another example is a recent reverse immunology approach proposed by Robins and coworkers, in which candidate MHC-binding nonamers are generated *in silico* based on single mutations identified by exome sequencing of melanoma cell lines obtained from patients and tested for recognition by autologous tumor-infiltrating lymphocytes [103].

Although reverse immunology strategies are less laborious than forward approaches, they often fail to confirm the processing and immunogenicity of the candidate MiHAs under natural conditions because they suffer from a high false discovery rate that can represent 95% of candidate peptides [103]. Currently, reverse immunology approaches reflect the limited knowledge and unpredictability of the antigen processing pathway, in which only 0.1 % of the produced peptides are indeed presented on the cell surface [104]. To improve discovery, **novel approaches** have combined bioinformatics predictions with experimental identification of HLA-presented MIPs and MiHAs by mass spectrometry without the need of T cell clones [105]. For example, van Veelen and coworkers have developed a database of all possible polymorphic proteins generated from all reported ns-SNPs that can be used for identification of polymorphic MIPs and potential MiHAs by mass spectrometry [106]. Although this strategy relying on naturally presented peptides, instead of predicted peptides, has allowed the identification of the LB-NISCH-1A MiHA [107], the use of a generalized database for MS-based peptide identification introduced another type of bias leading to 25% of false positive peptide identifications [105,107].

Table 1. Features of known MiHAs associated to HLA I and HLA II molecules

The polymorphic residues are highlighted in bold; the alternative amino acid residue is shown in parentheses.

MiHAg	HLA allele	Gene/ chromosome	Gene name/function	Peptide sequence	Tissue distribution	technique	Note	Polymorphism	Reference
ACC-4^{R}	A*3101	CTSH/15q25.1	Cathepsin H	ATLPLLCA**R(G)**	Uniquitous	cDNA library and linkage analysis	Exonic ns-SNP	rs2289702	[1]
ACC-5^{R}	A*3303	CTSH/15q25.1	Cathepsin H	WATLPLLCA**R(G)**	Ubiquitous	cDNA library and linkage analysis	Exonic ns-SNP	rs2289702	[1]
ACC-6	B*4402, B*4403	HMSD/18q21.3	Histocompatibility minor serpin domain containing	**MEIFIEVFSHF,** HSMD variant	Hematopoietic	cDNA library	Intronic SNP, alternative splicing	rs9945924	[2]
BCL2A1 (ACC-1$^{Y/C}$)	A*2402	BCL2A1/15q24.3	B-cell leukemia/ lymphoma 2 related protein A1	DYLQ**Y(C)**VLQI	Hematopoietic, selected solid tumors? AML, CML	WGAs	Exonic ns-SNP	rs1138357	[3,4]
BCL2A1 (ACC-2^{D})	B*4403	BCL2A1/15q24.3	B-cell leukemia/ lymphoma 2 related protein A1	KEFED**D**(G)IINW	Hematopoietic, selected solid tumors? AML, CML	WGAs	Exonic ns-SNP	rs3826007	[4]
C19orf48 (HwA-11)	A*0201	C19orf48/19q13	Uncharacterized protein-coding	CIPPD**S(T)**LLFPA	Ubiquitous, select solid tumours	cDNA library	Exonic ns-SNP in ARF	rs3745526	[5]
CD19^{L}	A1*05 B1*02 DQ	CD19/16p11.2	B lymphocyte antigen CD19	WEGEPPC**L(V)**P	Hematopoietic	Genome-wide zygosity-genotype analysis	Exonic ns-SNP	rs2904880	[6]
HA-1^{H}	B60 B*40012	HMHA1/19p13.3 (KIAA0223/19p13)	Histocompatibility (minor) HA-1, Rho-like GTPase-activating protein	KECVL**H(R)**DDL	Hematopoietic, Select solid tumors AML	Polymorphic peptide screening	Exonic ns-SNP	rs1801284	[7]
HA-1^{H}	A*0201 A*0206	HMHA1/19p13.3 (KIAA0223/19p13)	Histocompatibility (minor) HA-1, Rho-like GTPase-activating protein	VL**H(R)**DDLLEA	Hematopoietic, Select solid tumors (bronchial, cervix, breast, prostate carcinoma) AML	HPLC+MS, peptide binding assay	Exonic ns-SNP	rs1801284	[8-10]
HA-2^{v}	A*0201	MYOG1/7p12p13 MYOG1/7p13-p11.2 Myosin 1G	Unconvensional myosin 1G	YIGEVLVS**V(M)**	Hematopoietic AA	HPLC+MS	Exonic ns-SNP	rs61739531	[8,11]
HA-3^{T}	A*0101	*LBC* (oncogene) / *AKAP13*/15q24-25	Guanine nucleotide exchange factors for RhoA Lymphoid blast crisis oncogene A kinase (PRKA) anchor protein 13	V**T(M)**EPGTAQY	Ubiquitous (hematopoietic, keratinocytes, fibroblasts, PTECs, HUVECs, melanocytes) AML	HPLC+MS	Exonic ns-SNP	rs7162168	[12]
HA-8^{R}	A*0201	*KIAA0020*/9p24.2	Pumilio-family member	**R(P)**TLDKVLEV	Ubiquitous (hematopoietic, fibroblasts) CML	HPLC+MS	Exonic ns-SNP	rs2173904	[13]

HB-1$^{H/Y}$	B*4402 B*4403	*HMHB1*/5q32 *HMHB1*	Histocompatibility (minor) HB-1	EEKRGSL**H(Y)**VW	Hematopoietic, B-ALL, EBV-BLCLs	cDNA-expression cloning	NS cSNP	rs161557	[14-16]
HEATR1^{E}	B*0801	*HEATR1/1q43*	BAP28 protein	ISKERA**E(G)**AL	Ubiquitous	T cell clones	*In vitro* stimulation of naïve T cells	rs2275687	[17]
HER2_1170	A*0201	*ERBB2*/17q12	HER2 protein	**A(P)** in peptide	Ubiquitous	Reverse immunology	Exonic ns-SNP	N/A	[18]
HY-A1	A*0101	*DFFRY (USP9Y)/* Yq11	Ubiquitin specific peptidase 9	**IVDC*LTEMY** (*with cysteinylated cysteine)	Ubiquitous (hematopoietic, fibroblasts) AA	HPLC+MS	Polymorphism with chromosome X	N/A	[19,20]
HY-A2	A*0201	*SMCY(KDM5D)/* Yq11.1	Lysine (K)-specific demethylase 5D	**FIDSYIC*QV** (*with cysteinylated cysteine)	Ubiquitous (hematopoietic, fibroblasts)	HPLC+MS, revese immunology	Polymorphism with chromosome X	N/A	[21,22]
HY-B27	B*2705	*DDX3Y(DBY)/*Yq11	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3	**SRDSRGKPGY**	Hematopoietic	cDNA expression library	Polymorphism with chromosome X	N/A	[23]
HY-A33	A*3303	*TMSB4Y/*Yq11	Thymosin, beta 4	**EVLLRPGLHFR**	Ubiquitous	cDNA library	5' UTR of ORF and polymorphism with chromosome X	N/A	[24]
HY-B52	B*5201	*RPS4Y1/*Yp11.3	Ribosomal protein S4	**TIRYPDPVI**	Hematopoietic (leukocytes-PHA blasts, B-LCLs, B cells, breast carcinoma, hepatocellular carcinoma, colon adenocarcinoma, AML, ALL multiple myeloma) probably ubiquitous	cDNA expression cloning	Polymorphism with chromosome X	N/A	[25]
HY-B60	B60	*UTY/*Yq11.1	Ubiquitously transcribed tetratricopeptide repeat gene	**RESEESVSL**	Ubiquitous, hematopoietic cells, fibroblasts, AML, AA	cDNA-expression cloning	Polymorphism with chromosome X	N/A	[20]
HY-B7	B*0702	*SMCY(KDM5D)/* Yq11.1	Protein containing zinc finger domains	**SPSVDKARAEL**	Ubiquitous (hematopoietic cells) AA	HPLC+MS	Polymorphism with chromosome X	N/A	[26]
HY-B8	B8	*UTY/* Yq11.1	Ubiquitously transcribed tetratricopeptide repeat gene	**LPHNHTDL**	Ubiquitous, hematopoietic cells, fibroblasts, AML, AA	cDNA-expression cloning	Polymorphism with chromosome X	N/A	[27]
HY-DQ5	DDQB1*05	*DDX3Y(DBY)/*Yq11	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3	**HIENFSDIDMGE**	Ubiquitous (hematopoietic cells, fibroblasts) CML	cDNA-expression cloning	Polymorphism with chromosome X	N/A	[28,29]
HY-RPPS4Y	DR7	*RPS4Y*	Ribosomal protein S4	**TGKIINFIKFDTGNL**	Ubiquitous	T cell clones + reverse immunology	Polymorphism with chromosome X	N/A	[30]
HY-DRB1*1501	DRB1*1501	*DDX3Y(DBY)/*Yq11	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3	**SKGRYIPPHLR**	Ubiquitous (hematopoietic, fibroblasts)	T cell clones + reverse immunology	Polymorphism with chromosome X	N/A	[31]
HY-DRB3*0301	DRB3*0301	*RPS4Y1/*Yq11.3	Ribosomal protein S4, Protein synthesis	**VIKVNDTVQI**	Ubiquitous (hematopoietic cells, fibroblasts) AA	cDNA-expression cloning	Polymorphism with chromosome X	N/A	[12]
HY-UTY	A*2402	*UTY/*Yq11.1	Ubiquitously transcribed tetratricopeptide repeat gene	**YYNAFHWAI**	Ubiquitous, hematopoietic cells, fibroblasts, AML, AA	reverse immunology	Polymorphism with chromosome X	N/A	[32]

LAMA1	DRB*0301	*LAMA1/*6q21	Laminin alpha 1	LLILRAIP**K**GIRDKGAK	Ubiquitous	SNP genotyping + reverse immunology	Exonic ns-SNP	N/A	[33]
LB-ADIR-1^{F}	A*0201	*ADIR(TOR3A)/*1q25.2	torsin family 3, member A	SVAPALAL**F(S)**PA	Ubiquitous, highly expressed in activated HC including myeloma, select solid tumours	HPLC+MS +tetramer	s-SNP in normal ORF but exonic ns-SNP in ARF	rs2296377	[34]
LB-APOBEC3B-1K (no tetramer staining)	B*0702	*APOBEC3B/*22q13	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B	K(G)PQYHAEMCFL or K(G)PQYHAEMCF or K(G)PQYHAEMC	Hematopoietic	GWAs	Exonic ns-SNP	rs2076109	[35]
LB-ARHGDIB-1^{R}	B*0702	*ARHGDIB/*12p12.3	Rho GDP dissociation inhibitor (GDI) beta	LPRACW**R(P)**EA or LPRACW**R(P)**EAR or LPRACW**R(P)**EART	Hematopoietic	GWAs	Exonic ns SNP in ARF	rs4703	[35]
LB-BCAT2-1^{R}	B*0702	*BCAT2/*19q13	branched chain amino-acid transaminase 2, mitochondrial	QP**R(T)**RALLFVIL or QP**R(T)**RALLFVI or QP**R(T)**RALLFV or GSVQP**R(T)**RALL or GSVQP**R(T)**RAL	Ubiquitous	GWAs (no tetramer validation)	Exonic ns-SNP	rs11548193	[35]
LB-EBI3-1^{I} (no tetramer staining)	B*0702	*EBI3/*19p13.3	Epstein-Barr virus induced 3	RPRARYY**I(V)**QVA or RPRARYY**I(V)**QV or RPRARYY**I(V)**Q or AVRPRARYY**I(V)**	N/A	Whole genome association scanning	Exonic ns-SNP	rs4740	[35]
LB-ECGF-1^{H}	B*0702	*ECGF-1/*22q13	Thymidine phosphorylase	RP**H(R)**AIRRPLAL	Hematopoietic, low expression in other tissues	cDNA expression cloning	Exonic ns-SNP in an alternatively translated peptide, but syn SNP in normal ORF	N/A	[36]
LB-ERAP1-1R	B*0702	*ERAP1/*5q15	endoplasmic reticulum aminopeptidase 1	HP**R(P)**QEQIALLA or HP**R(P)**QEQIALL or HP**R(P)**QEQIAL	Ubiquitous	GWAs	Exonic ns-SNP	rs26653	[35]
LB-GEMIN4-1V	B*0702	*GEMIN4/*17p13	gem (nuclear organelle) associated protein 4	FPALRFVE**V(G)**	Hematopoietic	GWAs	Exonic ns-SNP	rs4968104	[35]
LB-LY75-1K	DRB1*1301	*LY75/2q24*	Scavenger receptor	Core epitope GITYRN**K(N)** SLM	Hematopoietic, dendritic cells, cortical thymic epithelial cells, AML	Bacterial cDNA library	Exonic ns-SNP	rs12692566	[37]
LB-MR1-1R	DRB3*0202	*MR1/*1q25.3	major histocompatibility complex, class I-related	YFRLGVSDPI**R(H)**G	Hematopoietic, AML	Bacterial cDNA library	Exonic ns-SNP	rs2236410	[37,38]
LB-MTHFD1-1Q	DRB1*0301	*MTHFD1/*14q24	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1	Core epitope SSIIAD**Q(R)** IALKL	Hematopoietic, AML	Bacterial cDNA library	Exonic ns-SNP	rs2236225	[37,38]
LB-NISCH-1A	A*0201	*NISCH/*3p21.1	Nischarin	N/A	Ubiquitous including B-LCL	MS+reverse immunology +tetramers	Exonic ns-SNP	N/A	[39]
LB-NUP133-1R	B*4001	*NUP133/*1q42.13	nucleoporin 133kDa	SEDLILC**R(Q)**L	Hematopoietic	GWAs	Exonic ns-SNP	rs1065674	[40]
LB-P14K2B-1S	DQB1*0603	*PI4K2B/*4p15.22	Phosphatidylinositol 4-kinase type II β gene	Core epitope SRSS**S(P)** AELDRSR	Ubiquitous	Bacterial cDNA library	Exonic ns-SNP	rs313549	[38]

LB-PDCD11-1F	B*0702	*PDCD11*/10q24.33	programmed cell death 11	GPDSSKT**F(L)**LCL or GPDSSKT**F(L)**L	Ubiquitous	GWAs	Exonic ns-SNP	rs2986014	[35]
LB-PRCP-1D	A*0201	*PRCP*/11q14	prolylcarboxypeptidase (angiotensinase C)	FMWDVAE**D(E)**LKA or FMWDVAE**D(E)**L	Ubiquitous	GWAs	Exonic ns-SNP	rs2298668	[35]
LB-PTK2B-1T	DRB3*0101	*PTK2B*/8p21.1	protein tyrosine kinase 2 beta	Core epitope VYMND**T(K)** SPL	Hematopoietic, AML	Bacterial cDNA library	Exonic ns-SNP	rs751019	[37,38]
LB-SON-1R	B*4001	*SON*/21q22.11	SON DNA binding protein	SETKQ**R(Q)**TVL	Hematopoietic	GWAs	Exonic ns-SNP	rs13047599	[40]
LB-SSR1-1S	A*0201	*SSR1*/6p24.3	signal sequence receptor, alpha	**S(L)**LAVAQDLT	Hematopoietic	GWAs	Exonic ns-SNP	rs10004	[35]
LB-SWAP70-1Q	B*4001	*SWAP70*/11p15	SWAP switching B-cell complex 70kDa subunit	MEQLE**Q(E)**LEL	Hematopoietic	GWAs	Exonic ns-SNP	rs415895	[40]
LB-TRIP10-EPC	B*4001	*TRIP10*/19p13.3	thyroid hormone receptor interactor 10	G**EP**QDL**C**TL	Ubiquitous	GWAs	3 SNPs in the 3' UTR, ns-SNPs in ARF	rs1049229 rs1049230 rs1049232	[40]
LB-WNK1-1I	A*0201	*WNK1*/12p13.3	WNK lysine deficient protein kinase 1	RTLSPE**I(M)**ITV or TLSPE**I(M)**ITV	Ubiquitous	GWAs	Exonic ns-SNP	rs12828016	[35]
LRH-1	B*0702	*P2X5*/17p13.2	purinergic receptor P2X, ligand-gated ion channel, 5	TP**NQRQNVC**	Hematopoietic (lymphoid)	Genetic-linkage analysis	Single nucleotide deletion causing a frameshift	rs5818907	[41]
PANE1 (CTL-7A7)	A*0301	*PANE1/ CENPM*/22q13.2	Centromere protein M	R(*)VWDLPGVLK	Hematopoietic (resting B cells and B-LCLs)	HPLC+MS	Exonic ns-SNP causing a stop codon	-----	[42]
SLC1A5[A]	B*4002	*SLC1A5*/19q13.3	solute carrier family 1 (neutral amino acid transporter), member 5	AE**A(P)**TANGGLAL	Hematopoietic	GWAs	Exonic ns-SNP	rs3027956	[43]
SLC19A1	DRB1*1501	*SLC19A1*/21q22.3	Transmembrane protein, transport of folate compounds	9 15-mer overlapping peptides including sequence **R(H)**LVCYLC	Ubiquitous, hematopoietic, lung, liver	Genome-wide zygosity-genotype correlation analysis	NS cSNP	rs1051266	[44]
SP110[R] (HwA-9)	A*0301	*SP110*/2q37.1	SP110 nuclear body protein	SLPR(G)GTSTPK **R(G)**GTSTPK	Hematopoietic, IFN-gamma inducible	cDNA library expression, HPLC+MS	Exonic ns-SNP resulting in alternate splicing (spliced together in a non-contiguos order by the proteasome)	rs1365776	[45]
UGT2B17	A*2902 B*4403 B*4402	*UGT2B17*/4q13.2	UDP-glucuronosyltransferase 2 family polypeptide B17	**AELLNIPFLY**	Ubiquitous, liver, colon, APCs, DCs, B-LCLs, hematopoietic? AML	cDNA-expression cloning, WGAs	Homozygous gene deletion	CNV	[46,47]
UGT2B17	A*0206	*UGT2B17*/4q13.2	UDP-glucuronosyltransferase 2 family polypeptide B17	**CVATMIFMI**	Ubiquitous	GWAs	Homozygous gene deletion	CNV	[43]
UTA2-1	A*0201	*C12orf35*/12p11.21	KIAA1551	QL**L(P)**NSVLTL	hematopoietic MM	genome-wide zygosity-genotype correlation analysis	NS cSNP	rs2166807	[48]
ZAPHIR	B7	*ZNF419*/19q13.43	zinc finger protein 419	IPR**D(G)**SWWVEL	RCC, B-LCLs	GWAs	Intronic SNP, alternative splicing	N/A	[49]
ZNF544	DRB1*0301	*ZNF544*/19q13.43	zinc finger protein 544	KQNSAFIN**D**EKNGADGK	Ubiquitous	SNP genotyping + reverse immunology	Exonic ns-SNP	N/A	[33]

4.6 References

1 Goulmy, E. *et al.* (1976) Alloimmunity to human H-Y. *Lancet* 2, 1206

2 Goulmy, E. *et al.* (1983) A minor transplantation antigen detected by MHC-restricted cytotoxic T lymphocytes during graft-versus-host disease. *Nature* 302, 159–161

3 Warren, E.H. *et al.* (2012) Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. *Blood* 120, 2796–2806

4 Roopenian, D.C. (1992) What are minor histocompatibility loci? A new look at an old question. *Immunology Today* 13, 7–10

5 Bleakley, M. and Riddell, S.R. (2011) Exploiting T cells specific for human minor histocompatibility antigens for therapy of leukemia. *Immunology and Cell Biology* 89, 396–407

6 Bleakley, M. and Riddell, S.R. (2004) Molecules and mechanisms of the graft-versus-leukaemia effect. *Nature Reviews* Cancer 4, 371–380

7 Roopenian, D. *et al.* (2002) The immunogenomics of minor histocompatibility antigens. *Immunol Rev* 190, 86–94

8 de Rijke, B. *et al.* (2005) A frameshift polymorphism in P2X5 elicits an allogeneic cytotoxic T lymphocyte response associated with remission of chronic myeloid leukemia. *Journal of Clinical Investigation* 115, 3506–3516

9 Brickner, A.G. *et al.* (2006) The PANE1 gene encodes a novel human minor histocompatibility antigen that is selectively expressed in B-lymphoid cells and B-CLL. *Blood* 107, 3779–3786

10 Murata, M. *et al.* (2003) A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. *The Journal of Experimental Medicine.* 197, 1279–1289

11 Brickner, A.G. *et al.* (2001) The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. *The Journal of Experimental Medicine.* 193, 195–206

12 Spierings, E. *et al.* (2003) The minor histocompatibility antigen HA-3 arises from differential proteasome-mediated cleavage of the lymphoid blast crisis (Lbc) oncoprotein. *Blood* 102, 621–629

13 Haan, den, J.M. *et al.* (1998) The minor histocompatibility antigen HA-1: a diallelic gene with a single amino acid polymorphism. *Science* 279, 1054–1057

14 Nicholls, S. *et al.* (2009) Secondary anchor polymorphism in the HA-1 minor histocompatibility antigen critically affects MHC stability and TCR recognition. *PNAS* 106, 3889–3894

15 Torikai, H. *et al.* (2006) The human cathepsin H gene encodes two novel minor histocompatibility antigen epitopes restricted by HLA-A*3101 and -A*3303.

*British Journ*al of Haematology 134, 406–416

16 Spierings, E. *et al.* (2009) Steric hindrance and fast dissociation explain the lack of immunogenicity of the minor histocompatibility HA-1Arg Null allele. *The. Journal of Immunology* 182, 4809–4816

17 Slager, E.H. *et al.* (2006) Identification of the angiogenic endothelial-cell growth factor-1/thymidine phosphorylase as a potential target for immunotherapy of cancer. *Blood* 107, 4954–4960

18 Van Bergen, C.A.M. *et al.* (2007) Multiple myeloma-reactive T cells recognize an activation-induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. *Blood* 109, 4089–4096

19 Ofran, Y. *et al.* (2010) Diverse patterns of T-cell response against multiple newly identified human Y chromosome-encoded minor histocompatibility epitopes. *Clinical Cancer* Research 16, 1642–1651

20 Terakura, S. *et al.* (2007) A single minor histocompatibility antigen encoded by UGT2B17 and presented by human leukocyte antigen-A*2902 and -B*4403. *Transplantation* 83, 1242–1248

21 Kamei, M. *et al.* (2009) HapMap scanning of novel human minor histocompatibility antigens. *Blood* 113, 5041–5048

22 Griffioen, M. *et al.* (2012) Identification of 4 novel HLA-B*40:01 restricted minor histocompatibility antigens and their potential as targets for graft-versus-leukemia reactivity. *Haematologica* 97, 1196–1204

23 Torikai, H. *et al.* (2004) A novel HLA-A*3303-restricted minor histocompatibility antigen encoded by an unconventional open reading frame of human TMSB4Y gene. *The.Journal of Immunology* 173, 7046–7054

24 Tykodi, S.S. *et al.* (2008) C19orf48 encodes a minor histocompatibility antigen recognized by CD8+ cytotoxic T cells from renal cell carcinoma patients. *Clinical Cancer* Research 14, 5260–5269

25 Kawase, T. *et al.* (2007) Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. *Blood* 110, 1055–1063

26 Broen, K. *et al.* (2011) A polymorphism in the splice donor site of ZNF419 results in the novel renal cell carcinoma-associated minor histocompatibility antigen ZAPHIR. *PLoS ONE* 6, e21699

27 Dolstra, H. *et al.* (1999) A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. *The Journal of Experimental Medicine.* 189, 301–308

28 Haan, den, J.M. *et al.* (1995) Identification of a graft versus host disease-associated human minor histocompatibility antigen. *Science* 268, 1476–1480

29 Pierce, R.A. *et al.* (2001) The HA-2 minor histocompatibility antigen is derived from a diallelic gene encoding a novel human class I myosin protein. *The.Jour-*

nal of Immunology 167, 3223–3230
30 Warren, E.H. *et al.* (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* 313, 1444–1447
31 Spierings, E. *et al.* (2004) Minor histocompatibility antigens--big in tumour therapy. *Trends in Immunology* 25, 56–60
32 Brickner, A.G. (2007) Mechanisms of minor histocompatibility antigen immunogenicity: the role of infinitesimal versus structurally profound polymorphisms. *Immunologic Research* 36, 33–41
33 Dolstra, H. *et al.* (2002) Bi-directional allelic recognition of the human minor histocompatibility antigen HB-1 by cytotoxic T lymphocytes. *European Journal of Immunology* 32, 2748–2758
34 Yadav, R. *et al.* (2003) The H4b minor histocompatibility antigen is caused by a combination of genetically determined and posttranslational modifications. *The.Journal of Immunology* 170, 5133–5142
35 Meadows, L. *et al.* (1997) The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. *Immunity* 6, 273–281
36 Pierce, R.A. *et al.* (1999) Cutting edge: the HLA-A*0101-restricted HY minor histocompatibility antigen originates from DFFRY and contains a cysteinylated cysteine residue as identified by a novel mass spectrometric technique. *The. Journal of Immunology* 163, 6360–6364
37 Falkenburg, J.H.F. *et al.* (2003) Minor histocompatibility antigens in human stem cell transplantation. *Experimental Hematology* 31, 743–751
38 Hambach, L. and Goulmy, E. (2005) Immunotherapy of cancer through targeting of minor histocompatibility antigens. *Current Opinion in Immunology* 17, 202–210
39 Turpeinen, H. *et al.* (2013) Minor histocompatibility antigens as determinants for graft-versus-host disease after allogeneic haematopoietic stem cell transplantation. *International Journal of Immunogenetics* DOI: 10.1111/iji.12051
40 Bleakley, M. *et al.* (2010) Leukemia-associated minor histocompatibility antigen discovery using T-cell clones isolated by in vitro stimulation of naive CD8+ T cells. *Blood* 115, 4923–4933
41 Spierings, E. *et al.* (2013) Multicenter Analyses Demonstrate Significant Clinical Effects of Minor Histocompatibility Antigens on GvHD and GvL after HLA-Matched Related and Unrelated Hematopoietic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation* 19, 1244–1253
42 Falkenburg, J.H.F. and Willemze, R. (2004) Minor histocompatibility antigens as targets of cellular immunotherapy in leukaemia. *Best Practice and Reseacrh Clinical Haematology* 17, 415–425

43 MacDonald, K.P. *et al.* (2013) Biology of graft-versus-host responses: recent insights. *Biology of Blood and Marrow Transplantation* 19, S10–4

44 Mutis, T. *et al.* (1999) Tetrameric HLA class I-minor histocompatibility antigen peptide complexes demonstrate minor histocompatibility antigen-specific cytotoxic T lymphocytes in patients with graft-versus-host disease. *Nature Medicine* 5, 839–842

45 Gratwohl, A. *et al.* (2001) Female donors influence transplant-related mortality and relapse incidence in male recipients of sibling blood and marrow transplants. *The Hematology Journal* 2, 363–370

46 Randolph, S.S.B. *et al.* (2004) Female donors contribute to a selective graft-versus-leukemia effect in male recipients of HLA-matched, related hematopoietic stem cell transplants. *Blood* 103, 347–352

47 Stern, M. *et al.* (2008) Female-versus-male alloreactivity as a model for minor histocompatibility antigens in hematopoietic stem cell transplantation. *American Journa*l of Transplantation 8, 2149–2157

48 Takami, A. *et al.* (2004) Expansion and activation of minor histocompatibility antigen HY-specific T cells associated with graft-versus-leukemia response. *Bone Marrow Transplant* 34, 703–709

49 Hobo, W. *et al.* (2013) Association of disparities in known minor histocompatibility antigens with relapse-free survival and graft-versus-host disease after allogeneic stem cell transplantation. *Biology of Blood and Marrow Transplantation* 19, 274–282

50 Dzierzak-Mietla, M. *et al.* (2012) Occurrence and Impact of Minor Histocompatibility Antigens' Disparities on Outcomes of Hematopoietic Stem Cell Transplantation from HLA-Matched Sibling Donors. *Bone Marrow Research* 2012, 257086

51 Linscheid, C. and Petroff, M.G. (2013) Minor histocompatibility antigens and the maternal immune response to the fetus during pregnancy. *American Journal of* Reproductive Immunology 69, 304–314

52 Verdijk, R.M. *et al.* (2004) Pregnancy induces minor histocompatibility antigen-specific cytotoxic T cells: implications for stem cell transplantation and immunotherapy. *Blood* 103, 1961–1964

53 van Halteren, A.G.S. *et al.* (2009) Naturally acquired tolerance and sensitization to minor histocompatibility antigens in healthy family members. *Blood* 114, 2263–2272

54 Feng, X. *et al.* (2008) Targeting minor histocompatibility antigens in graft versus tumor or graft versus leukemia responses. *Trends in Immunology* 29, 624–632

55 Akatsuka, Y. *et al.* (2007) Minor histocompatibility antigens as targets for im-

munotherapy using allogeneic immune reactions. *Cancer Science* 98, 1139–1146

56 Vincent, K. *et al.* (2011) Next-generation leukemia immunotherapy. *Blood* 118, 2951–2959

57 Warren, E.H. *et al.* (1998) Cytotoxic T-lymphocyte-defined human minor histocompatibility antigens with a restricted tissue distribution. *Blood* 91, 2197–2207

58 Akatsuka, Y. *et al.* (2003) Identification of a polymorphic gene, BCL2A1, encoding two novel hematopoietic lineage-specific minor histocompatibility antigens. *The Journal of Experimental Medicine.* 197, 1489–1500

59 Bonnet, D. *et al.* (1999) CD8(+) minor histocompatibility antigen-specific cytotoxic T lymphocyte clones eliminate human acute myeloid leukemia stem cells. *Proc.Natl.Acad.Sci.U.S.A* 96, 8639–8644

60 Marijt, W.A.E. *et al.* (2003) Hematopoiesis-restricted minor histocompatibility antigens HA-1- or HA-2-specific T cells can induce complete remissions of relapsed leukemia. *Proc.Natl.Acad.Sci.U.S.A* 100, 2742–2747

61 Lin, M.T. *et al.* (2001) Absence of statistically significant correlation between disparity for the minor histocompatibility antigen-HA-1 and outcome after allogeneic hematopoietic cell transplantation. *Blood* 98, 3172–3173

62 Spellman, S. *et al.* (2009) Effects of mismatching for minor histocompatibility antigens on clinical outcomes in HLA-matched, unrelated hematopoietic stem cell transplants. *Biology of Blood and Marrow Transplantation* 15, 856–863

63 Larsen, M.E. *et al.* (2010) Degree of Predicted Minor Histocompatibility Antigen Mismatch Correlates with Poorer Clinical Outcomes in Nonmyeloablative Allogeneic Hematopoietic Cell Transplantation. *Biology of Blood and Marrow Transplantation* 16, 1370–1381

64 Katagiri, T. *et al.* (2006) Mismatch of minor histocompatibility antigen contributes to a graft-versus-leukemia effect rather than to acute GVHD, resulting in long-term survival after HLA-identical stem cell transplantation in Japan. *Bone Marrow Transplant* 38, 681–686

65 Torikai, H. *et al.* (2007) The HLA-A*0201-restricted minor histocompatibility antigen HA-1H peptide can also be presented by another HLA-A2 subtype, A*0206. *Bone Marrow Transplant* 40, 165–174

66 Tait, B.D. *et al.* (2001) Clinical relevance of the minor histocompatibility antigen HA-1 in allogeneic bone marrow transplantation between HLA identical siblings. *Transplantation Proceedings* 33, 1760–1761

67 Jervis, S. *et al.* (2013) Increased severity of acute graft versus host disease as a result of differential expression following a homozygous gene deletion. *International Journal of Immunogenetics* 40, 116–119

68 McCarroll, S.A. *et al.* (2009) Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease. *Nature Genetics* 41, 1341–1344

69 Hambach, L. *et al.* (2006) Human cytotoxic T lymphocytes specific for a single minor histocompatibility antigen HA-1 are effective against human lymphoblastic leukaemia in NOD/scid mice. *Leukemia* 20, 371–374

70 Spaapen, R.M. *et al.* (2010) Eradication of medullary multiple myeloma by CD4+ cytotoxic human T lymphocytes directed at a single minor histocompatibility antigen. *Clinical Cancer* Research 16, 5481–5488

71 Pion, S. *et al.* (1995) Immunodominant minor histocompatibility antigens expressed by mouse leukemic cells can serve as effective targets for T cell immunotherapy. *Journal of Clinical Investigation* 95, 1561–1568

72 Perreault, C. *et al.* (1998) Immunodominant minor histocompatibility antigens: the major ones. *Immunology Today* 19, 69–74

73 Fontaine, P. *et al.* (2001) Adoptive transfer of minor histocompatibility antigen-specific T lymphocytes eradicates leukemia cells without causing graft-versus-host disease. *Nature Medicine* 7, 789–794

74 Meunier, M.-C. *et al.* (2005) T cells targeted against a single minor histocompatibility antigen can cure solid tumors. *Nature Medicine* 11, 1222–1229

75 Goulmy, E. (2006) Minor histocompatibility antigens: from transplantation problems to therapy of cancer. *Human Immunology* 67, 433–438

76 Oostvogels, R. *et al.* (2013) Towards effective and safe immunotherapy after allogeneic stem cell transplantation: identification of hematopoietic-specific minor histocompatibility antigen UTA2-1. *Leukemia* 27, 642–649

77 Warren, E.H. *et al.* (2010) Therapy of relapsed leukemia after allogeneic hematopoietic cell transplantation with T cells specific for minor histocompatibility antigens. *Blood* 115, 3869–3878

78 Spierings, E. *et al.* (2007) Phenotype Frequencies of Autosomal Minor Histocompatibility Antigens Display Significant Differences among Populations. *PLoS Genetics* 3, e103

79 Warren, E.H. *et al.* (2000) The human UTY gene encodes a novel HLA-B8-restricted H-Y antigen. *The.Journal of Immunology* 164, 2807–2814

80 Vogt, M.H. *et al.* (2000) UTY gene codes for an HLA-B60-restricted human male-specific minor histocompatibility antigen involved in stem cell graft rejection: characterization of the critical polymorphic amino acid residues for T-cell recognition. *Blood* 96, 3126–3132

81 Kawase, T. *et al.* (2008) Identification of human minor histocompatibility antigens based on genetic association with highly parallel genotyping of pooled DNA. *Blood* 111, 3286–3294

82 Griffioen, M. *et al.* (2008) Identification of phosphatidylinositol 4-kinase type II beta as HLA class II-restricted target in graft versus leukemia reactivity. *Proc. Natl.Acad.Sci.U.S.A* 105, 3837–3842

83 Voigt, A. *et al.* (2007) 20S proteasome-dependent generation of an IEpp89 murine cytomegalovirus-derived H-2Ld epitope from a recombinant protein. *Biochemical and Biophysical Research Communications* 355, 549–554

84 Wang, W. *et al.* (1995) Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science* 269, 1588–1590

85 Rotzschke, O. *et al.* (1990) Characterization of naturally occurring minor histocompatibility peptides including H-4 and H-Y. *Science* 249, 283–287

86 Warren, E.H. *et al.* (2002) Feasibility of using genetic linkage analysis to identify the genes encoding T cell-defined minor histocompatibility antigens. *Tissue Antigens* 59, 293–303

87 Akatsuka, Y. *et al.* (2003) Disparity for a newly identified minor histocompatibility antigen, HA-8, correlates with acute graft-versus-host disease after haematopoietic stem cell transplantation from an HLA-identical sibling. *British Journal of Haematology* 123, 671–675

88 Van Bergen, C.A.M. *et al.* (2010) High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. *Cancer Research* 70, 9073–9083

89 Spaapen, R.M. *et al.* (2008) Toward targeting B cell cancers with CD4+ CTLs: identification of a CD19-encoded minor histocompatibility antigen using a novel genome-wide analysis. *The Journal of Experimental Medicine.* 205, 2863–2872

90 Dolstra, H. *et al.* (1997) Recognition of a B cell leukemia-associated minor histocompatibility antigen by CTL. *The.Journal of Immunology* 158, 560–565

91 Stumpf, A.N. *et al.* (2009) Identification of 4 new HLA-DR-restricted minor histocompatibility antigens as hematopoietic targets in antitumor immunity. *Blood* 114, 3684–3692

92 Spaapen, R.M. *et al.* (2009) Rapid identification of clinical relevant minor histocompatibility antigens via genome-wide zygosity-genotype correlation analysis. *Clinical Cancer Research* 15, 7137–7143

93 Stone, B. *et al.* (2012) Identification of novel HLA class II target epitopes for generation of donor-specific T regulatory cells. *Clinical Imm*unology 145, 153–160

94 Consortium, T.I.H.3. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58

95 Popovic, J. *et al.* (2011) The only proposed T-cell epitope derived from the TEL-AML1 translocation is not naturally processed. *Blood* 118, 946–954

96 Sewell, A.K. (2012) Why must T cells be cross-reactive? *Nature Reiviews Immunology* 12, 669–677

97 Laugel, B. *et al.* (2007) Different T cell receptor affinity thresholds and CD8 coreceptor dependence govern cytotoxic T lymphocyte activation and tetramer binding properties. *The Journal of Biological Chemistry*. 282, 23799–23810

98 Chattopadhyay, P.K. *et al.* (2008) Techniques to improve the direct ex vivo detection of low frequency antigen-specific CD8 +T cells with peptide-major histocompatibility complex class I tetramers. *Cytometry* 73, 1001–1009

99 Hombrink, P. *et al.* (2011) High-throughput identification of potential minor histocompatibility antigens by MHC tetramer-based screening: feasibility and limitations. *PLoS ONE* 6, e22523

100 Feldhahn, M. *et al.* (2012) miHA-Match: computational detection of tissue-specific minor histocompatibility antigens. *Journal of Immunological Methods* 386, 94–100

101 Wenandy, L. *et al.* (2009) The 1170 A-P single-nucleotide polymorphism (SNP) in the Her-2/neu protein (HER2) as a minor histocompatibility antigen (mHag). *Leukemia* 23, 1926–1929

102 Mortensen, B.K. *et al.* (2012) Identification of a novel UTY-encoded minor histocompatibility antigen. *Scandinavian Journal of Immunology* 76, 141–150

103 Robbins, P.F. *et al.* (2013) Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nature Medicine* 19, 747–752

104 Yewdell, J.W. *et al.* (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature Reviews Immunology* 3, 952–961

105 Hassan, C. *et al.* (2013) The Human Leukocyte Antigen-presented Ligandome of B Lymphocytes. *Molecular and Cellular Proteomics* 12, 1829–1843

106 Nijveen, H. *et al.* (2010) HSPVdb—the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands. *Immunogenetics* 63, 143–153

107 Hombrink, P. *et al.* (2013) Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. *The.Journal of Immunology* 190, 3869–3877

CHAPTER 5

5 Impact of Genomic Polymorphisms on the Repertoire of Human MHC Class I-Associated Peptides

Diana Paola Granados[1,2,†], Dev Sriranganadane[1,3,†], Tariq Daouda[1,4,†], Antoine Zieger[1,4], Céline M. Laumont[1,2], Olivier Caron-Lizotte[1], Geneviève Boucher[1], Marie-Pierre Hardy[1], Patrick Gendron[1], Caroline Côté[1,2], Sébastien Lemieux[1,4], Pierre Thibault[1,3,*] & Claude Perreault[1,2,*]

[†] These authors contributed equally to this work

[1]Institute for Research in Immunology and Cancer, [2]Department of Medicine, [3]Department of Chemistry, and [4]Department of Informatics and Operational Research, Université de Montréal, 2950 Chemin de Polytechnique, Montréal, Québec, Canada H3T 1J4

[*] Correspondence should be addressed to :
Pierre Thibault or Claude Perreault

This article was submitted for publication to **Nature Communications** on September 19, 2013. The accepted (March 10 2014) and published (9 April 2014) version can be found on http://www.nature.com/ncomms (doi:10.1038/ncomms4600)

5.1 Authors' contributions

DPG*: Study design, conducting experiments (except for cytotoxicity assays and MS), data analysis, integration and discussion, preparation of figures and supplementary data, writing of first draft of the manuscript

DS*: Study design, conducting and analysis of MS experiments, data analysis and discussion, preparation of figures and supplementary data, writing of first draft of the manuscript

TD*: Study design, *In silico* translation of exome and transcriptome sequences to generate personalized proteomes (figure 1b), filtering of MHC I-peptides derived from polymorphic regions (supplementary data 4) and bootstrapping analysis (figure 6c)

AZ: Development of filtering method for discrimination of MHC I peptides from contaminants (figure 1c)

CML: Data analysis and general discussion

OCL: Development of bioinformatic tools for MS analysis

BG: Preparation of circos plot figures (figures 2 and 6a) and estimation of sequencing statistics (supplementary data 1)

MPH: Conducting cytotoxicity experiments (figure 5)

PG: Mapping of sequencing reads to reference genome (supplementary data 1)

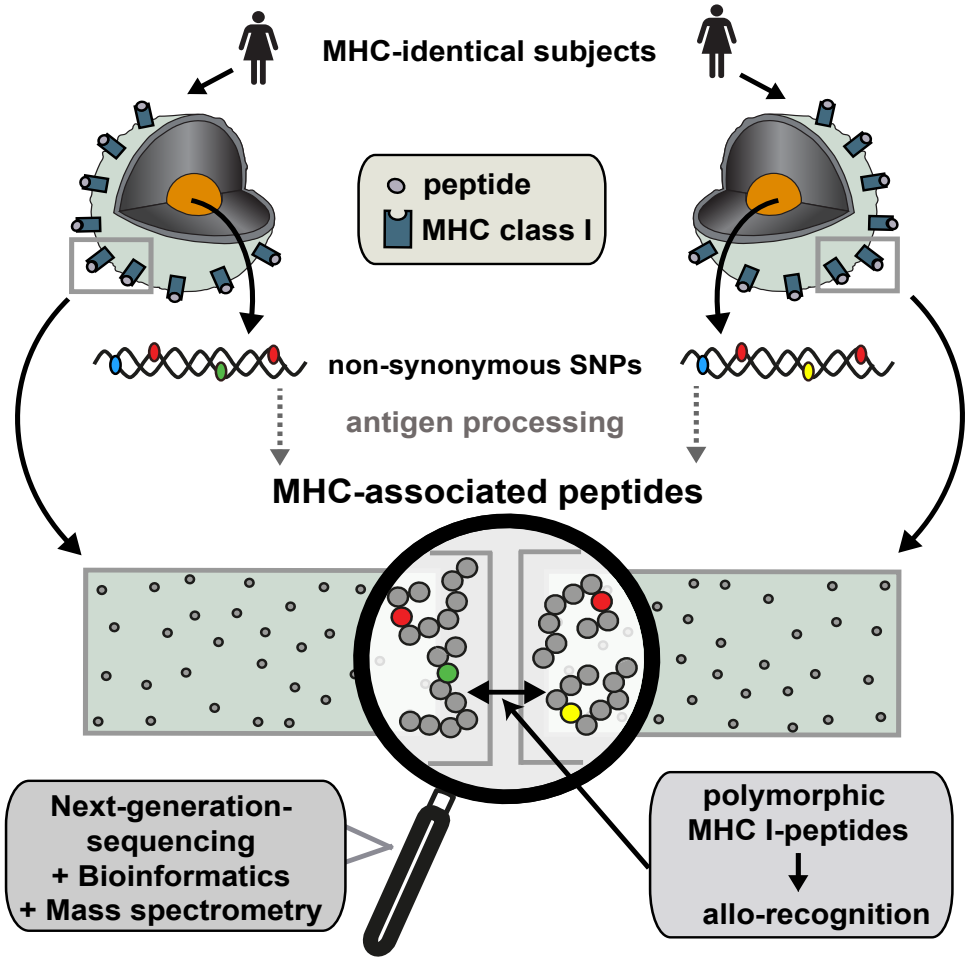
CC: Technical assistance in cell culture

SL: Study design, development of bioinformatics tools, statistical analysis and general discussion

PT: Study design, data analysis, general discussion and manuscript writing

CP: Study design, data analysis, general discussion and manuscript writing

5.2 Graphical abstract



5.3 Abstract

We developed a novel approach combining next-generation sequencing, bioinformatics and mass spectrometry to assess the impact of non-MHC polymorphisms on the repertoire of MHC I-associated peptides (MIPs). We compared the genomic landscape of MIPs eluted from B lymphoblasts of two non-twin MHC-identical siblings and discovered that 0.5% of non-synonymous single nucleotide variations were represented in the MIP repertoire. We identified 34 polymorphic MIPs which were encoded by biallelic loci and behaved as dominant or recessive traits. We determined that, at the population level, at least 536 polymorphic MIPs can be presented by the five HLA class I allotypes expressed by our subjects and that 12% of the MIP-coding exome is polymorphic. Our method provides fundamental insights into the relation between the genomic self and the immune self and accelerates the discovery of polymorphic MIPs (also known as minor histocompatibility antigens), which play a major role in allorecognition.

5.4 Introduction

Multicellular organisms distinguish self and genetically distinct members of the same species (nonself) based on recognition of polymorphic cell surface molecules referred to as "histocompatibility antigens"1. The capacity to distinguish self from allogeneic histo-incompatible "nonself" (allorecognition) is critical for multicellular life[2-4]. In *Botryllus schlosseri*, a member of the urochordates (the closest living sister group of vertebrates), self is defined by a single polymorphic gene, *BHF*5. The definition of self is more complex in vertebrates. Thus, the TCR of classic adaptive CD8 T cells recognizes MHC class I-associated peptides (MIPs), and the ensemble of MIPs presented on the surface of a cell (the "immunopeptidome") establishes its immunologic identity[6]. CD8 T cells are eminently self-referential and highly discriminant: they are selected on self-MIPs, sustained by self-MIPs, and must swiftly react when confronted with nonself MIPs interspersed in a sea of self-MIPs[7, 8]. Understanding the molecular definition of self for CD8 T cells has been made possible by high-throughput mass spectrometry (MS) analyses of MIPs[9-15]. Progress in this field has been heralded by the development of MS instruments whose sensitivity, dynamic range and mass accuracy are orders of magnitude superior to those of analyzers available a decade ago[16]. The immunopeptidome is shaped by numerous co- and post-translational events[9, 17, 18]. Accordingly, high-throughput MS studies have revealed that the immunopeptidome is highly complex and that its composition (i.e., the source of MIPs) cannot be inferred solely from transcript or protein abundance[10, 12, 15, 19-21].

The MHC I region contains two major classes of genes: modern classical MHC Ia genes (e.g., *HLA-A*, *HLA-B* and *HLA-C* in humans) and more ancient MHC Ib genes (e.g., *HLA-E* and *HLA-G*). MHC Ia molecules play a dominant role in adaptive immunity. They bind MIPs and are encoded by the most polymorphic genes known[22, 23]. Since MHC Ia allotypes display distinct peptide binding motifs, the HLA genotype has a major impact on the MIP repertoire[24]. Notably, almost all genetic polymorphisms in HLA Ia alleles are located in exons 2 and 3, which encode the MIP-binding pocket. These exons contain less than one millionth of the 3.3×109 bases in the haploid human genome[25]. Besides, the 1000 Genomes Project Consortium has identified 38 million single nucleotide

polymorphisms (SNP), 1.4 million short insertions and deletions, after comprehensive studies on 1,092 subjects[23]. This raises the fundamental question: what might be the impact of the numerous polymorphisms outside of the MHC on the MIP repertoire? In other words, to what extent do genomic polymorphisms translate into differences in the immunopeptidome?

Several MIPs have been found to derive from polymorphic genomic regions[26, 27]. For historical reasons, these polymorphic MIPs are referred to as minor histocompatibility antigens (MiHAs). MiHAs are a consequence of genetic variations that hinder MIP generation (e.g., gene deletion) or the structure of a MIP [e.g., high frequency non-synonymous SNPs (ns-SNPs) or less-well characterized non-synonymous single nucleotide variations (ns-SNVs)][25, 28, 29]. MiHAs are generally defined according to three criteria: they are present in some but not in all subjects bearing a given HLA allele, their presence/absence is linked to a well-defined genetic polymorphism, and they can elicit allo-immune T-cell responses[25, 28, 29]. Three decades of research have led to the discovery of about 35 human MiHAs encoded by autosomes and presented by HLA class I molecules[25, 29]. The discovery of each MiHA has been a major endeavor, if not a technical tour de force[30-35]. However, due to the lack of a suitable systems level approach, we ignore the global impact of non-MHC genomic polymorphisms on the immunopeptidome (i.e., what proportion of MIPs are MiHAs). Based on various theoretical premises, it has been speculated that the number of MiHAs expressed by an individual might be very low (less than ten) or very high (greater than 1,000)[25, 27]. In addition to its conceptual importance, the impact of genetic polymorphisms on the immunopeptidome is of considerable medical relevance because MiHAs are the targets of three allo-immune processes: graft rejection, graft-versus-host disease and graft-versus-tumor reaction[25, 36-41].

Systems-level molecular definition of the immunopeptidome can be achieved only by MS studies. However, since current MS approaches cannot reliably detect polymorphic peptides, they are inadequate for MiHA discovery[42]. Furthermore, since several steps of MIP processing cannot be modeled with available algorithms [e.g., formation of defective ribosomal products[43]], MiHA identification using prediction tools is a daunting task fraught with high false discovery rates[42]. To resolve this conundrum, we have developed a genoproteomic

strategy that hinges on a combination of next-generation sequencing and high-throughput MS peptide identification. Our personalized platform provides unprecedented insights into the genomic landscape of human MIPs and enables high-throughput identification of MiHAs and of their underlying genomic polymorphisms.

5.5 Results

5.5.1 Development of a combined MS, next-generation sequencing and bioinformatics approach for the identification of MIPs

To evaluate the impact of non-HLA genetic polymorphisms on the MIP repertoire, we analyzed the immunopeptidome of Epstein-Barr virus-transformed B lymphoblastoid cell lines (B-LCLs) from two non-twin HLA-identical siblings (Fig. 1a). The success of our endeavor hinged on two factors: the need to reliably identify MIPs encoded by polymorphic genomic regions and to maximize the coverage of the immunopeptidome (the number of unique MIPs identified).

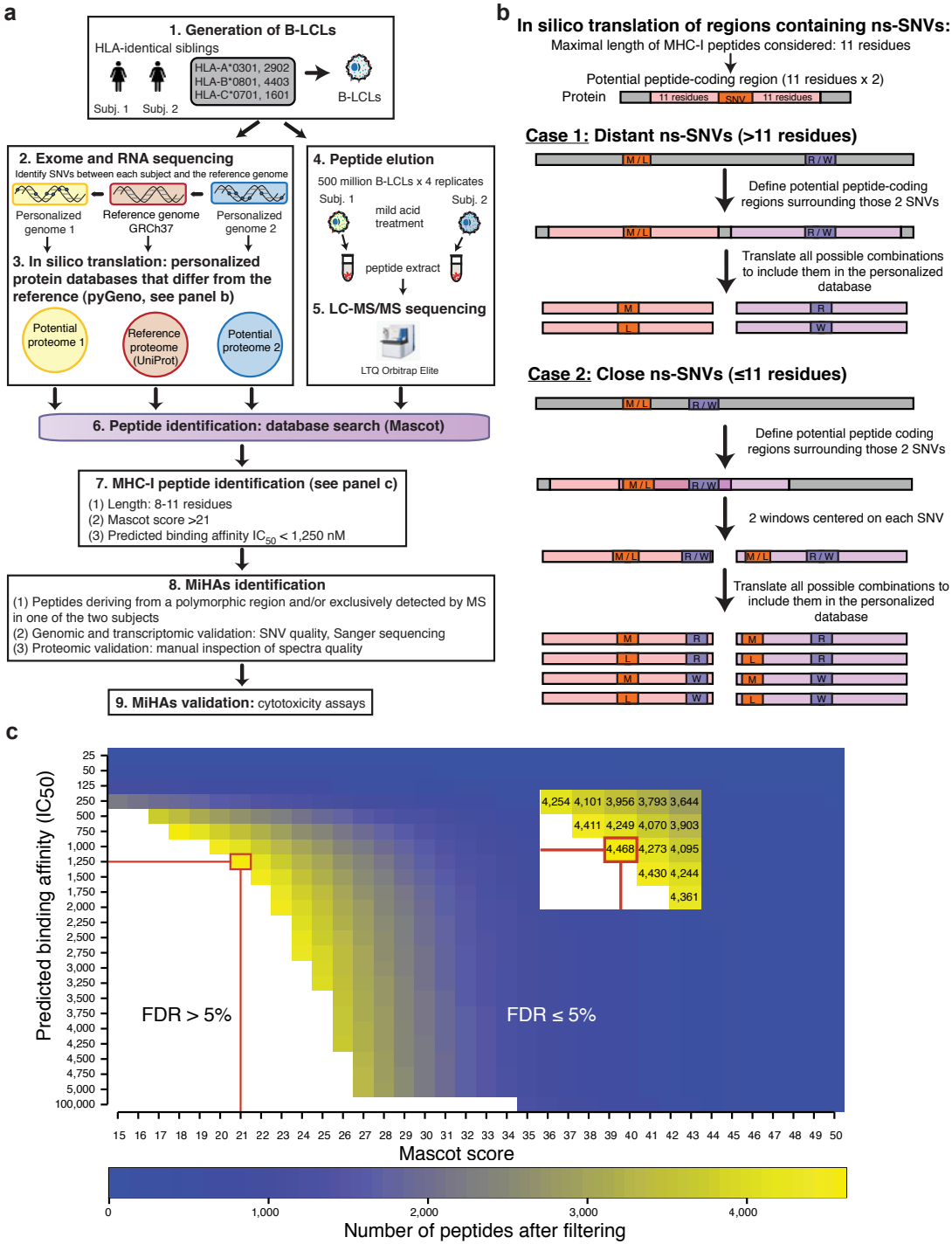


Figure 1. High-throughput genoproteomic strategy used for the identification of polymorphic MIPs on B-LCLs from 2 HLA-identical siblings

(**a**) General overview of the personalized approach, which combines next-generation sequencing, MS and bioinformatics. (**b**) Schematic representation of the combinatorial method used to translate *in silico* polymorphic regions containing ns-SNVs. (**c**) Com-

bining the predicted MHC binding affinity and Mascot score enables to discriminate between MIPs and contaminant peptides. The dataset of peptides identified with an FDR ≤ 5% was filtered according the Mascot score (which represents the confidence level of a peptide assignation), and the predicted MHC binding affinity. The red rectangle and lines indicate the combination of values (IC50 ≤ 1,250 nM and Mascot score ≥ 21) that allowed identifying the maximum number of MIPs with a 5% FDR threshold.

5.5.2 Detection of MIPs encoded by polymorphic genomic sequences using personalized proteomic databases

Large-scale MS-based analyses represent the sole approach enabling comprehensive molecular definition of the MIP repertoire[6, 15, 44]. However, standard high-throughput MS is blind to a whole universe of polymorphic peptides. Indeed, sequencing (or assignation) of peptides by tandem MS is done using engines (e.g., Mascot) that attempt to correlate tandem MS fragment ions from a sample under study with those predicted from available protein databases (e.g. UniProt). Unfortunately, most polymorphic peptides are absent from these databases and tandem MS spectra from unlisted polymorphic peptides will inevitably remain unassigned or misassigned. We reasoned that the most straightforward solution to this conundrum would be to use next generation sequencing data to create subject-specific proteomic databases that would serve as a reference for MS sequencing. We therefore sequenced both the exome and the transcriptome of B-LCLs from each subject in order to combine the benefits of both sequencing technologies and to cover as much as possible each individual's coding genome[45] (Fig. 1a). Next-generation sequencing data were then used to build *in silico* the proteome of B-LCLs from our subjects using the in-house developed python module pyGeno[24] (Fig. 1b). Following integration of exome and transcriptome sequencing, similar number of base pairs and proportions of the human exome were covered in both siblings **(**Fig. 2 track 3 blue vs. orange and Supplementary Data 1 available online**)**. Exome and transcriptome sequencing data of each subject were used to identify SNVs with respect to the reference genome (GRCh37.p2, NCBI). SNVs were then filtered according to their quality, combined into a single set and integrated at their respective position on the reference human genome to obtain two "personalized genomes", from which we extracted and translated every transcript (see "*In silico* generated proteomes"). The translations were then compiled in two

"personalized protein databases", one for each subject.

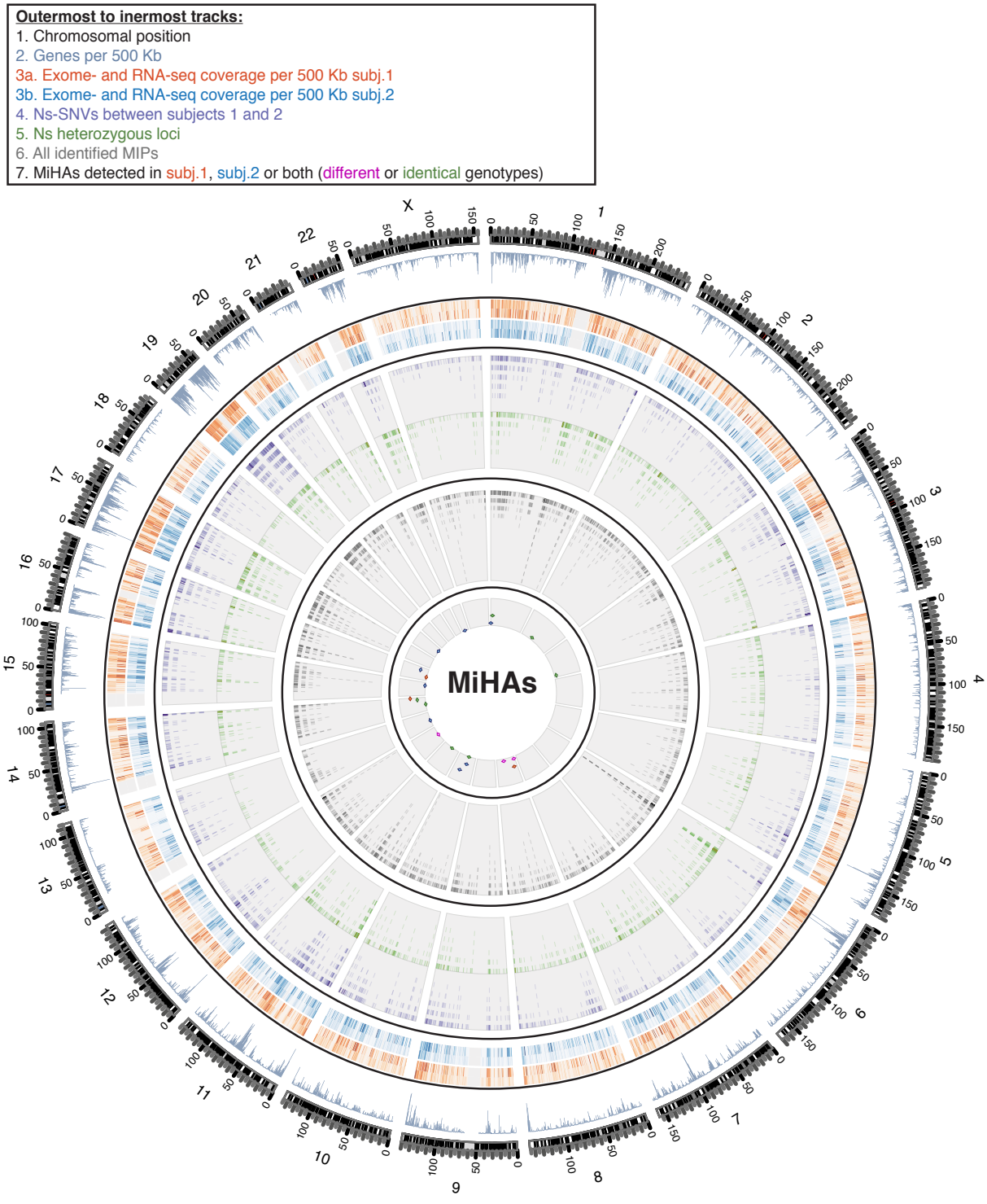


Figure 2. Integrative view of the genomic landscape of the MIP repertoire of HLA-identical siblings

Circos plot showing similar proportions of sequenced genomic and transcriptomic

regions in both siblings and the relatively small number of regions that give rise to MS-detected MIPs including polymorphic MiHAs among all sequenced genes. From outermost to innermost tracks: 1) ideogram indicating chromosomal positions for each chromosome, 2) histogram depicting the number of genes for 500-Kb windows, 3) heat map showing the fraction of bases of 500-Kb windows covered by exome (outer circle) or transcriptome (inner circle) sequencing of subjects 1 (orange) and 2 (blue), 4) tile graph of 4,833 ns-SNV between siblings (purple), 5) tile graph of 3,774 heterozygous loci where both alleles are shared by the two subjects and lead to non-synonymous amino acid changes (green), 6) tile graph representing genomic regions that give rise to 4,468 MIPs, 7) Each dot represents one single gene-encoded MiHA deriving from regions containing ns-SNVs and detected by MS in subjects 1 (orange), 2 (blue) or both (green).

5.5.3 Maximizing the level of coverage of the immunopeptidome

MIPs were eluted from the cell surface by mild acid elution performed on 4 biological replicates of 500 million cells for each subject. Eluted peptides were desalted and separated on strong cation exchange chromatography prior to LC-MS/MS analyses using high resolution precursor and product ion spectra. Compared to other methods such as MHC I immunoprecipitation, acid elution has the advantage of harvesting almost all MIPs, irrespective of their MHC binding affinity46. However, direct acid elution can increase the amount of non-MHC contaminant peptides that are recovered11. In order to maximize the sensitivity and specificity of MIP detection, we have therefore developed an analysis pipeline that relies on a combination of four parameters: i) the canonical MIP length of 8 to 11 amino acids, (ii) the predicted MHC binding affinity given by the NetMHCcons algorithm[47], (iii) the Mascot score, which reflects the quality of peptide assignation, and (iv) the false discovery rate (FDR), which indicates the proportion of decoy (false) vs. target (true) identifications (see Materials and methods). We found that for an FDR of 5%, the best coverage of the immunopeptidome was obtained by combining a Mascot score ≥ 21 and an MHC binding affinity ≤ 1,250 nM (Fig. 1c and Supplementary Fig. S1-S2).

Next, we compared the number of peptide identifications obtained by Mascot using the regular human protein database (UniProt) and personalized databases based on exome and transcriptome sequencing (Supplementary Fig. S3A).

We identified 4,468 unique MIPs from the 2 personalized databases (Supplementary Data 2). The numbers of MIPs identified with the reference database vs. personalized databases were similar with a 96% overlap (Supplementary Fig. S3A). Notably, replacement of reference with the personalized databases had no impact on the quality (Mascot score) of identified peptides (Supplementary Fig. S3B).

5.5.4 The MIP repertoire of HLA-identical siblings is similar but not identical

We have previously shown that the HLA genotype has a major impact on the MIP repertoire of MHC-mismatched individuals[24]. Here, we compared the MIP repertoire of HLA-identical siblings to evaluate the impact of non-HLA genetic polymorphisms on the immunopeptidome. In addition to having identical HLA genotypes, the two siblings showed similar expression levels of the *HLA-A*, *HLA-B* and *HLA-C* genes (Supplementary Data 3) and of the total amount of MHC class I molecules at the cell surface (Supplementary Fig. S4). Following mild acid elution of peptides of comparable efficacy between subjects (Supplementary Fig. S4), we identified a total of 4,468 MIPs encoded by genes from all chromosomes (Fig. 2 track 6 and Supplementary Data 2), detected in a variable number of biological replicates (Fig. 3A) and associated to HLA-A*0301, -A*2902, -B*0801, -B*4403 or –C*1601. Similar numbers of MIPs were identified from the two subjects (4,114 in subject 1 and 4,186 in subject 2). As expected, the majority of the MIPs (86%) were detected in both subjects (Fig. 3a). Most MIPs (75%) had a predicted binding affinity < 500 nM (Fig. 3b). We found no significant difference in the average binding affinity of 282 peptides exclusively detected in subject 1 vs. 351 peptides exclusively detected in subject 2 (Fig. 3b). Furthermore, the number of peptides predicted to bind each of the HLA molecules was similar between the 2 subjects, suggesting that both siblings have comparable surface expression of each of the 5 HLA allelic products tested (Fig. 3c). Collectively, these results show that the MIP repertoire of HLA-identical subjects is similar yet not identical.

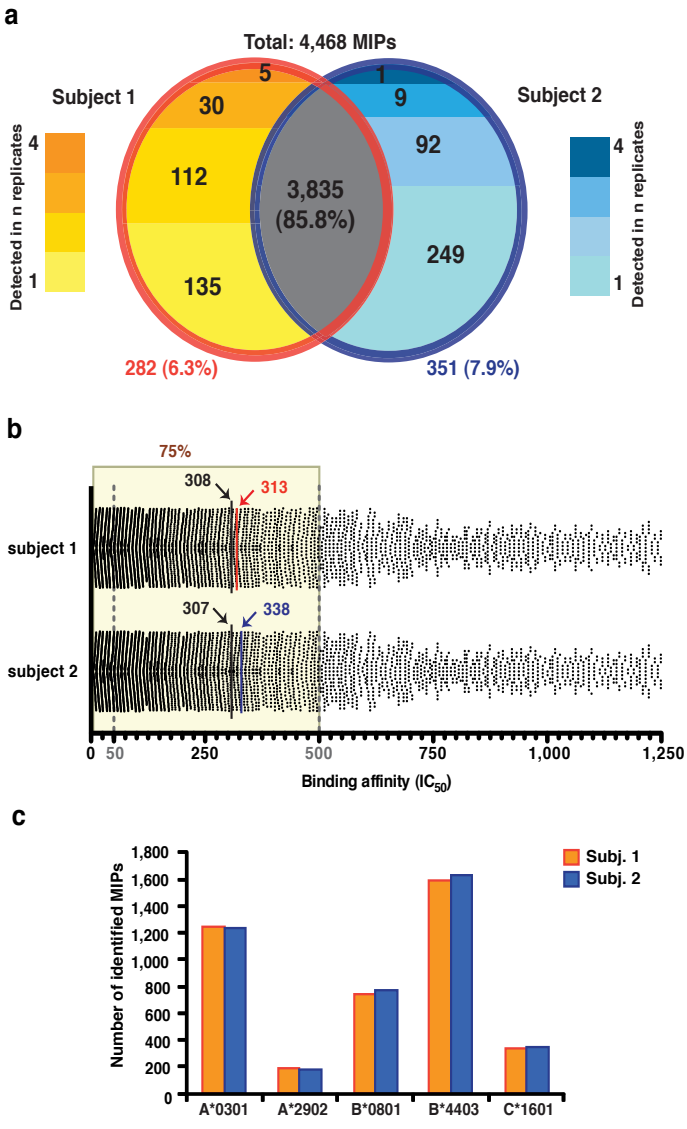


Figure 3. HLA-identical siblings present similar but not identical MIP repertoires
(**a**) Venn diagram showing that 86% of MIPs from HLA-identical siblings were detected in both subjects. A total of 4,468 MIPs were identified in the siblings after analysis of 8 biological samples (4 biological replicates per sibling). MIPs were detected in variable number of biological replicates. For peptides exclusively detected in one subject, the number of replicates in which the peptide was found is shown. The total numbers of MIPs exclusively detected in subject 1 or 2 are shown in red and blue, respectively. (**b**) Scatter plot showing that 75% of identified MIPs are predicted to bind their respective HLA molecules with an IC50 < 500 nM. The IC_{50} for 5 HLA alleles was calculated with the NetMHCcons algorithm. For each peptide (represented by dots), the best binding score for a specific allele was kept. The yellow boxes highlights 75% of all peptides. The black lines and values indicate the average binding affinity of all pep-

tides identified in each sibling. Red and blue lines and numbers represent the average binding affinity of 282 and 351 unshared peptides exclusively detected in subject 1 or 2, respectively. The predicted binding affinity of the two sets of unshared MIPs was statistically indistinguishable (P = 8.5 x 10^6 by 2-tailed Mann-Whitney test). (**c**) The number of peptides associated to each HLA molecule was similar between the 2 subjects.

5.5.5 Identification of MiHAs among unshared MIPs

MiHAs are typically encoded by bi-allelic loci[28, 29]. For each locus where two alleles are present in our subjects, three genotypes are possible: AA, AB and BB. At the peptidomic level, each allele can be dominant (generate a MIP) or recessive (a null allele that generates no MIP). Moreover, by comparing MIPs eluted from two HLA-identical individuals, dominant MiHAs can be separated into two groups: unshared MIPs (genotypes AA vs. BB, AB vs. AA or BB), or shared MIPs (AB vs. AB). If one allele is recessive (e.g., B), subjects can be similar at the peptidomic level (express the A MiHA) even though they have different genotypes (AA vs. AB) (Fig. 4).

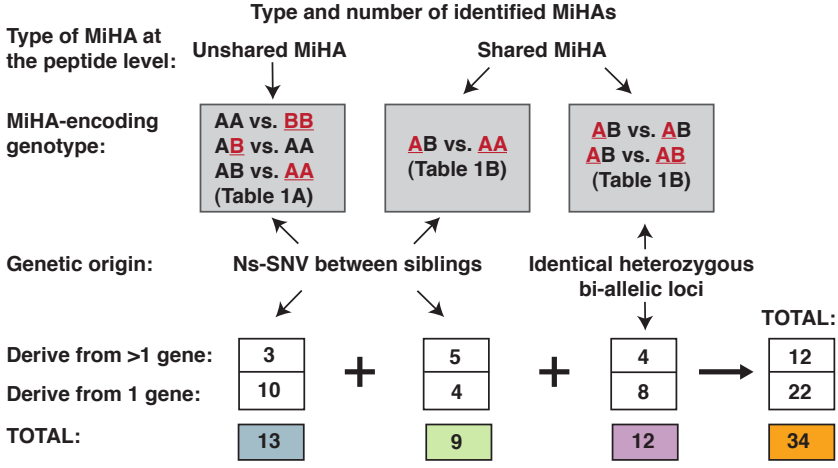


Figure 4. Overview of MiHAs identified following analysis of genomic and peptidomic data from our two subjects

In our search for MiHAs, we first performed in-depth analyses of MIPs detected in only one subject (unshared MIPs; Fig. 3a). Here the key finding was that out of 633 unshared MIPs, only 13 (2%) were encoded by genomic regions harboring ns-SNVs between the two subjects (Fig. 4 and Supplementary Data 2). The origin of 3 of these 13 MIPs was ambiguous (they could derive from several

genes), whereas the other ten MiHAs were assigned to a single gene (Fig. 4 and Table 1A). The genetic polymorphisms responsible for almost all MiHAs corresponded to SNPs reported in the dbSNP database[48] (Table 1A). Consistent with previous findings on human MiHAs[29], only one of the two possible variants was detected by MS for each MiHA locus (Table 1A). In other words, at the peptide level, one allele was dominant (generated a MIP) and one was recessive (generated no MIP) (Table 1A). In 4 out of 10 cases, absence of the variant MiHA at the cell surface could be explained by a decreased binding affinity of the variant for the corresponding HLA molecule (IC_{50} difference ≥ 2x). Nine of our best characterized MiHAs are novel, while one (KEFEDGIINW) has been previously reported49. Four MiHAs were detected by MS in the subject homozygous for the corresponding allele but not in the heterozygous subject (Table 1A, rows 7-10). This suggests that zygosity influences MiHA abundance and that low abundance MiHAs may fall below the MS detection threshold in heterozygous subjects. Consistent with this, the MS intensity for these four MiHAs was low in homozygous subjects (Supplementary Data 2). Six MiHAs were coded by an allele present only in one subject (Table 1A, rows 1-6), and were thus potentially immunogenic for the other sibling. We further validated the presence of the ns-SNV in the corresponding DNA and/or cDNA regions of these six MiHAs in both subjects by Sanger sequencing to discard errors in base calling or mapping (Supplementary Fig. S5). Then, we assessed the immunogenicity of four of these MiHAs by cytotoxicity assays. In all cases, *in vitro* generated MiHA-specific CTLs selectively killed allogeneic MiHA-positive B-LCLs (to similar levels as autologous B-LCLs pulsed with the MiHA) but not autologous MiHA-negative unpulsed B-LCLs (Fig. 5a).

Table 1. MiHAs resulting from ns-SNVs in the MIP-coding region and detected in (A) one of the two subjects or in (B) both subjects

Selected features of the MiHAs are shown, including the amino acid sequence, the subject (S) in which the MiHA was detected, the source gene, the HLA molecule for which the MiHA has the best predicted binding affinity (IC_{50}), the translated genotype shown in amino acids (AA), the alternative MiHA variant and its predicted HLA binding affinity (IC_{50}), and the dbSNP identification when the ns-SNV corresponds to a known SNP. The MS-detected and non-detected polymorphic residues are highlighted in bold underlined and bold italics, respectively. IC_{50} values of the non-detected MiHA variants are shown in italics when they show a fold difference ≥ 2 relative to the de-

tected MiHAs. Further features can be found in Supplementary data 2

	MiHA name	MiHA sequence	S	Gene symbol	HLA allele	IC_{50} nM	AA sub.1	AA sub.2	Alternative MiHA variant	IC_{50} nM	dbSNP
A	ITGAL-1^{T}	S**T**ALRLTAF	1	*ITGAL*	C*1601	306	TR	RR	S***R***ALRLTAF	*2,969*	rs2230433
	IGHV5-51-1^{V}	**V**IYPGDSDTRY	1	*IGHV5-51*	A*2902	27	VI	SS	***I/S***IYPGDSDTRY	19/31	rs199610746
	NQO1-1^{R}	AMYDKGPF**R**SK	2	*NQO1*	A*0301	12	WW	RW	AMYDKGPF***W***SK	11	rs1131341
	GRP-1^{R}	**R**ELPLVLL	2	*GRP*	B*4403	285	SS	RR	***S***ELPLVLL	159	rs1062557
	C13orf18-1^{R}	RVSLPTSP**R**	2	*C13orf18*	A*0301	235	GG	RR	RVSLPTSP***G***	*11,858*	rs1408184
	IGLV2-11-1^{HH}	SDVGG**HH**Y	2	*IGLV2-11*	A*2902	660	YY/NN	HH/HH	SDVGG***YN***Y	412	---
	R3HCC1-1^{H}	AENDFV**H**RI	1	*R3HCC1*	B*4403	61	HH	HR	AENDFV***R***RI	123	rs11546682
	NADK-1^{K}	AVHNGLGE**K**	2	*NADK*	A*0301	229	KN	KK	AVHNGLGE***N***	59	rs4751
	ACC-2^{G}	KEFED**G**IINW	2	*BCL2A1*	B*4403	49	GD	GG	KEFED***D***IINW	*24,349*	rs3826007
	KIF20B-1^{I}	QELETS**I**KKI	2	*KIF20B*	B*4403	288	IN	II	QELETS***N***KKI	*615*	rs12572012
B	MCPH1-1R	EEINLQ**R**NI	1,2	*MCPH1*	B*44:03	503	RR	RI	EEINLQ***I***NI	*212*	rs2083914
	MDM1-1I	V**I**QERVHSL	1,2	*MDM1*	B*08:01	61	IT	II	V***T***QERVHSL	*401*	rs962976
	FAM82B-1K	VMGNPGTF**K**	1,2	*FAM82B*	A*03:01	23	KN	KK	VMGNPGTF***N***	*15,374*	rs6980476
	TMEM132A-1A	AAADRVGP**A**A	1,2	*TMEM132A*	C*16:01	1,236	AP	AP	AAADRVGP***P***A	*1,203*	---
	MAGEF1-1A	ALAAK**A**LAR	1,2	*MAGEF1*	A*03:01	136	AS	AS	ALAAK***S***LAR	*109*	rs10937187
	TRIP11-1K	DVQ**K**KLMSL	1,2	*TRIP11*	B*08:01	216	KN	KN	DVQ***N***KLMSL	*534*	rs145868557
	IMMT-1S	KQ**S**ASQLQK	1,2	*IMMT*	A*03:01	65	SP	SP	KQ***P***ASQLQK	421	rs1050301
	DLGAP5-1H	KTY**H**VTPMTPR	1,2	*DLGAP5*	A*03:01	27	HQ	HQ	KTY***Q***VTPMTPR	48	rs8010791
	ZWINT-1R	QELD**G**VFQKL	1,2	*ZWINT*	B*44:03	366	RG	RG	QELD**R**VFQKL	197	rs2241666
	MIIP-1K	SEESAVP**K**RSW	1,2	*MIIP*	B*44:03	235	KE	KE	SEESAVP***E***RSW	245	rs2295283

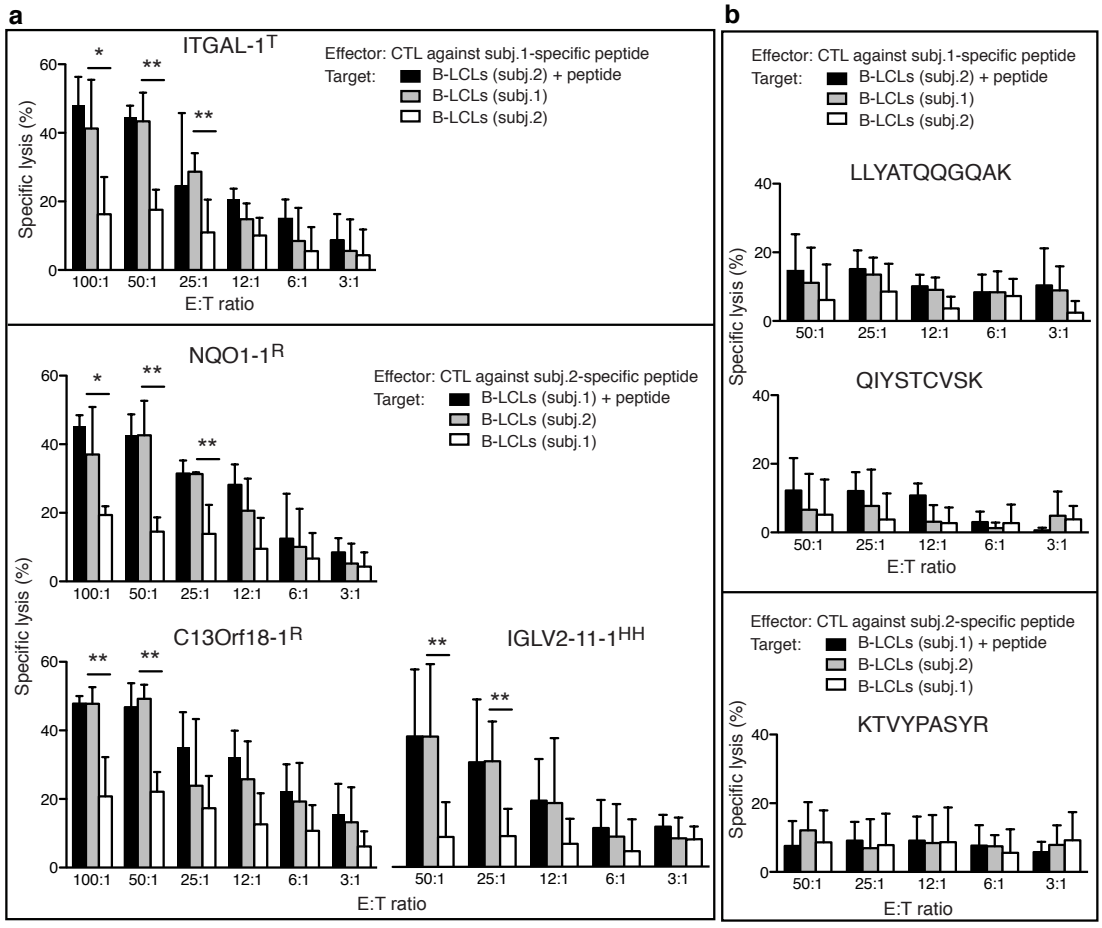


Figure 5. Unshared MIPs encoded by polymorphic loci are immunogenic
Frozen PBMCs from the MIP-negative subject were thawed and stimulated with autologous dendritic cells pulsed with an unshared MIP detected in the other individual. Primed cells were restimulated with irradiated autologous B-LCLs pulsed with the same peptide for another 7 days. Restimulated cells were tested for *in vitro* cytotoxicity activity against autologous B-LCLs pulsed with the relevant peptide (positive control, black), unpulsed autologous B-LCLs (negative control, white), or MIP-positive allogeneic B-LCLs (test, grey) at various effector-to-target (E:T) ratios. The minimal cytotoxic activity against unpulsed autologous B-LCLs is most likely due to recognition of EBV epitopes. Average and S.D. of three or four independent experiments are shown. Significant differences are indicated by * $P < 0.05$ or ** $P < 0.01$, 2-tailed Student *t*-test. (**a**) Unshared MIPs encoded by polymorphic loci. (**b**) Unshared MIPs encoded by non polymorphic loci.

What is the significance of the unshared MIPs derived from non-polymorphic regions but detected by MS in only one subject (n = 633-13 = 620; Fig. 3a)?

Could they be MiHAs whose presence is regulated by *cis*- or *trans*- acting polymorphisms (outside of the MIP-coding genomic sequence) that would affect MIP processing[25, 28]? The MS/MS spectra of each of these MIPs were manually validated and to further confirm the absence of these MIPs in one of the two subjects, we searched these MIPs in two additional biological replicates from each cell line. Most non polymorphic unshared MIPs were detected in only one or two replicates (Fig. 3a). This suggests that the presence of these MIPs was inconsistent, perhaps because of the shotgun nature of MS. Another possible explanation is that these MIPs derive from proteins expressed only in specific stages of the cell cycle11, and that they are not genuine MiHAs. However, 41 unshared MIPs could not be discarded so easily because they were detected in 3-6 replicates of one sibling and absent in 6 replicates of the other sibling. Except for 2 cases, exclusive detection of these MIPs in one of the siblings was not caused by differences in abundance of the MIP-source transcript (Supplementary Fig. S6a) or in the expression of the MIP-coding exon between subjects (Supplementary Fig. S6b), nor by differences in the expression of genes involved in the antigen processing and presentation pathway (Supplementary Data 3). We therefore evaluated the immunogenicity of the three most enticing unshared MIPs coded by non-polymorphic regions, i.e., those showing the highest MS intensity and reproducibility (Fig. 5b and Supplementary Data 2). We reasoned that if these MIPs were MiHAs, they should be immunogenic, even if their presence was dictated by unidentified polymorphisms outside of the MIP-coding genomic sequence[25, 28]. None of the tested MIPs could elicit the generation of cytotoxic T cells in the MIP-negative sibling. We therefore failed to discover a single MiHA among unshared MIPs coded by non-polymorphic regions. The most parsimonious explanation is that these MIPs were simply differentially expressed peptides whose abundance was below the MS detection threshold in B-LCLs from one subject. Though further studies may be warranted to understand interindividual differences in the abundance of specific MIPs, we predict that in most if not all cases, differentially expressed peptides coded by non-polymorphic regions are irrelevant as regards alloreactivity. Accordingly, we conclude that identification of MiHAs absolutely requires a combination of MS and genomic data. Reliance solely on MS detection of unshared MIPs would overestimate the number of MiHAs. In contrast, the use of personalized databases based on whole exome and transcriptome sequencing allows

to rapidly identifying genuine MiHAs coded by polymorphic loci.

5.5.6 The global imprint of ns-SNVs on the MIP repertoire

In order to assess the global imprint of ns-SNVs on the MIP repertoire, we asked the question: what proportion of ns-SNVs between our two subjects were located in MIP coding exomic sequences? By comparing the whole exome and RNA-seq data from our two HLA-identical siblings, we found a total of 4,833 ns-SNVs (Fig. 2**,** track 4). Overall, 26 of these ns-SNVs were located in regions coding for 22 MiHAs identified by MS, of which 14 originated from a single gene and are depicted in the Circos plot (Fig. 2 track 4 vs. 7 blue, orange and pink**)** and 8 have an ambiguous origin (e.g., immunoglobulin genes) (Fig. 4**,** n=3+5 and Supplementary Data 2). The 14 unambiguously assigned MiHAs were unshared (Table 1A) or shared (Table 1B rows 1-4) at the peptidomic level. Thus, from a genomic perspective, only 0.5% of all ns-SNVs (26/4,833) found between our subjects were represented in their MIP repertoire.

5.5.7 Identification of MiHAs among shared MIPs

Among 3,835 shared MIPs, 21 were encoded by bi-allelic loci and therefore represent MiHAs (Fig. 4**,** n=9+12). These shared MIPs would not be immunogenic for our subjects but would be immunogenic for subjects homozygous for the alternative allele. In 9 cases, one subject was homozygous for a dominant MiHA allele (AA) and the other subject was heterozygous for the dominant and a recessive allele (AB) (Fig. 4**)**. The origin of five of these 9 MiHAs was ambiguous (they could derive from several genes), whereas the other four MiHAs were assigned to a single gene (Table 1B**,** rows 1-4 and Fig. 4). The exome of our subjects shared 3,774 heterozygous loci (Fig. 2, track 5). Twelve MiHAs derived from such bi-allelic loci for which our subjects shared the same heterozygous genotype (AB). Eight of these twelve MiHAs could be unambiguously assigned to a single gene (Fig. 2 track 7 in green, Fig. 4 and Table 1B rows 5-11). The two alleles were co-dominant in one case, whereas only one allele was dominant (identified by MS) in all other cases. MIP dominance did not reflect allelic imbalance at the transcript level (data not shown). Nonetheless, in four cases of shared MiHAs, the product of the recessive allele was predicted to have a lower

MHC binding affinity than the product of the dominant allele (Table 1B).

5.5.8 Overall differences in the MIP repertoire of HLA-identical siblings

Comparison of genomic and proteomic data from our subjects led to the discovery of 34 MiHAs (Fig. 4), of which 22 were unambiguously assigned to a specific gene (Fig. 2, track 7 and Table 1). Out of 34 MiHAs, 13 were found in only one of the two subjects whereas 21 MiHAs were shared MIPs (Fig. 4). This means that out of 4,468 MIPs only 13 (0.3%) would be immunogenic for one of our subjects. If all unshared non-polymorphic MIPs are not immunogenic, as we observed with the 3 MIPs tested (Fig. 5b), this would mean that each subject would be tolerant to about 99.7% of the MIPs found on the B-LCLs of this sibling. The use of personalized databases for tandem MS sequencing was instrumental in the discovery of many MiHAs. Taking into account only the best characterized MiHAs, 6 out of the ten listed in Table 1A (unshared MIPs) and 5 out of the 12 listed in Table 1B (shared MIPs) would have been missed in the absence of personalized databases, because these 11 peptides were absent in the Uniprot database.

5.5.9 What proportion of MIP coding regions is polymorphic at the population level?

To address this question, we searched in the dbSNP database for validated ns-SNPs in the genomic sequences coding our 4,468 MIPs. We found that at the population level, 88% of our MIP coding sequences were invariant whereas 12% contained at least one ns-SNP: 670 ns-SNPs were found in the genomic region coding for 536 MIPs (Fig. 6a-b and Supplementary Data 4). Hence, at the population level, 536 MiHAs can be presented by the five HLA class I molecules studied herein: HLA-A*0301, -A*2902, -B*0801, -B*4403 and –C*1601. Further studies will be required to determine the number of dominant and recessive peptide variants encoded by these 536 MiHA loci.

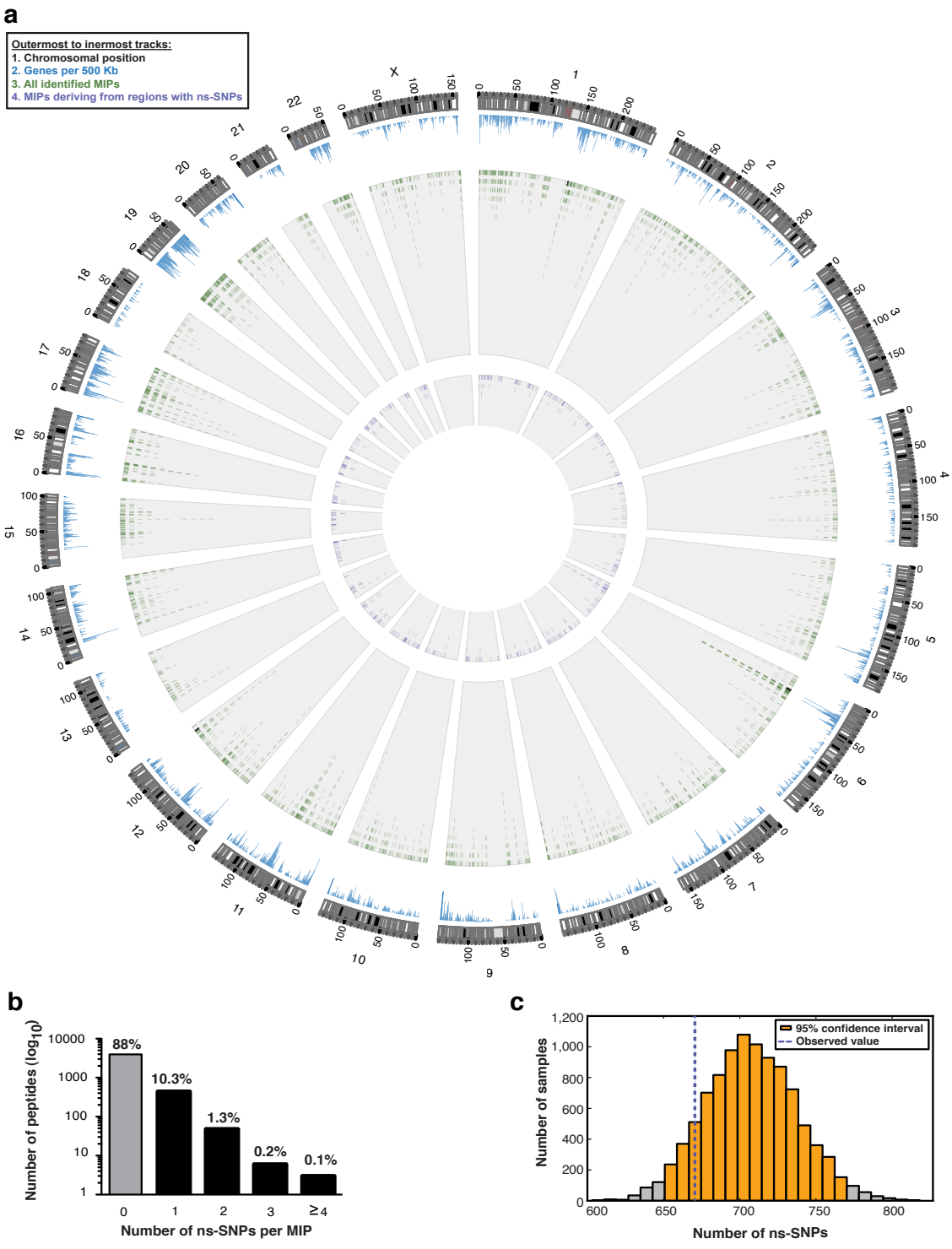


Figure 6. Frequency of ns-SNPs in the MIP coding exome

(**a**) Circos plot illustrates the relative proportion of polymorphic MIPs (n = 536) in the immunopeptidome and the genomic location of their coding loci. (**b**) Histogram showing the number and percentages of MIP coding regions containing ns-SNPs in

the global population. We used dbSNP to find validated ns-SNPs in the exomic sequences encoding the 4,468 MIPs identified in our subjects. In the case of MIPs deriving from multiple source regions, the average number of ns-SNPs of all possible MIP source regions was calculated. (**c**) The 4,468 MIPs of our subjects were encoded by 13,404 nucleotides. We performed 10,000 random samplings of 4,468 exomic sequences (containing a total of 13,404 nucleotides) from the human reference exome (Ensemble GRCh37.65). In all samplings, the frequency of exomic sequences coding for 8-,9-,10- and 11-mers was identical to the frequency found in the 4,468 MIP coding sequences from our subjects. The histogram depicts the distribution of validated ns-SNPs (dbSNP) in exomic sequences from the global population found in 10,000 random samplings of the whole exome. The average number of ns-SNPs of all random samplings was 708 (s.d. 30.4, 95% confidence interval: 650-768 shown in orange). The blue dotted line shows the number of ns-SNPs (n = 670) in the exomic sequences coding for the MIPs detected in our subjects.

5.5.10 Is there a bias in favor or against ns-SNPs in MIP coding regions?

To address this question, we wished to compare, in the global population, the frequency of ns-SNPs in the whole exome vs. the frequency in the 4,468 exomic sequences coding for the MIPs identified herein. To this end, we designed a bootstrap procedure (10,000 iterations) based on random samplings of 4,468 peptide-coding regions (13,404 base pairs/sampling) from the human reference exome (Ensemble GRCh37.65). For each sampling, we then calculated the number of validated ns-SNPs reported in dbSNP (Fig. 6c). In all samplings, the frequency of exomic sequences coding for 8-,9-,10- and 11-mers was identical to the frequency found in the 4,468 MIP coding sequences from our subjects. The number of ns-SNPs in the MIP coding exome (n = 670) fell in the range of ns-SNPs found in 10,000 random samplings of the whole exome (average = 708; 95% confidence interval 650-768). We therefore conclude that the MIP coding exome reflects the frequency of ns-SNPs in the human genome.

5.6 Discussion

MS is the sole method that enables direct identification of MIPs and large scale analyses of the MIP repertoire[6, 20]. Indirect predictions based on reverse immunology approaches are fraught with false discovery rates that may reach 95%[50, 51]. Currently, MS sequencing has been largely limited to peptides represented

in the reference UniProt database. Our work demonstrates that the universe of peptides identified by MS can be expanded and refined by using personalized databases that include whole exome and transcriptome sequencing data.

As well stated by J. Yewdell et al. “Despite the fact that quantitative aspects of systems are critical to their understanding, they are frequently ignored”[52]. In line with this concept, our data provide the answer to a longstanding question: what is the proportion of invariant vs. polymorphic MIPs presented by MHC molecules? In other words, to what extent do non-MHC genomic polymorphisms enhance the interindividual variability of the immunopeptidome? We found that, at the population level, at least one ns-SNP is found in 12% of exomic sequences coding the MIPs presented by five common HLA class I allotypes. That about 88% of the immunopeptidome of HLA-identical siblings is invariant in the global population illustrates the overwhelming importance of the HLA genotype in defining the content of the MHC class I immunopeptidome.

In depth analyses of genomic and proteomic data revealed that about 0.5% of ns-SNVs between the exome of our subjects were represented in their MIP repertoire. Consequently, 13 MIPs were unique to one subject and might elicit allogeneic T-cell responses from his sibling, as demonstrated for four of them. Integration of personalized genomic and proteomic data was absolutely essential for identification of these rare polymorphic MIPs interspersed among thousands of non-polymorphic MIPs. Since the MIP repertoire is molded by the transcriptome, some MIPs are ubiquitous and others are cell lineage-specific[11, 53]. Accordingly, various cell types present non-identical MIP repertoires. MIPs derive mostly from transcripts expressed at medium to high levels (as opposed to very low or low levels), and about 8,500 transcripts are expressed at medium to high levels in B-LCLs[24]. We therefore posit that, at the organismal level, the total number of unshared MiHAs between two HLA-identical siblings would be about 2.5-fold the number found in B cells, assuming a total number of 21,000 human transcripts (i.e., 13 x (21,000/8,500) = 32). Unrelated individuals share fewer gene sequences than siblings. As a consequence, it has been calculated that the frequency of unshared MiHAs is increased by about 1.8-fold in unrelated (HLA-matched) subjects relative to siblings[29]. Thus, two

unrelated HLA-identical subjects would display about 58 unshared MiHAs. Of note, these numbers might increase with more sensitive MS instruments and better sequencing coverage of difficult regions (e.g. GC-rich).

All MHC antigens are dominant. Our data show that this is not the case for MiHAs. Out of 21 MiHA loci (Table 1), 20 had one dominant (MIP generating) and one recessive (no MIP generated) allele. This observation is clearly consistent with population analyses of ten well characterized autosomal MiHA loci: only one locus had two dominant alleles[29]. The absence of MIP could be explained by a decreased MHC binding affinity of peptides coded by eight of our 21 recessive alleles. We infer that for the other 13 recessive alleles, the absence of MIP must be due to interference of the polymorphism with some step in MIP processing that precedes MHC binding (e.g., cleavage by the proteasome or other proteases)[18, 43]. With tens of thousands of proteins, mammalian cells are the most complex entity in the antigenic universe faced by our immune system54. Theoretical estimates suggest that the immunopeptidome contains 0.1% of the 9-mer sequences present in the proteome6. Few peptides win the fierce competition for inclusion in the immunopeptidome. Thus, if we consider MiHAs coded by dominant alleles, as winners, it follows that in 20 out of 21 cases (the recessive alleles), a single amino acid substitution transformed winners into losers. This is an eloquent reminder that we cannot predict the molecular composition of the immunopeptidome based on our limited understanding of the complexity of the MIP processing pathway.

Allogeneic hematopoietic cell transplantation has led to the discovery of the allogeneic graft-versus-leukemia (GVL) effect, which remains the most widely effective strategy for cancer immunotherapy in humans. GVL is mediated mainly, if not exclusively, by donor T cells that recognize host MiHAs. In line with recent progress in the field of cell therapy, MiHAs are therefore attractive targets for adoptive T-cell immunotherapy of cancer, particularly hematologic cancers[36-41]. However, because of the low number of molecularly defined human MiHAs, less than 30% of patients would currently be eligible for immunotherapy targeted to specific MiHAs[55]. Our report reveals a strategy for high-throughput MiHA discovery that could greatly accelerate the development of MiHA-targeted immunotherapy.

Our genoproteomic method combining next generation sequencing and MS shows how it is possible to accurately identify by MS any peptide, provided that its source DNA or RNA has been sequenced. This notion opens new avenues in systems immunology and should be invaluable for exploration of several "black holes" in the immunopeptidome. One particularly important black hole is the "cancer immunome"[56]. Compelling evidence suggests that the most immunogenic antigens present on cancer cells are mutant peptides derived from the numerous mutations found in neoplastic cells[57-59]. However, tumor-specific mutant peptides (alike MiHAs) are not detected by standard large scale MS approaches. We posit that our method should enable discovery of tumor-specific peptides (the product of somatic mutations) with the same accuracy as MiHAs (the product of germline genetic polymorphisms). Accordingly, our next priority will be to use this method to explore the impact of the cancer mutations on the immunopeptidome of cancer cells.

5.7 Methods

5.7.1 Cell culture and HLA typing

This study was approved by the Comité d'Éthique de la Recherche de l'Hôpital Maisonneuve-Rosemont and all subjects provided written informed consent. Peripheral blood mononuclear cells (PBMCs) were isolated from blood samples of 2 non-twin HLA-identical Caucasian female siblings. B-LCLs were derived from PBMCs with Ficoll-Paque Plus (Amersham) as described60. High-resolution HLA genotyping was performed at the Maisonneuve-Rosemont Hospital. The two siblings are HLA-A*03:01,*29:02; B*08:01,*44:03; C*07:01,*16:01; DRB1*03:01,*07:01.

5.7.2 RNA extraction and preparation of transcriptome libraries

Total RNA was isolated from 5 million B-LCLs using RNeasy mini kit including DNase I treatment (Qiagen) according to the manufacturer's instructions. Total RNA was quantified using the NanoDrop 2000 (Thermo Scientific) and RNA quality was assessed with the 2100 Bioanalyzer (Agilent Technologies). Transcriptome libraries were generated from 1 μg of total RNA using the TruSeq

RNA Sample Prep Kit (v2) (Illumina) following the manufacturer's protocol. Briefly, poly-A mRNA was purified using poly-T oligo-attached magnetic beads using two rounds of purification. During the second elution of the poly-A RNA, the RNA was fragmented and primed for cDNA synthesis. Reverse transcription of the first strand was performed using random primers and SuperScript II (InvitroGene). A second round of reverse transcription was done to generate a double-stranded cDNA, which was then purified using Agencourt AMpure XP PCR purification system (Beckman Coulter). End repair of fragmented cDNA, adenylation of the 3' ends and ligation of adaptors were completed following the manufacturer's protocol. Enrichment of DNA fragments containing adapter molecules on both ends was done using 15 cycles of PCR amplification and the Illumina PCR mix and primers cocktail.

5.7.3 DNA extraction, preparation of genomic DNA libraries and exome enrichment

Genomic DNA was extracted from 5 million B-LCLs using the PureLink Genomic DNA Mini Kit (Invitrogen) according to the manufacturer's instructions. DNA was quantified and quality-assessed using the NanoDrop 2000 (Thermo Scientific). Genomic libraries were constructed from 1μg of genomic DNA using the TruSeq DNA Sample Preparation Kit (v2) (Illumina) following the manufacturer's protocol. We used 500 ng of DNA-Seq libraries for hybrid selection-based exome enrichment with the TruSeq exome enrichment kit (Illumina) according to the manufacturer's instructions.

5.7.4 Whole transcriptome sequencing (RNA-Seq) and exome sequencing

Paired-end (2 x 100 bp) sequencing was performed using the Illumina HiSeq2000 machine running TruSeq v3 chemistry. Cluster density was targeted at around 600-800k clusters/mm2. Two RNA-Seq or four exomes libraries were sequenced per lane (8 lanes per slide). More than 53 and 50 million of mega bases (Mb) of annotated exons were covered (≥ 5 reads) in subjects 1 and 2, representing 76-81% of the human annotated exome **(**Supplementary Data 1**)**. Sequence data were mapped to the human reference genome (hg19) using the Casava 1.8.1 and the Eland v2e mapping softwares (Illumina). First, the *.bcl files were con-

verted into compressed FASTQ files, following by demultiplexing of separate multiplexed sequence runs by index. Single reads were aligned to the human reference genome using the multiseed and gapped alignment method. Multiseed alignment works by aligning the first seed of 32 bases and consecutive seeds separately. Gapped alignment extends each candidate alignment to the full length of the read and allows for gaps up to 10 bases. The following criteria were applied: i) a read contains at least one seed that matches with at most 2 mismatches without gaps and ii) gaps were allowed for the whole read, as long as they correct at least five mismatches downstream. For each candidate alignment a probability score, which is based on the sequencing base quality values and the positions of the mismatches, was calculated. The alignment score of a read, which is expressed on the Phred scale, was computed from the probability scores of the candidate alignments. The best alignment for a given read corresponded to the candidate alignment with the highest probability score and was kept if the alignment score exceeded a threshold. Reads that mapped at 2 or more locations (multireads) were not included in further analyses. For the exome paired-end libraries, the best scoring alignments for each half of the pair were computed and compared to find the best paired-read alignments according to the estimated insert size distribution. In the case of RNAseq libraries, an additional alignment was performed against splice junctions and contaminants (mitochondrial and ribosomal RNA). Sequences mapping to contaminants were discarded whereas reads uniquely mapping to splice junctions were kept and converted back to genome coordinates. The Integrative Genomics Viewer v2.0 (http://www.broadinstitute.org/igv/)61 was used to visualize the mapped reads.

5.7.5 Quantification of transcript expression

We used two methods to estimate and compare transcript expression between subjects. In the first method, the Casava 1.8.1 software (Illumina) was used to estimate gene or exon expression levels (RNA-seq) measured as RPKM (i.e. Read Per Kilobases of exon model per Million mapped reads) using the following formula: Gene or exon RPKM = 109 x Cb/Nb L, where Cb is the number of bases that fall on the feature, Nb is the total number of mapped bases and L is the length of the feature in base pairs. We also used the DESeq package[62],

which is based on raw counts, to compare transcript expression.

5.7.6 Identification of single nucleotide variations and read counting

Variant call, indel detection and read counting were done using the Casava 1.8.1 software (Illumina). Reads were re-aligned around candidate indels to improve the quality of variant calls and site coverage summaries. Casava was also used to retrieve all SNVs observed between the reference genome (GRCh37.p2, NCBI) and the sequenced transcriptome and exome of our subjects. For each called SNV, Casava retrieves and calculates various statistics including its position, the reference base, the number of base calls for each nucleotide, the most probable genotype (max_gt), and a Q-value expressing the probability of the most probable genotype (Qmax_gt). The Q-value is a quality score that measures the probability that a base is called incorrectly. This information was loaded into an in-house python module, pyGeno[24], for further processing and filtering (see "*In silico* generated proteomes" section).

5.7.7 In silico generated proteomes (personalized databases)

As the transcriptome is functionally closer to the proteome than the genome, we used the transcriptome sequencing data as the main information to construct personalized databases. To further increase the sequencing coverage of our subjects and thus improve the personalized databases, we integrated the exome sequencing data to the transcriptome sequencing data. For every SNV found by transcriptome sequencing, we retained the most probable genotype if the Q-value (Qmax_gt) was ≥20, which corresponds to a 1% error rate (a higher quality score indicates a smaller probability of error). We also included the genotypes of SNVs that were only found by exome sequencing and that had a Q-value ≥20. Lastly, we included all bases of SNVs called by both the transcriptome and exome sequencing, regardless of the Q-Value. The retained genotypes of all SNVs were then integrated in the reference genome (GRCh37.p2) at their right position to construct a "personalized genome" for each subject. These personalized genomes were used to extract all transcripts reported in the Ensembl gene set (GRCh37.65) for all chromosomes except for the Y chromosome and mitochondrial DNA. These transcripts were then *in silico*

translated into proteins using the reading frame specified in the Ensembl gene set. Considering that the vast majority of MIPs have a maximum length of 11 amino acids, we established a window of 21 amino acids centered at each heterozygous ns-SNV. When a window contained more than one SNV, we translated *in silico* all possible combinations and included them in the personalized databases (Fig. 1b). Finally, we compiled all translation products into two databases (one for each subject) that were used for the identification of MIPs (see "MS/MS sequencing and peptide clustering" section). Both resulting databases had a similar size, in terms of number of residues (36,007,210 in subject 1 and 36,010,026 in subject 2) and number of entries (95,806 in subject 1 and 95,687 in subject 2). Moreover their size is comparable to the size of the reference UniProt human database used (43,384,120 residues and 75,530 entries).

5.7.7 MS/MS sequencing and peptide clustering

Based on our previous studies on MS data reproducibility across technical and biological replicates11, we prepared four biological replicates of 5 x 10^8 exponentially growing B-LCLs from each subject. MIPs were released by mild acid treatment, desalted on an HLB cartridge 30cc, filtered with a 3000Da cut-off membrane and separated into seven fractions by cation exchange chromatography using an off-line 1100 series binary LC system (Agilent Technologies) as previously described11, 12. Fractions containing MIPs were resuspended in 0.2% formic acid and analyzed by LC-MS/MS using an Eksigent LC system coupled to a LTQ-Orbitrap ELITE mass spectrometer (Thermo Electron). Peptides were separated on a custom C18 reversed phase column (150 μm i.d. X 100 mm, Jupiter Proteo 4 μm, Phenomenex) using a flow rate of 600 nL/min and a linear gradient of 3-60% aqueous ACN (0.2% formic acid) in 120 mins. Full mass spectra were acquired with the Orbitrap analyzer operated at a resolving power of 30 000 (at m/z 400). Mass calibration used an internal lock mass (protonated (Si(CH3)2O))6; *m/z* 445.120029) and mass accuracy of peptide measurements was within 5 ppm. MS/MS spectra were acquired at higher energy collisional dissociation (HCD) with a normalized collision energy of 35%. Up to 6 precursor ions were accumulated to a target value of 50000 with a maximum injection time of 300 ms and fragment ions were transferred to the Orbitrap analyzer operating at a resolution of 15000 at *m/z* 400.

Mass spectra were analyzed using Xcalibur software and peak lists were generated using Mascot distiller Version 2.3.2 (http://www.matrixscience.com). Database searches were performed against UniProt Human database (43,384,120 residues, released on April 2, 2013), databases specific to subjects 1 and 2 (34,976,580 and 34,990,381 residues, respectively, see "*in silico* generated proteome" section) and EBV_B95.8 database (40,946 residues), using Mascot (Version 2.3.2, Matrix Science). To calculate the false discovery rate (FDR), we performed a Mascot search against a concatenated target/decoy database using the human UniProt or subject-specific databases. The target represents the forward sequences and the decoy its reverse counterparts. Mass tolerances for precursor and fragment ions were set to 5 ppm and 0.02 Da, respectively. Searches were performed without enzyme specificity with variable modifications for cysteinylation (Cys), phosphorylation (Ser, Thr and Tyr), oxidation (Met) and deamidation (Asn, Gln). Raw data files were converted to peptide maps comprising *m/z* values, charge state, retention time and intensity for all detected ions above a threshold of 8,000 counts using in-house software (Proteoprofile)[12]. Peptide maps corresponding to all identified peptide ions were aligned together to correlate their abundances across sample sets and replicates. The MS/MS spectra of unshared MIPs were validated manually.

5.7.8 Identification of MIPs

MIP identification was based on four criteria: i) the canonical MIP length of 8 to 11 amino acids, (ii) the predicted MHC binding affinity given by the NetMHCcons algorithm[47], (iii) the Mascot score, which reflects the quality of peptide assignation, and (iv) the false discovery rate (FDR), which indicates the proportion of decoy (false) vs. target (true) identifications. First, we evaluated the correlation between these parameters. We found a strong correlation (0.88) between FDR values <60% and MHC binding affinity values ≤ 1,750 nM for all 8-11mers (Supplementary Fig. S1). Indeed, the proportion of peptides with an MHC binding affinity ≤ 1,750 nM increases as the FDR decreases (Supplementary Fig. S2a**).** This correlation was specific to MIPs, since no correlation was found for random peptides (Supplementary Fig. S1 and S2b). These results show that low FDR values allow enrichment of high affinity peptides (MHC binding affinity ≤

1,750 nM) and thus of MIPs. However, the drawback of using a stringent low FDR as the main filter is that the total number of identifications considerably decreases (Supplementary Fig. S2a**)** as well as the proportion of small peptides (8-9 mers) identified (Supplementary Fig. S2c**)**. Accordingly, the relative proportion of peptides found in target vs. decoy decreased with increasing peptide length63, in accordance with the notion that short peptides such as MIPs generally require higher Mascot scores to achieve a low FDR. Moreover, the tandem MS fragment ions of MIPs are less predictable and evenly distributed than those of tryptic peptides which further complicate their assignment by database search engines such as Mascot. To set a more suitable Mascot score threshold for high-throughput MIP detection, we evaluated the relation between the Mascot score and the predicted binding affinity for all 8-11 mer peptides identified with an FDR ≤ 5% (Fig. 1c). Then, we calculated the number of MIPs identified with all combinations of Mascot score and predicted binding affinity. We found that the highest number of MIP identifications was obtained by combining a Mascot score ≥ 21 and an MHC binding affinity ≤1,250 nM at a 5% FDR (Fig. 1c).

5.7.9 ns-SNPs found in MIP coding genomic regions in the population

For each MIP, we retrieved the coordinates of the peptide-coding DNA region. These coordinates were then used to extract both the corresponding reference sequence and all non-synonymous validated SNPs reported by dbSNP (Build 137) for that region. For MIPs deriving from multiple source regions, the number of ns-SNPs reported, corresponds to that of the MIP source region possessing the maximal number of ns-SNPs.

5.7.10 Random peptide sampling

We constructed a genome-wide index. To do so, we indexed every coding sequences reported in the Ensembl gene set (GRCh37.65), except for those located in the Y chromosome or the mitochondrial DNA, into a segment tree. Next, we kept only the first layer of the tree and removed the gaps between the indexed regions, effectively transforming the tree into a coding DNA sequence list, which was used for the random peptide sampling. For each of the 4,468

identified peptides, a random peptide of the same length and that fell entirely into a single coding DNA sequence, was chosen. Next, for each randomly selected peptide, we counted the number of ns-SNPs reported in dbSNP137 (validated and missense). The distribution was obtained after repeating the sampling of 4,468 random peptides 10,000 times.

5.7.11 PCR and Sanger sequencing

PCR amplification of the MiHA-encoding DNA and cDNA regions was performed with the Phusion ® High-Fidelity PCR kit (New England BioLabs). For each candidate, 1-2 pairs of sequencing primers were designed manually and with the PrimerQuest software (Integrated DNA Technologies) (Supplementary Table S1), and were synthesized by Sigma. PCR products were purified with the PureLink Quick Gel Extraction Kit (Invitrogen). Sanger sequencing was performed on candidate DNA and cDNA at the IRIC's Genomics Platform. Sequencing results were visualized with the Sequencher software v4.7 (Gene Codes Corporation).

5.7.12 Cytotoxicity assays

Dendritic cells (DCs) were generated from frozen PBMCs, as previously described64. To generate cytotoxic T cells, autologous DCs were irradiated (4,000 cGy), loaded with 2 μM of peptide and cultured for 7 days with freshly thawed autologous PBMCs at a DC:T cell ratio of 1:10. From day 7, responder T cells were restimulated for 7 additional days with irradiated autologous B-LCLs pulsed with the same peptide (B-LCL:T cell ratio 1:5). Expanding T cells were cultured in RPMI 1640 (Invitrogen) containing 10% human serum (Sigma-Aldrich) and L-glutamine. IL-2 (50 U/ml) was added for the last 5 days of the culture. Cytotoxicity assays were performed as described12, with minor modifications. Briefly, B-LCLs were labeled with carboxyfluorescein succinimidyl ester (CFSE) (Invitrogen), extensively washed, irradiated (4,000 cGy) and then used as targets in cytotoxicity assays. Target cells were plated in 96-well U-bottom plates at 5,000 cells/well. Effector cells were added at different effector-to-target ratios in a final volume of 200 μl/well. Plates were centrifuged and incubated for 18h-20h at 37°C. Flow cytometry analysis was performed using

a LSRII cytometer with a high throughput sampler device (BD Biosciences). The percentage of specific lysis was calculated as follows: [(number of CFSE+ cells remaining after incubation with unpulsed target cells – number of CFSE+ cells remaining after incubation with peptide-pulsed target cells) / number of CFSE+ cells remaining after incubation with unpulsed target cells] x100.

5.7.13 Statistical analysis and data visualization

The 2-tailed Student *t*-test was used to identify differentially expressed MIPs and MiHAs that induced cytotoxicity. The 2-tailed Mann-Whitney test was used to compare the MHC binding affinity of unshared MIPs. Differentially expressed transcripts were identified with the DESeq package that uses a model based on the negative binomial distribution62. The Spearman correlation was used to evaluate the relation between differences in MIP abundance and differences in MIP-coding gene or exon expression. The genomic location of identified MIPs including MiHAs and the RNA-seq and exome sequencing coverage were visualized with the Circos software[65] .

5.8 References

1. Rosengarten,R.D. & Nicotra,M.L. Model systems of invertebrate allorecognition. *Current Biology* **21**, R82-R92 (2011).
2. Leinders-Zufall,T. *et al.* MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1033-1037 (2004).
3. Parham,P. & Moffett,A. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nature Reviews Immunology* **13**, 133-144 (2013).
4. Siddle,H.V. *et al.* Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proc. Natl. Acad. Sci. U. S. A* **110**, 5103-5108 (2013).
5. Voskoboynik,A. *et al.* Identification of a colonial chordate histocompatibility gene. *Science* **341**, 384-387 (2013).
6. de Verteuil,D., Granados,D.P., Thibault,P., & Perreault,C. Origin and plasticity of MHC I-associated self peptides. *Autoimmunity Reviews* **11**, 627-635 (2012).
7. Davis,M.M. *et al.* T cells as a self-referential, sensory organ. *Annual Review of Immunology .* **25**, 681-695 (2007).
8. Gilchuk,P. *et al.* Discovering naturally processed antigenic determinants that confer protective T cell immunity. *Journal of Clinical Investigation* **123**, 1976-1987 (2013).
9. Zarling,A.L. *et al.* Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc. Natl. Acad. Sci. U. S. A* **103**, 14889-14894 (2006).
10. Lemmel,C. *et al.* Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling. *Nature Biotechn*ology **22**, 450-454 (2004).
11. Fortier,M.H. *et al.* The MHC class I peptide repertoire is molded by the transcriptome. *The Journal of Experimental Medicine* **205**, 595-610 (2008).
12. Caron,E. *et al.* The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Molecular Systems Biology* **7**, 533 (2011).
13. Illing,P.T. *et al.* Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature* **486**, 554-558 (2012).
14. Croft,N.P. *et al.* Kinetics of antigen expression and epitope presentation during virus infection. *PLoS Pathogens* **9**, e1003129 (2013).
15. Mester,G., Hoffmann,V., & Stevanovic,S. Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. *Cellular and Molecular Life Sciences* **68**, 1521-1532 (2011).
16. Bensimon,A., Heck,A.J., & Aebersold,R. Mass spectrometry-based proteomics and network biology. *Annual Review of Biochemestry* **81**, 379-405 (2012).

17. Yewdell,J.W. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. *Trends in Immun*ology **32**, 548-558 (2011).
18. Neefjes,J., Jongsma,M.L.M., Paul,P., & Bakke,O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823-836 (2011).
19. Milner,E., Barnea,E., Beer,I., & Admon,A. The turnover kinetics of MHC peptides of human cancer cells. *Mol. Cell. Proteomics* **5**, 357-365 (2006).
20. Adamopoulou,E. *et al.* Exploring the MHC-peptide matrix of central tolerance in the human thymus. *Nat Commun.* **4**, 2039 (2013).
21. Weinzierl,A.O. *et al.* Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol. Cell. Proteomics* **6**, 102-113 (2007).
22. Petersdorf,E.W. & Hansen,J.A. New advances in hematopoietic cell transplantation. *Curr. Opin. Hematol.* **15**, 549-554 (2008).
23. The 1000 Genomes Project Consortium An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 55-65 (2012).
24. Granados,D.P. *et al.* MHC I-associated peptides preferentially derive from transcripts bearing miRNA recognition elements. *Blood* **119**, e181-e191 (2012).
25. Warren,E.H. *et al.* Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. *Blood* **120**, 2796-2806 (2012).
26. Wallny,H.J. & Rammensee,H.G. Identification of classical minor histocompatibility antigen as cell-derived peptide. *Nature* **343**, 275-278 (1990).
27. Simpson,E., Roopenian,D., & Goulmy,E. Much ado about minor histocompatibility antigens. *Immunol Today* **19**, 108-112 (1998).
28. Roopenian,D., Choi,E.Y., & Brown,A. The immunogenomics of minor histocompatibility antigens. *Immunol. Rev.* **190**, 86-94 (2002).
29. Spierings,E. *et al.* Phenotype frequencies of autosomal minor histocompatibility antigens display significant differences among populations. *PLoS. Genet.* **3**, e103 (2007).
30. Morse,M.C. *et al.* The COI mitochondrial gene encodes a minor histocompatibility antigen presented by H2-M3. *J. Immunol.* **156**, 3301-3307 (1996).
31. Wang,W. *et al.* Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science* **269**, 1588-1590 (1995).
32. den Haan,J.M. *et al.* The minor histocompatibility antigen HA-1: a diallelic gene with a single amino acid polymorphism. *Science* **279**, 1054-1057 (1998).
33. Simpson,E. & Roopenian,D. Minor histocompatibility antigens. *Curr. Opin. Immunol.* **9**, 655-661 (1997).
34. Zuberi,A.R., Christianson,G.J., Mendoza,L.M., Shastri,N., & Roopenian,D.C. Positional cloning and molecular characterization of an immunodominant cytotoxic

determinant of the mouse *H3* minor histocompatibility complex. *Immunity* **9**, 687-698 (1998).
35. Klein,C.A. *et al.* The hematopoietic system-specific minor histocompatibility antigen HA-1 shows aberrant expression in epithelial cancer cells. *The Journal of Experimental Medicine* **196**, 359-368 (2002).
36. Fontaine,P. *et al.* Adoptive transfer of T lymphocytes targeted to a single immunodominant minor histocompatibility antigen eradicates leukemia cells without causing graft-versus-host disease. *Nat. Med.* **7**, 789-794 (2001).
37. Spierings,E., Wieles,B., & Goulmy,E. Minor histocompatibility antigens - big in tumour therapy. *Trends Immunol.* **25**, 56-60 (2004).
38. Bleakley,M. & Riddell,S.R. Molecules and mechanisms of the graft-versus-leukaemia effect. *Nat. Rev. Cancer* **4**, 371-380 (2004).
39. Meunier,M.C. *et al.* T cells targeted against a single minor histocompatibility antigen can cure solid tumors. *Nat. Med.* **11**, 1222-1229 (2005).
40. Vincent,K., Roy,D.C., & Perreault,C. Next-generation leukemia immunotherapy. *Blood* **118**, 2951-2959 (2011).
41. Warren,E.H. *et al.* Therapy of relapsed leukemia after allogeneic hematopoietic cell transplant with T cells specific for minor histocompatibility antigens. *Blood* **115**, 3869-3878 (2010).
42. Hombrink,P. *et al.* Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. *J. Immunol.* **190**, 3869-3877 (2013).
43. Yewdell,J.W., Reits,E., & Neefjes,J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature Rev. Immunol.* **3**, 952-961 (2003).
44. Perreault,C. The origin and role of MHC class I-associated self-peptides. *Prog. Mol Biol. Transl. Sci.* **92**, 41-60 (2010).
45. Wang,Z., Gerstein,M., & Snyder,M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57-63 (2009).
46. Gebreselassie,D., Spiegel,H., & Vukmanovic,S. Sampling of major histocompatibility complex class I-associated peptidome suggests relatively looser global association of HLA-B*5101 with peptides. *Hum. Immunol.* **67**, 894-906 (2006).
47. Karosiene,E., Lundegaard,C., Lund,O., & Nielsen,M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* **64**, 177-186 (2012).
48. Sherry,S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308-311 (2001).
49. Akatsuka,Y. *et al.* Identification of a polymorphic gene, BCL2A1, encoding two novel hematopoietic lineage-specific minor histocompatibility antigens. *The Journal of Experimental Medicine* **197**, 1489-1500 (2003).
50. Popovic,J. *et al.* The only proposed T-cell epitope derived from the TEL-AML1

translocation is not naturally processed. *Blood* **118**, 946-954 (2011).
51. Robbins,P.F. *et al.* Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nature Medicine* **19**, 747-752 (2013).
52. Princiotta,M.F. *et al.* Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343-354 (2003).
53. de Verteuil,D. *et al.* Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. *Molecular and Cellular Proteomics* **9**, 2034-2047 (2010).
54. Malarkannan,S. *et al.* Differences that matter: major cytotoxic T cell–stimulating minor histocompatibility antigens. *Immunity* **13**, 333-344 (2000).
55. Bleakley,M. *et al.* Leukemia-associated minor histocompatibility antigen discovery using T-cell clones isolated by in vitro stimulation of naive CD8+ T cells. *Blood* **115**, 4923-4933 (2010).
56. Kroemer,G. & Zitvogel,L. Can the exome and the immunome converge on the design of efficient cancer vaccines? *Oncoimmunology.* **1**, 579-580 (2012).
57. Lennerz,V. *et al.* The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. *Proc. Natl. Acad. Sci. U. S. A* **102**, 16013-16018 (2005).
58. Heemskerk,B., Kvistborg,P., & Schumacher,T.N. The cancer antigenome. *EMBO J.* **32**, 194-203 (2013).
59. Zitvogel,L., Galluzzi,L., Smyth,M.J., & Kroemer,G. Mechanism of action of conventional and targeted anticancer therapies: reinstating immunosurveillance. *Immunity* **39**, 74-88 (2013).
60. Tosato,G. & Cohen,J.I. Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines. *Curr. Protoc. Immunol.* **Chapter 7**, Unit 7.22 (2007).
61. Robinson,J.T. *et al.* Integrative genomics viewer. *Nature Biotechn*ology **29**, 24-26 (2011).
62. Anders,S. & Huber,W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
63. Elias,J.E. & Gygi,S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207-214 (2007).
64. Bollard,C.M. *et al.* Complete responses of relapsed lymphoma following genetic modification of tumor-antigen presenting cells and T-lymphocyte transfer. *Blood* **110**, 2838-2845 (2007).
65. Krzywinski,M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639-1645 (2009).

5.9 Acknowledgments

We thank Wafaa Yahyaoui for technical assistance, and the personnel of the following IRIC core facilities: genomics, proteomics, bioinformatics and flow cytometry. We are thankful to Pierre Chagnon and Brian Wilhem for advice and thoughtful comments on exome and transcriptome sequencing. We also thank our blood donors. This work was supported by grants from the Fonds d'Innovation Pfizer-Fonds de Recherche Santé Québec and the Canadian Cancer Society Research Institute (grant # 701564). D.P.G. is supported by a studentship from the Canadian Institutes of Health Research. C.P. and P.T. hold Canada Research Chairs in Immunobiology, and Proteomics and Bioanalytical Spectrometry, respectively. IRIC is supported in part by the Canada Foundation for Innovation, and the Fonds de Recherche Santé Québec.

5.10 Author contributions

D.P.G. and D.S. designed the study, performed experiments, analyzed data, prepared the figures and wrote the first draft of the manuscript. T.D. designed the study, developed pyGeno, performed bioinformatic analyses and contributed to the writing. O.C. and A.Z. developed bioinformatics tools for the analysis of MS data and prepared figures. C.L. performed analyses and prepared a figure. C.C. and M.P.H. performed experiments. G.B. prepared Circos figures and bioinformatics analyses. P.G. performed sequencing mapping and analysis. S.L. designed the study and discussed statistical analyses and results. P.T. and C.P. designed the study, analyzed data, discussed results and wrote the manuscript. All authors edited and approved the final manuscript.

5.11 Additional information

MS/MS spectra: PASS00270 (PeptideAtlas, http://www.peptideatlas.org/), MHC-I peptide sequences: under revision in Immune Epitope Database (http://www.iedb.org/), RNA-seq data: GSE48918 (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/), exome data PRJNA210790 (NCBI Sequence Read Archive, http://www.ncbi.nlm.nih.gov/sra).

5.13 Competing Financial Interests

Université de Montréal has filed a patent related to the research presented in this manuscript.

5.14 Supplementary Information

Impact of Genomic Polymorphisms on the Repertoire of Human MHC Class I-Associated Peptides

Diana Paola Granados, Dev Sriranganadane, Tariq Daouda, Antoine Zieger, Céline M. Laumont, Olivier Caron-Lizotte, Geneviève Boucher, Marie-Pierre Hardy, Patrick Gendron, Caroline Côté, Sébastien Lemieux, Pierre Thibault* & Claude Perreault*

*Corresponding authors

Supplementary Figure S1. Global false discovery rate and predicted binding affinity allow discrimination between MIPs and contaminant (non-MIPs) peptides.

Supplementary Figure S2. The global false discovery rate allows enrichment of MIPs and affects the proportion of small (8-9mers) and long peptides (10-11mers) identified.

Supplementary Figure S3. Comparison of MIPs identified using UniProt vs. personalized protein databases built with next-generation-sequencing data.

Supplementary Figure S4. Quantification of surface HLA-ABC before and after peptide elution.

Supplementary Figure S5. Validation of 6 MiHA-coding sequences by Sanger sequencing.

Supplementary Figure S6. Differential expression of non-polymorphic MIPs

does not correlate with differences in MIP-coding genes or exons between subjects.

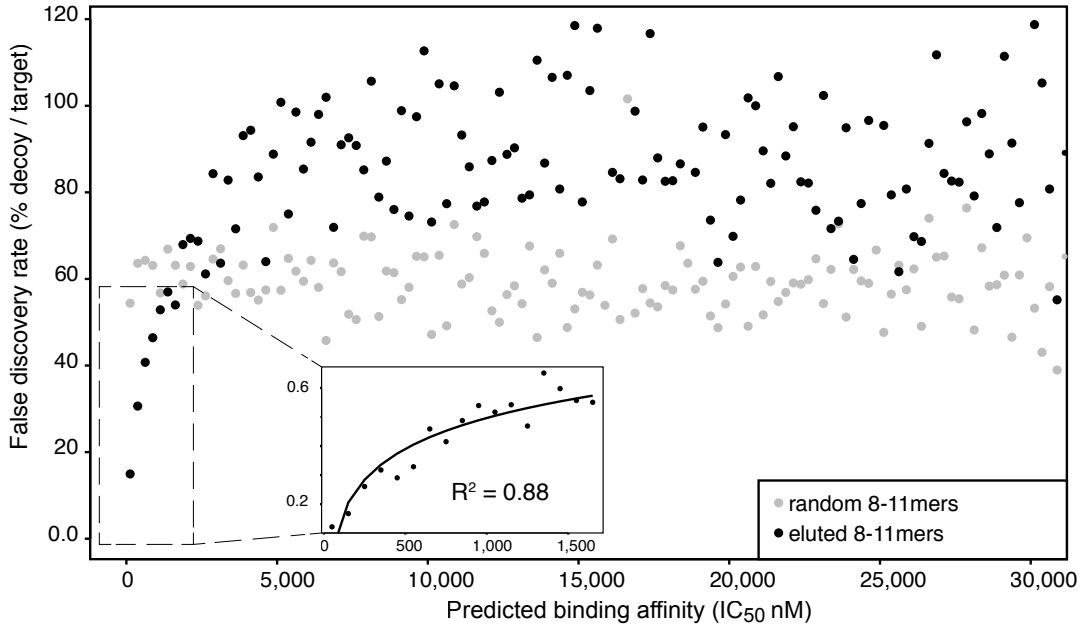
Supplementary Table S1. Exome and transcriptome sequencing and mapping statistics.

Supplementary Table S2. Identified MIPs and their genomic and proteomic features.

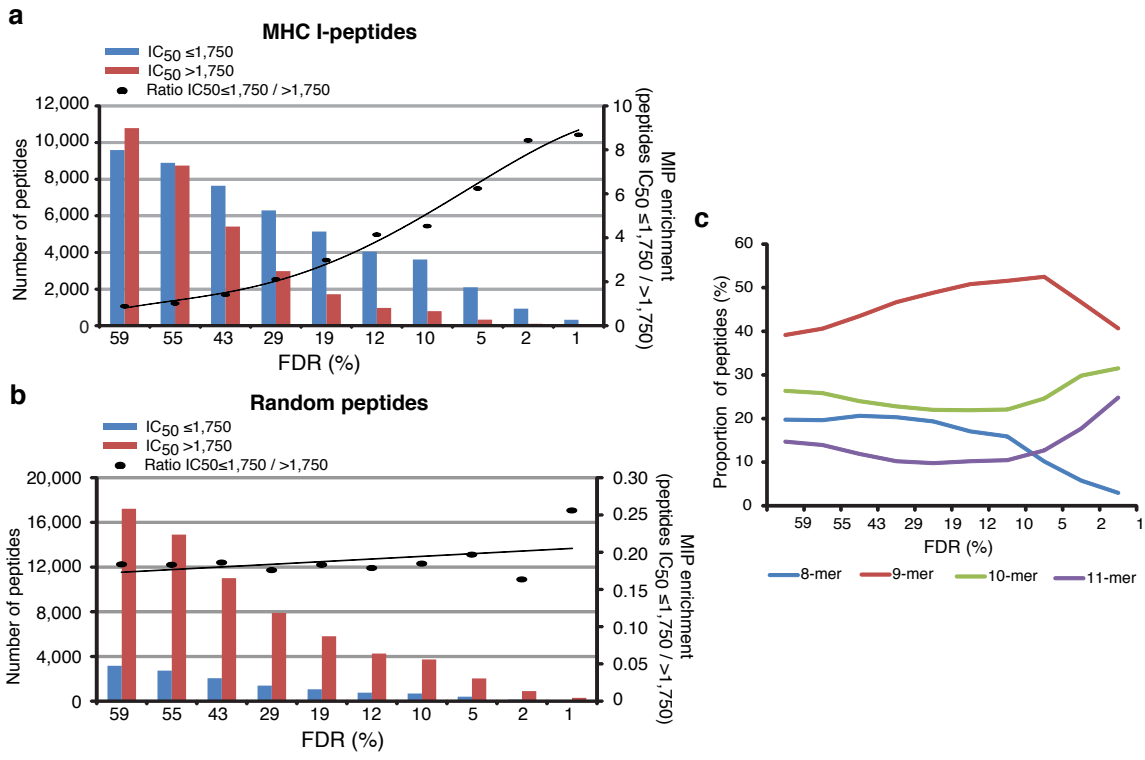
Supplementary Table S3. Transcript abundance in B-LCLs from subjects 1 and 2.

Supplementary Table S4. Identified MIPs coded by genomic regions with one or more validated ns-SNPs (dbSNP).

Supplementary Table S1. List of primers for PCR amplification and Sanger sequencing.

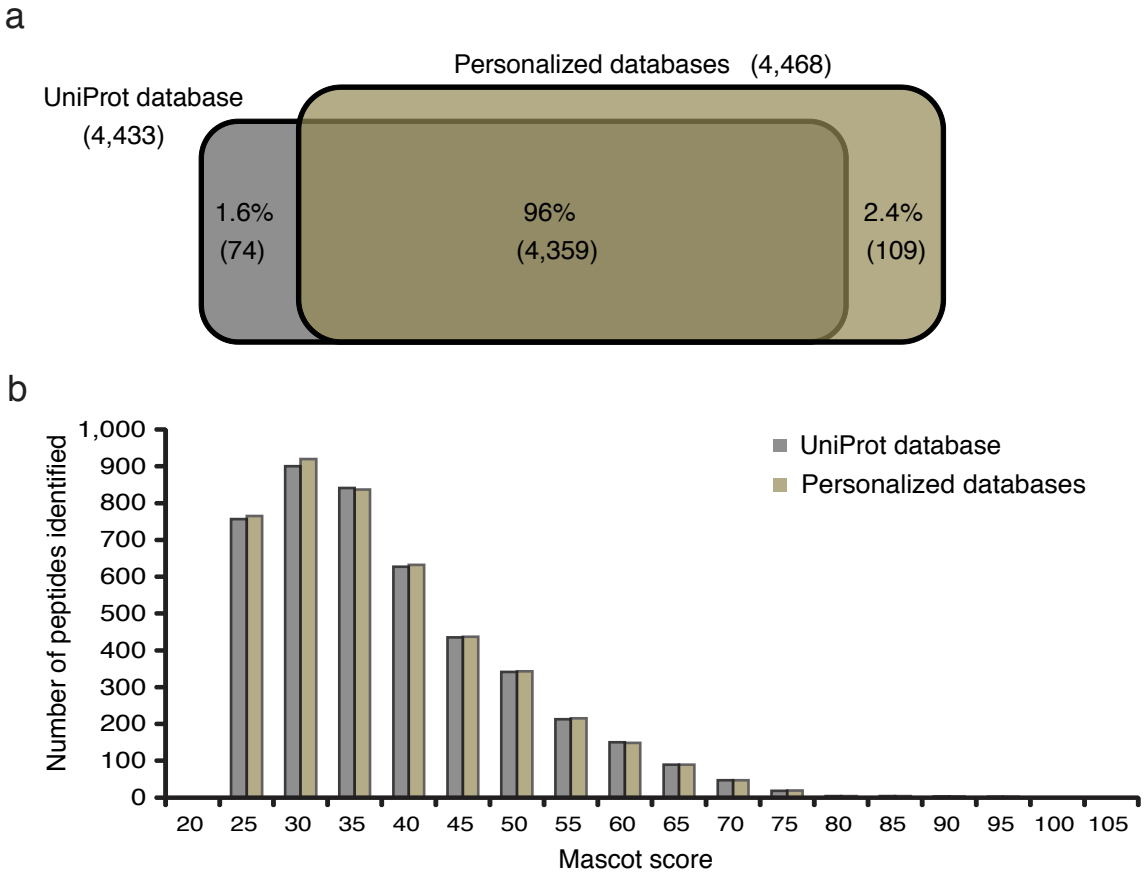


Supplementary Figure S1. Global false discovery rate (FDR) and predicted binding affinity allow discrimination between MIPs and contaminant (non-MIP) peptides
Predicted binding affinity to the relevant HLA molecules of 20,380 eluted peptides (black) and 20,380 random peptides generated from the personalized protein databases of subjects 1 and 2 (grey) was calculated using NetMHCcons. Binding score categories were generated by intervals of 500 nM (IC50). Each dot represents the mean predicted binding affinity for peptides in a given bin. For each category, the level of accuracy in the peptide identification is shown as the FDR. The FDR (database of 20,319 target/12,011 decoy peptides identified by Mascot and defined by the clustering) was calculated for the eluted and the random peptides. The inset shows a high correlation between FDR values < 60% and PBA values < 1,750 nM. The distribution of random peptides shows no correlation between the predicted binding affinity and the FDR.



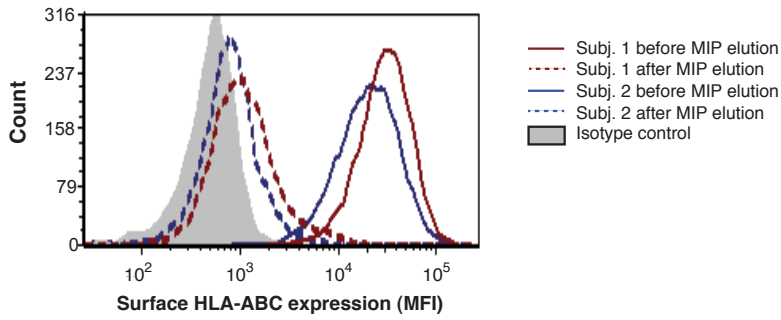
Supplementary Figure S2. The global false discovery rate (FDR) allows enrichment of MIPs and affects the proportion of small (8-9mers) and long peptides (10-11mers) identified

a) We calculated the predicted binding affinity (IC50) for the 8-11mer peptides obtained after applying different FDR thresholds. FDR values were calculated from a dataset of 20,319 eluted peptides (target) and 12,011 reverse peptide versions (decoy). Without filtering any 8-11mer peptide identified by Mascot, the FDR value corresponds to 59% (maximal FDR). Bars show the number of peptides with a predicted binding affinity ≤ 1,750 nM (blue) or > 1,750 nM (red). The second y-axis shows the enrichment in MHC I-peptides calculated as the ratio of peptides with a predicted binding affinity ≤1,750 / >1,750 nM. b) We performed the same analysis by randomly generating the same amount of peptides for each FDR threshold. The figure shows that lower FDR thresholds increase the proportion of eluted peptides with high predicted binding affinity (a) but have no impact on random peptides (b). c) Proportion of 8 – 11mers identified by applying different FDR thresholds. Low FDR values favor the identification of long peptides and disfavor the identification of short peptides.



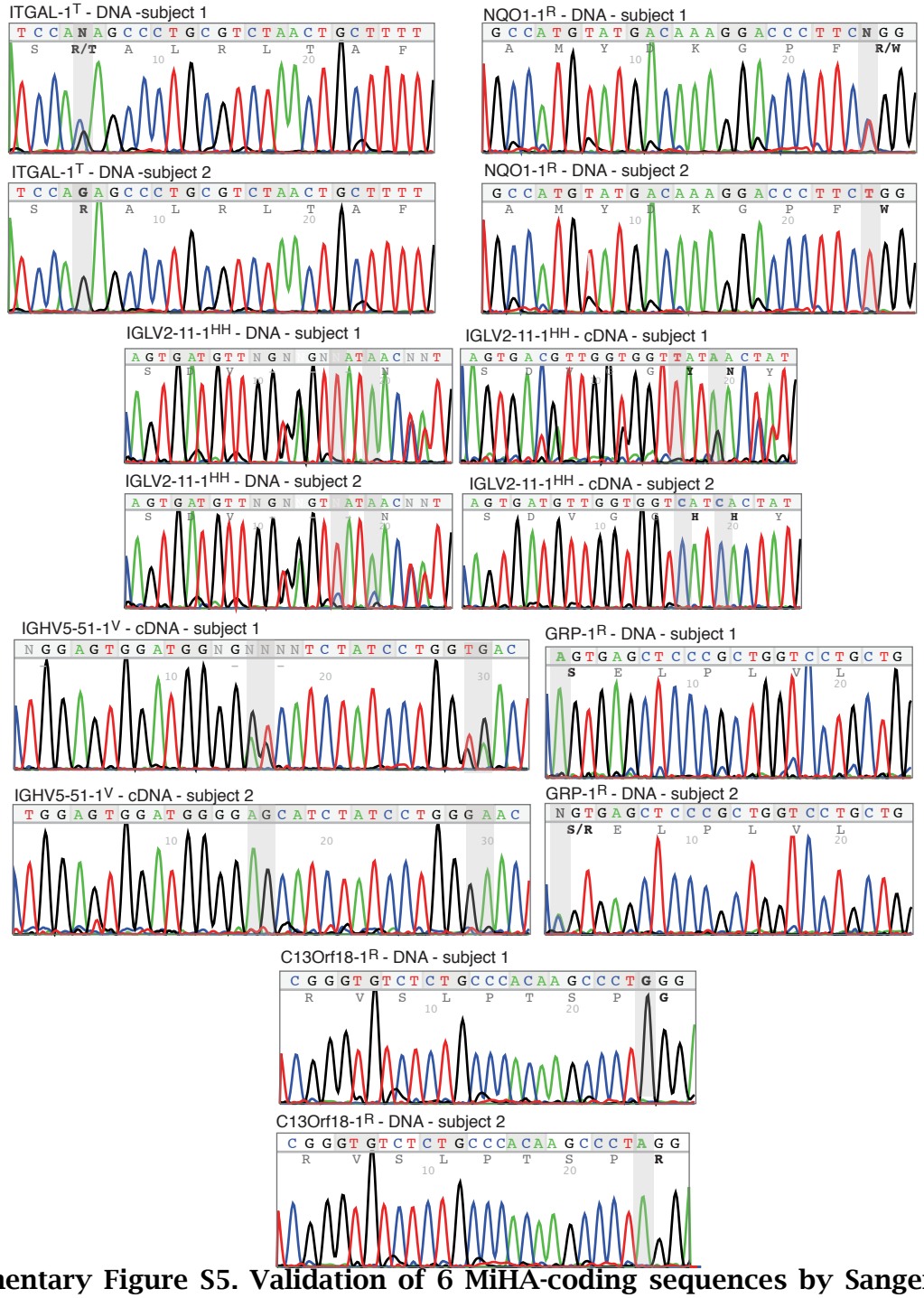
Supplementary Figure S3. Comparison of MIPs identified using UniProt vs. personalized databases built with next generation sequencing data

(a) Venn diagram comparing the number and percentage of unique and common MIP sequences found using UniProt vs. personalized databases. (b) Mascot score of MIPs identified with each database.



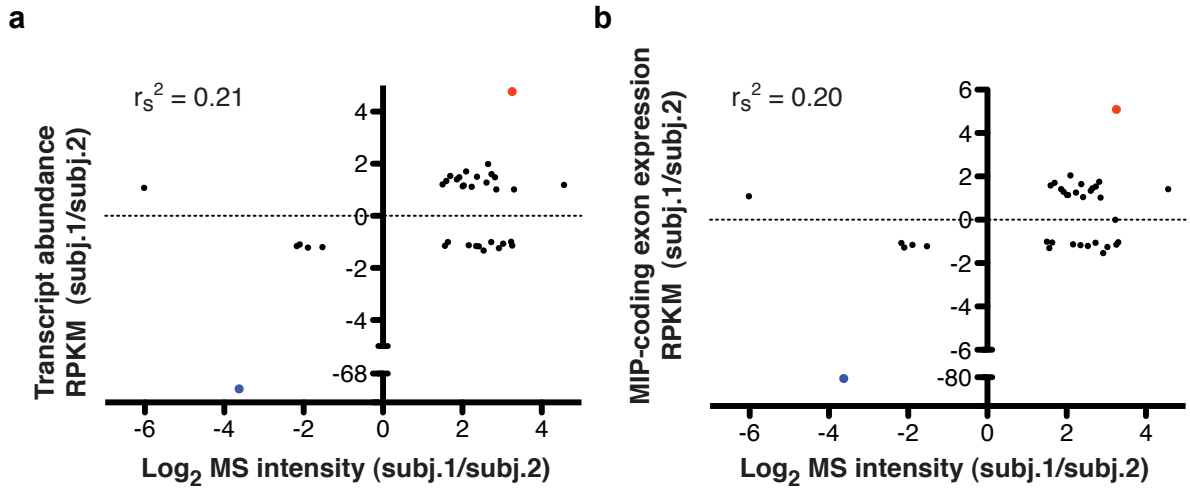
Supplementary Figure S4. Quantification of surface HLA-ABC before and after peptide elution

B-LCLs from subjects 1 and 2 were stained with PE anti human HLA-ABC monoclonal antibody or the corresponding isotypic control and the mean fluorescence intensity (MFI) was analyzed by flow cytometry before or after mild acid elution of peptides. The histogram shows similar levels of HLA-ABC surface expression before and after peptide elution in representative samples of subjects 1 and 2.



Supplementary Figure S5. Validation of 6 MiHA-coding sequences by Sanger sequencing

Chromatograms obtained after Sanger sequencing of PCR-amplified DNA or cDNA encoding 6 MiHAs. The primers used are shown in Supplementary Table S1. Polymorphic loci are highlighted in grey. The IGLV2-11-1^{HH} MiHA results from 2 nucleotide changes at the transcript level in subject 2.



Supplementary Figure S6. Differential expression of non-polymorphic MIPs does not correlate with differences in MIP-coding genes or exons between subjects
For 41 non-polymorphic MIPs that were exclusively detected in one sibling, we calculated the fold difference in intensity of the MIP and the fold difference in expression of the underlying MIP-coding gene (a) or exon (b) measured in Reads Per Kilobase per Million mapped Reads (RPKM). In only 2 cases, MIP abundance differences reflect MIP-coding transcript differences (dots in red and blue). The calculated Spearman r value (rs2) shows no correlation.

5.15 Supplementary Data

Supplementary data 1-4 are available upon request.

CHAPTER 6

6 Discussion

6.1 The influence of specific HLA allelic products on the MHC I immunopeptidome

A subject who is heterozygous for *HLA-A*, *HLA-B* and *HLA-C* genes would express 6 different types of MHC I molecules on the cell surface. Our analyses of the MHC I immunopeptidome of B-LCLs from HLA-disparate families confirmed that the HLA genotype ultimately defines the immunopeptidome of a cell. This result was expected based on the specificities of the various binding motifs [1] and is in accordance with a recent MS-based study of soluble plasma MIPs showing that two individuals with no shared HLA alleles present little more than 10% of the same peptides [2].

Different MHC I molecules not only possess distinct binding motifs, but also interact in different ways with various components of the PLC and TAP [3-5] and this probably influences in some extent the set of peptides that bind each MHC I allelic product. Accordingly, it has been reported that 2 different HLA allelic products can compete in the ER for the presentation of a viral peptide [6]. In line with this, ~9% of MHC I peptides identified in our studies showed predicted promiscuous binding to two or more HLA allotypes of the same subject. Since we elute all the peptides from the cell surface, our method does not enable to determine whether promiscuous peptides were bound to one or all predicted HLA molecules. Nevertheless, we found that most promiscuous MHC I peptides were predicted to bind strongly to one HLA and weakly to another HLA allelic product. We also remarked the presence of overlapping peptides deriving from the same or from different proteins. Based on these observations, one can expect some level of "competition" in the ER between the HLA allotypes for the pool of peptides. For example, let's suppose that the same N-terminal extended precursor peptide "ABCDEFGHIJ" binds "HLA#1" (weakly) and "HLA#2" (strongly) in the ER and that both HLA molecules are expressed at the same level, then the binding affinity and/or other unknown factors should determine whether "HLA#1", "HLA#2" or both HLA molecules will bind and present the peptide and how much of it. Moreover, the presented peptide could be the same or slightly different (e.g. CDEFGHIJ or ABCDEF-

GHIJ) depending on the peptide optimization done in the ER which is adapted according to the HLA molecule [7-9], similar to what was found for a viral peptide [6]. The scenario should be probably different for a subject who only possesses “HLA#2” and lacks “HLA#1”. Would the peptide be less or more abundant? I presume that the repertoire and quantity of peptides presented by a given MHC I molecule will be influenced in some extent by the other HLA alleles expressed by the individual. To test this hypothesis, the immunopeptidome of various individuals who share only 1 HLA allelic product could be compared. Notably, this type of analysis is only possible with high-throughput studies as the one presented here. If this concept is true, then we should not only analyze the peptide coding region and its predicted binding affinity for one given MHC to predict whether a peptide will be presented by a subject, but take into account possible interactions with all other MHC allelic products of the individual.

6.2 The MHC I immunopeptidome is cell-specific and subject–specific

Since the HLA genotype is the most determinant factor in defining the immunopeptidome of a person, one would expect to find very similar immunopeptidomes in HLA-identical subjects. Accordingly, our analyses of the immunopeptidome of HLA-identical siblings showed that 86% of the identified peptides were shared. The shotgun nature of the MS [10] or inter-replicate variations in gene expression could explain the exclusive detection of certain peptides in only one subject. However, 41 out of 4,468 (0.9%) peptides were constantly present in one subject and absent across 6 replicates of the other subject, suggesting that those peptides represent real differences in the immunopeptidomes of these siblings. Of note, the exclusive detection of 13 out of 41 unshared peptides in one of the siblings could be explained by ns-SNVs in the peptide-coding region affecting the peptide binding or perhaps its processing. We ignore the mechanism underlying the exclusive expression of the remaining peptides but it could be due to genomic polymorphisms elsewhere than in the peptide-coding region or to posttranslational or epigenetic mechanisms. Our results show that non-HLA inter-individual differences account for at least 0.9% differences in the repertoire of MHC I peptides of HLA-identical siblings. Hence, our results show that the MHC I immunopeptidome of an in-

dividual is shaped by its genome and ultimately defined by the HLA genotype.

In line with this, we found that the MHC I-peptide repertoires and transcripts source of peptides were different in HLA-disparate families. However, these different sets of transcripts/proteins were functionally interconnected and implicated in the same biologic pathways. Moreover, several immune-specific and B cell-specific functional categories were overrepresented in the set of proteins encoding MHC I peptides in B-LCLs. In contrast, comparison of the immunopeptidome of two different mouse cell types (DCs and thymocytes) showed that their peptide repertoires originated from transcripts with distinct biologic functions that were specific to each cell type (appendix 2) [11]. Hence, collectively these results show that the immunopeptidome is subject-specific regarding the nature of the peptides and cell-specific regarding the functional origin of those peptides.

The identification of the source proteins from which the peptides originate can give us some cues about the factors that govern peptide processing. Moreover, it can stimulate new questions about how the immune system is able to distinguish self from non-self. For instance, it would be very interesting to analyze proteins that generate many different peptides and proteins that generate many copies of the same peptide. Which biological or chemical properties make of them a preferential or efficient source of MHC I peptides? Previous independent studies have found that for specific proteins, some properties such as the degree of hydrophobicity [12,13], the presence of transmembrane domains [12], the presence of degradation signals [14,15], the tendency to misfold [14,16] and the intracellular localization [14,17,18], can enhance peptide presentation. However, other studies have found that some of these features do not affect antigen presentation from other proteins [15]. These studies have been done in particular cellular models and have been observed for specific proteins. We can now explore big datasets such as the ones presented here to determine whether these observations can be generalized to other proteins source of peptides and discover other features that are common to source proteins.

6.3 Influence of the transcriptome on the MHC I immunopeptidome

Our study in human cells helped resolve previous contradicting results regarding the relationship between the transcriptome and the immunopeptidome [19-22]. We found that the MHC I immunopeptidome derived preferentially from abundant transcripts, although transcripts expressed at low levels generated MHC I peptides. This result confirms previous results from our laboratory in mouse thymocytes [21] and is in a agreement with bioinformatic analysis of the expression profile of transcripts coding for MHC I peptides [22]. On the contrary, the transcriptome-immunopeptidome correlation was absent in studies comparing changes in peptide and transcript abundance induced by neoplastic transformation or metabolic stress [20,21,23]. This suggest that under steady-state conditions, the MHC I immunopeptidome is biased toward peptides that derive from abundant transcripts, but that this tendency is absent in stressed cells. Accordingly, we have previously shown that ER stress and the unfolded protein response impaired the presentation of MHC I-associated peptides owing to post-transcriptional mechanisms: inhibition of overall protein synthesis and reduced supply of peptides (appendix 1). Moreover, we showed that the impact was more severe when a model protein source of peptides was localized in the cytosol than in the ER, suggesting that stress could affect in different extents proteins source of peptides depending on their location. Another study from our laboratory has shown that changes in the immunopeptidome induced by metabolic stress (inhibition of the mTOR pathway) were mainly caused by co- and/or post-translational mechanisms [23]. The plasticity of the immunopeptidome could reflect not only metabolic changes but also changes in the mechanisms responsible for the generation of peptides under particular conditions. This could be further investigated in other types of cells and under other types of stresses.

In a related observation, among the functional categories and pathways that were enriched in transcripts source of MIPs, we found terms associated to protein synthesis and transcription. Notably, Toung and coworkers found an enrichment of equivalent GO-term categories in medium and high-expressing genes in an RNA-sequencing analysis of human B-LCLs [24]. Thus, the overrepresentation of some functional categories such as protein synthesis and

transcription in the immunopeptidome of B-LCLs could be reflecting those transcripts source of peptides that are abundant.

Since transcript abundance is the net result of ongoing transcription and transcript degradation [25], it would be necessary to dissociate the effect of active transcription vs. low rate of degradation of transcripts source peptides on the immunopeptidome. We showed that the enrichment for highly abundant transcripts as a source of MHC I peptides is accentuated in the transcriptional hotspot comprised in the 6p21 chromosomal region. This results points towards transcription rate as the major player but further studies are needed to prove this hypothesis (see perspectives). In line with this, it would be of great interest to test whether other transcriptional hotspots would be a preferential source of MHC I peptides or whether this is a unique property of the 6p21 region. A high transcription rate comes with an increased number of errors during transcription and perhaps our results reflect a selective mechanism for choosing RNAs for translation into DRiPs based on quality and not quantity.

6.4 The immunopeptidome preferentially derives from transcripts bearing microRNA response elements (MREs)

We discovered that MIPs derive preferentially from transcripts bearing MREs. This relationship was observed in different cell types of mouse and human origin and hence it appears to be a generalized phenomenon. We further confirmed that the microRNAs (miRNAs) that were predicted to target transcripts source of MHC I peptides were indeed expressed at higher levels in our cells compared to non-lymphoid cell lines.

miRNAs regulate gene expression posttranscriptionally by base-pairing to the mRNA 3'-UTR leading to mRNA decay and/or repression of protein synthesis by mechanisms that are not fully understood [26]. Although we did not explore how mRNA targeting by miRNAs could favor MHC I-peptide presentation, we hypothesize that miRNAs could represent major regulators of the DRiP rate via destabilization of mRNAs. miRNAs could generate DRiPs during the targeted mRNA degradation step and/or during translation inhibition. This idea is supported by previous studies implicating RNA regulation in DRiP genera-

tion [27]. It was reported that a small interfering RNA (siRNA) targeted downstream of a reporter peptide results in production of 5'mRNA products in the absence of full-length mRNA. These products were translated into truncated proteins resulting in enhanced antigen presentation *in vitro* and increased tumor destruction *in vivo* [27]. miRNAs can also repress translation in different ways: inhibition of translation initiation, slow down or inhibition of translation elongation, ribosome drop-off leading to premature termination of translation and co-translational degradation of the nascent polypeptide [26,28-30]. All these events can potentially generate DRiPs. In fact, it has been shown that mRNAs carrying premature stop codons that prevent the production of full-length proteins via the nonsense-mediated decay pathway can produce MHC I peptides by a noncanonical mRNA translation process [31,32]. Similarly, miRNA targeting could increase the number of pioneer translation products produced during translation initiation.

One could argue that if the ribosomes would pre-terminate before reaching the 3' of the ORF or if 5'mRNA products are produced, there should be enrichment for peptides from the N-terminus of the protein (i.e. the 5' of the mRNA coding sequence) in the immunopeptidome. Yewdell and coworkers have recently analyzed the locations of a large set of known viral epitopes within their proteins and found a significant bias in the central region of the proteins [33]. On the contrary, recent bioinformatics analyses made in our laboratory show that the C-terminal portion of proteins are underrepresented in the immunopeptidome of B-LCLs (T. Daouda, unpublished data), which would be in line with our results. Of note, viral infection induce a variety of metabolic changes including ER stress leading to global attenuation of protein synthesis [34,35]. In response, viruses often use internal ribosomal entry sites (IRES) allowing translation initiation in the middle of the mRNA [36]. Therefore, the location of the peptide in the protein source could be different for viral antigens than for self-peptides and this could explain the discrepancies in the location bias found between Yewdell's work and our laboratory.

6.5 Which factors determine the number of MHC I peptides generated by a transcript?

Our results show that the number of MHC I peptides generated by a source transcript can be influenced by at least two factors: the transcript abundance and the presence of MREs in the transcript.

A priori our results seem contradictory, since recognition by miRNAs can lead to translation inhibition and/or to the degradation of the targeted mRNA and probably to decreased transcript abundance. However, it is important to consider that transcript abundance reflects the net effect of ongoing transcription and mRNA degradation. For instance, increased mRNA degradation will be masked by a high transcription rate. Moreover, some mRNA targets of miRNAs can be translationally repressed with little if not change in mRNA abundance [29].

One possibility is that transcript abundance and the presence of MREs in the transcript are related factors. In this scenario, lowly abundant mRNAs that generate MHC I peptides could do so because they are highly targeted by miRNAs (and presumably resulting in a high DRiP rate). If true, we would expect to find more miRNA targets among lowly abundant transcripts that give rise to peptides. We tested this hypothesis and found similar proportions of miRNA targets among low and high abundant transcripts (data not shown). Although this preliminary analysis was based on what is predicted to be targeted and not what is indeed targeted by miRNAs and thus can be further improved with experimental data, it suggests that predicted miRNA targeting cannot explain *per se* how low abundance transcripts generate MHC I peptides. Perhaps, other mechanisms, such as nonsense-mediated decay pathway [37-39] could explain this observation.

Another possibility is that transcript abundance and miRNA targeting constitute features of a "special" set of transcripts that are a preferred source of peptides. This set of transcripts could represent those that are bound by "immunoribosomes" [38,40] in a selective mechanism for choosing RNAs for translation into DRiPs and rapid peptide generation.

6.6 A novel personalized approach for the identification of MHC I-associated peptides including MiHAs

Current high-throughput MS approaches rely on traditional search engines (e.g. Mascot) that match tandem mass spectra sequences against hypothetical spectra generated from known protein sequence databases (e.g. UniProt Human database) [41-43]. This algorithmic strategy limits peptide identification only to reference proteins and is blind to variations in the amino-acid sequence [44]. Hence, conventional MS is inadequate for MiHA detection because it cannot detect polymorphic peptides. Previous efforts have been made in the development of databases suitable for identification of polymorphic peptides [44-46]. However, the inclusion of all reported protein variants considerable increases database size rendering it impractical and leads to the assignment of peptides to sequences that are not necessary present in the analyzed sample (false positives). To overcome these obstacles, we developed a novel strategy that hinges on high-throughput identification of MHC I peptides by tandem MS analyses with a personalized proteome database generated from *in silico* translation of combined exome and transcriptome sequencing data. This approach enable us to identify 4,468 MHC I peptides including 34 MiHAs. This personalized method allows a more precise MS-based identification of any peptide whose DNA or RNA coding region has been sequenced. Although the comparison of the immunopeptidome identified with a reference database vs. the personalized database showed that the number and nature of identified peptides is very similar (96% overlap), we demonstrated that ~50% of identified MiHAs would have been missed in the absence of a personalized database. Moreover, the use of both exome and transcriptome sequencing allows the validation of ns-SNVs in the peptide-coding region with 2 independent approaches.

Both forward and reverse immunology strategies have been used for the identification of MHC peptides including MiHAs (chapter 4). Previous reverse immunology approaches have combined genome sequencing and bioinformatics. A recent study compared ns-SNP differences between HLA-identical siblings using a microarray-based SNP genotyping assay for 10,000 validated ns-SNPs [47]. The protein coding variations unique to one of the siblings were used to predict epitopes that could bind HLA-DRB1*0301 and 10 of these candidates

were tested *in vitro* for DRB1*0301 binding and stimulation of CD4+ T cell responses. Interestingly, only 2 out of 10 candidates could induce CD4+ T cell responses. This means that only 20% of predicted peptide candidates based on genomic differences and binding predictions were truly MiHAs. Similarly, Robins and coworkers synthesized candidate MHC-binding nonamers that were predicted *in silico* based on single mutations identified by exome sequencing of patient-derived melanoma cell lines and tested the peptides for recognition by autologous tumor-infiltrating lymphocytes [48]. They found that only 5% of the predicted peptides were indeed recognized by the tumor infiltrating lymphocytes. These two examples illustrate that reverse immunology approaches based on genomic sequencing and bioinformatic predictions suffer from a high false discovery rate reflecting our limited knowledge and the unpredictability of the antigen-processing pathway. We identified 4,833 ns-SNVs between these siblings of which only 0.5% (26) were detected in the coding region of 22 MiHAs. Of note, identified MiHAs behave as recessive or dominant traits at the cell surface and hence only one of the two possible variants was detected. Moreover, we showed that in several cases the amino acid change considerably decreased the binding affinity of the peptide for the HLA molecule. Although more sensitive MS instruments could enhance MHC I peptide identification, these results suggest that only a tiny fraction of all potential MiHAs that could be generated from each ns-SNV, reach the cell surface. We further examined 6 MiHAs that were coded by an allele present only in one subject and found that all 6 MiHAs induced selective killing by MiHA-specific CTLs (i.e. were immunogenic). Our approach demonstrates the necessity of using not only genomics and bioinformatics, but also proteomics data to identify those “winner peptides” that are successfully processed and presented in a given cell type and thus decrease the number of false discoveries.

There are currently 47 reported MiHAs (table 1, chapter4). Our method enabled the discovery of 34 MiHAs and 536 potential MiHAs in the general population, by analyzing one single couple of HLA-identical siblings. Hence, this strategy could greatly accelerate the development of MiHA-targeted immunotherapy. Importantly, our approach can be further improved at various steps.

6.7 Proposed improvements to our personalized approach: the database

Our approach focused on the identification of MiHAs caused by ns-SNPs in the peptide-coding region. Nevertheless, any genetic polymorphism that qualitatively or quantitatively affects the display of self-peptides at the cell surface could in principle give rise to MiHA disparities. Other type of genomic polymorphisms such as indels (insertions/deletions), gene deletions and copy number variants (CNVs) could be considered.

Gene deletions can be quite frequent without causing any phenotype, especially for genes with high degree of homology [49,50]. For example, the frequency of the homozygous deletion of the *UGT2B17* gene can be more than 80% in some populations [51]. Deletion of the *UGT2B17* gene is responsible for the generation of at least 2 known MiHAs (and probably many other unidentified MiHAs) presented by 3 different HLA molecules. Y chromosome genes can also encode many HY MiHAs, as illustrated by *KDM5D* and *UTY* genes encoding more than 6 distinct HY MiHAs [52]. HY MiHAs could be identified in the future by studying sister-brother siblings. Also, CNVs and short indels, which are quite frequent in the human genome [53,54], could be detected from the exome data using algorithms such as Splitread [50]. Finally, mitochondrial DNA could be sequenced and incorporated in the database. Indeed, peptides of mitochondrial origin can be presented by MHC I molecules and mitochondrial MiHAs have been identified in mice [55,56]. The drawback of including all these types of genetic variations is a considerable increase in database size, since all possible protein combinations are predicted *in silico* from genomic variation. This combinatorial problem gets bigger in highly polymorphic regions where various SNPs (and perhaps indels and CNVs) are contained in a short frame. One alternative would be to search for genetic linkage information in order to reduce the number of combinations.

The *in silico* translation of exome and transcriptome sequences was done according to the canonical ORF. However, it is widely known that cryptic MHC I peptides including MiHAs can derive from products of ARFs [57,58]. Hence, it would be relevant to include all 3 forward and 3 reverse reading frames in

the database to discover not only MiHAs but also cryptic peptides in general. Of note, MiHAs deriving from ARFs have seldom been described [59,60]. This might reflect their scarcity or most probably the difficulty of identification. In line with this, MHC I peptides including MiHAs can also arise from translation of "untranslated" 3'- and 5'-UTRs and from intronic regions [60-62]. In our RNA-seq experiments, a significant proportion of sequenced reads (~10%) mapped to introns. This information as well as sequences from UTR were not analyzed but could be considered in the construction of the database. Interestingly, we identified a set of 700 peptides by MS that the search engine (Mascot) couldn't match to any sequence in our personalized proteomes (data not shown). One possibility is that these peptides correspond to erroneous MS sequencing or to peptides that do not fragment well in the MS and thus have a bad identification quality. Although, a manual inspection of the sequences will be needed to discard this possibility, another explanation will be an "atypical" origin of those peptides. They could represent any type of sequence not contained in the reference proteome: 5' or 3' UTRs, translation products of ARFs or posttranslational protein splicing events [63-66]. The identification of this latter category of peptides is currently the most challenging. We currently ignore how frequently these events occur and what proportion of the MHC I immunopeptidome originates from those non-conventional sources. Their existence suggest that the MHC I immunopeptidome is probable more diverse and larger than anticipated and that many MHC I peptides including MiHAs cannot be predicted merely on genomic polymorphisms.

6.8 Proposed improvements to our personalized approach: peptide elution and MS analyses

The identification of MHC I peptides by MS is relatively challenging given the complexity of the peptide mixture [19,42,67]. This complexity is even higher in samples obtained with mild-acid elution composed of thousands of MHC I and non-MHC I peptides, most of them with homogeneous size [67]. Because MHC I peptides are relatively short, their fragmentation spectra frequently do not allow unambiguous matches to database sequences [43]. Moreover, the fact of having mostly only 1 MHC I peptide representative of each protein in one sample renders protein identification more difficult [42]. Finally, most current MS

approaches have been developed to favor the identification of peptides resulting from trypsin digestion. Tryptic peptides are characterized by the presence of a basic residue (lysine or arginine) at the C-terminus, which favors peptide fragmentation and leads to better MS/MS spectra [67]. MHC I peptides do not necessarily show these features and are fragmented randomly depending on their amino acid composition [67].

A quality score such as the Mascot score is commonly used in proteomics to estimate the quality of the MS/MS identification. However, the score is directly correlated with the peptide length [68], which means that short peptides such as MHC I-peptides would tend to have lower scores despite good spectra quality. We demonstrated that the Mascot score can be combined with the predicted binding affinity to adapt this MS quality measure to MHC I peptides. Our study suggest that the frequently used Mascot score threshold of 35 is too stringent for MHC I peptides, as approximately half of the identified MHC I peptides including MiHAs had Mascot scores <35.

These challenges merit further development of MS algorithms that predict the characteristics of MHC I peptides to enhance their identification. Also, novel MS approaches not relying on proteomic databases such as *de novo* sequencing [69] could enable the identification of MHC I peptides not encoded in the genome such as peptides resulting from posttranslational protein splicing events [66].

We identified four MiHAs in the subject homozygous for the corresponding allele but not in the heterozygous subject. This suggests that zygosity influences MiHA abundance and that low abundance MiHAs may fall below the MS detection threshold in heterozygous subjects. Consistent with this, the MS intensity for these four MiHAs was low in homozygous subjects. Similarly, we assessed the immunogenicity of three MHC I peptides that had identical peptide coding sequences in both subjects but were detected in one sibling and absent in all replicate samples from the other sibling. However, none of these three peptides could elicit CTLs in the MHC I-peptide-negative sibling. This suggests that these peptides were differentially expressed in both siblings and that its abundance was below the MS threshold in one subject. We presently ignore the

number of peptides whose abundance at the cell surface is below the MS detection limit. We believe that we are currently detecting “the top of the iceberg” and that many more peptides are probably presented by MHC I molecules, but they might escape detection because of low abundance and/or relatively low sensitivity in the electrospray ionization process. It becomes necessary to accurately determine the detection threshold of our current MS technique. The concentration of peptides in the ER is supposed to be the rate-limiting step for MHC I-peptide presentation [70]. With the development of more sensitive instruments we will probably be able to precisely answer the question: what is the total amount of peptide complexes that is exposed at the cell surface? Also, we need to apply absolute quantitative approaches to estimate the total number of copies of each specific complex.

MS also allowed the identification of 652 peptides with post-translational modifications including oxidation, deamidation, cysteinylation and phosphorylation that merit further investigation. Although we know that post-translational modifications can affect the processing, presentation or recognition of certain peptides by CD8 T cells [71], our current knowledge of the biological significance of these types of peptides is very limited due to their rarity. In fact, we ignore whether post-translationally modified MHC I peptides are indeed rare or simply not targeted by current MS protocols. Our method to process eluted MHC I peptides could be better adapted to identify post-translationally modified peptides. In the future, it could be interesting to explore and compare the proportion of post-translational modifications in the immunopeptidome and in the proteome. Since most (but not all) post-translational modifications occur post translation, an underrepresentation of posttranslationally modified peptides could give some cues about the preferred source of MHC I peptides. If DRiPs are the major source of MHC I peptides then one would expect to find less post-translational modifications in these peptides in comparison to the proportion found in the proteome. However, it would be also necessary to precisely determine when and where the post-translational modification occurs.

6.9 Impact of genomic polymorphisms on the MHC I immunopeptidome

We studied the effect of non-MHC SNPs on the MHC I immunopeptidome. We

concentrated our analyses to ns-SNPs in the peptide-coding region. Nevertheless, ns-SNPs in the vicinity of the peptide coding region could produce changes in the MHC I immunopeptidome of HLA-identical subjects. For instance, MiHA disparities can be generated by ns-SNPs in the peptide flanking region that affect peptide translocation by TAP or its degradation by the proteasome [72-74]. In addition, polymorphisms can affect the expression of genes *in trans* [75] and this could quantitatively and qualitatively affect the MHC I-peptide repertoire. Polymorphisms in non-coding regions could not only modulate gene expression but also affect alternative splicing, RNA editing and generate cryptic translation products. For example, polymorphisms in MREs can affect recognition by miRNAs with a consequent impact on target transcripts [76].

We concentrated our analyses to SNPs that change the amino acid sequence. Nevertheless, synonymous SNPs (s-SNPs) in the mRNA source of peptides or *in trans* could also affect the immunopeptidome. “Silent” or synonymous codon positions, define mRNA secondary structure and stability and affect the rate of translation, folding and post-translational modifications of nascent polypeptides [77,78]. Since the number of s-SNVs that can be found in the peptide-coding regions of two individuals might lie in the order of thousands, bioinformatic tools could be used to filter those s-SNVs that are predicted to affect mRNA abundance or the corresponding protein. Further biochemical studies will also be needed to understand how s-SNPs affect the immunopeptidome, although this would not be an easy task.

Our results suggest that zygosity influences MiHA abundance, as evidenced by the detection of four MiHAs in the subject homozygous for the corresponding allele but not in the heterozygous subject. However, we ignore whether differences in peptide abundance caused by zygosity could lead to immune responses. Neoplastic transformation can provoke differential expression of peptides or so called tumor-associated antigens, which can be immunogenic [21]. Similarly, it would be of great interest to test whether allelic dosage could be sufficient to cause immunogenicity.

We found that at least 41 MHC I peptides were exclusively detected by MS in one sibling and constantly absent in the other sibling yet the peptide-coding se-

quence was identical in the 2 individuals. With the aim of explaining the origin of these peptide differences, we analyzed various possible mechanisms. First, we did not find any correlation between the exclusive detection of a peptide in a subject and an increased expression of the transcript source of peptide in the same subject. Second, we compared the level of transcript expression of the main known players in the antigen processing pathway between siblings and found no significant differences. Although transcript abundance is a very rough estimate of protein levels, our results would discard quantitative differences in molecules involved in the antigen processing to explain peptide discrepancies in this sibling. Nevertheless, not only quantitative but also qualitative differences in molecules involved in the antigen processing could induce changes in the immunopeptidome. As described in the first chapter, many proteins participate in the antigen processing and presentation pathway and many others are probably unknown. Polymorphisms in one or more of these molecules, especially those in active sites or affecting the conformation of enzymes could affect their interaction with the peptide or their function. For example, ns-SNP in *ERAP1* can affect the protein structure, its cleavage capacity and thus the generation of some peptides [79]. Accordingly, it has been shown that some ERAP1 mutants have a ~40% lower peptidase activity *in vitro* [80] and defects in basic enzymatic properties [81]. In another example, Andrés and coworkers described a polymorphism inducing destruction of ERAP2 mRNA by nonsense-mediated decay leading to very low ERAP2 protein levels and reduction in MHC I surface expression in B-LCLs [82]. Based on these studies and since ERAP1 and ERAP2 are the main aminopeptidases responsible for trimming the N-terminus of peptides, we search for non-synonymous polymorphisms between these subjects. We found 2 ns-SNPs that have been shown to affect the enzymatic activity of ERAP1 (data not shown) and therefore, I presume that polymorphisms in molecules involved in the antigen processing pathway such as ERAP could indirectly change the repertoire of the MHC I peptides.

6.10 MHC I peptides encoded by conserved and polymorphic genomic regions

We explored whether MHC I-associated peptides derive preferentially from in-

variant vs. polymorphic genomics regions. This yet unexplored question is not only biologically important but also very pertinent to clinical transplantation: can we expect MiHAs to be derived from allogeneic polymorphisms throughout the human genome, or are person-to-person differences in select proteins more subject to presentation? Our results showed that the MHC I immunopeptidome is not enriched with polymorphic or invariant peptides. This result was supported by two independent analyses: i) a comparison of the frequency of SNPs/bp in the peptide-coding DNA sequence and in the human exome (chapter 3) and ii) a bootstrapping procedure to compare the number of ns-SNPs (reported in db-SNP) found in 10,000 random samplings of 4,468 hypothetical peptide-coding regions from the human reference exome vs. the number of ns-SNPs found in the regions coding for 4,468 identified peptides (chapter 5). We found that at the population level, 88% of peptide-coding sequences identified were invariant whereas 12% contained at least one ns-SNP: 670 ns-SNPs were found in the genomic region coding for 536 MHC I polymorphic peptides. Our results were confirmed by a recent study in which a similar proportion (10%) of polymorphic MHC I peptides were identified with a immunoprecipitation-based mass spectrometry approach [83].

6.11 Identification of potential MiHAs in the general population

Our approach can be used to discover potential MiHAs in the general population by identifying MHC I peptides that derive from known polymorphic regions using databases such as dbSNP. To choose the best potential MiHAs, one could further select those MHC I peptides whose coding region contains ns-SNPs that have balanced allele frequencies in a population to increase the chance of observing a disparity between donors and recipients. Moreover, the expression of genes source of MiHAs could be verified in different cells and tissues to favor those whose expression is restricted to hematopoietic cells and cancer cells and thereby favor GvL or GvT effect and avoid GvHD. Gene expression could be tested *in silico* using publicly available databases for different cells and tissues such as BioGPS (http://biogps.org/). Moreover, available RNA-seq data could be interrogated to have a more precise idea of the expression of MiHA-encoding transcripts in small subpopulations in each tissue. Importantly, lowly expressed transcripts cannot be totally neglected, as our

results showed that low abundant transcripts can generate MHC I peptides. Thus, although a preliminary *in silico* inspection can be very informative, proteomic confirmation of the presence or absence of a given MiHA in a specific tissue or cell type would be preferable. It would be also important to verify whether the potential MiHAs are encoded by a single gene. Lastly, identification of MiHAs that are promiscuous binders might have a broader application as they could be naturally presented by more than two HLA alleles [84].

6.12 How many human MiHA exist?

The 1000 Genomes Project Consortium has recently reported 38 million SNPs, 1.4 million short indels and more than 14,000 larger deletions in humans [54]. In average, one individual human genome presents 3.6 million SNPs, 344 thousand indels and 717 large deletions[54]. Because in principle, any genetic polymorphism that qualitatively or quantitatively affects the display of self-peptides at the cell surface could give rise to MiHA disparities, these numbers would suggest that a large number of MiHAs mismatches would occur even despite selection for identity at the MHC in a clinical setting. Accordingly, we compared the exome and transcriptome of two HLA-identical siblings and found 4,833 ns-SNVs, most of which corresponded to reported SNPs. This number lies in the range of ns-SNVs found by other groups comparing the exomes of 12 unrelated individuals (~3,600-6,400) [47,85]. Since only a fraction of the proteome is sampled in the immunopeptidome owing to HLA binding selectivity and destruction during antigen processing, only a small fraction of the genetic disparities between two HLA-identical individuals is expected to be reflected in differences in the MHC I peptide repertoire. Our results reflected this selectivity since only 26 (0.5%) were detected in the coding region of 22 MiHAs, of which 13 MiHAs were encoded by an unshared allele.

Theoretical estimates of the number of MiHAs between MHC-identical subjects range from 15 to thousands [86-88]. We addressed this issue experimentally and identified 13 MiHAs disparities in B-LCLs from two HLA-identical siblings. This means that out of 4,468 MHC I peptides present on B-LCLs, 13 (0.3%) would be immunogenic for one of these siblings. This number could vary depending on the genes expressed by a given cell type and the genetic similarity

of the 2 individuals. The number of MiHAs between HLA-identical subjects might be probably much more greater since we restricted our analyses to MiHAs resulting from ns-SNPs in the peptide-coding region. Indels, CNV as well as s-SNPs and ns-SNPs inside and outside the peptide-coding region might considerably increase this number. Moreover, more sensitive MS instruments and better sequencing coverage could enable identification of more MHC I peptides (and thus MiHAs) as well as more polymorphisms, respectively.

Perspectives

Study of the MHC I immunopeptidome from a systems biology perspective

Our work illustrates the utility of systems biology approaches to study the composition of the MHC I immunopeptidome as well as the "rules" or general patterns underlying its biogenesis. In line with this, the *Human Immunopeptidome Project* has been recently proposed with the aim of analyzing the full repertoire of HLA-associated peptides in health and disease and understanding the molecular mechanisms involved in its formation [89].

In ecological sciences the biodiversity index of an ecosystem is often estimated to monitor changes or perturbations in the ecosystem's functionality [90,91]. In immunopeptidomics, an analogous definition of a diversity index would be: a quantitative measure that reflects how many different peptides compose the immunopeptidome of a given cell under specific conditions, and simultaneously takes into account how evenly the number of copies per peptide are distributed. The value of a diversity index would increase both when the number of peptides increases and when evenness increases. Thus, for a given number of peptides, the value of a diversity index would be maximized when all peptides are equally abundant. With the development of more quantitative MS approaches, it would be feasible to calculate the diversity index of the immunopeptidome and compare and study its biological significance (if any) under steady-state conditions and following intrinsic and extrinsic perturbations for different cells or tissues.

Since multiple events are involved in the generation of MHCI peptides including MiHAs, I personally believe that approaches combining the analysis of various functional genomic levels (genomic, transcriptomic, proteomic, degradomic, peptidomic) could be very meaningful. Inclusion of bioinformatics tools and methods was crucial in our work and will be indispensable in this kind of studies. Although our results contributed to a better understanding of the genome-immunopeptidome and transcriptome-immunopeptidome relationships, further experiments are needed to dissociate the contribution of actively

transcribed genes vs. that of very stable transcripts to the immunopeptidome. For example, RNA Polymerase II CHIP-seq transcription profiling could be used to measure transcription rates as a function of RNA Polymerase II occupancy across the genome, enabling the measurement of transcription rates without the influence of RNA half-life. Similarly, to systematically test the concept that the immunopeptidome samples what is being translated as opposed to what has been translated, genome-wide measurements of occupancy and density of ribosomes on mRNA [92-94] could be performed simultaneously with analysis of the immunopeptidome. It would be also of great interest to study the global contribution of the degradome to the immunopeptidome using quantitative proteomics [95]. Finally, the impact of epigenetic regulation on the MHC I immunopeptidome has never been addressed. Studies of the immunopeptidome and the "methylolme" of identical twins could reveal some surprises. These studies could be well complemented with molecular assays to validate data-driven hypotheses.

Model to unravel the factors that control which MHC I peptides are displayed and in which quantity

Defining the nature of endogenous peptides contributes not only to the understanding of MHC I antigen presentation but to other central aspects of cell biology such as gene expression, protein synthesis and degradation. Our findings can contribute to the optimal design of CD8+ T cell vaccines for tumors and pathogens and could be also applied in the autoimmunity field. One interesting model to study the intrinsic factors that promote MHC I peptide generation are medullary thymic epithelial cells (mTECs). mTECs have the unique ability to transcribe otherwise tissue-restricted genes through the action of the transcription factor AIRE as well as AIRE-independent factors [96]. Accordingly, RNA-seq experiments performed in our laboratory have recently shown that more genes are expressed in medullary thymic epithelial cells leading to increased average transcript abundance in comparison to other cell populations [97]. This promiscuous gene expression leads to presentation of tissue-restricted antigens by mTECs and is instrumental during negative selection of T lymphocytes [96]. One would predict a high "diversity index" in the immunopeptidome of mTECs, i.e. the presence of a huge variety of peptides with

little variation in the number of copies per peptide. Furthermore, these cells would need an efficient mechanism for MHC I peptide generation or a specialized DRiP apparatus. This apparatus could be a highly active miRNAome or a specific set of miRNAs that efficiently act on transcripts leading to DRiPs formation and facilitating the generation of MHC I peptides. Accordingly, recent studies have shown that AIRE can control the expression of miRNAs that characterize mature mTECs [98,99]. These features made of mTECs a very interesting model to study the biogenesis of the immunopeptidome.

MiHAs in personalized therapy

We have shown that the immunopeptidome is subject-specific and cell-specific. In this context, I envisage personalized therapies, where information of genetic differences between donor and recipients are combined with proteomic studies for specific cell types and tissues to predict the outcome of transplantation in different donor-recipient scenarios and thereby facilitate selection of the most optimal donor. However, some remaining obstacles need urgently to be bypassed before translation to the clinics. The first one is the identification of cryptic MiHAs generated by non-conventional processes and not encoded in the genome. *De novo* sequencing MS approaches seem promising in this regard [69]. The second challenge is the development of methods to discover MHC II-bound MiHAs. To date, very few MHC II-associated MiHAs are known despite renewed efforts in their identification. Because MHC II molecules are not expressed on most non-hematopoietic tissues under non-inflammatory conditions, MHC II-associated MiHAs are less likely to cause GvHD [100]. These antigens could be very useful if presented on leukemic or cancer cells expressing MHC II. Lastly, MS-based identification still requires huge amount of starting material (500 mill cells for the strategy based on mild acid elution). This amount is even higher for cells that express low surface levels of MHC and for immunoprecipitation-based MS approaches. We are still far from a very sensitive technique that will allow quantification of peptides from small populations without the need of *in vitro* expansion and immortalization methods, such as infection with EBV. More sensitive approaches will definitely allow the characterization of the immunopeptidome in many more cell types and the application of approaches such as ours in the clinic.

Conclusion

We employed systems biology approaches to unravel the composition and cellular origin of the self MHC I-associated peptide repertoire presented by B-LCLs derived from 2 pairs of HLA-identical siblings.

First, we found that HLA-different subjects present different immunopeptidomes derived from different sets of proteins. These proteins were functionally interconnected, implicated in the same biological pathways and conveyed to the cell surface a cell type-specific signature. Our results showed that the human MHC I immunopeptidome is subject-specific and cell-specific.

Second, our analyses revealed that the MHC I immunopeptidome i) showed no bias toward conserved vs. polymorphic genomic sequences, ii) derived preferentially from abundant transcripts and iii) was enriched in transcripts bearing MREs. Furthermore, while the MHC I immunopeptidome of HLA-disparate subjects is coded by different sets of transcripts, these transcripts are regulated by mostly similar miRNAs. Our data support an emerging model in which the generation of MIPs by a transcript depends on its abundance and its regulation by miRNAs.

Finally, we developed a novel personalized approach combining mass-spectrometry, next-generation sequencing and bioinformatics for high-throughput identification of MHC I peptides including MiHAs caused by ns-SNPs in the peptide-coding region. We discovered 34 MiHAs in a pair of HLA-identical siblings and 536 potential MiHAs in the general population. Our results showed that MiHAs, which are encoded by biallelic loci, behaved as dominant or recessive traits at the cell surface. Comparison of the genomic landscape of the MHC I-peptide repertoires identified in these siblings revealed that i) 0.5% of ns-SNVs were represented in the immunopeptidome and ii) 0.3% of the MHC I-peptide repertoire would be immunogenic for one of the siblings.

Our results are relevant for the design of peptide-based immunotherapeutics. In addition, our MHC I-peptide identification strategy could greatly accelerate the development of MiHA-targeted immunotherapy. Finally, our exome, tran-

scriptome, miRNAome and immunopeptidome datasets of B-LCLs represent a new resource for other scientists interested in inter-individual differences and therefore is of great valuable for the scientific community.

References

1. Lin, H. *et al.* (2008) Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunolo*gy 9, 8
2. Bassani-Strenberg, M. *et al.* (2010) Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc.Natl.Acad.Sci.U.S.A* 107, 18769-18776
3. Zernich, D. (2004) Natural HLA Class I Polymorphism Controls the Pathway of Antigen Presentation and Susceptibility to Viral Evasion. *The Journal of Experimental Medicine* 200, 13–24
4. Sieker, F. *et al.* (2008) Differential tapasin dependence of MHC class I molecules correlates with conformational changes upon peptide dissociation: A molecular dynamics simulation study. *Molecular Immunology* 45, 3714–3722
5. Neefjes, J.J. *et al.* (1993) Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science* 261, 769–771
6. Tussey, L.G. *et al.* (1995) Different MHC class I alleles compete for presentation of overlapping viral epitopes. *Immunity* 3, 65–77
7. Turnquist, H.R. *et al.* (2002) Disparate binding of chaperone proteins by HLA-A subtypes. *Immunogenetics* 53, 830–834
8. Turnquist, H.R. *et al.* (2000) HLA-B polymorphism affects interactions with multiple endoplasmic reticulum proteins. *European Journal of Immunology* 30, 3021–3028
9. H, H.W. *et al.* (2002) HLA Class I Polymorphism Has a Dual Impact on Ligand Binding and Chaperone Interaction. *Human Immunology* 63, 248–255
10. Pachl, F. *et al.* (2013) Characterization of a high field Orbitrap mass spectrometer for proteome analysis. *Proteomics* 13, 2552–2562
11. de Verteuil, D. *et al.* (2010) Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. *Molecular and Cellular Proteomics* 9, 2034–2047
12. Huang, L. *et al.* (2011) Hydrophobicity as a driver of MHC class I antigen processing. *The EMBO Journal* 30, 1634–1644
13. Istrail, S. *et al.* (2004) Comparative immunopeptidomics of humans and

their pathogens. *Proc.Natl.Acad.Sci.U.S.A* 101, 13268–13272
14. Caron, E. *et al.* (2005) The structure and location of SIMP/STT3B account for its prominent imprint on the MHC I immunopeptidome. *International Immunology* 17, 1583–1596
15. Golovina, T.N. *et al.* (2005) The impact of misfolding versus targeted degradation on the efficiency of the MHC class I-restricted antigen processing. *The.Journal of Immunology* 174, 2763–2769
16. Ostankovitch, M. *et al.* (2005) Regulated Folding of Tyrosinase in the Endoplasmic Reticulum Demonstrates That Misfolded Full-Length Proteins Are Efficient Substrates for Class I Processing and Presentation. *The. Journal of Immunology* 174, 2544–2551
17. Kang, T.H. *et al.* (2006) Enhancement of dendritic cell-based vaccine potency by targeting antigen to endosomal/lysosomal compartments. *Immunology Letters* 106, 126–134
18. Lacaille, V.G. and Androlewicz, M. (2000) Targeting of HIV-1 Nef to the Centrosome: Implications for Antigen Processing. *Traffic* 1, 884–891
19. Mester, G. *et al.* (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. *Cellular and Molecular Life Sciences* 68, 1521–1532
20. Weinzierl, A.O. *et al.* (2006) Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface. *Molecular & Cellular Proteomics* 6, 102–112
21. Fortier, M.H. *et al.* (2008) The MHC class I peptide repertoire is molded by the transcriptome. *Journal of Experimental Medicine* 205, 595–610
22. Juncker, A.S. *et al.* (2009) Systematic Characterisation of Cellular Localisation and Expression Profiles of Proteins Containing MHC Ligands. *PLoS ONE* 4, e7448
23. Caron, E. *et al.* (2011) The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Molecular Systems Biology* 7, 533
24. Toung, J.M. *et al.* (2011) RNA-sequence analysis of human B-cells. *Genome Research* 21, 991–998
25. Perez-Ortin, J.E. *et al.* (2013) Eukaryotic mRNA Decay: Methodologies, Pathways, and Links to Other Stages of Gene Expression. *Journal of Molecular Biology* 425, 3750–3775

26.Fabian, M.R. *et al.* (2010) Regulation of mRNA translation and stability by microRNAs. *Annual Review of Biochemestry* 79, 351–379
27.Gu, W. *et al.* (2009) Both treated and untreated tumors are eliminated by short hairpin RNA-based induction of target-specific immune responses. *Proc.Natl.Acad.Sci.U.S.A* 106, 8314–8319
28.Filipowicz, W. *et al.* (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews Genetics* 2008, 102–114
29.Hendrickson, D.G. *et al.* (2009) Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA. *Plos.Biology* 7, e1000238
30.Huntzinger, E. and Izaurralde, E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics* 12, 99–110
31.Apcher, S. *et al.* (2011) Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. *Proc.Natl.Acad.Sci.U.S.A of the National Academy of Sciences* 108, 11572–11577
32.Apcher, S. *et al.* (2012) The role of mRNA translation in direct MHC class I antigen presentation. *Current Opinion in Immunology* 24, 71–76
33.Kim, Y. *et al.* (2013) Positional bias of MHC class I restricted T-cell epitopes in viral antigens is likely due to a bias in conservation. *Plos Comput ational Biology* 9, e1002884
34.Gleimer, M. and Parham, P. (2003) Stress Management: MHC Class I and Class I-like Molecules as Reporters of Cellular Stress. *Immunity* 19, 469–477
35.Marciniak, S.J. and Ron, D. (2006) Endoplasmic Reticulum Stress Signaling in Disease. *Physiological Reviews* 86, 1133–1149
36.Baird, S.D. (2006) Searching for IRES. *RNA* 12, 1755–1785
37.Schweingruber, C. *et al.* (2013) Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica and Biophysica Acta* 1829, 612–623
38.Yewdell, J.W. and Nicchitta, C.V. (2006) The DRiP hypothesis decennial: support, controversy, refinement and extension. *Trends in Immunology* 27, 368–373

39. Pastor, F. *et al.* (2010) Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. *Nature* 465, 227–230
40. Dolan, B.P. *et al.* (2011) Distinct pathways generate peptides from defective ribosomal products for CD8+ T cell immunosurveillance. *The.Journal of Immunology* 186, 2065–2072
41. Walther, T.C. and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology. *The Journal of Cell Biology* 190, 491–500
42. Purcell, A.W. and Gorman, J.J. (2004) Immunoproteomics: Mass spectrometry-based methods to study the targets of the immune response. *Molecular and Cellular Proteomics* 3, 193–208
43. Engelhard, V.H. (2012) The contributions of mass spectrometry to understanding of immune recognition by T lymphocytes. *International Journal of Mass Spectrometry* 259, 32–39
44. Edwards, N.J. (2007) Novel peptide identification from tandem mass spectra using ESTs and sequence database compression. *Molecular Systems Biology* 3, 102
45. Schandorff, S. *et al.* (2007) A mass spectrometry-friendly database for cSNP identification. *Nature Methods* 4, 465–466
46. Nijveen, H. *et al.* (2010) HSPVdb—the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands. *Immunogenetics* 63, 143–153
47. Stone, B. *et al.* (2012) Identification of novel HLA class II target epitopes for generation of donor-specific T regulatory cells. *Clinical Imm*unology 145, 153–160
48. Robbins, P.F. *et al.* (2013) Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nature Medicine* 19, 747–752
49. McCarroll, S.A. *et al.* (2005) Common deletion polymorphisms in the human genome. *Nature Publishing Group* 38, 86–92
50. Karakoc, E. *et al.* (2012) Detection of structural variants and indels within exome data. *Nature Methods* 9, 176–178
51. Terakura, S. *et al.* (2005) A UGT2B17-positive donor is a risk factor for higher transplant-related mortality and lower survival after bone marrow transplantation. *British Journal of Haematology* 129, 221–228
52. Ofran, Y. *et al.* (2010) Diverse patterns of T-cell response against mul-

tiple newly identified human Y chromosome-encoded minor histocompatibility epitopes. *Clinical Cancer* Research 16, 1642–1651
53. Redon, R. *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444, 444–454
54. Abecasis, G.R. *et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65
55. Loveland, B. *et al.* (1990) Maternally transmitted histocompatibility antigen of mice: a hydrophobic peptide of a mitochondrially encoded protein. *Cell* 60, 971–980
56. Morse, M.C. *et al.* (1996) The COI mitochondrial gene encodes a minor histocompatibility antigen presented by H2-M3. *The.Journal of Immunology* 156, 3301–3307
57. Starck, S.R. and Shastri, N. (2011) Non-conventional sources of peptides presented by MHC class I. *Cellular and Molecular Life Sciences* 68, 1471–1479
58. Bullock, T.N. and Eisenlohr, L.C. (1996) Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. *The Journal of Experimental Medicine.* 184, 1319–1329
59. Dolstra, H. *et al.* (1999) A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. *The Journal of Experimental Medicine.* 189, 301–308
60. Torikai, H. *et al.* (2004) A novel HLA-A*3303-restricted minor histocompatibility antigen encoded by an unconventional open reading frame of human TMSB4Y gene. *The.Journal of Immunology* 173, 7046–7054
61. Griffioen, M. *et al.* (2012) Identification of 4 novel HLA-B*40:01 restricted minor histocompatibility antigens and their potential as targets for graft-versus-leukemia reactivity. *Haematologica* 97, 1196–1204
62. Tykodi, S.S. *et al.* (2008) C19orf48 encodes a minor histocompatibility antigen recognized by CD8+ cytotoxic T cells from renal cell carcinoma patients. *Clinical Cancer Research* 14, 5260–5269
63. Warren, E.H. *et al.* (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* 313, 1444–1447
64. Hanada, K.-I. *et al.* (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 427, 252–

256
65. Vigneron, N. (2004) An Antigenic Peptide Produced by Peptide Splicing in the Proteasome. *Science* 304, 587–590
66. Hanada, K.-I. and Yang, J.C. (2005) Novel biochemistry: post-translational protein splicing and other lessons from the school of antigen processing. *Journal of Molecular Medicine* 83, 420–428
67. Fortier, M.-H. (2009), Développement de méthodes analytiques pour la protéomique et l'identification de peptides MHC I issus de cellules leucémiques. Université de Montréal
68. Elias, J.E. and Gygi, S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* 4, 207–214
69. Allmer, J. (2011) Algorithms for the de novo sequencing of peptides from tandem mass spectra. *Expert Review of Proteomics* 8, 645–657
70. Neefjes, J. *et al.* (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology* 11, 823–836
71. Engelhard, V.H. *et al.* (2006) Post-translational modifications of naturally processed MHC-binding epitopes. *Current Opinion in Immunology* 18, 92–97
72. Craiu, A. *et al.* (1997) Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc.Natl.Acad.Sci.U.S.A* 94, 10850–10855
73. Nussbaum, A.K. *et al.* (1998) Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc.Natl.Acad. Sci.U.S.A* 95, 12504–12509
74. Brickner, A.G. *et al.* (2001) The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. *The Journal of Experimental Medicine.* 193, 195–206
75. Small, K.S. *et al.* (2011) Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nature Genetics* 43, 561–564
76. Gartner, J.J. *et al.* (2013) Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc.Natl.Acad.Sci.U.S.A* 110, 13481–13486

77. Shabalina, S.A. *et al.* (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res* 41, 2073–2094
78. Gingold, H. and Pilpel, Y. (2011) Determinants of translation efficiency and accuracy. *Molecular Systems Biology* 7, 1–13
79. van Endert, P. (2011) Post-proteasomal and proteasome-independent generation of MHC class I ligands. *Cellular and Molecular Life* Sciences 68, 1553–1567
80. Evans, D.M. *et al.* (2011) Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nature Genetics* 43, 761–767
81. Evnouchidou, I. *et al.* (2011) Cutting Edge: Coding Single Nucleotide Polymorphisms of Endoplasmic Reticulum Aminopeptidase 1 Can Affect Antigenic Peptide Generation In Vitro by Influencing Basic Enzymatic Properties of the Enzyme. *The Journal of Immunology* 186, 1909–1913
82. Andrés, A.M. *et al.* (2010) Balancing Selection Maintains a Form of ERAP2 that Undergoes Nonsense-Mediated Decay and Affects Antigen Presentation. *PLoS Genetics* 6, e1001157
83. Hassan, C. *et al.* (2013) The Human Leukocyte Antigen-presented Ligandome of B Lymphocytes. *Molecular and Cellular Proteomics* 12, 1829–1843
84. Mommaas, B. *et al.* (2002) Identification of a novel HLA-B60-restricted T cell epitope of the minor histocompatibility antigen HA-1 locus. *The. Journal of Immunology* 169, 3131–3136
85. Ng, S.B. *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276
86. Perreault, C. *et al.* (1990) Minor histocompatibility antigens. *Blood* 76, 1269–1280
87. Roopenian, D. *et al.* (2002) The immunogenomics of minor histocompatibility antigens. *Immunological Reviews* 190, 86–94
88. Lindahl, K.F. (1991) Minor histocompatibility antigens. *Trends Genetics* 7, 219–224
89. Admon, A. and Bassani-Sternberg, M. (2011) The Human Immunopeptidome Project, a suggestion for yet another postgenome next big thing. *Molecular and Cellular Proteomics* 10, O111.011833–O111.011833

90. Hooper, D.U. and Vitousek, P.M. 06-Aug-(1997), The Effects of Plant Composition and Diversity on Ecosystem Processes. *Science*. [Online]. Available: http://www.sciencemag.org/content/277/5330/1302.short. [Accessed: 06-Oct-2013]
91. Naeem, S. *et al.* (1994) Declining biodiversity can alter the performance of ecosystems. *Nature* 368, 734–737
92. Ingolia, N.T. *et al.* (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223
93. Ingolia, N.T. *et al.* (2013) Genome-wide annotation and quantitation of translation by ribosome profiling. *Current Protocols in* Molecular Biology 4, Unit 4.18
94. Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
95. Larance, M. *et al.* (2013) Global Subcellular Characterization of Protein Degradation Using Quantitative Proteomics. *Molecular and Cellular Proteomics* 12, 638–650
96. Klein, L. *et al.* (2009) Antigen presentation in the thymus for positive selection and central tolerance induction. *Nature Reviews Immunology* 9, 833–844
97. St-Pierre, C. *et al.* (2013) Transcriptome sequencing of neonatal thymic epithelial cells. *Scientific Reports* 3, 1860
98. Ucar, O. *et al.* (2013) An evolutionarily conserved mutual interdependence between Aire and microRNAs in promiscuous gene expression. *European Journal of Immunology* 43, 1769–1778
99. Macedo, C. *et al.* (2013) Autoimmune regulator (Aire) controls the expression of microRNAs in medullary thymic epithelial cells. *Immunobiology* 218, 554–560
100. Bleakley, M. and Riddell, S.R. (2011) Exploiting T cells specific for human minor histocompatibility antigens for therapy of leukemia. *Immunology and Cell Biology* 89, 396–407

APPENDIXES

APPENDIX 1

ER stress affects processing of MHC class I-associated peptides

Diana P. Granados[1], Pierre-Luc Tanguay[2], Marie-Pierre Hardy[1], Étienne Caron[1], Danielle de Verteuil[1], Sylvain Meloche[2], Claude Perreault[1]*

1Department of Medicine, Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Canada, 2Department of Pharmacology, Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Canada

* Corresponding author

This article was submitted for publication to **BMC Immunology** on November 28, 2008, accepted on February 16, 2009 and published online on February 16, 2009.

BMC Immunology, Volume 10:10 (2009)

Abstract

Background. Viral infection and neoplastic transformation trigger endoplasmic reticulum (ER) stress. Thus, a large proportion of the cells that must be recognized by the immune system are stressed cells. Cells respond to ER stress by launching the unfolded protein response (UPR). The UPR regulates the two key processes that control major histocompatibility complex class I (MHC I)-peptide presentation: protein synthesis and degradation. We therefore asked whether and how the UPR impinges on MHC I-peptide presentation.

Results. We evaluated the impact of the UPR on global MHC I expression and on presentation of the H2Kb-associated SIINFEKL peptide. EL4 cells stably transfected with vectors coding hen egg lysozyme (HEL)-SIINFEKL protein variants were stressed with palmitate or exposed to glucose deprivation. UPR decreased surface expression of MHC I but did not affect MHC I mRNA level nor the total amount of intracellular MHC I proteins. Impaired MHC I-peptide presentation was due mainly to reduced supply of peptides owing to an inhibition of overall protein synthesis. Consequently, generation of H2Kb-SIINFEKL complexes was curtailed during ER stress, illustrating how generation of MHC I peptide ligands is tightly coupled to ongoing protein synthesis. Notably, the UPR-induced decline of MHC I-peptide presentation was more severe when the protein source of peptides was localized in the cytosol than in the ER. This difference was not due to changes in the translation rates of the precursor proteins but to increased stability of the cytosolic protein during ER stress.

Conclusion. Our results demonstrate that ER stress impairs MHC I-peptide presentation, and that it differentially regulates expression of ER- vs. cytosol-derived peptides. Furthermore, this work illustrates how ER stress, a typical feature of infected and malignant cells, can impinge on cues for adaptive immune recognition.

Background

The ultimate role of the immune system in host defense is to eliminate infected and transformed cells [1, 2]. A fundamental feature of infected and neoplastic cells is that they are stressed cells [3-5]. In line with this, the innate immune system uses receptors such as NKG2D to recognize stressed cells [4, 6, 7]. One key question, however, is whether cellular stress can influence recognition of transformed or infected cells by the adaptive immune system [4, 8].

The single feature uniting different stress stimuli (heat shock, hypoxia, viral replication, abnormal proteins, starvation or transformation) is that they all ultimately lead to accumulation of unfolded or misfolded proteins in the lumen of the ER [4, 5]. Infection and neoplastic transformation increase protein translation and thereby the folding demand on the ER [9, 10]. This is particularly true for cells submitted to hypoxia, nutrient deprivation or low pH in poorly vascularized bulky tumors, metastases and sites of inflammation [11, 12]. Moreover, acquisition of numerous mutations during tumor progression leads to accumulation of abnormal proteins with an increased propensity to misfolding that further raises the ER folding burden [3, 13].

The ER responds to the accumulation of unfolded proteins by activating intracellular signal transduction pathways, collectively called the unfolded protein response (UPR) [14, 15]. The UPR is a highly conserved adaptive response that allows survival to limited stress but leads to apoptosis in the presence of overwhelming stress [16, 17]. Mammalian UPR acts through three main transducers (PERK, ATF6 and IRE1) that are activated by dissociation of the master chaperone BiP/GRP78 [5, 15]. Activation of PERK leads to phosphorylation of the translation initiation factor eIF2α and attenuation of cap-dependent translation [18]. The endonuclease activity of IRE1 generates a frameshift splice variant of XBP-1 encoding an active transcription factor that activates genes involved in protein degradation and controls the transcription of chaperones [19-21]. Targets of the cleaved active form of ATF6 include the chaperones BiP and GRP94, and the transcription factors XBP-1 and CHOP [17, 19]. Activation of these UPR transducers has pervasive effects on cellular protein economy: i) attenuation of protein translation, ii) increased degradation of ER proteins

by ER-associated degradation (ERAD), iii) transcriptional activation of genes involved in the folding machinery of the ER and iv) increased degradation of ER-localized mRNAs [14, 22].

Presentation of MHC I-associated peptides to CD8 T cells is tightly linked to protein economy. MHC I peptides are preferentially generated from newly synthesized but rapidly degraded polypeptides relative to slowly degraded proteins [23, 24]. Following proteasomal degradation, peptides are translocated into the ER where they undergo N-terminal trimming, loading onto MHC I/β2-microglobulin (β2m) heterodimers and export at the cell surface [25-29]. Since the UPR regulates the two key processes that shape MHC I peptide processing (protein translation and degradation) we reasoned that ER stress should impinge on MHC I peptide presentation. We addressed this question and found that MHC I presentation was impaired during ER stress induced by palmitate or glucose starvation. Moreover, ER stress differentially affected presentation of peptides derived from a protein localized in the ER vs. the cytosol.

Results

Engineering of Kb-SIINFEKL stable transfectant cell lines

Evidence suggests that subcellular localization of a protein (e.g., in the cytosol vs. the secretory pathway) may influence MHC I presentation of peptides derived from that specific protein [30-32]. Moreover, the UPR is primarily orchestrated to decrease protein overload in the ER [14, 15]. We therefore wished to determine whether the UPR would differentially affect MHC I presentation of peptides derived from a precursor protein located in the cytosol versus the ER. To this end, we created stable EL4 transfectant cell lines expressing a chimeric protein located either in the ER or the cytoplasm (Figure 1A). We selected the EL4 thymoma cell line as a model because it expresses relatively high levels of MHC I [33] which allows us to assess changes of MHC I abundance over a wide dynamic range. To create the chimeric constructs, a minigene coding for the SIINFEKL peptide was fused to previously described plasmids encoding hen egg lysozyme (HEL) targeted to the ER or the cytosol [34, 35] (see methods). The ovalbumin-derived SIINFEKL peptide is presented by H2Kb and cell sur-

face expression of Kb-SIINFEKL complexes was assessed by staining with the 25-D1.16 monoclonal antibody [36]. As shown in Figure 1B, EL4 stably transfected clones, denoted EL4/HEL-ER-SIINFEKL and EL4/HEL-Cyto-SIINFEKL, can process and present SIINFEKL derived from an ER-localized or a cytosolic chimeric protein, respectively. These two clones, which display similar amounts of Kb-SIINFEKL at the cell surface, were used in further experiments.

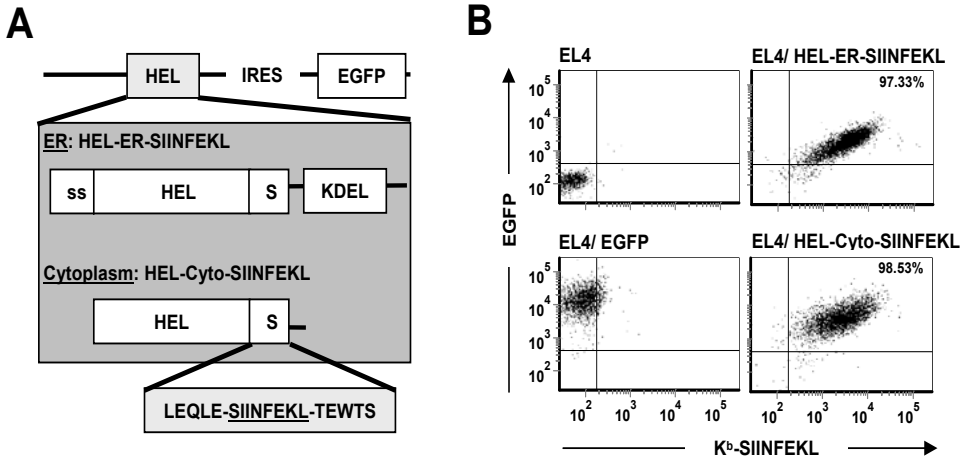


Figure 1. EL4 stable transfectants express the SIINFEKL peptide derived from HEL targeted to the ER or to the cytosol

(A) Schematic representation of the constructs used to generate EL4 stable transfectants. Modified coding sequences of HEL [34, 35] in frame with the region coding for the ovalbumin-derived peptide SIINFEKL and its flanking region were cloned into the pIRES-EGFP2 vector. HEL-ER-SIINFEKL possesses HEL N-terminal signal sequence (ss) and the ER-retention signal KDEL and targets HEL to the ER; HEL-Cyto-SIINFEKL lacks the N-terminal signal sequence and the ER-retention signal and targets HEL to the cytoplasm (see materials and methods). (B) EL4 stable transfectants express Kb-SIINFEKL at the cell surface. EL4 cells were transfected with the pIRES-EGFP2 vector encoding HEL-ER-SIINFEKL or HEL-Cyto-SIINFEKL. Stable transfectants were selected by repeated cycles of FACS of EGFP-positive cells combined with drug resistance (1000 µg/ml of G418). Cells were stained with 25-D1.16 monoclonal antibody, recognizing the Kb-SIINFEKL complex, followed by staining with APC-conjugated anti-mouse IgG1 as secondary antibody. Depicted in the graphs are EGFP and Kb-SIINFEKL MFI values of untransfected EL4 cells (*upper left*), EGFP-transfected cells (*lower left*) and the two representative clones that were used in further studies: EL4/HEL-ER-SIINFEKL (*upper right*) and EL4/HEL-Cyto-SIINFEKL (*lower right*). Percentages represent the proportion of cells expressing EGFP and Kb-SIINFEKL.

UPR activation impairs MHC I surface expression

Various pharmacological agents are widely used to induce ER stress. For instance, tunicamycin and dithiothreitol are known to cause ER stress by preventing N-linked glycosylation or disrupting disulfide bond formation in the ER, respectively [37, 38]. However, since MHC I proteins are glycosylated and contain disulfide bonds, we surmised that tunicamycin and dithiothreitol would directly hinder the assembly of MHC I molecules. We elected to use more physiological ER stress stimuli that should have less drastic effects on the synthesis of MHC I molecules: palmitate and glucose starvation. Palmitate is a saturated fatty acid recently shown to cause ER stress by disrupting mainly the structure and integrity of the ER [39-41]. Palmitate is abundant in the 'high fat Western diet', which renders this type of stress more physiological [42]. Glucose starvation is a common condition present for instance in vascularized bulky tumors and metastases, and is also a prototypical and strong inducer of ER stress [43].

Activation of the UPR was monitored by quantitative real-time reverse transcriptase polymerase chain reaction (RT-qPCR) analysis of BiP, CHOP and the normal and spliced XBP-1 transcripts, which are known to be induced during ER stress [19, 44]. As expected, treatment of both EL4 transfectants, EL4/HEL-Cyto-SIINFEKL and EL4/HEL-ER-SIINFEKL (data not shown), with palmitate for 18 hours induced a mild UPR that was similar in both EL4 cell lines and of lesser magnitude than that induced by tunicamycin stimulation (Figure 2A). Similarly, we monitored UPR induction in EL4 cell lines grown in high glucose (4.5 mg/ml), low glucose (1 mg/ml) or no glucose-containing medium for different time durations (Figure 2B). BiP, XBP-1 and CHOP transcripts were significantly induced in both EL4 cell lines when they were completely deprived of glucose for 18 or 24 hours, indicating activation of the UPR under these conditions. However, none of these UPR markers were upregulated in cells grown in low glucose-containing medium, suggesting that 1 mg/ml of glucose is sufficient to keep the homeostasis of the ER in EL4 cells. The notable point here is that glucose starvation for 18-24h induced a robust UPR that seemed to be of greater magnitude than that induced by palmitate (Figures 2A and B). Thus, ER stress induced by palmitate treatment or glucose starvation activates

the UPR in EL4 cells, albeit to different extents.

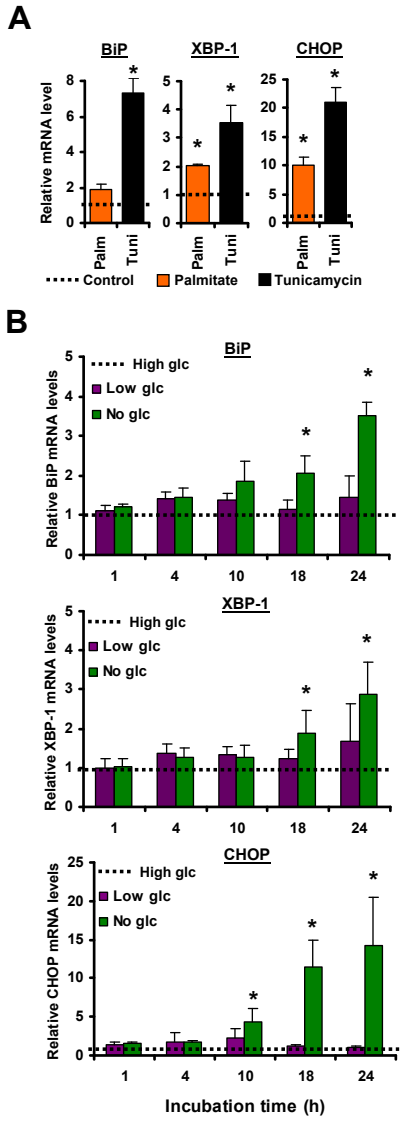


Figure 2. Induction of ER stress in EL4 cells

(A) UPR activation induced by palmitate treatment. EL4/HEL-Cyto-SIINFEKL cells were either non-treated or treated with 0.25 mM of palmitate or 2.5 µg/ml of tunicamycin for 18 hours. BiP, XBP-1 and CHOP mRNA levels were analyzed by RT-qPCR. Expression levels were normalized to the endogenous control gene β-actin. Transcript levels of treated cells were compared with basal mRNA values of untreated cells (*dotted line*), which were set to 1. (B) UPR activation induced by glucose deprivation. EL4 stable cell lines were incubated in DMEM medium lacking glucose or containing low glucose (1 mg/ml) or high glucose (4.5 mg/ml) for different durations. BiP, XBP-1 and CHOP mRNA levels were analyzed by RT-qPCR. Expression levels were normalized to the endogenous control gene β-actin. Transcript levels of cells incubated under low

(purple) or no glucose *(green)* were compared to levels of cells grown in high glucose medium *(dotted line),* which were set to 1. Similar results were obtained with EL4/HEL-ER-SIINFEKL cells (data not shown). Bars represent the mean and SD from three independent experiments performed in triplicate. $*P < 0.05$ when comparing untreated with palmitate- or tunicamycin-treated cells, or high glucose with low glucose or no glucose conditions.

To evaluate the effect of the UPR on MHC I expression, we quantified by flow cytometry surface levels of H2Kb and H2Db in both EL4 cell lines submitted to ER stress (Figure 3). Cells in later apoptotic stages were excluded from the analysis by gating on propidium iodide-negative cells. Activation of the UPR with palmitate reduced cell surface expression of H2Db and H2Kb by 30-40% in both cell lines (Figure 3A). Likewise, we evaluated whether UPR induced by glucose deprivation also affected MHC I surface expression. EL4 stable cell lines were incubated in medium lacking glucose or containing low glucose (1 mg/ml) or high glucose (4.5 mg/ml) for 18 hours and MHC I surface levels were measured by flow cytometry (Figure 3B). MHC I expression was impaired in cells grown both in low or no glucose conditions, albeit to a different extent. Cells that were completely deprived of glucose expressed only 25-30% of normal H2Kb and H2Db levels, similar to the decline produced by tunicamycin (not shown). On the contrary, cells incubated in low glucose medium were less affected since around 70–90% of normal H2Kb and H2Db levels were detected. Of note, a glucose dose of only 1 mg/ml was sufficient to raise MHC I levels by around three-fold (compare no-glucose with low glucose conditions in Figure 3B). As observed in the case of palmitate treatment, glucose starvation caused a similar downregulation of MHC I in the two stable cell lines (Figure 3B). These results show that ER stress induced by glucose deprivation or palmitate treatment causes decreased expression of surface MHC I molecules in EL4 cells.

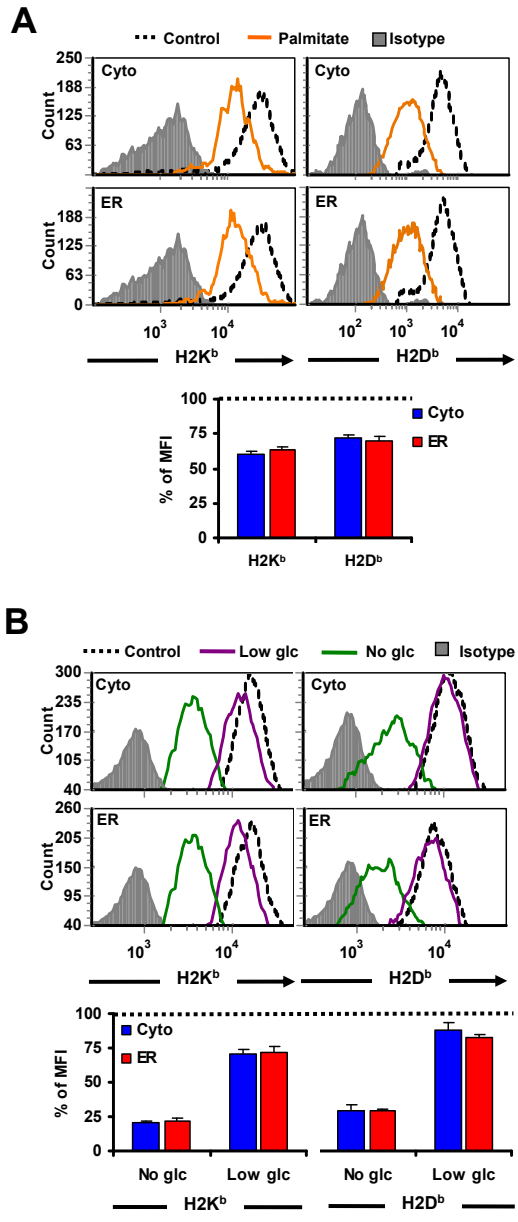


Figure 3. ER stress impairs MHC I surface expression

(A) Decreased MHC I surface expression induced by palmitate treatment. EL4 cells were either non-treated (*dotted line*) or treated with 0.25 mM of palmitate (*orange*) for 18 hours. EL4/HEL-Cyto-SIINFEKL (*top*) and EL4/HEL-ER-SIINFEKL *(bottom*) cells were stained with antibodies against H2Kb, H2Db or the corresponding isotypic control and analyzed by flow cytometry. Representative histograms of one of three independent experiments are depicted. Bars represent % of MFI intensity in treated EL4/HEL-Cyto-SIINFEKL (*blue*) and EL4/HEL-ER-SIINFEKL *(red*) cells relative to untreated cells (*dotted line*). Differences between untreated and treated cells are all significant ($P <$

0.05). (B) Decreased MHC I surface expression induced by glucose deprivation. EL4 cells were incubated in medium lacking glucose (*green*) or containing low glucose (1 mg/ml) (*purple*) or high glucose (4.5 mg/ml) (*dotted line*) for 18 hours and analyzed as in *A*. Bars represent % of MFI intensity in glucose-deprived EL4/HEL-Cyto-SIINFEKL (*blue*) and EL4/HEL-ER-SIINFEKL *(red*) cells relative to untreated cells (*dotted line*). Bars represent the mean and SD from three independent experiments performed in triplicate. Differences between control and glucose-deprived cells are all significant ($P < 0.05$).

Posttranscriptional mechanism(s) cause decreased expression of surface MHC I molecules during ER stress

Since the UPR blocks transcription of numerous genes and can provoke premature degradation of mRNAs encoding secreted or membrane proteins [22], we investigated whether decreased MHC I surface expression was due to downregulation of MHC I transcripts. Using RT-qPCR, we found that mRNA expression levels of H2Kb, H2Db and β2m were unaffected in glucose-deprived or palmitate-treated cells (Figure 4A). In fact, the abundance of the β2m transcript, whose protein is essential for the formation of stable MHC I-peptide complexes, tended to increase in stressed cells relative to control cells (although this increase was not statistically significant). We therefore conclude that UPR induced with palmitate or glucose starvation leads to posttranscriptional attenuation of cell surface MHC I molecules.

To test whether diminished MHC I upon ER stress occurred only at the cell surface, we quantified total MHC I protein amount from whole lysates of cells previously treated with palmitate or deprived of glucose. We found that none of these conditions affected the steady state level of MHC I (Figure 4B). Nevertheless, one of the consequences of UPR activation is attenuation of protein synthesis [18]. Thus, we tested whether the UPR could impact on synthesis of MHC I in metabolically-labeled EL4 cells previously subjected to glucose deprivation or palmitate treatment for 18 hours. We found that glucose starvation, and to a much lesser extent palmitate, curtailed the synthesis of new MHC I molecules by around 40% and 5%, respectively (Figure 4C). Of note, the MHC I band in the no-glucose condition migrated faster than the bands in the control and the palmitate conditions. This effect is likely due to incomplete

glycosylation of MHC I molecules in glucose-deprived cells. Thus, under our experimental conditions, ER stress did not affect the level of MHC I transcripts nor the total amount of MHC I protein (Figure 4A and B), but decreased the synthesis of new MHC I molecules (Figure 4C) and the amount of MHC I molecules at the cell surface (Figure 3).

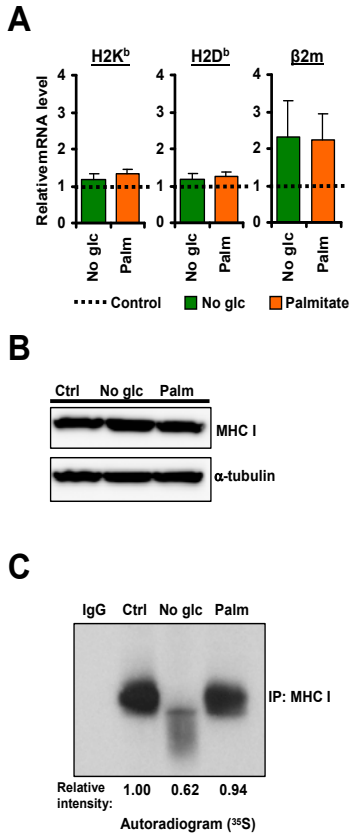


Figure 4. ER stress impairs cell surface MHC I expression through posttranscriptional mechanism(s)

EL4 cells were incubated in DMEM control medium containing glucose (4.5 mg/ml), or in medium lacking glucose or supplemented with 0.25 mM of palmitate for 18 hours. (A) ER stress does not decrease MHC I mRNA levels. H2Kb, H2Db and β2m mRNA levels were assessed and analyzed by RT-qPCR. Expression levels were normalized to the endogenous control gene β-actin. Transcript levels of glucose-starved (*green*) or palmitate-treated (*orange*) cells were compared with basal mRNA values of control cells (*dotted line*), which were set to 1. Bars represent the mean and SD from three independent experiments performed in triplicate. No significant differences were detected between untreated and treated cells ($P < 0.05$). (B) ER stress does not affect total MHC I protein amount. MHC I proteins from whole cell lysates were detected by Western

blot with anti-MHC I antibodies. α-tubulin was used as loading control. A representative image of three independent experiments is shown. (C) ER stress differentially affects synthesis of MHC I. EL4 cells were incubated in control conditions, deprived of glucose or treated with 0.25 mM of palmitate for 17 hours and pulse-labeled with [35S]methionine/ [35S]cysteine for 1 hour. Cell extracts were lysed and subjected to immunoprecipitation with anti-MHC I antibody or the corresponding isotypic antibody. Immunoprecipitated proteins were separated by SDS-PAGE and analyzed by fluorography. One representative experiment out of two is shown.

Decreased overall protein synthesis hinders MHC I-peptide presentation during ER stress

To understand how ER stress decreased cell surface expression of MHC I proteins, we evaluated the impact of ER stress on surface expression of a variety of glycoproteins (Figure 5). As shown before, palmitate treatment and glucose starvation severely impacted MHC I surface level. In contrast, surface expression of glycoproteins CD32, CD45.2, TCR-β and CD5 (Ly1) was minimally or not affected. These results show that the deleterious impact of the UPR is more severe on surface MHC I expression than on other glycoproteins. This suggests that reduction in the amount of cell surface MHC I molecules during ER stress cannot be attributed solely to defective MHC I synthesis. That contention is further supported by two elements. First, a 5% decline of MHC I synthesis in palmitate-treated cells (Figure 4C) is not commensurate with a 30-40% reduction of MHC I molecules at the cell surface (Figure 3A). Second, the total amount of MHC I proteins was not affected in stressed cells (Figure 4B), suggesting that MHC I proteins were relatively stable and that they did not reach the cell surface because they were not properly loaded with their peptide cargo and were therefore retained in the ER.

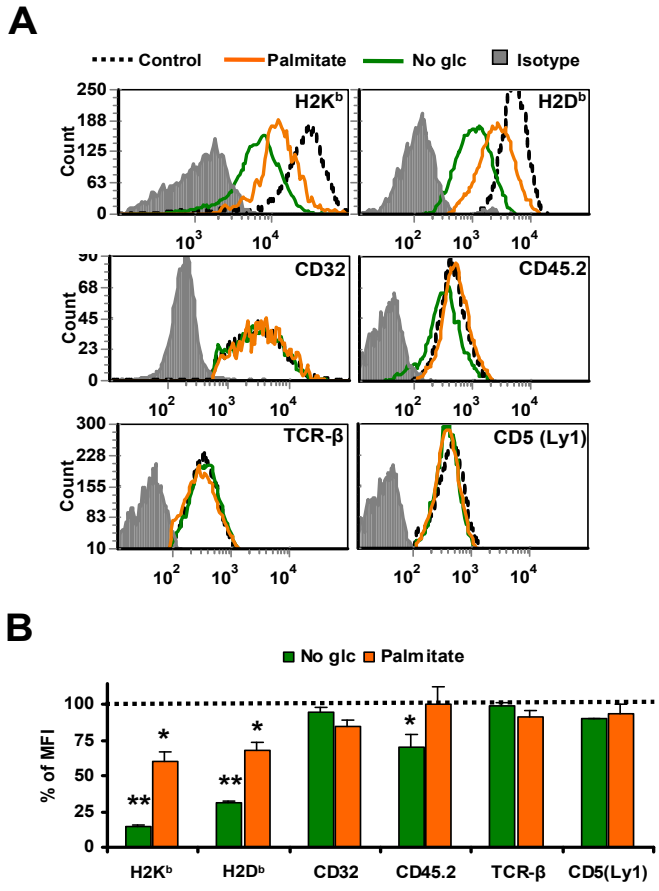


Figure 5. Differential effects of ER stress on surface expression of various glycosylated proteins

EL4 cells were incubated in high glucose (4.5 mg/ml) medium (*dotted line*) or in medium lacking glucose (*green*) or supplemented with 2.5 mM of palmitate (*orange*) for 18 hours. (A) Effect of glucose starvation or palmitate treatment on surface expression of glycosylated proteins. Surface expression of H2Kb, H2Db, CD32, CD45.2, TCR-β and CD5 (Ly1) was determined by flow cytometry analysis. Representative histograms of one of three independent experiments are depicted. (B) Comparative effect of palmitate treatment and glucose starvation on surface expression of glycosylated proteins. Bars represent % of MFI intensity in glucose-starved (*green*) or palmitate-treated (*orange*) cells relative to control cells (*dotted line*). Bars represent the mean and SD from three independent experiments. **P* <0.05 and ***P*<0.01 when comparing no glucose or palmitate with control conditions.

Assembly and presentation of MHC I-peptide complexes at the cell surface requires peptide delivery to the ER [28, 45]. Since MHC I binding peptides derive mostly from recently synthesized proteins [23], we investigated whether

glucose starvation and palmitate treatment attenuated protein translation. To test this idea we determined the rate of global protein synthesis in ER-stressed EL4 cells by measuring the rate of [3H]leucine incorporation. Translation was severely compromised in cells deprived of glucose, which showed a 75% decline in the rate of protein synthesis (Figure 6A). The impact on protein synthesis was comparable to that observed with the translation inhibitor cycloheximide (Figure 6A). Protein synthesis was less attenuated in palmitate-treated cells, but yet decreased by approximately 25%. Of note, ER stress produced similar inhibition of protein synthesis in EL4/HEL-Cyto-SIINFEKL and EL4/HEL-ER-SIINFEKL cells.

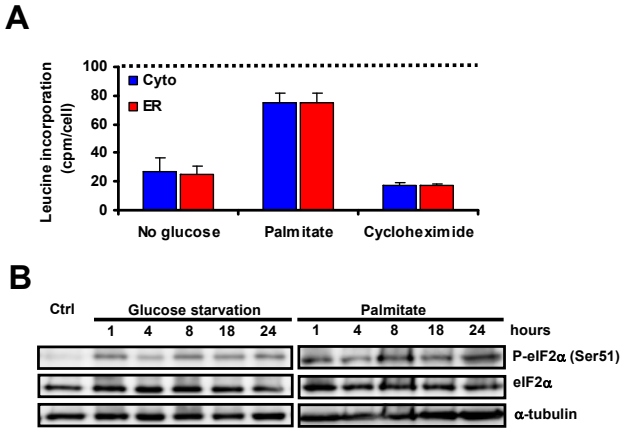


Figure 6. ER stress inhibits protein synthesis through phosphorylation of eIF2α in EL4 stable cell lines

(A) Decreased overall rate of protein synthesis upon ER stress. EL4 stable cell lines were deprived of glucose, treated with 0.25 mM of palmitate, 100 µg/ml of cycloheximide or cultured under control conditions for 19 hours. [3H]Leucine (10µCi/mL) was added during the last hour. The rate of protein synthesis was measured by [3H] leucine incorporation. The results are expressed as the % of [3H]leucine incorporation per cell relative to control cells (*dotted line*) in EL4/HEL-Cyto-SIINFEKL (*blue*) and EL4/HEL-ER-SIINFEKL *(red*) cells. Bars depict the mean and SD of one representative experiment performed in triplicate. Differences between untreated and treated cells are all significant ($P < 0.05$). (B) Phosphorylation of eIF2α. EL4 cells were deprived of glucose or treated with 0.25 mM of palmitate for different durations over a 24-h period. Total cell lysates were immunoblotted against phosphorylated eIF2α (Ser51) or total eIF2α. α-tubulin was used as loading control. One representative experiment out of three is shown.

Following UPR signaling, inhibition of cap-dependent translation occurs via phosphorylation of Ser51 of the translation initiation factor eIF2α by activated PERK [18]. In line with this, we detected a rapid phosphorylation of eIF2α in EL4 cells after only 1 hour of glucose deprivation or treatment with palmitate (Figure 6B). This phosphorylated form persisted for 24 hours in both cases. These results show that eIF2α-mediated inhibition of protein synthesis occurs during glucose starvation or palmitate treatment and support the idea that impaired surface MHC I expression is caused by an inadequate peptide supply.

Differential cell surface presentation of ER- vs. cytosol-derived peptide by MHC I molecules during ER stress

In the next series of experiments, we studied the impact of ER stress on MHC I-peptide presentation, using the SIINFEKL peptide as a model. Kb-SIINFEKL surface expression was quantified by flow cytometry in EL4 stable cell lines submitted to ER stress by glucose deprivation or palmitate treatment for 18 hours. We found that abundance of cell surface Kb-SIINFEKL decreased by more than 40% in cells that were completely deprived of glucose relative to control cells (Figure 7A). Similarly, Kb-SIINFEKL complexes were diminished by 20% or more in the presence of palmitate and by 10% in cells grown in the presence of low glucose. Thus, consistent with what we observed in the case of surface MHC I molecules, MHC I-peptide presentation is reduced during ER stress.

In addition, we found that although Kb-SIINFEKL expression was reduced in both cell lines upon ER stress, EL4/HEL-ER-SIINFEKL cells presented significantly more complexes than EL4/HEL-Cyto-SIINFEKL cells (Figure 7A). This difference occurred during complete glucose starvation, or treatment with palmitate (Figure 7A) or tunicamycin (not shown), but not when the glucose concentration was low, suggesting that it is UPR-specific (Figure 7A). Of note, both cell lines displayed similar amounts of Kb-SIINFEKL complexes under normal conditions (Figure 1B and Figure 7B top). We wish to emphasize that differences in abundance of Kb-SIINFEKL among the two types of EL4 transfectants during ER stress (Figure 7A) cannot be ascribed to an overall difference in expression of H2Kb at the cell surface (Figure 3A and B). We therefore conclude that during ER stress, diminution of Kb-SIINFEKL presentation was more drastic when the peptide derived from a protein localized in the cytosol than

from an ER-retained protein.

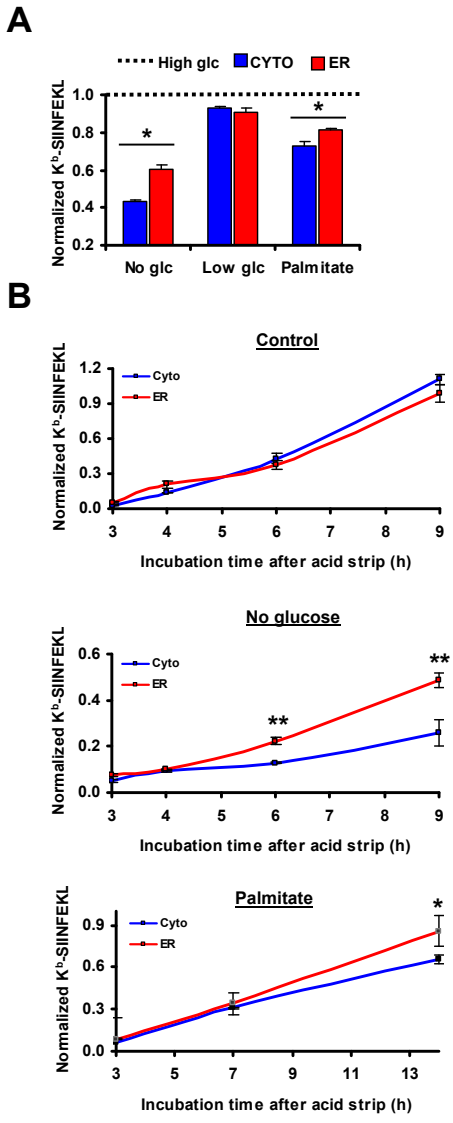


Figure 7. Increased presentation of SIINFEKL peptide derived from ER-localized relative to cytosolic HEL protein during ER stress

(A) ER stress differentially affects surface expression of Kb-SIINFEKL complexes. EL4 stable cell lines were incubated in medium lacking glucose or containing low glucose (1 mg/ml) or high glucose (4.5 mg/ml) or supplemented with palmitate (2.5 mM) for 18 hours. Kb-SIINFEKL abundance was assessed with the 25-D1.16 monoclonal antibody and APC-conjugated anti-mouse IgG1 antibody. Graph represents MFI values of glucose-deprived EL4/HEL-Cyto-SIINFEKL (*blue*) or EL4/HEL-ER-SIINFEKL (*red*) cells normalized to values of control cells, which were set to 1 (*dotted line*). (B) ER stress differentially affects surface expression of newly generated Kb-SIINFEKL complexes. EL4 stable cell lines were incubated under control conditions (*top*), deprived of glu-

cose (*middle*) or treated with 0.25 mM of palmitate (*bottom*) for 18 hours. Existent MHC-I complexes were eluted by acid strip and expression of new Kb-SIINFEKL complexes was assessed as in *A* at the indicated times. MFI values of unstripped cells incubated under normal conditions and representing normal level of Kb-SIINFEKL in each cell line were used to normalize MFI values of stripped cells. Bars represent the mean and SD from three independent experiments performed in triplicate. $*P<0.05$ and $**P<0.01$ when comparing normalized Kb-SIINFEKL expression in EL4/HEL-Cyto-SIINFEKL with that of EL4/HEL-ER-SIINFEKL.

Cell surface Kb-SIINFEKL complexes have been shown to be very stable [46]. We therefore postulated that monitoring Kb-SIINFEKL in the aforementioned experimental conditions might lead us to underestimate the impact of ER stress on exportation of "new" MHC I-peptide complexes at the cell surface. Thus, in the next series of experiments, we took advantage of the fact that cell surface MHC I-peptide complexes can be disrupted by mild acid elution at pH 3.3 [47-49]. EL4 stable cell lines were submitted or not to ER stress, then existent Kb-SIINFEKL complexes were acid stripped and generation of new complexes was measured at different time points. We reasoned that in this way we could directly assess the effect of the UPR on the generation of new Kb-SIINFEKL complexes. In control conditions, cells rapidly re-expressed Kb-SIINFEKL and initial control levels were reached 9 hours after acid stripping (Figure 7B, top). Notably, EL4/HEL-ER-SIINFEKL and EL4/HEL-Cyto-SIINFEKL cell lines showed similar kinetics. In contrast, stressed cells were not able to reach basal amount of Kb-SIINFEKL after acid strip (Figure 7B, middle and bottom). This effect was more striking in glucose-starved than in palmitate-treated cells, consistent to what we observed for MHC I expression (Figure 5B). Remarkably, EL4/HEL-ER-SIINFEKL cells generated significantly more cell surface Kb-SIINFEKL complexes than EL4/HEL-Cyto-SIINFEKL during ER stress (Figure 7B, middle and bottom). It should be noted that it was not possible to measure generation of complexes at time points later than 9 hours after acid strip, since at this time cells had already been stressed for 24 hours and cell death became a confounding variable. We conclude that ER stress decreases presentation of both existent (Figure 7A) and newly generated (Figure 7B) Kb-SIINFEKL complexes and that it differentially affected abundance of SIINFEKL derived from an ER- vs. a cytosol-localized protein.

Changes in stability of cytosolic and ER-retained HEL during ER stress

As mentioned above, newly synthesized proteins are the major substrates for MHC I processing. In addition, it has been shown that the protein synthesis machinery of the cytosol and ER compartments is under distinct regulatory control during the UPR [50]. Thus the differential effect of ER stress on presentation of ER- or cytosol-derived SIINFEKL could be due to changes in the translation rates of the source proteins. We explored this possibility and compared the synthesis rate of HEL-ER and HEL-Cyto in EL4 stable cell lines under normal conditions and during glucose starvation in metabolic labeling experiments. The rate of synthesis of cytosolic HEL and ER-retained HEL was not affected by glucose deprivation (Figure 8A). Hence, the different abundance of SIINFEKL at the surface of these cell lines during ER stress is not due to changes in the rate of synthesis of the precursor proteins

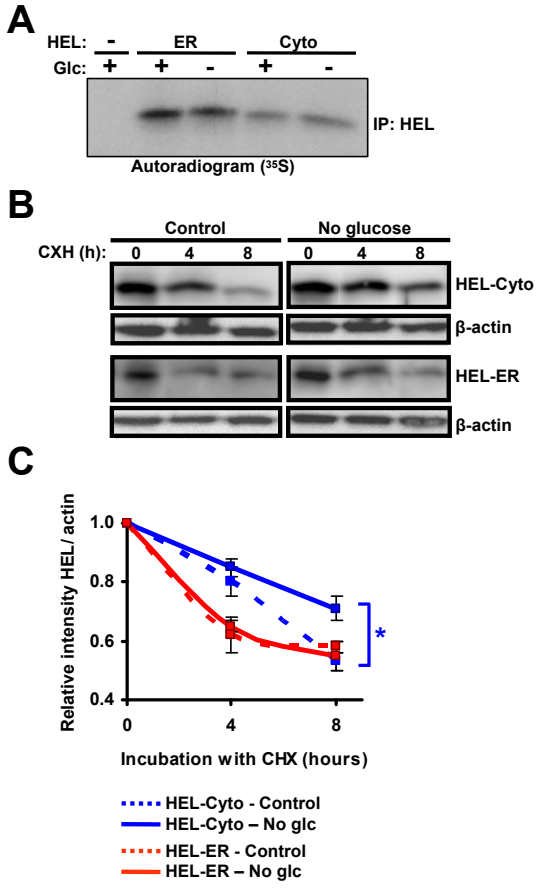


Figure 8. Stability of cytosolic HEL and ER-retained HEL during ER stress
(A) Rate of synthesis of cytosolic HEL and ER-retained HEL. EL4/HEL-ER-SIINFEKL and EL4/HEL-Cyto-SIINFEKL cells were incubated in control conditions (4.5 mg/ml) or deprived of glucose for 17 hours and pulse-labeled with [35S]methionine/ [35S]cysteine for 1 hour. Cell extracts were lysed and subjected to immunoprecipitation with anti-HEL antibody. Immunoprecipitated proteins were separated by SDS-PAGE and analyzed by fluorography. One representative experiment out of two is shown. (B) Stability of cytosolic HEL and ER-retained HEL. EL4/HEL-ER-SIINFEKL and EL4/HEL-Cyto-SIINFEKL cell lines were incubated in control conditions (4.5 mg/ml) or deprived of glucose for 18 hours. Then, 100 μg/ml of cycloheximide were added to inhibit protein synthesis and cell lysates taken at different times were immunoblotted against HEL or β-actin (used as loading control). One representative immunoblot out of three is shown. (C) Graph represents relative intensities of HEL (means and SD) from three independent experiments. $*P < 0.05$ when comparing control vs. no glucose conditions.

MHC I-peptide presentation not only relies on protein synthesis but also on protein degradation. Therefore, we explored whether the stability of these proteins could be differentially affected during ER stress. EL4 stable cell lines were deprived or not of glucose for 18 hours and then treated with cycloheximide to inhibit protein synthesis. The protein levels of cytosolic HEL and ER-retained HEL were assessed by Western blot thereafter. We observed an increased stability of cytosolic HEL in the absence of glucose compared to control conditions (Figure 8B and C). In contrast, the stability of ER-retained HEL was the same in control conditions and during glucose starvation. These results suggest that reduced presentation of SIINFEKL by H2Kb when the peptide derives from the cytosolic protein compared to the ER-retained protein is due to increased stability of the cytosolic protein during ER stress.

Discussion

The ER stands at the crossroad of two fundamental cellular processes: MHC I antigen presentation and UPR activation during ER stress. The UPR regulates protein synthesis and degradation, chaperoning and decay of ER mRNAs [14, 15]. Thus, it has enormous potential to impinge on MHC I antigen processing which relies on all these processes. Here, we assessed the effect of ER stress on the final outcome of antigen processing and presentation: MHC I-peptide abundance. We demonstrated that ER stress induced by tunicamycin, palmitate or

glucose deprivation, decreases peptide presentation by MHC I molecules. This finding is consistent with prior studies reporting reduced MHC I surface levels in human cells expressing a mutant HFE protein or overexpressing transcriptionally active isoforms of UPR-activated transcription factors ATF-6 and XBP-1 [51, 52]. Recently, reduced expression of MHC I molecules was also observed in antigen presenting cells during palmitate treatment [53]. Thus, diminution of MHC I surface expression upon UPR activation appears to be a generalized phenomenon occurring during ER stress induced by a variety of stimuli (pharmacological agents, mutant proteins, glucose starvation and saturated fatty acid).

Since the UPR provokes the degradation of ER-localized mRNAs [22], accelerated decay of MHC I mRNA might have been responsible for the reduction of cell surface MHC I expression. However, the presence of normal levels of MHC I and β2m transcripts allowed us to exclude this possibility. During ER stress, transducers of the UPR seek to decrease the ER burden by suppressing translation initiation through phosphorylation of eIF2α by activated PERK [14, 15]. We demonstrated that inhibition of protein synthesis and phosphorylation of eIF2α did occur in EL4 cells treated with palmitate or deprived of glucose. Of note, the effect of these two treatments on phosphorylation of eIF2α was similar, yet inhibition of overall protein synthesis was more severe in glucose-deprived than in palmitate-treated cells. We presume that this discrepancy was due to brisk inhibition of the mammalian target of rapamycin (mTOR) pathway during glucose starvation [54]. Inhibition of mTOR blocks phosphorylation of p70 ribosomal S6 kinase and eukaryotic initiation factor 4E binding protein 1 and thereby leads to inhibition of protein synthesis. Given the dramatic inhibition of protein synthesis during glucose starvation, it was notable that the translation rate of the two HEL variants was not affected. That feature of our HEL variants is not unique as there are several proteins whose synthesis is unaffected during ER stress [9, 54].

We found that ER stress-induced inhibition of overall protein synthesis curtails the synthesis of new MHC I molecules. Nevertheless, we do not believe that decreased synthesis of MHC I proteins per se was a leading factor responsible for decreased levels of MHC I molecules at the cell surface. Our assertion

is based on three lines of evidence: i) a 5% decline of MHC I synthesis in palmitate-treated cells was not commensurate with a drop of 30-40% of surface MHC I, ii) during ER stress, cell surface levels of MHC I proteins were decreased much more than those of other glycoproteins that must also pass through the same maturation process and quality control in order to be exported at the cell surface, and iii) the total amount of intracellular MHC I proteins was not decreased during stress suggesting that MHC I molecules did not reach the cell surface mainly because they were sequestered in the ER. In addition, de Almeida et al. showed that a partial UPR signaling induced by overexpression of ATF-6 or XBP-1 in the absence of genuine stress stimulus also resulted in decreased MHC I surface expression [52]. MHC I heavy chains and β2m are present in excess in the ER. The limiting factor in the assembly and presentation of MHC I-peptide complexes is peptide delivery to the ER [28, 45]. Moreover, peptides presented by MHC I molecules derive mainly from proteins that are degraded a few seconds or minutes after their synthesis as opposed to stable proteins with a slow turnover. Thus, generation of MHC I peptide ligands is tightly coupled to ongoing protein synthesis and inhibition of translation rapidly decreases the amount of cell surface MHC I-peptide complexes [55]. Our favorite hypothesis is therefore that decreased MHC I presentation during ER stress is due mainly, albeit not exclusively, to restriction of peptide availability. Given that MHC I molecules preferentially sample polypeptides that are being actively translated [55], we posit that global attenuation of protein synthesis caused by palmitate and glucose starvation limits the amount of a vast repertoire of peptides available for insertion in MHC I molecules. Nevertheless, we do not exclude the possibility that defective synthesis of MHC I and other possible mechanisms such as inappropriate loading of peptides, contribute to diminution of MHC I-peptide presentation. This would be mainly the case of peptides deriving from proteins whose synthesis is not curtailed upon ER stress. For instance, our results show that ER stress diminished presentation of Kb-SIINFEKL complexes even though the synthesis of the proteins source of this particular peptide (HEL variants) was not affected.

A main conclusion of our work is that ER stress-induced attenuation of MHC I-peptide presentation is more severe when the source protein is localized in the cytosol than in the ER. The difference between proteins in these two cell

compartments was UPR-specific because it did not occur in the low glucose condition in which no UPR markers were significantly induced. Our cell lines expressing HEL-Cyto-SIINFEKL and HEL-ER-SIINFEKL displayed identical responses to palmitate treatment or glucose starvation. The two cell lines showed similar upregulation of UPR markers and equivalent reduction in cell surface levels of H2Kb and H2Db during ER stress. Despite the fact that the translation rates and degradation profiles normally differ in both cell lines, they displayed similar levels of Kb-SIINFEKL complexes under steady-state conditions. On the contrary, presentation of Kb-SIINFEKL complexes was differentially affected in these cell lines during ER stress. Only 1-2 out of every 10,000 peptides generated by the proteasome bind to MHC I molecules [28]. Our data therefore beg the question: how would an ER-retained protein generate more peptides than a cytosolic protein during ER stress? We showed that this difference was not due to variations in the translation rate of each precursor protein during ER stress. This suggests that differences in peptide presentation resulted from discrepancies in the degradation of ER vs. cytosolic proteins during ER stress. UPR transducers specifically enhance degradation of proteins in the secretory pathway in order to decrease the ER folding load. During ER stress, cotranslational protein translocation is inhibited and newly-synthesized ER proteins are triaged for degradation (ERAD) [38, 56, 57]. Furthermore, retrotranslocation of ER-resident proteins in the cytosol for proteasomal degradation is enhanced [58]. Based on this, we expected to see an increased degradation of the ER-retained HEL variant during ER stress. However, the stability of the ER-retained protein remained unchanged while the stability of the cytosolic HEL variant increased during ER stress. The most parsimonious explanation for the latter findings would be that during ER stress, proteasomes focus primarily on degradation of ER as opposed to cytosolic proteins. This would be consistent with the fact that the primary role of the UPR is to decrease the folding burden in the stressed ER. We therefore propose that regulation of proteasomal degradation during ER stress leads to a reduction in MHC I peptide ligands generated from cytosolic precursors. Further studies will be needed to determine whether this concept can be generalized to other proteins and other MHC I-associated peptides.

What might be the impact of the UPR on immune recognition of infected and

neoplastic cells? Paradoxically, if the decreased generation of MHC I-peptide complexes results mainly from inhibition of translation, it could facilitate recognition of virus-infected cells. Phosphorylation of eIF2α hampers canonical cap-dependent translation initiation which regulates synthesis of 95-98% of cellular mRNAs [9]. However, some viruses can use internal ribosomal entry sites in their 5' noncoding region to initiate cap-independent translation [9, 59]. Thus, by preferentially repressing presentation of self peptides, the UPR could facilitate recognition of viral peptides (the needle in the haystack [60]). The potential impact of the UPR on recognition of neoplastic cells is not inherently obvious. On the one hand, by repressing production of MHC I-peptide complexes, the UPR may hinder presentation of tumor antigens to CD8 T cells. Indeed, generation of optimal CD8 T cell responses is promoted by high epitope density on antigen presenting cells [61, 62]. However, an elegant study by Schwab *et al.* has shown that upon induction of eIF2α phosphorylation by ER stress, cells can generate MHC I-associated peptides derived from cryptic translational reading frames [63]. Expression of such cryptic peptides by neoplastic cells might trigger recognition of stressed cells by CD8 T lymphocytes. Finally, a high fat diet rich in saturated fatty acids such as palmitate, could potentiate the conditions of ER stress found in tumour cells and lessen even more MHC I-peptide presentation. In fact, obesity has been associated with increased susceptibility to infection and impaired immune responses [53, 64]. We anticipate that high-throughput sequencing of MHC I-associated peptides [33] will be necessary to comprehensively evaluate how ER stress molds the peptide repertoire (in terms of both abundance and diversity), and to gain further insights into the global impact of the UPR on recognition of stressed cells by CD8 T lymphocytes.

Conclusions

Our work shows that ER stress impinges on the MHC I peptide repertoire in two ways: by decreasing overall MHC I-peptide presentation and by changing the relative contribution of ER- vs. cytosol-proteins to the MHC I peptide repertoire. Since ER stress is a characteristic feature of infection and malignancy, dysregulation of MHC I-peptide presentation could have major implications in the recognition of infected and transformed cells by CD8 T lymphocytes.

Methods

Cell lines

EL4 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (GIBCO Burlington, ON, Canada) supplemented with 5% fetal bovine serum (FBS) and antibiotics. EL4 stable transfectants were grown in the same medium supplemented with 1000 µg/ml of G418.

DNA constructs

pHYK/HEL-ER/myc and pCMV/HEL-Cyto/myc plasmids encoding ER-retained or cytoplasmic HEL, respectively, were provided by S. Ostrand-Rosenberg (University of Maryland, Baltimore, USA). The pHYR/HEL-ER plasmid contains the HEL gene (that includes a signal sequence) fused to the ER-retention signal KDEL, whereas pCMV/HEL-Cyto codes for HEL with a modified N-terminus and lacks ER-retention signal. These plasmids have successfully been shown to target HEL to the ER or to the cytosol [34, 35]. pHYK/HEL-ER and pCMV/HEL-Cyto were sequenced to ascertain correct sequence and reading frame. Fragments coding for HEL-ER or HEL-Cyto were fused by PCR to the region coding for the ovalbumin-derived peptide SIINFEKL, flanked by a sequence of 18 bp (LEQLE-SIINFEKL-TEWTS, here referred to as SIINFEKL) to ensure proteasome- and TAP-dependent peptide processing [65, 66]. PCR amplification products were subcloned into the pPCR-Script Amp cloning vector (Stratagene, Cedar Creek, TX, USA). HEL-ER-SIINFEKL or HEL-Cyto-SIINFEKL were excised and cloned into the bicistronic pIRES-EGFP2 vector (Clontech, Mountain View, CA, USA) to generate pIRES-EGFP2/HEL-ER-SIINFEKL and pIRES-EGFP2/HEL-Cyto-SIINFEKL (Figure 1A). Both constructs were sequenced to ascertain correct sequence and reading frame.

Stable transfectants

EL4/HEL-ER-SIINFEKL and EL4/HEL-Cyto-SIINFEKL were generated by transfecting EL4 cells with the appropriate HEL-containing pIRES-EGFP2 plasmid. Transfections were done with Lipofectamine LTX Reagent (Invitrogen, Burlington,

ON, Canada) as instructed by the manufacturer. 24 hours after transfection, single cells expressing the brightest signal of EGFP were sorted by fluorescence-activated cell sorting (FACS) on a FACSAria cell sorter (BD Biosciences, Mississauga, ON, Canada). Stable transfected clones were further selected by drug resistance (1000 µg/ml of G418) in combination with repeated cycles of FACS of EGFP-positive cells. Clones expressing similar levels of Kb-SIINFEKL at the cells surface were selected for use in further experiments.

Stress induction

ER stress was induced by incubating cells in fresh medium containing 0.25 mM of palmitate or 2.5 µg/ml of tunicamycin (Sigma-Aldrich, St. Louis, MO, USA) for the indicated times. Palmitate was prepared as described previously [67] and delivered as a complex with fatty acid-free BSA. Glucose starvation was induced by culturing cells in glucose and sodium pyruvate-free or in low glucose (1000 mg/L) DMEM medium (GIBCO) supplemented with 5 % dialyzed FBS and antibiotics for the indicated times. Control cells were grown in high glucose DMEM medium, containing 4500 mg/L of glucose and 110mg/L of sodium pyruvate supplemented with 5% FBS and antibiotics.

Flow cytometry

MHC I molecules at the cell surface were stained with biotin-conjugated anti-H2Kb (clone AF6-88.5) and biotin-conjugated anti-H2Db (clone KH95), followed by PeCy5 or APC-conjugated streptavidin. Other cell surface glycosylated proteins were stained with FITC-conjugated anti-CD45.2, FITC-conjugated anti-CD5 (Ly1), APC-conjugated anti-TCR-β and PE-conjugated anti-CD32. All antibodies were purchased from BD Biosciences. Kb-SIINFEKL levels were determined with the 25-D1.16 antibody [36] followed by staining with APC-conjugated anti-mouse IgG1 (Clone X56). Propidium iodide (BD Biosciences) was used to exclude cells in later apoptotic stages from the analysis. Cells were analyzed on a BD LSR II flow cytometer using FACSDiva (BS Biosciences) and FCS Express softwares (De Novo Software, Los Angeles, CA, USA) [68, 69].

Acid strip assay

MHC I-peptide complexes were eluted with acid treatment as previously described [47-49]. Briefly, cells (~5 x 105) were resuspended in 0.2 ml of citrate phosphate buffer at pH 3.3 (0.131 M citric acid/0.066 M Na2HPO4, NaCl 150mM) for 1 minute, neutralized with appropriate medium pH 7.4 and either reincubated in fresh medium or stained for flow cytometry analysis.

RNA extraction, reverse transcription and RT-qPCR

Total RNA was isolated with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Purified RNA was reverse transcribed using the High Capacity cDNA reverse transcription Kit with random primers (Applied Biosystems, Foster City, CA, USA) as described by the manufacturer. A reference RNA (Stratagene, La Jolla, CA, USA) was also transcribed in cDNA. Expression level of target genes was determined using primer and probe sets from Universal ProbeLibrary (https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp) or Applied Biosystems (ABI Gene Expression Assays or SYBR green PCR Master Mix, http://www.appliedbiosystems.com/). Primer sequences are given in Additional file 1. RT-qPCR assay for XBP-1 was designed to amplify both the normal and spliced forms. Pre-developed TaqMan® assays for β-actin were used as endogenous controls. RT-qPCR analyses were performed as described using a PRISM® 7900HT Sequence Detection System (Applied Biosystems) [70]. The relative quantification of target genes was determined by using the rrCT (cycle threshold) method. Relative expression (RQ) was calculated using the Sequence Detection System (SDS) 2.2.2 software (Applied Biosystems) and the formula RQ = 2-rrCT.

Protein synthesis and metabolic labeling

To measure protein synthesis, EL4 cell lines were cultured in presence or absence of glucose (4.5 mg/ml) for 18 hours. [3H]Leucine (10μCi/mL) was added during the last hour. Cells were washed twice with ice-cold PBS and fixed for 30 minutes on ice with 10% TCA. Cells were then rinsed with water and lysed with 0.1N NaOH. Radioactivity incorporation was determined with a liquid scintillation analyzer Tri-CArb 2800TR (Perkin Elmer).

In vivo biosynthetic labeling experiments were carried out as described previously [71]. Briefly, to evaluate the rate of synthesis of HEL and MHC I, EL4 cell lines were grown in control conditions or in the presence of 2.5 mM of palmitate or in the absence of glucose for 17 hours. After this period, 107 cells per condition were starved of methionine and cysteine for 30 min. 35S-labeled methionine and cysteine (220µCi/mL) were then added for 1 hour. Cells were harvested and lysed in Triton X-100 buffer (50mM Tris pH7.5, 150mM NaCl, 1% Triton X-100, 1mM EDTA, 40mM β-glycerophosphate) supplemented with complete protease inhibitor cocktail (Roche Molecular Biochemicals, Laval, QC, Canada) and phosphatase inhibitors (1 mM Na3VO4 and 5 mM NaF). Immunoprecipitation of ER-retained HEL or cytosolic HEL and MHC I were performed using anti-HEL antibody purchased from Affinity BioReagents (Golden, CO, USA) or anti-H2Kb or anti-H2Db hybridoma culture supernatans antibody [49], according to the method described previously [71]. Proteins were separated by SDS-PAGE and labeled proteins were detected by fluorography.

Immunoblotting

EL4 cell lines were cultured under control conditions or submitted to glucose deprivation or palmitate treatment (0.25 mM) for the indicated times. When indicated, 100 µg/mL of cycloheximide (Sigma-Aldrich) was used for various durations to measure the stability of HEL variants. Cells were harvested and lysed in Triton X-100 buffer. The lysates were cleared by centrifugation and the protein content was measured by the Bradford method (Biorad, Mississauga, ON, Canada). Samples were resolved by SDS-PAGE and immunoblotted with the following antibodies: anti-β-actin (AC-15) from Sigma-Aldrich, anti-HEL from Affinity BioReagents, anti-MHC class I (2G5) from Santa Cruz Biotechnology Inc. (Santa Cruz, CA, USA), anti-α-tubulin, anti-phospho-eIF2α (Ser51), anti-eIF2α and horseradish peroxidase (HRP)-conjugated anti-rabbit IgG from Cell Signaling Technology (Beverly, MA, USA), and HRP-conjugated goat anti-mouse IgG from BD Pharmigen (San Diego, CA, USA). Chemiluminescent signal was detected using a LAS3000 imaging system (Fujifilm, Tokyo, Japan) and quantification of band intensities was done using the Multi Gauge v3.0 (Fujifilm) and the ImageQuaNT v5.0 (Molecular Dynamics, Sunnyvale, CA, USA) softwares.

Statistical analysis

The means of normally distributed data were compared using the Student *t* test, with a *P* value of < 0.05 considered significant. Data are presented as the mean and SD. Whenever the results are expressed as a percentage of control, the statistical analysis was performed on the actual value.

Abbreviations

APC, antigen presenting cell; β2m, β2-microglobulin; ER, endoplasmic reticulum; ERAD, ER-associated degradation; FACS, fluorescence-activated cell sorting; HEL, hen egg lysozyme; MHC I, major histocompatibility complex class I; mTOR, mammalian target of rapamycin; RT-qPCR, quantitative real-time reverse transcriptase polymerase chain reaction; UPR, unfolded protein response.

Authors' contributions

DPG designed the study, carried out experiments and analyzed the data. PLT designed and carried out metabolic labeling experiments. MPH and DDV participated in flow cytometry experiments. EC participated in molecular cloning experiments. SM designed biochemical experiments. CP conceived and designed the study. DPG and CP drafted the manuscript, and all authors edited and approved the final manuscript.

Acknowledgements

We are grateful to Danièle Gagné for advice and assistance with flow cytometry and cell sorting, Pierre Chagnon for help and assistance with RT-qPCR and Caroline Côté for technical assistance with RNA extraction. We thank Dr. S. Ostrand-Rosenberg for kindly providing us with plasmids encoding cytoplasmic or ER-retained HEL. This work was supported by funds from the Canadian Cancer Society through the National Cancer Institute of Canada. The Institute for Research in Immunology and Cancer is supported by the Canada Foundation for Innovation, the FRSQ and the Networks of Centres of Excel-

lence through the Center of Excellence for Commercialization and Research program. DPG was supported by training grants from Foreign Affairs Canada and the Cole Foundation. PLT and EC were supported by training grants from the Canadian Institutes of Health Research. DDV was supported by the FRSQ. CP and SM hold Canada Research Chairs in Immunobiology and Cellular Signalling, respectively.

References

1. P Wong, EG Pamer: CD8 T cell responses to infectious pathogens. *Annu Rev Immunol* 2003, 21:29-70.
2. L Zitvogel, A Tesniere, G Kroemer: Cancer despite immunosurveillance: immunoselection and immunosubversion. *Nat Rev Immunol* 2006, 6:715-27.
3. BK Shin, H Wang, AM Yim, F Le Naour, F Brichory, JH Jang, R Zhao, E Puravs, J Tra, CW Michael, et al: Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J Biol Chem* 2003, 278:7607-16.
4. M Gleimer, P Parham: Stress management: MHC class I and class I-like molecules as reporters of cellular stress. *Immunity* 2003, 19:469-77.
5. SJ Marciniak, D Ron: Endoplasmic reticulum stress signaling in disease. *Physiol Rev* 2006, 86:1133-49.
6. K Wiemann, HW Mittrucker, U Feger, SA Welte, WM Yokoyama, T Spies, HG Rammensee, A Steinle: Systemic NKG2D down-regulation impairs NK and CD8 T cell responses in vivo. *J Immunol* 2005, 175:720-9.
7. DH Raulet: Roles of the NKG2D immunoreceptor and its ligands. *Nat Rev Immunol* 2003, 3:781-90.
8. HD Hickman-Miller, WH Hildebrand: The immune response under stress: the role of HSP-derived peptides. *Trends Immunol* 2004, 25:427-33.
9. M Holcik, N Sonenberg: Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol* 2005, 6:318-27.
10. Y Mamane, E Petroulakis, O LeBacquer, N Sonenberg: mTOR, translation initiation and cancer. *Oncogene* 2006, 25:6416-22.
11. M Bi, C Naczki, M Koritzinsky, D Fels, J Blais, N Hu, H Harding, I Novoa, M Varia, J Raleigh, et al: ER stress-regulated translation increases tolerance to extreme hypoxia and promotes tumor growth. *EMBO J* 2005, 24:3470-81.
12. M Moenner, O Pluquet, M Bouchecareilh, E Chevet: Integrated endoplasmic reticulum stress responses in cancer. *Cancer Res* 2007, 67:10631-4.
13. DR Carrasco, K Sukhdeo, M Protopopova, R Sinha, M Enos, DE Carrasco, M Zheng, M Mani, J Henderson, GS Pinkus, et al: The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. *Cancer Cell* 2007, 11:349-60.
14. M Schroder, RJ Kaufman: The mammalian unfolded protein response. *Annu*

Rev Biochem 2005, 74:739-89.
15. D Ron, P Walter: Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol* 2007, 8:519-29.
16. DT Rutkowski, SM Arnold, CN Miller, J Wu, J Li, KM Gunnison, K Mori, AA Sadighi Akha, D Raden, RJ Kaufman: Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. *PLoS Biol* 2006, 4:e374.
17. E Szegezdi, SE Logue, AM Gorman, A Samali: Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Rep* 2006, 7:880-5.
18. HP Harding, Y Zhang, A Bertolotti, H Zeng, D Ron: Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol Cell* 2000, 5:897-904.
19. H Yoshida, T Matsui, A Yamamoto, T Okada, K Mori: XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* 2001, 107:881-91.
20. H Yoshida, T Matsui, N Hosokawa, RJ Kaufman, K Nagata, K Mori: A time-dependent phase shift in the mammalian unfolded protein response. *Dev Cell* 2003, 4:265-71.
21. AH Lee, NN Iwakoshi, LH Glimcher: XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol Cell Biol* 2003, 23:7448-59.
22. J Hollien, JS Weissman: Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science* 2006, 313:104-7.
23. JW Yewdell, CV Nicchitta: The DRiP hypothesis decennial: support, controversy, refinement and extension. *Trends Immunol* 2006, 27:368-73.
24. LC Eisenlohr, L Huang, TN Golovina: Rethinking peptide supply to MHC class I molecules. *Nat Rev Immunol* 2007, 7:403-10.
25. HG Rammensee, K Falk, O Rotzschke: Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol* 1993, 11:213-44.
26. MT Heemels, H Ploegh: Generation, translocation, and presentation of MHC class I-restricted peptides. *Annu Rev Biochem* 1995, 64:463-91.
27. E Pamer, P Cresswell: Mechanisms of MHC class I-restricted antigen processing. *Annu Rev Immunol* 1998, 16:323-58.
28. JW Yewdell, E Reits, J Neefjes: Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat Rev Immunol* 2003, 3:952-61.

29.N Shastri, S Cardinaud, SR Schwab, T Serwold, J Kunisawa: All the peptides that fit: the beginning, the middle, and the end of the MHC class I antigen-processing pathway. *Immunol Rev* 2005, 207:31-41.
30.TN Golovina, EJ Wherry, TN Bullock, LC Eisenlohr: Efficient and qualitatively distinct MHC class I-restricted presentation of antigen targeted to the endoplasmic reticulum. *J Immunol* 2002, 168:2667-75.
31.JA Leifert, MP Rodriguez-Carreno, F Rodriguez, JL Whitton: Targeting plasmid-encoded proteins to the antigen presentation pathways. *Immunol Rev* 2004, 199:40-53.
32.E Caron, R Charbonneau, G Huppe, S Brochu, C Perreault: The structure and location of SIMP/STT3B account for its prominent imprint on the MHC I immunopeptidome. *Int Immunol* 2005, 17:1583-96.
33.MH Fortier, E Caron, MP Hardy, G Voisin, S Lemieux, C Perreault, P Thibault: The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med* 2008, 205:595-610.
34.TD Armstrong, VK Clements, BK Martin, JP Ting, S Ostrand-Rosenberg: Major histocompatibility complex class II-transfected tumor cells present endogenous antigen and are potent inducers of tumor-specific immunity. *Proc Natl Acad Sci U S A* 1997, 94:6886-91.
35.L Qi, JM Rojas, S Ostrand-Rosenberg: Tumor cells present MHC class II-restricted nuclear and mitochondrial antigens and are the predominant antigen presenting cells in vivo. *J Immunol* 2000, 165:5451-61.
36.A Porgador, JW Yewdell, Y Deng, JR Bennink, RN Germain: Localization, quantitation, and in situ detection of specific peptide-MHC class I complexes using a monoclonal antibody. *Immunity* 1997, 6:715-26.
37.JI Murray, ML Whitfield, ND Trinklein, RM Myers, PO Brown, D Botstein: Diverse and specific gene expression responses to stresses in cultured human cells. *Mol Biol Cell* 2004, 15:2361-74.
38.SW Kang, NS Rane, SJ Kim, JL Garrison, J Taunton, RS Hegde: Substrate-specific translocational attenuation during ER stress defines a pre-emptive quality control pathway. *Cell* 2006, 127:999-1013.
39.KD Jeffrey, EU Alejandro, DS Luciani, TB Kalynyak, X Hu, H Li, Y Lin, RR Townsend, KS Polonsky, JD Johnson: Carboxypeptidase E mediates palmitate-induced beta-cell ER stress and apoptosis. *Proc Natl Acad Sci U S A* 2008, 105:8452-7.

40.W Guo, S Wong, W Xie, T Lei, Z Luo: Palmitate modulates intracellular signaling, induces endoplasmic reticulum stress, and causes apoptosis in mouse 3T3-L1 and rat primary preadipocytes. *Am J Physiol Endocrinol Metab* 2007, 293:E576-86.
41.NM Borradaile, X Han, JD Harp, SE Gale, DS Ory, JE Schaffer: Disruption of endoplasmic reticulum structure and integrity in lipotoxic cell death. *J Lipid Res* 2006, 47:2726-37.
42.GA Bray, BM Popkin: Dietary fat intake does affect obesity! *Am J Clin Nutr* 1998, 68:1157-73.
43.U Ozcan, L Ozcan, E Yilmaz, K Duvel, M Sahin, BD Manning, GS Hotamisligil: Loss of the tuberous sclerosis complex tumor suppressors triggers the unfolded protein response to regulate insulin signaling and apoptosis. *Mol Cell* 2008, 29:541-51.
44.AS Lee: The ER chaperone and signaling regulator GRP78/BiP as a monitor of endoplasmic reticulum stress. *Methods* 2005, 35:373-81.
45.JJ Neefjes, F Momburg, GJ Hammerling: Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science* 1993, 261:769-71.
46.NA Kukutsch, S Rossner, JM Austyn, G Schuler, MB Lutz: Formation and kinetics of MHC class I-ovalbumin peptide complexes on immature and mature murine dendritic cells. *J Invest Dermatol* 2000, 115:449-53.
47.S Sugawara, T Abo, K Kumagai: A simple method to eliminate the antigenicity of surface class I MHC molecules from the membrane of viable cells by acid treatment at pH 3. *J Immunol Methods* 1987, 100:83-90.
48.WJ Storkus, HJ Zeh, 3rd, RD Salter, MT Lotze: Identification of T-cell epitopes: rapid isolation of class I-presented peptides from viable cells by mild acid elution. *J Immunother Emphasis Tumor Immunol* 1993, 14:94-103.
49.C Perreault, J Jutras, DC Roy, JG Filep, S Brochu: Identification of an immunodominant mouse minor histocompatibility antigen (MiHA). T cell response to a single dominant MiHA causes graft-versus-host disease. *J Clin Invest* 1996, 98:622-8.
50.SB Stephens, RD Dodd, JW Brewer, PJ Lager, JD Keene, CV Nicchitta: Stable ribosome binding to the endoplasmic reticulum enables compartment-specific regulation of mRNA translation. *Mol Biol Cell* 2005, 16:5819-31.
51.SF de Almeida, IF Carvalho, CS Cardoso, JV Cordeiro, JE Azevedo, J Neefjes,

M de Sousa: HFE cross-talks with the MHC class I antigen presentation pathway. *Blood* 2005, 106:971-7.
52. SF de Almeida, JV Fleming, JE Azevedo, M Carmo-Fonseca, M de Sousa: Stimulation of an unfolded protein response impairs MHC class I expression. *J Immunol* 2007, 178:3612-9.
53. SR Shaikh, D Mitchell, E Carroll, M Li, J Schneck, M Edidin: Differential effects of a saturated and a monounsaturated fatty acid on MHC class I antigen presentation. *Scand J Immunol* 2008, 68:30-42.
54. N Hay, N Sonenberg: Upstream and downstream of mTOR. *Genes Dev.* 2004, 18:1926-1945.
55. SB Qian, E Reits, J Neefjes, JM Deslich, JR Bennink, JW Yewdell: Tight linkage between translation and MHC class I peptide ligand generation implies specialized antigen processing for defective ribosomal products. *J Immunol* 2006, 177:227-33.
56. S Oyadomari, C Yun, EA Fisher, N Kreglinger, G Kreibich, M Oyadomari, HP Harding, AG Goodman, H Harant, JL Garrison, et al: Cotranslocational degradation protects the stressed endoplasmic reticulum from protein overload. *Cell* 2006, 126:727-39.
57. M Shenkman, S Tolchinsky, GZ Lederkremer: ER stress induces alternative nonproteasomal degradation of ER proteins but not of cytosolic ones. *Cell Stress Chaperones* 2007, 12:373-83.
58. Y Oda, T Okada, H Yoshida, RJ Kaufman, K Nagata, K Mori: Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *J Cell Biol* 2006, 172:383-93.
59. KD Tardif, K Mori, A Siddiqui: Hepatitis C virus subgenomic replicons induce endoplasmic reticulum stress activating an intracellular signaling pathway. *J Virol* 2002, 76:7453-9.
60. JW Yewdell: Plumbing the sources of endogenous MHC class I peptide ligands. *Curr Opin Immunol* 2007, 19:79-86.
61. EJ Wherry, KA Puorro, A Porgador, LC Eisenlohr: The induction of virus-specific CTL as a function of increasing epitope expression: responses rise steadily until excessively high levels of epitope are attained. *J Immunol* 1999, 163:3735-45.
62. SE Henrickson, TR Mempel, IB Mazo, B Liu, MN Artyomov, H Zheng, A Peixoto, MP Flynn, B Senman, T Junt, et al: T cell sensing of antigen dose governs

interactive behavior with dendritic cells and sets a threshold for T cell activation. *Nat Immunol* 2008, 9:282-91.
63. SR Schwab, JA Shugart, T Horng, S Malarkannan, N Shastri: Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* 2004, 2:e366.
64. ME Falagas, M Kompoti: Obesity and infection. *Lancet Infect Dis* 2006, 6:438-46.
65. AK Nussbaum, TP Dick, W Keilholz, M Schirle, S Stevanovic, K Dietz, W Heinemeyer, M Groll, DH Wolf, R Huber, et al: Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc Natl Acad Sci U S A* 1998, 95:12504-9.
66. A Craiu, T Akopian, A Goldberg, KL Rock: Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc Natl Acad Sci U S A* 1997, 94:10850-5.
67. E Diakogiannaki, NG Morgan: Differential regulation of the ER stress response by long-chain fatty acids in the pancreatic beta-cell. *Biochem Soc Trans* 2008, 36:959-62.
68. MC Meunier, JS Delisle, J Bergeron, V Rineau, C Baron, C Perreault: T cells targeted against a single minor histocompatibility antigen can cure solid tumors. *Nature Medicine* 2005, 11:1222-9.
69. ME Blais, S Brochu, M Giroux, MP Belanger, G Dulude, RP Sekaly, C Perreault: Why T cells of thymic versus extrathymic origin are functionally different. *J Immunol* 2008, 180:2299-312.
70. C Baron, R Somogyi, LD Greller, V Rineau, P Wilkinson, CR Cho, MJ Cameron, DJ Kelvin, P Chagnon, DC Roy, et al: Prediction of graft-versus-host disease in humans by donor gene-expression profiling. *PLoS Med* 2007, 4:e23.
71. MJ Servant, P Coulombe, B Turgeon, S Meloche: Differential regulation of p27(Kip1) expression by mitogenic and hypertrophic factors: Involvement of transcriptional and posttranscriptional mechanisms. *J Cell Biol* 2000, 148:543-56.

APPENDIX 2

I participated in the analysis of the MHC I immunopeptidomes from mouse DCs and thymocytes. The results were included in a published article:

Deletion of Immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules

Danielle de Verteuil, Tara L. Muratore-Schroeder, Diana P. Granados, *et. al.* **Molecular & Cellular Proteomics**, Volume 9:2034-2047 (2010)

Comparison of both sets of peptides revealed that the MHC I peptide-repertoire (MIP) of DCs conceals a unique signature.

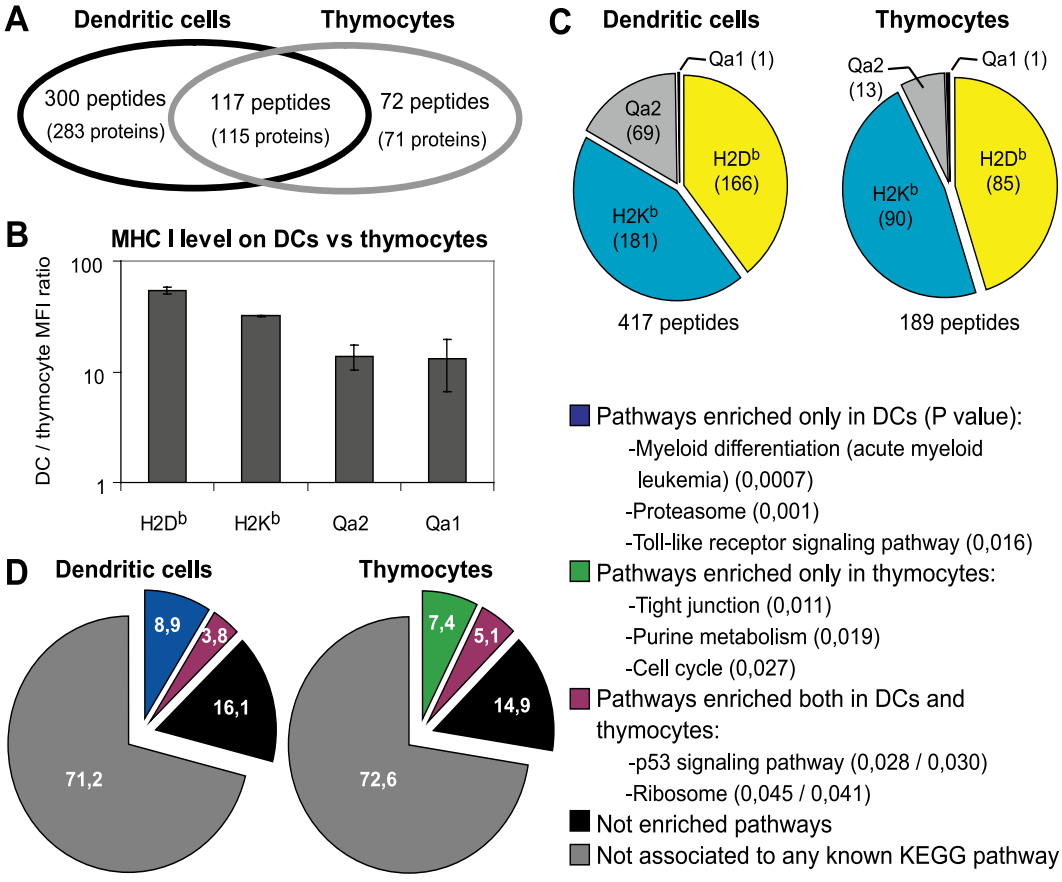


FIG. 2. **The MIP repertoire conceals a cell–type specific signature.** (*A*) Venn diagram representation of the relation between MIPs (and their source proteins) eluted from C57BL/6 thymocytes and DCs. (*B*) Cell surface expression of MHC I allelic products was evaluated by flow cytometry. Histogram shows the DC/thymocyte mean fluorescence intensity ratio for $H2D^b$ ($p = 6 \times 10^{-5}$; Student *t* test), $H2K^b$ ($p = 6 \times 10^{-7}$), Qa1 ($p = 7 \times 10^{-6}$) and Qa2 ($p = 3 \times 10^{-2}$) (mean ± S.D. of triplicate experiments). (*C*) Proportion of peptides associated to different MHC I allelic products in DCs and thymocytes. (*D*) Pie charts represent the relations between peptide source genes (389 for DCs, 186 for thymocytes) and KEGG pathways. Examples of pathways enriched in the gene datasets are depicted (with *p* value for enrichment in parentheses).