Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Jonas Dieter Förster

born in: Mosbach, Germany

Oral examination: 12th December 2023

Mass spectrometry-based identification of HPV16 target epitopes for therapeutic vaccine design

Referees: Prof. Dr. Benedikt Brors

PD Dr. Dr. Angelika Riemer

The work described in this thesis was performed from May 2018 to October 2023 under scientific supervision of PD Dr. Dr. Angelika Riemer in the Division of Immunotherapy and Immunoprevention at the German Cancer Research Center (DKFZ) in Heidelberg, Germany.

Peer-reviewed publications based on this work

(Salek & Förster et al., 2022)

Salek, M.*, Förster, J. D.*, Lehmann, W.-D., & Riemer, A. B. (2022). Light contamination in stable isotope-labelled internal peptide standards is frequent and a potential source of false discovery and quantitation error in proteomics. Analytical and bioanalytical chemistry, 414(8), 2545-2552. https://doi.org/10.1007/s00216-022-03931-w

(Mohan et al., 2022)

Mohan, N., Wellach, K., Özerdem, C., Veits, N., Förster, J. D., Foehr, S., Bonsack, M., & Riemer, A. B. (2022). Effects of hypoxia on antigen presentation and T cell-based immune recognition of HPV16-transformed cells. Frontiers in Immunology, 13, 918528-918528. https://doi.org/10.3389/fimmu.2022.918528

(Schifflers et al., 2023)

Schifflers, C., Zottnick, S., Förster, J. D., Kruse, S., Yang, R., Wiethoff, H., Bozza, M., Hoppe-Seyler, K., Heikenwälder, M., Harbottle, R. P., Michiels, C., & Riemer, A. B. (2023). Development of an Orthotopic HPV16-Dependent Base of Tongue Tumor Model in MHC-Humanized Mice. Pathogens, 12(2), 1-13. https://doi.org/10.3390/pathogens12020188

Pre-prints based on this work

(Salek & Förster et al., 2023)

Salek, M.*, Förster, J. D.*, Becker, J. P., Meyer, M., Charoentong, P., Lyu, Y., Lindner, K., Lotsch, C., Volkmar, M., Momburg, F., Poschke, I., Fröhling, S., Schmitz, M., Offringa, R., Platten, M., Jäger, D., Zörnig, I., & Riemer, A. B. (2023). optiPRM: A targeted immunopeptidomics LC-MS workflow with ultra-high sensitivity for the detection of mutation-derived tumor neoepitopes from limited input material. bioRxiv, 2023.2008.2022.554248-552023.554208.554222.554248. https://doi.org/10.1101/2023.08.22.554248

Patent applications based on this work

Salek, M., Förster, J. D., & Riemer, A. B. (2023). An automated system for providing at least one sample for electrospray ionization in a mass spectrometer system. WO2023104771A2

Riemer, A. B., Bonsack M., Förster J. D., Becker J. P., (2023) Validated HPV16-derived stimulation Peptides. EP23159742.8

Riemer, A. B., Bonsack M., Förster J. D., Becker J. P., (2023) Therapeutic HPV vaccines based on validated target epitopes. EP23159739.4

Riemer, A. B., Bonsack M., Förster J. D., Becker J. P., (2023) A therapeutic HPV vaccine based on validated target epitopes. EP23159734.5

* contributed equally

Abstract

Human papillomavirus (HPV) is the most common sexually transmitted infectious agent in humans. Infections are often asymptomatic and resolve spontaneously, but a subset of persisting infections can develop into warts, precancers, and eventually cancers in women and men. The so-called high-risk HPV (hrHPV) types are a major cause of cancer responsible for ~4.5 % of all annual cancer cases worldwide. Among hrHPVs, HPV16 is the most prevalent type and causes ~60 % of invasive cervical cancer and ~85 % of all other HPV-associated cancers. The primary oncoproteins of HPV16, E6 and E7, represent ideal targets for immunotherapy as they drive the malignant transformation of infected cells, are constitutively expressed, and are immunologically foreign. However, the development of a therapeutic vaccine has proven to be challenging, likely due to the low immunogenicity of persistent HPV16 infections. A vaccine must be highly immunogenic and elicit a strong antigen-specific cytotoxic response. To identify the most potent epitopes for immunotherapies, the research group set the aim of establishing an epitome map for the HPV16 oncoproteins E6 and E7 that specifies truly human leucocyte antigen (HLA)-presented and immunogenic epitopes. A major part of this effort is the detection of HLA-presented peptides on the surface of naturally HPV16-transformed cell lines by immunopeptidomics.

In this thesis, I developed a targeted immunopeptidomics workflow for the sensitive and highly specific detection of HLA class I-presented peptides. To this end, I developed effective quality control measures for the wet-lab technique and implemented chemical modification steps for peptide oxidation and alkylation that allow for optimal detection of the chemically diverse peptides. I optimized the liquid chromatography-mass spectrometry (LC-MS) acquisition of the peptides so that each target could be acquired with high sensitivity. In part, this was achieved by optimizing the LC gradient and tuning the scheduling of the MS acquisition in a manner that minimizes sensitivity-limiting parallel acquisition. Additionally, I showed that tuning the MS acquisition parameters enhances the sensitivity of peptide detection in complex samples. I characterized all targets in detail to allow for effective fine-tuning on a per-peptide basis. This included a newly developed approach for optimization of collision energies on a per-precursor basis, which I found to be especially effective for immunopeptidomics. The resulting method is scalable for targeting of hundreds of peptides, while allowing for maximum sensitivity where extra validation is required.

I applied the established method for the immunopeptidomics part of epitome map project. This involved the analysis of 20 cell lines covering six selected HLA supertypes (A01, A02, A03_A11, A24, B07 and B15), which are groups of HLA allotypes that bind similar peptides. Altogether, 239 distinct peptides were targeted. The HLA-restrictions of all peptide detections were validated through the complementary use of a newly developed untargeted immunopeptidomics analysis. This enabled me to detect the presentation of 25 HPV16-derived peptides on naturally HPV-transformed cell lines, 23 of which had not been MS-detected before. The associated HLA-restrictions cover all six of the targeted HLA supertypes, which I confirmed to allow for a robust and extensive predicted vaccine population coverage of more than 99 %.

These data complete the immunopeptidomics part of the HPV16 E6 and E7 epitome map. Ongoing experiments of the complementary immunogenicity assessment functionally characterize the peptides regarding their immunogenicity, immunodominance and frequency of memory responses in peripheral blood mononuclear cells from healthy donors. The most promising epitopes, as informed by these combined efforts, will be tested for their capability to induce specific killing of HPV-transformed cells. Taken together, this will inform the selection of epitopes for inclusion in epitope-specific vaccine formulations for future clinical trials. Ultimately, the presented findings will contribute significantly to the rational design of a therapeutic HPV16 vaccine applicable for a large part of the population.

Zusammenfassung

Das Humane Papillomvirus (HPV) ist das am häufigsten sexuell übertragene menschliche Pathogen. Die Infektionen verlaufen häufig asymptomatisch und heilen spontan ab, während eine Minderheit der Infektionen persistiert und zu Gewebetransformationen mit Bildung von Genitalwarzen, Präkanzerosen und letztendlich Karzinomen in Frauen und Männern führen kann. Die so genannten Hochrisiko-HPV-Typen (hrHPV) sind ein häufiger Grund für die Entwicklung von Karzinomen und sind verantwortlich für ca. 4,5 % der gesamten jährlichen Karzinominzidenz weltweit. Unter den hrHPV hat HPV16 die größte Prävalenz und verursacht ca. 60 % der invasiven Zervixkarzinome und ca. 85 % aller anderen HPV-assoziierten Karzinome. Die primären Onkoproteine von HPV16, E6 und E7, sind ein ideales Ziel für Immuntherapien, da sie die maligne Transformation der infizierten Zellen vorantreiben, konstitutiv exprimiert werden und immunologisch fremd sind. Trotzdem hat sich die Entwicklung eines therapeutischen Vakzins als schwierig erwiesen, wahrscheinlich aufgrund der geringen Immunogenität von persistierenden HPV16 Infektionen. Zu den Grundvoraussetzungen eines wirksamen therapeutischen Vakzins gehören eine hohe Immunogenität und das Auslösen einer starken zytotoxischen Immunantwort. Um die idealen Epitope für zukünftige Immuntherapien zu identifizieren, hat sich die Forschungsgruppe zum Ziel gesetzt eine Kartierung der Epitope der HPV16 Onkoproteine E6 und E7 vorzunehmen, welche alle tatsächlich humanes Leukozytenantigen (HLA)-präsentierten und immunogenen Epitope abbildet. Ein wichtiger Teil dieses Vorhabens ist die Detektion von HLA-präsentierten Peptiden auf den Oberflächen von natürlich HPV16 transformierten Zelllinien mit den Methoden der Immunpeptidomforschung (*Immunopeptidomics*).

In dieser Arbeit habe ich ein Protokoll für zielgerichtete *Immunopeptidomics* zur sensitiven und hochspezifischen Detektion von HLA Klasse I-präsentierten Peptiden entwickelt. Hierzu habe ich effektive Qualitätssicherungsmaßnahmen für die *Wet-lab* Methodik entwickelt und chemische Modifikationen der Peptide durch Oxidation und Alkylierung implementiert, welche eine optimale Detektion der chemisch diversen Peptide erlaubt. Ich habe die Flüssigchromatographie-Massenspektrometrie (LC-MS) Akquisition der Peptide so optimiert, dass jedes Zielpeptid mit einer hohen Sensitivität erkannt werden kann. Ein Teil dieser Entwicklung war die Optimierung des LC Gradienten und die Anpassung der zeitlichen Parameter der MS-Akquisition, sodass sensitivitätslimitierende parallele Akquisition von Peptiden minimiert wurde. Darüber hinauszeigte ich, dass eine Anpassung der MS-Akquisitionsparameter zur Verbesserung der Detektionssensitivität von Peptiden in komplexen Proben führt. Ich habe alle Zielpeptide charakterisiert, um eine effektive Optimierung der Methodik für jedes einzelne Peptid zu ermöglichen. Hierbei habe ich einen neuen Ansatz für die Optimierung der Kollisionsenergien für die einzelnen Präkursoren entwickelt, und festgestellt, dass dies besonders für *Immunopeptidomics* vorteilhaft ist. Die entwickelte Methode ist skalierbar für die Untersuchung von hunderten Peptiden und ermöglicht gleichzeitig eine maximale Sensitivität, wo eine zusätzliche Validierung von Peptiden notwendig ist.

Ich habe die entwickelte Methode für den *Immunopeptidomics*-Teil des Epitomkartierungsprojektes eingesetzt. Hierzu wurden 20 Zelllinien, die sechs ausgewählte HLA-Supertypen (A01, A02, A03_11, A24, B07 und B15) abdecken, untersucht. HLA-Supertypen sind Gruppen von HLA-Allotypen, die ähnliche Peptide binden können. Insgesamt wurden 239 verschiedene Peptide gezielt gesucht. Die HLA-Restriktion aller detektierten Peptide wurde durch eine ergänzende neu entwickelte nicht-zielgerichtete *Immunopeptidomics* validiert. Es gelang der Nachweis der Präsentation von 25 aus HPV16 stammenden Peptiden auf natürlich HPV-transformierten Zelllinien, von denen 23 Peptide bisher noch nicht mittels MS detektiert worden waren. Die assoziierten HLA-Restriktionen decken alle sechs der anvisierten HLA-Supertypen ab, was eine robuste und umfassende Populationsabdeckung eines entsprechenden Vakzins von über 99 % erwarten lässt.

Diese Daten komplettieren den *Immunopeptidomics*-Teil der Kartierung des HPV16 E6 und E7 Epitoms. In parallel laufenden Experimenten zur Beurteilung der Immunogenität werden die Peptide aktuell im Hinblick auf die intrinsische Immunogenität, Immundominanz und die Häufigkeit der Gedächtnisantworten in peripheren mononukleären Blutzellen von gesunden Spendern charakterisiert. Die vielversprechendsten Peptide in diesem gemeinsamen Projekt werden bezüglich ihrer Fähigkeit, spezifische Zytotoxizität gegen HPV-transformierte Zellen zu induzieren, untersucht. Zusammengenommen kann so eine Selektion von Epitopen erfolgen, die die Aufnahme in epitopspezifische Vakzinformulierungen für zukünftige klinische Studien bestimmt. Schlussendlich sind die hier beschriebenen Ergebnisse ein erheblicher Beitrag zum rationalen Design eines therapeutischen HPV16-Vakzins, dass in einem Großteil der Bevölkerung anwendbar ist.

Acknowledgements

Hiermit möchte ich mich bei all jenen bedanken, die mich in den letzten Jahren bei der Erstellung dieser Doktorarbeit, mit allen Höhen und Tiefen, unterstützt haben.

Zunächst möchte ich mich bei PD Dr. Dr. **Angelika Riemer** dafür bedanken, dass sie mir ermöglicht hat, in ihrer Arbeitsgruppe zu promovieren. Ich schätze es sehr, dass du mir diese Chance gegeben hast. Du hast mir ein Umfeld geboten, in dem sich für mich unheimlich viele tolle Möglichkeiten ergaben mich weiterzuentwickeln und mich einzubringen. Dabei habe ich auch viel Vertrauen verspürt, wenn Dinge einmal nicht so einfach waren.

Ich möchte mich bei Dr. **Jonas Becker** bedanken, der viel frischen Elan gebracht und damit erst den Abschluss dieses Projekts im vollen Umfang ermöglicht hat. Vielen Dank dir, für den fruchtbaren Austausch, die super Zeit, besonders auch abseits des Campus, und auch die Hilfe beim Korrekturlesen der Arbeit.

Mein Dank gebührt auch Dr. **Mogjib Salek**, da er mich während eines Großteils dieses Projekts betreut hat. Trotz aller Widrigkeiten die das Projekt bot, konnte ich viel von dir lernen. Niemand sonst hat so klar auch Probleme in meinen Ansätzen benannt und mich so weitergebracht.

Ich möchte auch **Rebecca Köhler** besonders danken, die einen Großteil der Experimente im Labor durchgeführt hat. Nur Dank deinem besonderen Interesse für unser Vorhaben und deinem Einsatz konnte ich diese Arbeit abschließen. Ich möchte mich in diesem Zusammenhang auch bei **Alexandra Klevenz** und **Nika Vučković** bedanken, die viele der Ziellinien kultiviert haben sowie bei **Sophia Föhr**, die sich vielfach eingebracht hat.

Bei Prof. Dr. **Benedikt Brors** möchte ich mich für seine Rolle als Fakultätsgutachter bedanken und für seine wertvollen Ratschläge. Ebenso danke ich Dr. **Jürgen Pahle**, der meinen Werdegang schon vor dieser Arbeit begleitet hat und mir damit diesen Weg erst eröffnet hat, immer ein offenes Ohr für mich hatte, und sich als TAC Mitglied eingebracht hat. Dr. **Thomas Ruppert** hat sich ebenfalls als TAC Mitglied eingebracht, und ich möchte ihm dafür danken, dass er entscheidende Impulse durch seinen reichhaltigen Erfahrungsschatz zur Massenspektrometrie eingebracht hat.

Ich möchte mich herzlich bei der gesamten Arbeitsgruppe am DKFZ bedanken, sowohl vielen ehemaligen als auch aktuellen Mitgliedern. Bei **Sophia Föhr**, die mir oft geholfen hat, pragmatisch zu bleiben, und für viel gute Laune gesorgt hat. Bei Dr. **Maria Bonsack**, die mir durch ihre herzliche Art sehr geholfen hat, in das Projekt und die Gruppe hineinzufinden. Auch bei Dr. **Samantha Zottnick**, **Kathrin Wellach**, Dr. **Nitya Mohan**, **Nika Vučković**, Dr. **Christoph Schifflers**, **Ann-Katrin Schlosser**, Dr. **Sebastian Kruse,** die alle Teil waren eines tollen Teams von Doktoranden. Ich werde mich lange an die vielen schönen Momente zurückerinnern können, die wir zusammen hatten.

Dr. **Roman Sakson** möchte ich danken, da ich mit ihm ganz besonders viel, und mit besonderem Enthusiasmus, gemeinsam über Massenspektrometrie lernen konnte.

Ich möchte auch den Mitgliedern der **Proteomics Core Facility** am DKFZ danken. Besonders die kompetente Leitung durch Dr. **Dominic Helm** hat dem Projekt Vortrieb verschafft.

Meiner Freundin **Julia** danke ich dafür, dass sie mir immer liebevoll den Rücken gestärkt hat und dass wir dadurch zusammen diese herausfordernde Zeit gemeistert haben. Für all die schönen Tage und Momente von denen wir so unendlich viele hatten in dieser Zeit danke ich dir von ganzem Herzen.

Schließlich danke ich meiner Familie — meiner Mutter **Stephanie**, meinen Brüdern **Paul** und **Philipp** und seiner Frau **Theresa —** für den vielen Zuspruch und für die vielen tollen Erlebnisse abseits der Arbeit.

Table of contents

Abstract I
Zusammenfassung III
Acknowledgements V
Table of contents VII
Abbreviations and acronyms XI
List of figures XV
List of tables XVII
1 Introduction 1
1.1 Human papillomavirus 1
1.1.1 HPV and cancer 1
1.1.2 Classification 1
1.1.3 Genome and proteins 1
1.1.4 Life cycle 2
1.1.5 Epidemiology and disease burden 3
1.2 The human immune system and HPV 4
1.2.1 Innate versus adaptive immune system 4
1.2.2 Antigen presentation 4
1.2.3 T cells 8
1.2.4 HPV immune evasion and persistence 9
1.3 HPV prevention and treatment 10
1.3.1 Prophylactic vaccination 10
1.3.2 Conventional treatment 11
1.3.3 Immunotherapeutic approaches 12
1.4 Immunopeptidomics 14
1.4.1 Introduction 14
1.4.2 HLA immunoprecipitation 14
1.4.3 Mass spectrometry-based proteomics 15
1.4.4 Peptide HLA binding prediction 20
1.5 Aims 22
1.5.1 Overall aim of the research group 22
1.5.2 Generation of an HPV16 E6 and E7 epitome map 22
1.5.3 Aims of this thesis 22
2 Materials and methods 23
2.1 Materials 23
2.1.1 General 23
2.1.2 HLA immunoprecipitation 33
2.1.3 LC-MS 35

2.1.4 Software ... 40
2.2 Methods ... 42
2.2.1 Cell culture ... 42
2.2.2 Synthetic peptides ... 43
2.2.3 HLA immunoprecipitation ... 43
2.2.4 LC-MS ... 46
2.2.5 Direct infusion MS (DI-MS) ... 49
2.2.6 Data analysis ... 50
2.2.7 Patient tumor genotyping ... 53
2.2.8 ELISpot test for donor PBMC reactivity against synthetic peptide ... 54
3 Results ... 55
3.1 Optimization of the HLA immunoprecipitation protocol (Wet-lab) ... 55
3.1.1 Establishment of the HLA immunoprecipitation protocols ... 55
3.1.2 Charge state-resolved total ion current ... 55
3.1.3 Metallothionein-2 contamination ... 56
3.1.4 Methionine oxidation ... 57
3.1.5 Cysteine alkylation ... 59
3.2 Optimization of peptide LC-MS detection ... 60
3.2.1 Detection diagnostics ... 60
3.2.2 Retention time shifts ... 63
3.2.3 MS acquisition scheduling ... 64
3.2.4 LC gradient ... 66
3.2.5 MS scan parameter tuning ... 67
3.2.6 Collision energy optimization ... 72
3.2.7 Stable isotope labelling concerns ... 75
3.2.8 Acquisition sequence and controls ... 77
3.3 HLA supertype population coverage ... 79
3.4 Immunopeptidomics experiments for the HPV16 Epitome map ... 80
3.4.1 Outline ... 80
3.4.2 Untargeted immunopeptidomics LC-MS acquisition ... 82
3.4.3 Targeted HPV16 peptide detection by cell line ... 87
3.4.4 Targeted detections by peptide ... 92
3.4.5 Summary ... 121
3.5 Targeted HPV16 epitope detection in patient tumor sample ... 129
4 Discussion ... 131
4.1 Peptide detections ... 131
4.2 MS parameter optimization ... 136
4.3 Comparing HLA binders, MS detections and peptide immunogenicity ... 137
4.4 Cumulative population coverage ... 138
4.5 Conclusions and outlook ... 138

5 References 141
Appendix A IP table A-1
Appendix B Cell line results B-1

Abbreviations and acronyms

°C	degree Celsius (unit temperature)
AA	amino acid
AB	antibody
ACK	ammonium-chloride-potassium
AFND	allele frequency net database
AGC	automatic gain control
APC	antigen presenting cell
APD	advanced peak determination
β2m	β2-microglobulin
BLAST	basic local alignment search tool
bp	base pairs (unit genome size)
BSA	bovine serum albumin
CALR	calreticulin
CANX	calreticulin
CD4	cluster of differentiation 4
CD8	cluster of differentiation 8
csv	comma separated values (file format)
CTL	cytotoxic CD8+ T lymphocyte
CxCa	cervical cancer
Da	dalton (unit atomic mass)
DDA	data-dependent acquisition
DIA	data-independent acquisition
DKFZ	Deutsches Krebsforschungszentrum, German Cancer Research Center
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
e.g.	exempli gratia, “for example”
EDTA	ethylenediaminetetraacetic acid
EL	eluted ligand
ELISpot	enzyme-linked immunospot
ER	endoplasmic reticulum
ESI	electrospray ionization
FDR	false discovery rate
G	gram (unit mass)
h	hour (unit time)
HCD	higher energy collisional dissociation
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HIC	high-income countries
HLA	human leukocyte antigen
HLA IP	the immunopeptidomics experimental process that produces purified HLA ligands
HNSCC	head and neck squamous cell carcinoma
HPV	Human papillomavirus
HPV16	HPV type 16
hrHPV	high-risk HPV
HSIL	high-grade lesions squamous intraepithelial lesion
i.e.	id est, “that is”, “namely”

IAA	iodoacetamide
IC50	half maximal inhibitory concentration
IEDB	immune epitope database
IFN	interferon
IL	interleukin
IP	immunoprecipitation
l	liter (unit volume)
LC	liquid chromatography
LPS	lipopolysaccharide
m	meter (unit length)
M	molarity, also mol/l (unit concentration of substance)
m/z	mass-to-charge (unit)
MAC	multiple allele code
MHC	major histocompatibility complex
min	minute (unit time)
mol	mole (unit amount of substance)
MS	mass spectrometry
NCE	normalized collision energy
NSA	normalized spectral contrast angle
PAMP	pathogen-associated molecular pattern
pAPC	professional antigen presenting cell
PBMC	peripheral blood mononuclear cell
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PLC	peptide-loading complex
PRM	parallel reaction monitoring
PSM	peptide-spectrum match
PTM	post-translational modifications
QC	quality control
QTRAP	quadrupole linear ion trap
RF	radio frequency
RNA	ribonucleic acid
RPLC	reversed-phase liquid chromatography
RPMI	Roswell Park Memorial Institute
RT	retention time
s	second (unit time)
SD	standard deviation
SFU	spot forming units
SIL	stable isotope-labeled
SPE	solid-phase extraction
SRM	single reaction monitoring
TAP	transporter associated with antigen processing
TCEP	tris(2-carboxyethyl)phosphine
TCR	T cell receptor
TIC	total ion current
TIL	tumor infiltrating lymphocyte
TLR	toll-like receptor
TLR	Toll-like receptor
vs.	versus

WGS	whole-genome sequencing
×g	g multiple of standard gravity of 9.8 m/s2 (g-force)
XIC	extracted ion chromatogram

List of figures

Figure 1. HPV16 genome. 2
Figure 2. The life cycle of HPV. 3
Figure 3. Overview of HLA class I antigen processing and presentation. 6
Figure 4. HLA immunoprecipitation and LC-MS analysis. 14
Figure 5. Schematic of the Orbitrap Exploris 480 MS used for this thesis. 17
Figure 6. Peptide fragmentation products. 19
Figure 7. MS acquisition schemes used in this thesis. 20
Figure 8. Charge state-resolved total ion current chromatograms. 56
Figure 9. GammaBind Plus Sepharose beads elute Metallothionein-2 57
Figure 10. Oxidation of methionine-containing peptides in an HLA IP eluate 58
Figure 11. Alkylation of cysteine-containing peptides in an HLA IP eluate. 59
Figure 12. Peptide detection diagnostics 61
Figure 13. HLA-presented peptide identification at limit of detection. 62
Figure 14. HLA IP retention time shifts and alignment using standard peptides. 63
Figure 15. Systematic RT shifts as a function of peptide load on the LC system. 64
Figure 16. MS acquisition scheduling concerns. 65
Figure 17. The two main LC gradients used in this thesis. 66
Figure 18. Retention time windows of selected synthetic peptides of the JPTRP standard 67
Figure 19. Test series on the effect of MS resolution on detection sensitivity. 68
Figure 20. Test series on the effect of maximum MS2 injection time on detection sensitivity. 69
Figure 21. Test series on the effect of the precursor isolation window on detection sensitivity. 71
Figure 22. Collision energy optimization for target peptides. 73
Figure 23. Further examples of NCE optimization. 74
Figure 24. General evaluation of the NCE optimization benefit. 75
Figure 25. Light contamination with synthetic SIL peptides 75
Figure 26. Calibration of SIL spike-in for detected peptides. 76
Figure 27. Targeted HLA IP acquisition sequence and controls. 78
Figure 28. HLA supertype combination metrics. 80
Figure 29. Evaluation of untargeted immunopeptidomics experiments. 83
Figure 30. Effect of HLA IP antibody selection on detected peptides with cell line CaSki. 85
Figure 31. Effect of HLA IP antibody selection on detected peptides with cell line SNU-1005 86
Figure 32. Cell line CaSki result compilation 88
Figure 33. Cell line W12-20861 result compilation. 89
Figure 34. Cell line SCC090 result compilation. 90
Figure 35. Cell line SCC152 result compilation. 91
Figure 36. E7/11–19 YMLDLQPET targeted detections 92
Figure 37. E7/11–20 YMLDLQPETT targeted detections. 93
Figure 38. E7/49–57 RAHYNIVTF targeted detections for supertype B07 94
Figure 39. E7/49–57 RAHYNIVTF targeted detections for supertype B15 95
Figure 40. E7/49–57 RAHYNIVTF targeted detections for supertype A24 96
Figure 41. E7/5–13 TPTLHEYML targeted detections 97
Figure 42. E6/15–22 RPRKLPQL targeted detections. 98
Figure 43. E6/29–38 TIHDIILECV targeted detections. 99
Figure 44. E6/33–41 IILECVYCK targeted detections 100
Figure 45. E6/37–46 CVYCKQQLLR targeted detections for supertype A11. 101

Figure 46. E6/37–46 CVYCKQQLLR targeted detections for supertype A03. 102
Figure 47. E6/38–45 VYCKQQLL targeted detections. 103
Figure 48. E6/41–50 KQQLLRREVY targeted detections. 104
Figure 49. E6/42–50 QQLLRREVY targeted detections. 105
Figure 50. E6/53–61 AFRDLCIVY targeted detections. 106
Figure 51. E6/80–88 ISEYRHYCY targeted detections. 107
Figure 52. E6/89–99 SVYGTTLEQQY targeted detections. 108
Figure 53. E6/93–101 TTLEQQYNK targeted detections. 109
Figure 54. E6/125–133 HLDKKQRFH targeted detections. 110
Figure 55. E7/2–11 HGDTPTLHEY targeted detections. 111
Figure 56. E7/15–23 LQPETTDLY targeted detections. 112
Figure 57. E7/15–24 LQPETTDLYC targeted detections. 113
Figure 58. E7/15–25 LQPETTDLYCY targeted detections. 114
Figure 59. E7/43–52 GQAEPDRAHY targeted detections. 115
Figure 60. E7/66–74 RLCVQSTHV targeted detections. 116
Figure 61. E7/77–86 RTLEDLLMGT targeted detections. 117
Figure 62. E7/89–97 IVCPICSQK targeted detections. 118
Figure 63. E6/8–17 MFQDPQERPR untargeted detection and targeted validation. 119
Figure 64. E7/45–52 AEPDRAHY untargeted detection and targeted rejection. 120
Figure 65. Distribution of experimentally verified binders, MS-detected peptides and immunogenic peptides along the sequence of HPV16 E6. 125
Figure 66. Pile-up plot of the overlapping peptide sequences for the candidate peptides and MS-detected peptides of E6. 126
Figure 67. Distribution of experimentally verified binders, MS-detected peptides and immunogenic peptides along the sequence of HPV16 E7. 127
Figure 68. Pile-up plot of the overlapping peptide sequences for the candidate peptides and MS-detected peptides of E7. 128
Figure 69. Result CxCa versus HNSCC association. 128
Figure 70. Blast results for patient WGS reads that were assigned to HPV16. 129
Figure 71. Immunopeptidomics results for patient tumor sample. 130

Appendix Figure 1. Cell line MRI-H-196 result compilation. B-1
Appendix Figure 2. Cell line SNU-17 result compilation. B-2
Appendix Figure 3. Cell line SNU-703 result compilation. B-2
Appendix Figure 4. Cell line SNU-902 result compilation. B-3
Appendix Figure 5. Cell line SNU-1000 result compilation. B-4
Appendix Figure 6. Cell line SNU-1005 result compilation. B-4
Appendix Figure 7. Cell line SNU-1299 result compilation. B-5
Appendix Figure 8. Cell line 866 result compilation. B-6
Appendix Figure 9. Cell line 879 result compilation. B-6
Appendix Figure 10. Cell line Goerke result compilation B-7
Appendix Figure 11. Cell line DoTc2-4510 result compilation B-7
Appendix Figure 12. Cell line SiHa result compilation. B-8
Appendix Figure 13. Cell line SCC154 result compilation. B-8
Appendix Figure 14. Cell line UD-SCC2 result compilation. B-9
Appendix Figure 15. Cell line UM-SCC-47 result compilation. B-9
Appendix Figure 16. Cell line UM-SCC-104 result compilation. B-10
Appendix Figure 17. Cell line C33A result compilation. B-10

List of tables

Table 1. Allotypic peptides acquired as controls.....49
Table 2. HLA supertype population and all world coverage. 79
Table 3. Cell lines targeted for the epitome map.....81
Table 4. Peptide detections from untargeted HLA IP acquisitions. 86
Table 5. Summary of MS-detected peptides and associated literature. 121
Table 6. Summary of the epitome map outcome. 124

Appendix Table 1. Properties of the targeted IPs for the epitome map. A-1

1 Introduction

1.1 Human papillomavirus

1.1.1 HPV and cancer

Papillomaviruses are a diverse group of small DNA viruses that are highly species-specific and infect epithelia in vertebrates ranging from fish to mammals (McBride, 2022). Human papillomavirus (HPV) is the most common sexually transmitted infectious agent in humans. Infections are often asymptomatic and resolve spontaneously but a subset of persisting infections can develop into warts, precancers, and eventually cancers in women and men (Hall et al., 2021). A causal link between HPV and cervical cancer (CxCa) was first discovered by Prof. Harald zur Hausen and his team in the early 1980s for which he was later awarded the Nobel Prize in Physiology or Medicine 2008 (Dürst et al., 1983; Boshart et al., 1984). By now, HPV has additionally been associated with less common anogenital cancers, such as cancer of the vulva, vagina, penis, and anus and also oropharyngeal cancers (Hall et al., 2021).

1.1.2 Classification

About 450 distinct types of the HPV family of viruses have been identified (McBride, 2022). Five different phylogenic genera are distinguished: Alpha-, Beta-, Gamma-, Mu- and Nu-papillomaviruses (McBride, 2022). Most HPV types are considered non-oncogenic or 'low-risk' and cause no lesions or only cutaneous warts, benign low-grade cervical cell abnormalities, anogenital warts, and respiratory tract papillomas. The Alphapapillomaviruses are of special interest as they contain HPV types that infect the oral and genital mucosal epithelium. This is the subset that contains the oncogenic or 'high-risk' HPV types (hrHPV). Almost all HPV-associated cancers in men and 72 % of HPV-associated cancers in total are caused by high-risk types 16 and 18 (de Martel et al., 2020). Further important high-risk types are 31, 33, 45, 52 and 58 that together account for further 15 % of cervical cancers and 11 % of all HPV-associated cancers (Hall et al., 2021). In total, 12 types HPV are considered high-risk: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 with a further 13 being considered to be a probable/possible carcinogen (Bruni et al., 2023b). HPV type 16 (HPV16) is the most frequently detected HPV and causes about ~60 % of invasive cervical cancer worldwide and ~85 % of all other HPV-associated cancers (Schiffman et al., 2016). Individual HPV types are further classified into variant lineages and sublineages, where variant lineages show from 1 % to 10 % nucleotide sequence difference in their L1 open reading frame and additionally distinguished sublineages show a further difference of 0.5 % to 1 % across the genome (Burk et al., 2013).

1.1.3 Genome and proteins

HPV is a small double-stranded DNA virus with a genome size of around 8,000 base pairs (bps). The genome of HPV16 is shown in Figure 1. It contains three domains: The early genes E1, E2, E4, E5, E6, and E7 in the early coding region, the late genes L1 and L2 in the late coding region and the upstream regulatory region (URR) or long coding region. The early genes encode the proteins that mainly drive viral DNA replication, DNA synthesis and amplification. The late genes encode proteins that form the viral capsid. The URR is a noncoding region containing the origin of replication and regulatory elements such as transcription factor-binding sites. The primary oncogenes of interest are E6 and E7 which are known to cause dysregulation of cell cycle progression, apoptosis and other crucial processes (Moody & Laimins, 2010).

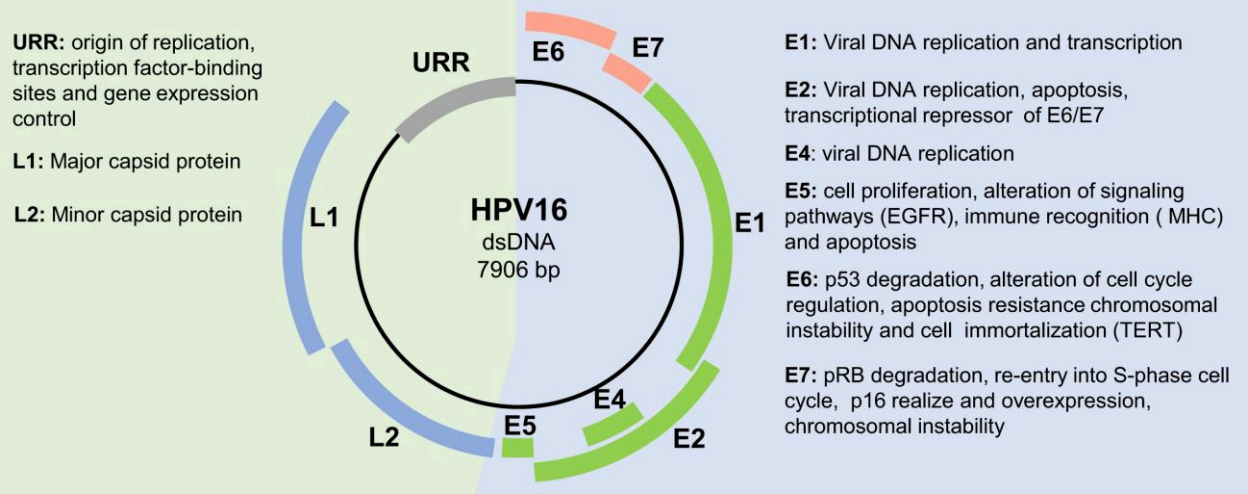


Figure 1. HPV16 genome. E1–E5 (green) are early genes. E6 and E7 are the two main oncogenes. L1 and L2 (blue) are late genes. URR (gray) is the upstream regulatory region. Reprinted with permission from de Sanjosé et al. (2018).

1.1.4 Life cycle

HPV infection begins with infection of basal epithelial cells (Figure 2). This layer of proliferating cells is protected by several layers of differentiated epithelial cells and infection occurs where they are accessible e.g. through microlesions in the skin or genital and oropharyngeal mucosae (Chabeda et al., 2018). The virion binds to heparan sulfate proteoglycans (HPSGs) on the basement membrane which introduces conformational changes to the capsid. This exposes the N-terminus of L2 for cleavage by extracellular furin, which reveals epitopes of L2 that are crucial for infection. The virus binds to cellular receptors and is taken up into the host cell by an endocytosis mechanism with similarities to micropinocytosis. The viral capsid uncoats in the endosome and seems to be retained in HPV-harboring vesicles until the breakdown of the nuclear envelope during mitosis allows for nuclear entry. Next, expression of the early stage genes is initiated (DiGiuseppe et al., 2017). Early proteins E1 and E2 enable virus DNA replication through the host machinery, roughly producing a low level of ca. 50–100 episomal copies per cell (Nelson & Mirabello, 2023).

The infection persists intracellularly in the basal keratinocytes which are driven into S phase through the combined action of proteins E5, E6 and E7. E5 promotes exit from G_0–G_1 phases by upregulating the epidermal growth factor receptor (EGFR) signaling pathway. E6 promotes cell survival and de-differentiation of cells by triggering proteasomal degradation of the pro-apoptotic factor p53 and interacting with many PDZ domains of host proteins (Ruesch & Laimins, 1998; Marx et al., 2017). It also upregulates MYC and telomerase expression to promote cell survival (Roden & Stern, 2018). E7 most notably binds retinoblastoma protein (pRb) to block its function and target it for proteasomal degradation. This helps to overcome the G_1 restriction point. In effect, cell differentiation is delayed and hyperproliferation of infected keratinocytes is enabled (Moody & Laimins, 2010).

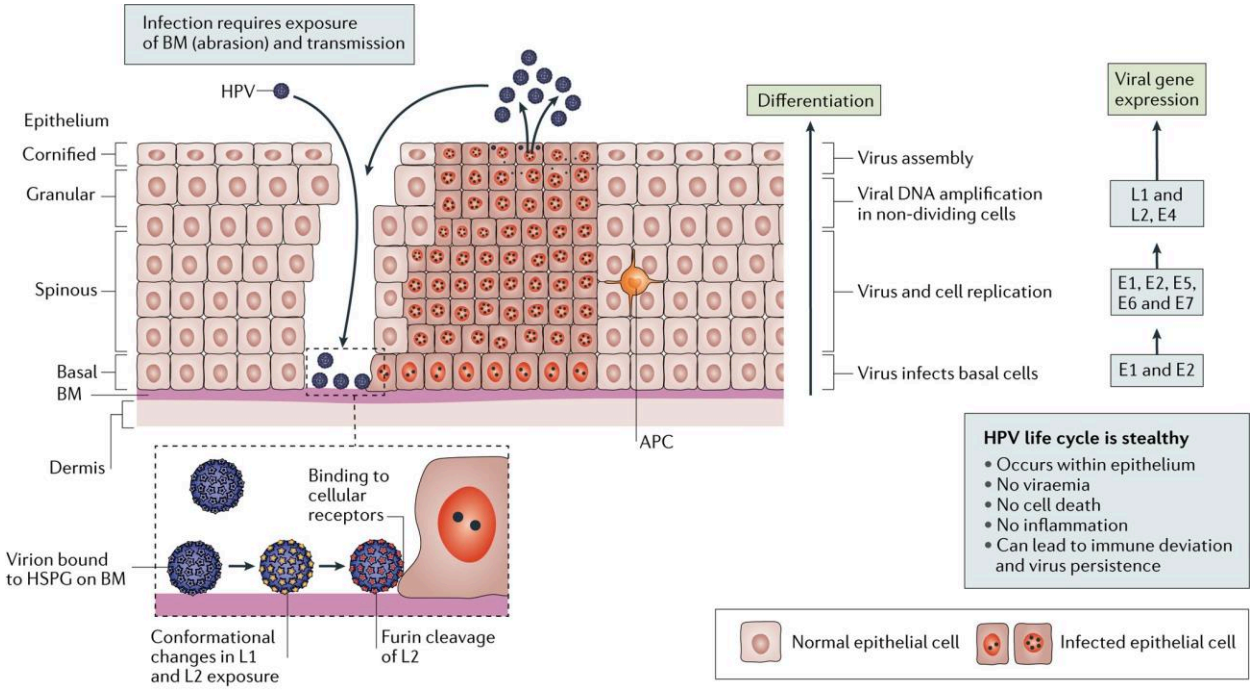


Figure 2. The life cycle of HPV. Infection occurs where virions access the basement membrane (BM) and the basal cell layer of the epithelium. The virus genome is replicated and passed on with the proliferating basal cells. As cells differentiate towards the upper layers of the epithelium, viral gene expression changes to facilitate DNA amplification and eventual virus assembly. Virions are only released when cells of the upper layers of the epithelium are shed. APC is an antigen presenting cell. Reprinted with permission from Roden and Stern (2018).

As infected cells divide and progressively differentiate towards the upper layers of the epithelium, low copy numbers of the viral genome are maintained. With terminal differentiation in the upper epithelial layers, different viral genes are expressed. First E4 is activated, followed by L1 and L2. E4 drives the generation of high copy numbers of the viral DNA. The expression of the capsid proteins leads to the assembly of virions, which are released as the infected cells in the upper epithelial layers are shed. In this way, the life cycle of the virus follows the differentiation of the infected cells and does not cause cell death (Roden & Stern, 2018).

1.1.5 Epidemiology and disease burden

Transmission of HPV occurs through intimate skin-to-skin contact with an infected person. From studying newly acquired HPV infections it has been shown that first infection commonly occurs soon after first sexual activity. A prospective study of college women found a 40 % incidence of infection by 24 months after first sexual intercourse, 10 % of which were with HPV16. Overall, between 50–79 % of women in high-income countries (HIC) have a lifetime risk of acquiring a genital HPV infection. Globally, among women of all age groups with normal cervical cytology, the prevalence of HPV was in 2019 estimated to be 9.9 % (Prudden et al., 2022).

Cervical cancer, which is almost exclusively caused by oncogenic HPV viruses is globally the fourth leading cause of cancer deaths for women (Okunade, 2020). It is estimated that 604,000 women were diagnosed with and 342,000 women died of cervical cancer in 2020 (Sung et al., 2021). However, HPV-driven cervical cancer is slow to develop and generally takes 15–20 years to develop after infection (Prudden et al., 2022). Due to this and the availability of prophylactic vaccination, screening and treatment options in HIC, almost 90 % of death associated with cervical cancer are by now in women in low- and middle-income countries (LMICs) (Prudden et al., 2022). Beyond cervical cancer in women, HPV also causes genital and oropharyngeal cancer in both sexes as well as penile cancer

in men (Roden & Stern, 2018). Head and neck squamous cell carcinoma (HNSCC) is increasingly attributed to infection with hrHPV and are on the rise globally (Johnson et al., 2020). A recent analysis showed that almost one in three men worldwide carry a genital HPV infection and at around one in five men are infected with an hrHPV type (Bruni et al., 2023a). Overall, HPV causes ~4.5 % of all cancer cases worldwide (de Martel et al., 2017; McBride, 2022).

1.2 The human immune system and HPV

1.2.1 Innate versus adaptive immune system

The human immune system is a complex network of cells, tissues and molecules that act to protect the body from pathogens and other influences recognized as non-self or 'foreign'. Broadly, it can be divided into two main branches: The innate immune system and the adaptive immune system. In the following, I will introduce in detail the aspects of most relevance for this thesis.

The innate immune system is a highly conserved first line of defense that is present from birth. It provides swift non-specific protection against a wide range of pathogens. Key factors for the innate immune system include physical barriers like skin and mucus, chemical factors such as proteins and enzymes of the complement system, and cellular defenses through numerous cell types such as macrophages, neutrophils and natural killer cells. Recognition of pathogens through conserved molecular patterns known as pathogen-associated molecular patterns (PAMPs) is enabled by pattern recognition receptors (PRRs) such as toll-like receptors (TLRs) on immune cells. This triggers a cascade of immune responses, including the release of pro-inflammatory cytokines and chemokines which recruit immune cells to sites of infection. For antiviral protection, infected and innate immune cells trigger an antiviral state through secretion of type I interferons (IFNs) and recruit professional antigen presenting cells (pAPCs). A key factor of the ensuing immune response is that pAPCs then capture so-called antigens from pathogens, which are the molecules that are recognized as foreign. pAPCs process and present these to cells of the adaptive immune system (Murphy et al., 2022).

In contrast to the innate immune system, the adaptive immune system elicits highly specific responses. While the initial adaptive immune response to novel challenges is far slower, immunological memory enables it to provide long-term and highly effective protection against known pathogens. Adaptive immunity is mediated by lymphocytes, the two main types of which are B cells and T cells. B cells enable humoral immunity through the production of antibodies that bind antigens and thereby neutralize pathogens or toxins or facilitate their clearance by other immune cells. Humoral immunity is particularly effective against extracellular pathogens such as bacteria and viruses that are feely circulating or intermittently present outside of infected cells. In contrast, T cells are the drivers of cell-mediated immunity. This type of immunity is crucial for protection against intracellular pathogens such as viruses that have infected host cells or cancer cells. It is enabled by the recognition of antigens on cell surfaces by T cell receptors (TCRs) (Murphy et al., 2022).

1.2.2 Antigen presentation

Antigen presentation is a crucial process for the cellular adaptive immune system that allows T cells to recognize and respond to entities that are recognized as foreign. Antigen presentation is performed by major histocompatibility complex (MHC) molecules which are expressed on the cell surface of most vertebrate cells. Two classes of MHC molecules are distinguished. In brief, MHC class I present short peptides, derived from intracellular polypeptides, on the surface of antigen presenting cells (APCs), which are all nucleated cells. For initiation of cellular immune responses, MHC class I presentation on pAPCs is required. MHC class II present slightly longer peptides exclusively on pAPCs, which are derived from extracellular sources that have been processed within endosomes or lysosomes (Dersh et al., 2021). In humans, MHCs are named human leukocyte antigens (HLAs). Here,

the focus will be on HLA class I antigen presentation which is essential for immune responses against intracellular pathogens such as HPV.

1.2.2.1 HLA class I structure and peptide binding

HLA class I molecules are heterodimers composed of a heavy α chain and a light β2-microglobulin (β2m) chain. The heavy chain is encoded by the respective HLA gene and consists of the three domains α1, α2 and α3. α1 and α2 form the crucial peptide-binding groove. β2m is an invariant chain that associates non-covalently and is encoded by the B2M gene (Murphy et al., 2022).

HLA peptide binding is based on the interaction of peptides with the binding-groove of the HLA. Six distinct binding pockets A–F within the groove have been identified (Saper et al., 1991). Of these, pockets B and F are the key determinants for peptide binding (Chujoh et al., 1998; Nguyen et al., 2021). Stable peptide binding is enabled by the presence of so-called anchor residues at the correct position of the peptide amino acid (AA) sequence that fit into these binding pockets, and the formation of hydrogen bonds between side-chains of the binding groove and the backbone of the peptide (Wieczorek et al., 2017). In addition, the ends of the groove possess conserved tyrosine residues that interact with the N- and C-termini of the peptide, which is thought to be the driving factor for the typical 8- to 10-mer length distribution of the binders (Bouvier & Wiley, 1994). Much longer peptides of up to 15 AAs are also commonly reported, with many of these understood to form bulges out of the binding groove such that anchor residues are conserved towards the termini (Trolle et al., 2016; Gfeller & Bassani-Sternberg, 2018). These interactions depend on the precise geometry, charge distribution and hydrophobicity of the binding groove and the pockets, which is the source of the variability in peptide binding across different HLA molecules (Wieczorek et al., 2017).

1.2.2.2 HLA class I alleles and supertypes

The HLA loci are some of the most polymorphic regions in the genome of vertebrates. The loci for the HLA class I genes are HLA-A, HLA-B and HLA-C, which exhibit highly overlapping functions. In the human population, thousands of distinct HLA alleles are present at significant frequencies and over 24,000 individual class I alleles have been described (Dersh et al., 2021; Barker et al., 2023). The encoded allotypes bind distinct repertoires of peptides with preferences for certain peptide sequence motifs but also share significant and variable overlaps with other allotypes (Falk et al., 1991; Sidney et al., 2008). Due to the availability of the three loci on chromosome 6, an individual possesses two haplotypes of three alleles and can express combinations of up to six HLA allotypes. Reflecting their significant role in immunity, various allele disease-phenotype associations are known and thousands of disease-associated HLA single-nucleotide polymorphisms have been identified (Milius et al., 2013).

To distinguish the various HLA class I alleles that encode different proteins, a consistent naming scheme has been defined (e.g. HLA-A*02:01). The scheme starts with the HLA prefix, the letter of the locus and an asterisk. This is followed by a sequence of digits separated by colons. The first two digits name the broad allele family. This is followed by a colon and a running sequence of further digits that get assigned upon discovery of the allele. Further groups of digits are sometimes specified but do not affect the protein sequence (Marsh et al., 2010).

The investigation of HLA structural features, especially the relevance of the B and F pockets, brought about the HLA supertype concept, which was first introduced in the 1990s (del Guercio et al., 1995; Sidney et al., 1996; Sette & Sidney, 1999). Through analysis of key residues in the B and F pockets, it was determined what kind of anchor residues would be favored. Where similarities were determined, HLA alleles could be grouped into supertypes. A comprehensive effort by Sidney et al. (2008), classified just over 760 HLA types into 12 supertypes. The supertype concept provided a

rationale for some of the first peptide-based vaccine developments because a peptide binding to one member of a supertype grouping would presumably bind to most of the other member alleles (Khan et al., 2006; Nguyen et al., 2021). Beyond this, the concept has successfully been used in clinical and immunological studies where HLA supertypes served as correlates or predictors of disease protection or severity (Nguyen et al., 2021).

1.2.2.3 Antigen processing

HLA-presented peptides are derived from source polypeptides that undergo a multi-step process for eventual presentation on the cell surface. Figure 3 illustrates many of the key aspects of this pathway.

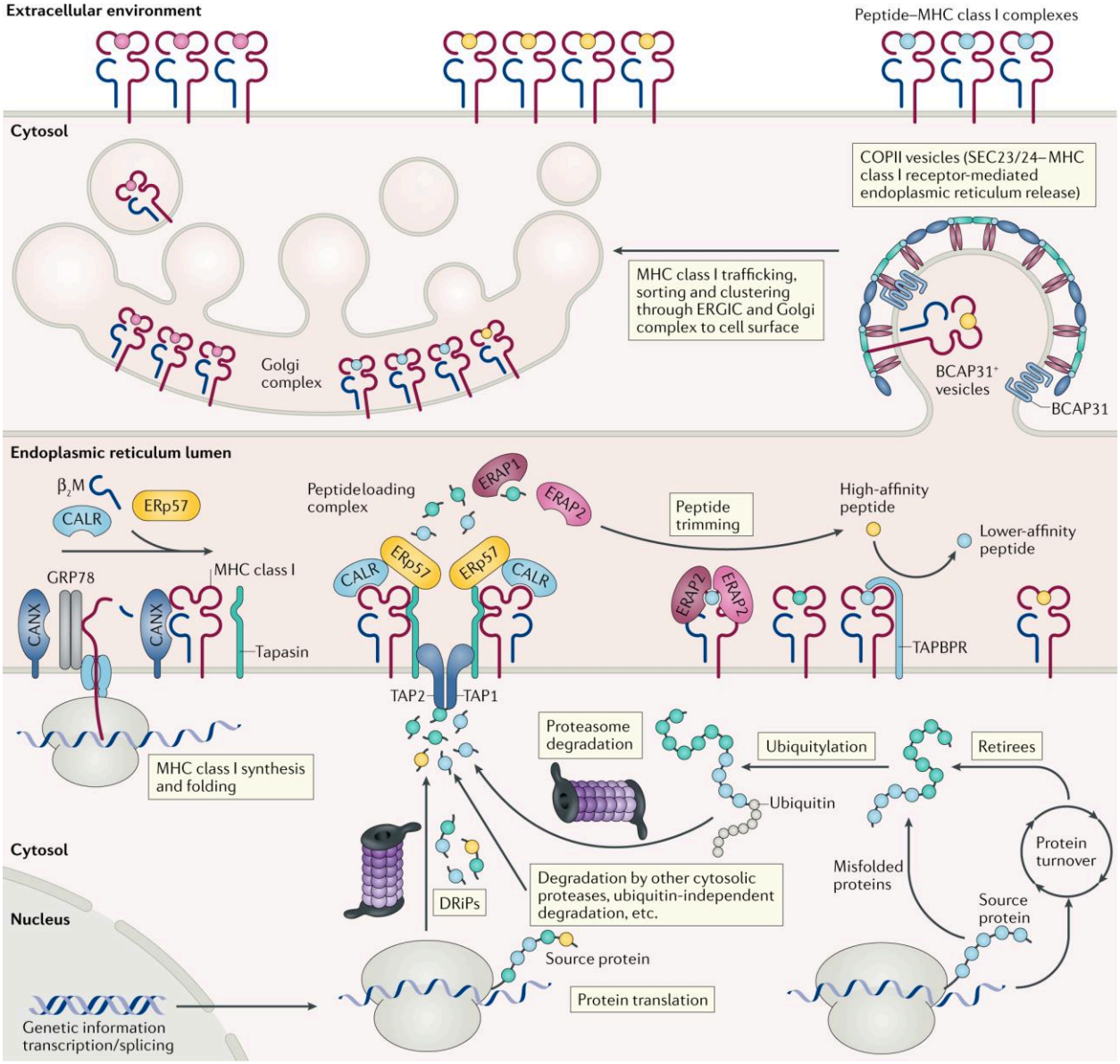


Figure 3. Overview of HLA class I antigen processing and presentation. The canonical pathway to peptide presentation is through proteasomal degradation of intracellular proteins with several further sources contributing. HLA (here termed MHC) class I heterodimers are assembled in the endoplasmic reticulum (ER). The peptide-loading complex (PLC) facilitates direct loading of peptides transported into the ER by transporter associated with antigen processing (TAP). Peptide trimming and replacement of low-affinity peptides continue outside of the fully formed PLC. Suitable HLA:peptide complexes are trafficked by vesicles via the Golgi complex towards the cell surface for antigen presentation. For abbreviations, see main text. Reprinted and adapted with permission from Dersh et al. (2021).

Antigen processing and presentation is a complex process under active research. The canonical pathway to presentation is through proteasomal degradation of intracellular proteins. In this process, misfolded or 'retired' proteins, marked through ubiquitylation, are broken down into smaller peptide fragments. The generated peptides are then transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) proteins. TAP binds and transports peptides of lengths of up to 16 AAs with a reported preference for peptides of lengths 8–12 AAs (Koopmann et al., 1996). It joins the peptide-loading complex (PLC) that further consists of HLA heterodimers, the oxidoreductase Endoplasmic Reticulum Protein 57 (ERp57) and the chaperones calreticulin (CALR) and tapasin (Blees et al., 2017). Prior assembly of the HLA molecule is chaperoned by calnexin (CANX) and is initially associated with β2m in what is an unstable heterodimer that is stabilized by tapasin. In the PLC, TAP and HLA are in close proximity to enhance loading of peptides. A two-step model for peptide loading suggests that low-affinity peptides may initially bind and subsequently be replaced by high-affinity peptides that sufficiently stabilize the HLA heterodimer. Tapasin plays a critical role in this as it promotes the dissociation of low-affinity peptides (van Hateren & Elliott, 2023). Peptides in the ER lumen may be further trimmed by ER-resident peptidases ERAP1 and ERAP2, which remove AAs from peptide N-termini and can thereby improve affinity for HLA molecules (Compagnone et al., 2019). The protein TAP-binding protein related (TAPBPR) is implicated in quality control of the HLA:peptide pairing in a similar fashion to tapasin but acts independent of the PLC (van Hateren & Elliott, 2023). Stable complexes are then loaded into B cell receptor-associated protein 31 (BCAP31)-positive COPII vesicles which facilitate transport to the Golgi apparatus. BCAP31 mediates HLA class I export from the ER and may also be involved in their quality control (Abe et al., 2009). From the Golgi apparatus, the complexes are ultimately transported to the cell surface. On the cell surface, peptides in the context of the binding HLA allotype act as epitopes that can be recognized by T cells (Dersh et al., 2021).

Beyond the canonical proteasomal degradation of retired proteins, several further sources of HLA epitopes have been described. Defective ribosomal products (DRiPs) and short-lived proteins (SLiPs) are both typically short lived and can be processed for presentation. DRiPs are defective translation products that are non-functional and SLiPs are a loose category of proteins that are retired quickly (Dersh et al., 2021). They are often intrinsically disordered and degraded in the absence of a stabilizing partner. DRiPs and SLiPs have been found to substantially contribute to the immunopeptidome and are of particular interest as they may explain the rapid presentation of peptides from viral proteins which are otherwise metabolically stable (Yewdell et al., 1996; Townsend et al., 1989; Dersh et al., 2021). One described source of DRiPs is translation initiation at near-cognate start codons such as CUG (as opposed to AUG), preferentially from 3' and 5' untranslated regions of mRNA or alternate reading frames. Such non-canonical initiation may be driven by stressors such as viral infection. Another source of peptides seems to be short translation products which can be processed by non-proteasomal proteases (Dersh et al., 2021). The importance of these alternative sources for HLA epitopes is still actively being debated (Admon, 2023; Joyce & Ternette, 2021).

In addition, it has been found that different HLA allotypes are dependent to varying degrees on proteasomal degradation and ubiquitylation. Even further, the dependence of different alleles on tapasin and TAPBPR for peptide loading seems to vary widely (van Hateren & Elliott, 2023). These mechanisms likely all serve to further broaden the peptide repertoire and thereby enhance the potential of cell-mediated immunity (Dersh et al., 2021).

1.2.3 T cells

Where HLA-presented peptides are recognized as foreign, T cells are the immune cells that enable this distinction. They carry T cell receptors (TCR), which are highly variable receptors that can detect specific HLA-bound peptides in the context of their HLA molecule (Murphy et al., 2022).

1.2.3.1 T cell receptors

The high variability of TCRs arises from a process called V(D)J recombination which occurs in developing T cells. In this process, three gene segments called variable (V), diversity (D) and joining (J) are rearranged and joined to form a functional TCR gene. A large repertoire of V, D and J gene segments are available for inclusion. It is further augmented through a process called junctional diversity, where nucleotides are added or deleted at the junctions between the rearranged gene segments. One alpha and one beta chain, generated through VJ and VDJ recombination, respectively, then form a complete TCR. This random process generates a vast and highly diverse repertoire of TCR clonotypes with unique antigen binding specificities (Murphy et al., 2022). It is estimated that it can give rise to between 10^{15} and 10^{20} possible TCR clonotypes but it is unclear what the actual realized diversity of a human repertoire is (Laydon et al., 2015).

1.2.3.2 T cell maturation

During T cell maturation, each T cell, after successful recombination of the TCR genes, undergoes positive and negative selection in the thymus. Through this process, central tolerance is achieved, which ensures that mature T cells are not self-reactive. Initially, T cells start out as double positive for the co-receptors cluster of differentiation 4 (CD4) and CD8. In a first positive selection step, they are tested for their TCRs' capability to bind with sufficient affinity to self-derived HLA:peptide complexes on the surface of thymic cortical epithelial cells. T cells that do not show this capability are neglected and die. Selected T cells lose the expression of either CD8 or CD4 and thereby differentiate into either $CD8^+$ or $CD4^+$ T cells based on whether their TCR interacts with HLA class I or class II, respectively. In a second negative selection step, T cells that strongly bind to self-derived epitopes in the thymus receive a signal that triggers apoptosis. This step is crucial to prevent excessive autoreactivity. The cells that pass this test mature to become naïve T cells, but only a small percentage of about 2 % of T cells will reach this stage. Mature T cells leave the thymus and enter the secondary lymphoid tissues through the blood stream, where they encounter antigens and initiate immune responses (Murphy et al., 2022).

1.2.3.3 T cell types

Many different types of T cells with distinct functions are known.

Helper T cells are $CD4^+$ T cells that recognize antigens presented by HLA class II on pAPCs. They play a central role in coordinating immune responses. Some of the different subsets of helper T cells are T_H1, T_H2, T_H17 and T follicular helper (T_{FH}) cells. T_H1 cells produce interferon-gamma (IFNγ) and thereby help to initiate an inflammatory response and to activate macrophages. In brief, T_H2 cells produce interleukin-4 (IL-4), IL-5 and IL-13 and help to eradicate helminths. T_H17 produce IL-17 and IL-22 and are involved in inflammation and defense against extracellular pathogens. T_{FH} primarily produce IL-21 and help to promote B cell antibody production (Murphy et al., 2022). Regulatory T cells (T_{reg}) are $CD4^+$ T cells that play a critical role in maintaining immune tolerance and preventing excessive immune responses. They are subclassified through their expression of the transcription factor Foxp3 and suppress the activity of other immune cells, including other T cells. Memory T cells are long-lived cells that persist after the initial immune response to an antigen. They enable rapid and enhanced immune response upon re-exposure to the same antigen. Both $CD8^+$ and $CD4^+$ memory T cells exist. Cytotoxic $CD8^+$ T lymphocytes (CTLs) recognize antigens presented by HLA class I on

cell surfaces. They are highly specialized and effective in killing of cells infected by intracellular viruses or cells undergoing malignant cell transformation.

1.2.3.4 Cytotoxic T cells

Naïve CTLs go through different stages, directed by recognition of their specific target antigen on pAPC, which induces priming and eventual activation of the full effector function that leads to specific killing of target cells, production of cytokines and clonal expansion. A key step in this process is the formation of the immunological synapse in response to binding of the TCR to a specific HLA:peptide complex with sufficient affinity. Multiple TCRs may engage with target peptide in close proximity and the interaction is stabilized by adhesion molecules such as LFA-1 on the CTL and ICAM-1 on the target cell. The active synapse can induce apoptosis through engagement of death receptors such as Fas (CD95) and TRAIL on the target cell. The synapse also activates further processes in the T cell, which reorganizes its cytoskeleton to polarize its cytotoxic granules towards to the target cell and eventually releases them in a direct attack. The granules contain a cocktail of proteins that serve different roles to ensure apoptosis. Perforins form pores in the target cell's membrane, allowing the entry of granzymes and other cytotoxic molecules. Granzymes are a family of serine proteases. Once inside the target cell, granzymes initiate the caspase cascade for apoptosis. CTLs can recycle to successively kill further target cells over the course of the effector phase of the immune response (Murphy et al., 2022). Once the effector phase subsides, most CTLs eventually die but it is thought that about 5–10 % differentiate into memory cells that are maintained long-term (Omilusik & Goldrath, 2017).

CTL activation has been found to be a highly sensitive process with claims of detection of as little as one matching HLA:peptide complex on the surface of a target cell, though activation may require on the order of ten complexes (Irvine et al., 2002). They have the capability to migrate towards sites of infection and infiltrate tissues through the use of tissue-homing receptors that direct their exit from the bloodstream through interaction with respective ligands on peripheral tissue vasculature (Brinkman et al., 2013).

While CTLs cells act as the main effectors of cell-mediated immunity, it is important to note that effective cell-mediated immune response requires concerted efforts by $CD8^+$ and $CD4^+$ T cells and by pAPCs (Dersh et al., 2021). Naïve CTLs are typically primed by pAPCs, through co-stimulation with molecules such as CD80 and CD86 that interact with CD28 receptors on T cells. This often involves pAPC uptake of exogenous antigen and presentation on HLA class I in a process called cross-presentation (Colbert et al., 2020). These processes are highly relevant for vaccination (Nelde et al., 2021).

1.2.4 HPV immune evasion and persistence

Most HPV infections are cleared by the immune system within two years. The viral infection can trigger innate immune responses through the recognition of PAMPs, such as the immunogenic viral capsid proteins L1 and L2 or the dsDNA viral genome (Hibma, 2013). The resulting immune response and the recruitment of pAPCs can trigger a cell-mediated adaptive immune response through presentation of HPV-derived peptides to $CD4^+$ T cells via HLA class II but most likely also to $CD8^+$ T cells via cross-presentation on HLA class I (Hibma, 2013). Eventual elimination of the HPV infection is considered to be driven by an effective T cell response involving CTLs and T_H1 cells that is directed against proteins E6, E7 and E2 (Woo et al., 2008; Schiffman et al., 2016). Still, a median duration of 6 months for clearance of infections is much slower than what is observed with most other viruses and about 10 % of cervical infections can become chronic (Grabowska & Riemer, 2012; McBride, 2022).

The exact mechanisms that determine whether an infection is cleared or persists are not fully understood. However, various mechanisms have been identified that allow HPV to evade the host

immune response and establish lo ng term persistence within the infected epithelium (Grabowska & Riemer, 2012). One mechanism for immune evasion are the generally low levels of expression of the early genes in the lower levels of the epithelium that is mediated by the protein E2 (Steinbach & Riemer, 2018). In the upper layers, where the immunogenic capsid proteins are expressed, early gene expression is tightly regulated such that the antigens E6 and E7 are not exposed (Roden & Stern, 2018). Steinbach and Riemer (2018) summarized that the virus affects antigen processing of infected cells in a variety of ways, ranging from suppression of immunoproteasome subunits, suppression of TAP, overexpression of ERAP1 and suppression of HLA surface expression. In effect, the immunopeptidome is significantly altered which affects the effective generation of viral epitopes. The viral oncogenes E6 and E7 play a crucial role in immune evasion. Apart from their important function in keeping the cell cycle active and preventing apoptosis (see section 1.1.4), E6 and E7 also have been found to play a crucial role in immune evasion. They suppress TLR9 expression and more generally, hrHPVs have been found to suppress TLR3. The JAK-STAT pathway is targeted by silencing of STAT1 by both E6 and E7. It has been found that reduced numbers of Langerhans cells, which are the pAPCs of the squamous epithelium, are present at sites of hrHPV infection. This is attributed to the downregulation of both, CC-chemokine ligand 20 (CCL20) and E-cadherin, by E6 and E7. CCL20 is an important factor in recruitment of Langerhans cells and E-cadherin enables binding of Langerhans cells to keratinocytes (Schiffman et al., 2016). These and several further mechanisms such as suppression of cytosolic DNA and RNA sensing serve to suppress innate immune reactions and IFN production, which in turn compromise and/or delay the adaptive immune response (Gusho & Laimins, 2021). These mechanisms allow the virus to persist within the epithelium, where it only replicates very slowly, without causing systemic inflammation. The entire life cycle is completed without directly causing cell death or systemic viraemia (Roden & Stern, 2018).

HPV is also prone to integration into the host's chromosomal DNA. This process often disrupts the E2 gene during linearization of the genome due to which its repressive effects on E6 and E7 are lost. The resulting overexpression compromises host DNA repair and leads to genomic instability (Jeon & Lambert, 1995). In turn, this makes E6 and E7 ideal targets for immunotherapy (Steinbach & Riemer, 2018).

While HPV itself does not directly cause cancer, the expression of the viral oncoproteins and the resulting genomic instability and dysregulated cell proliferation that provide a head start in immune evasion put infected cell on a path towards malignant transformation over the span of 1–2 decades (Roden & Stern, 2018). Due to this, persistent HPV infection is highly implicated in development of cancer.

1.3 HPV prevention and treatment

1.3.1 Prophylactic vaccination

Prophylactic HPV vaccines were developed to prevent HPV infections and thus reduce the risk of HPV-associated cancers. All current prophylactic HPV vaccines are based on the major capsid protein L1. The L1 proteins self-assembles into virus-like particles that are devoid of viral genome and induce a potent immune response, achieving high levels of neutralizing antibodies. Antibodies can bind HPV virions before they infect epithelial cells, which triggers elimination by native immune cells. This confers immunity to initial infection with HPV because the steps from binding to the basement membrane of the epithelium to uptake of the virion into epithelial cells take on the order of several hours. During this time, antibodies effectively neutralize the virus (Schiller & Lowy, 2018).

Currently, four prophylactic HPV vaccines are available. The first vaccine became available in 2006 and is a quadrivalent vaccine that offers protection against HPV 6, 11, 16 and 18. The second vaccine

became available in 2007 and protects only against the two most prevalent high-risk HPV types 16 and 18. The third vaccine became available in 2014 and is a nonavalent vaccine that protects against HPV 6, 11, 16, 18, 31, 33, 45, 52 and 58. The last vaccine to have been licensed is another bivalent vaccine for HPV16 and 18 (Prudden et al., 2022). Only limited cross-protection against further HPV types is achieved by the prophylactic vaccines (Malagón et al., 2012). Together, they all contribute to the global vaccine supply and further prophylactic vaccines are in development.

The major limitation of preventative vaccines is that they have no therapeutic effect on persisting HPV infection or lesions. Persistent HPV infections are not affected by the antibody response because even though they can trigger cell-mediated immune responses, the L1 antigen is not expressed in the infected basal epithelial cells (Roden & Stern, 2018). Therefore, it is projected that even though prophylactic HPV vaccination for adolescent girls will significantly reduce the incidence of cervical cancer in the future, substantial effects will likely not be seen for 30–40 years (Brisson et al., 2020). This is partially due to the fact that by now only about 13 % of young girls are fully vaccinated globally (Prudden et al., 2022). It is increasingly recognized that vaccination of boys and adolescent men can play a role in reducing the incidence of HPV-driven cancer. Men are an important reservoir of genital HPV infections and are themselves at risk of HPV-related morbidity. By 2022, 45 countries implemented vaccination for boys (Bruni et al., 2023a).

1.3.2 Conventional treatment

Screening tests allow for the detection of squamous intraepithelial lesions of the cervix that arise from persistent infection. These are distinguished as low-grade squamous intraepithelial lesions (LSIL) and high-grade squamous intraepithelial lesions (HSIL). HSIL are considered cancer precursors and, if undetected and untreated, they can progress over years or decades to cervical cancer. HSIL have also been identified as precursors to anal and vulvar and vaginal cancers. For oropharyngeal cancer no such precursor is known (Hall et al., 2021). For cervical cancer, consistent screening and histological diagnosis which indicates the appropriate treatment has been implemented highly successfully in high-income countries (HICs) and has reduced mortality considerably. The most widely known screening option is the Papanicolaou (Pap) test (Papanicolaou & Traut, 1997). It is a cytology-based test which is performed by inspecting a sample from the cervix (a cervical smear) under a microscope for the presence of abnormal cells. It has largely been replaced by Liquid Based Cytology. Newer screening methods such as HPV DNA or mRNA detection offer sensitive and cost-effective alternatives. However, up until now, screening measures have proven difficult to implement to the same effect in low- and middle-income countries (LMIC) (Prudden et al., 2022). Furthermore, as no precursor lesions are known for HPV-associated HNSCCs, there are no screening tests available (Johnson et al., 2020).

Treatment indications for detected HSIL lesions depend on the characteristics of the lesions. Small lesions that are entirely visible on the ectocervix may be ablated via thermal ablation, either by freezing or heating the tissue. Other lesions can be surgically removed by removing the entire abnormal transformation zone (Prudden et al., 2022). Treatment of suspected cases of cervical cancer is based on the stage of the disease. Early stages of the cancer can be treated with surgery or radiotherapy, including the option for chemotherapy. These measures have high success rates with long-term survival and cure rates of around 80 % but are only enabled by early diagnosis and high-quality treatment. However, they come with several drawbacks as they can cause radiation-induced damages to surrounding tissues and can affect fertility or sexual functions (Cohen et al., 2019). Disease progression to late stage metastatic cancer drastically reduces efficacy of standard of care radio- and chemotherapies (Yao et al., 2022).

1.3.3 Immunotherapeutic approaches

Immunotherapeutic approaches are intended to harness the body's immune system to detect and eliminate disease. Aside from non-specific approaches like immune checkpoint blockade, antigen-specific approaches are becoming more and more relevant in recent years (Waldman et al., 2020). Targeted immunotherapy aims to modulate the immune system's response to specific antigens such as virus- or cancer-derived HLA-presented epitopes. HPV-driven malignancies are in principle well suited for such immunotherapeutic approaches due to the obligatory presence of the known viral antigens E6 and E7 (Khallouf et al., 2014).

Today, immune checkpoint blockade treatments have already been approved for the treatment of certain HPV-associated cancers, and targeted immunotherapies such as therapeutic vaccines and TCR-based cellular therapies are being developed (Lee & Allen, 2021). In TCR-based therapies which use genetically engineered T cells, selected TCRs or chimeric antigen receptors (CARs) can be used to direct action against specific antigens (Norberg & Hinrichs, 2023). In contrast, the vaccination approach is intended to induce a cell-mediated immune response in the form of natural antigen-specific cytotoxic T cells and T helper cells (Peri et al., 2023).

Therapeutic HPV vaccines offer the potential to clear persistent HPV infection, cause regression of precursor lesions and cure HPV associated cancer. Furthermore, the induction of immunological memory is expected to protect the vaccinated individual against possible latent infection leading to tumor recurrence and against novel infections. Such a treatment could address the gaps currently left by prophylactic vaccination and screening programs.

No therapeutic vaccines are currently approved but several candidates are in development. A large variety of approaches have been used for the development of such vaccines, such as live vector-, nucleic acid-, protein- and peptide-based vaccines (Lee & Allen, 2021). Predominantly, they are designed to induce immune reactions against the proteins E6 and E7 and target advanced cervical cancer or HSIL (Prudden et al., 2022). To ensure safety where complete E6 and E7 are being used as immunogens, disabling mutations can be used, or even a 'shuffled' sequence of the protein, to abrogate binding to the main target proteins p53 and Rbs (Öhlschläger et al., 2006; Khallouf et al., 2014). While these are typically designed to minimally affect the derived epitope repertoire, they may still affect the production of viable epitopes.

Live vector-based vaccines for HPV utilize bacterial or viral vectors. These vectors are highly immunogenic themselves due to their pathogenic nature. The vectors used in these vaccines are engineered to lose the ability to replicate, reducing safety concerns. They can also be designed to deliver the immunogenic payload directly to pAPCs, *e.g.* through the use of lentiviral vectors. However, there may be pre-existing immunity towards the vector, which can limit vaccine efficacy in the general population as the vectors may be eliminated quickly (Lee & Allen, 2021; Khallouf et al., 2014).

Nucleic acid-based vaccines include DNA-based vaccines and RNA-based vaccines. These vaccines rely on the delivery of nucleic acids into cells, where they are translated and processed into antigens that can induce an immune response. Both types are relatively easy to produce and do not induce anti-vector immune responses, which means that even though they may be directly immunogenic, multiple applications will still be able to trigger responses specifically against the encoded antigens (Chabeda et al., 2018). Nucleic acid-based vaccines require delivery methods such as electroporation to enhance their uptake by cells. RNA vaccines have the advantage of not needing to cross the nuclear membrane, which may improve their transfection efficiency. Recently, mRNA-based approaches that are delivered via lipid nanoparticles have shown the capability to control HPV-associated tumors in mice (Ramos da Silva et al., 2023). They have attracted widespread interest and are expected to

receive increased attention in development of HPV vaccines due to their successful application in the COVID-19 pandemic (Polack et al., 2020; Baden et al., 2021). Nucleic acid-based vaccines allow for considerable flexibility in the encoding of target antigens or some variation of individual epitopes. In addition, proteins expressed from nucleic acid-based vaccines have to undergo natural antigen processing and presentation. (Lee & Allen, 2021; Hancock et al., 2018).

Protein-based vaccines for HPV utilize proteins derived from the virus to induce an immune response. In principle, these vaccines can contain all antigenic HLA epitopes and are not HLA-restricted. E7 is most frequently used for protein vaccines because of the two oncoproteins, it shows higher expression and is more conserved (Khallouf et al., 2014). Overall, the proteins have low immunogenicity. To enhance immunogenicity and presentation via HLA class I, protein-based vaccines employ various strategies. One approach is the creation of fusion proteins. These include fusions of viral proteins E6 and/or E7 with more immunogenic proteins such as L2 or with PAMPs to engage pAPCs (Khallouf et al., 2014). This can be achieved by fusing the antigen to molecules that specifically bind to receptors on DCs. Another strategy is the use of adjuvants with the vaccine formulation. These are substances that enhance the immune response (Lee & Allen, 2021).

Peptide-based vaccines for HPV utilize specific peptides derived from HPV antigens to induce an immune response against the virus. Peptide-based vaccines have an excellent safety profile and are easy to produce. However, depending on the approach they may be limited by their immunogenicity and HLA restriction. The use of strategies to enhance immunogenicity is crucial with peptide-based vaccines to be able to elicit a sufficient immune response and consequent immunological memory. Adjuvants are commonly used to improve immunogenicity as well as further approaches like lipopeptide conjugation for recognition via TLRs and direct loading of dendritic cells (Purcell et al., 2007). The use of synthetic long peptides can help to target multiple HLA types. This is achieved by using longer stretches of overlapping peptides between 15–35 amino acids in length which may accommodate multiple HLA restrictions (Khallouf et al., 2014). This means they require processing by pAPCs for HLA-presentation via HLA class II and for HLA class I cross-presentation to induce the desired response. A disadvantage of this approach is that processing of the long peptides does not follow the same pathway as the presentation of HPV epitopes on infected or cancerous cells. This might result in immune responses that are not specific to the target (Purcell et al., 2007). The vaccination with short peptides offers an alternative that can be precisely directed against selected epitopes. This allows to select the most promising epitopes, taking into account their immunogenicity and presence during the targeted disease stage. Each peptide used for this approach is effective only according to its HLA restriction but the use of multiple peptides and favoring of promiscuous HLA binders can help overcome this limitation (Khallouf et al., 2014; Lee & Allen, 2021).

Antigen selection is a critical step in the development of therapeutic HPV vaccines. This is emphasized through the promising developments stemming from the first HPV epitope for which HLA-presentation on cancer cells was validated via mass spectrometry (Riemer et al., 2010). With the epitope E7/11–19, a therapeutic vaccination strategy was successfully employed in an MHC-humanized E6/E7 mouse tumor model (Kruse et al., 2018). Furthermore, a therapeutic vaccine developed against the epitope, which is A02 restricted, is being tested in a clinical trial (DPX-E7, NCT02865135). An epitope-specific TCR has been isolated and enables T cells to kill cancer cell lines *in vitro* as well as established tumors in a mouse model (Jin et al., 2018). In the subsequent clinical study, TCR-based cellular therapy with this receptor led to robust tumor regression in 6 of 12 patients (Nagarsheth et al., 2021).

Despite these indications of HLA-presentation of viral epitopes, the development of a therapeutic vaccine has proven to be challenging, likely due to the low immunogenicity of persistent HPV infections. A vaccine must be highly immunogenic and elicit a strong antigen-specific cytotoxic

response. Where HSIL or cancer have undergone immune selection, a dominance of regulatory T cells and an immune-repressive environment pose additional challenges (Prudden et al., 2022).

1.4 Immunopeptidomics

1.4.1 Introduction

Immunopeptidomics is the study of the immunopeptidome, which is the repertoire of HLA-presented peptides of a given set of cells. Mass spectrometry-based immunopeptidomics has been pioneered in the 1990s and early 2000s in studies only where up to tens of peptides were identified (Falk et al., 1991; Hunt et al., 1992). Ever since, the field has been rapidly evolving due to improvements in experimental techniques and advancements in mass spectrometry that enable the detection of tens of thousands of peptides per experiment. It has been the critical technique that has revolutionized our understanding of the immunopeptidome and many of its crucial aspects such as the dynamics of HLA restriction and peptide generation, editing, loading and cell-surface presentation (Shapiro & Bassani-Sternberg, 2023). By now, the impact of the field goes beyond basic research as it has paved the way for novel approaches in several fields. It enables personalized cancer immunotherapies by identifying tumor-specific HLA-presented peptides that can be specifically targeted while sparing healthy tissues. In autoimmune diseases, the identification of self-peptides provided novel insight into the role of aberrant antigen presentation in autoimmune pathogenesis. In therapeutic vaccination against pathogens such as HPV, the technique allows to specifically identify the pathogen-derived peptides that are being presented by infected cells or pAPCs. This offers a new objective basis for vaccine design (Yewdell, 2022).

The most established technique for immunopeptidomics is purification of HLA:peptide complexes via immunoprecipitation with subsequent elution of peptides by acid treatment and eventual acquisition via liquid chromatography mass spectrometry (LC-MS) (Bassani-Sternberg, 2018; Chong et al., 2018; Purcell et al., 2019). The experiment is illustrated in Figure 4.

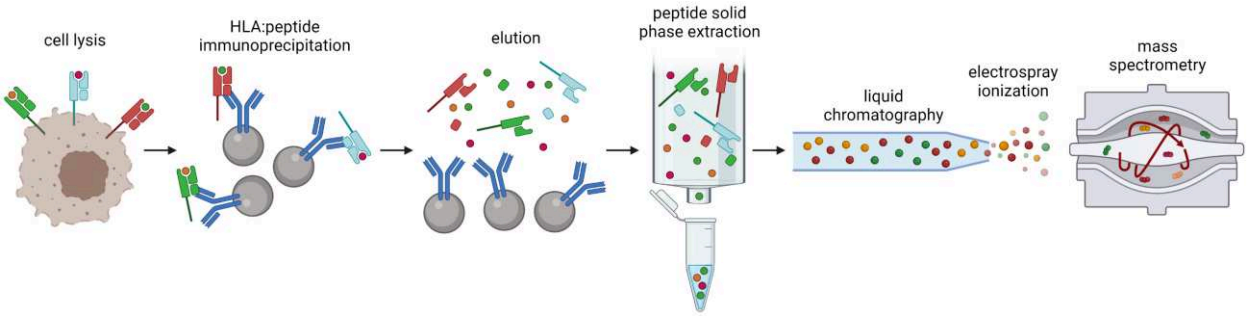


Figure 4. HLA immunoprecipitation and LC-MS analysis. Individual peptides are represented as colored spheres in the binding pockets of the HLA molecules. Different coloring of the HLA molecules signifies different allotypes. Source material is lysed and in an immunoprecipitation step, HLA:peptide complexes are bound by specific antibodies that are crosslinked to agarose beads. The anti-pan-HLA class I antibody W6/32 effectively binds most HLA class I molecules. After several washing steps, mild acid treatment is used for elution. This elutes bound peptides but also dissociates the α and β2m chains of the HLA heterodimer. Solid phase extraction is used to purify and concentrate eluted peptides. Analysis is performed by combined LC separation, electrospray ionization and acquisition in the mass spectrometer.

1.4.2 HLA immunoprecipitation

HLA immunoprecipitation (HLA IP) is the technique of choice to isolate HLA-presented peptides from the complex background of peptides, proteins and other molecules present in whole-cell lysate and make them accessible to analysis via LC-MS. It involves the specific capture and purification of HLA molecules, which carry with them the bound peptides. In the context of immunopeptidomics, the

term HLA IP commonly refers to the full process that produces purified HLA ligands and will be used as such in this thesis.

The process typically begins with the lysis of cells or tissues using mild detergents to solubilize membrane-bound HLA molecules. Immunoprecipitation is performed by incubation of the lysates with specific antibodies that are coupled on solid supports such as protein A or protein G Sepharose beads. Especially the pan-HLA class I specific antibody W6/32 has found widespread use due to its binding of a conserved epitope on the HLA class I alpha chain shared among HLA A, B and C alleles (Kuznetsov et al., 2020). After incubation, the captured complexes are separated from the non-specifically bound components through repeated washing steps. To elute the bound peptides, various methods are employed. One common approach is the use of mild acid elution, where a low pH elution buffer disrupts the peptide-MHC interactions and releases the peptides. This step typically also elutes the α chain and the β2m molecule from the antibody/bead complex. Other contaminants such as the detergents used for lysis and some salts may still be present. This necessitates a further step to purify and concentrate the HLA-associated peptides, which is most commonly solid-phase extraction (SPE). In this step, a sorbent such as C18 hydrophobic groups binds peptides and proteins of sufficient hydrophobic character, while water-soluble hydrophilic components are washed out. Elution may then be performed with organic solvents of limited concentration to favor the elution of small peptides as opposed to larger, more hydrophobic proteins. The resulting solution of concentrated and purified HLA-presented peptides is then ready for analysis (Kuznetsov et al., 2020).

The technique has some inadequacies. It is estimated that associated with the process approach 90–99 % for some peptides, that the immunoprecipitation step is not quantitative and that the peptide yields can vary drastically between laboratories due to the challenging execution (Hassan et al., 2014). To reach a suitable peptide yield, often 100 million to 10 billon cells are used as input material (Kuznetsov et al., 2020). While these facts indicate that the method has not yet fully matured, the variance of the results is also partially owed to the challenging execution.

1.4.3 Mass spectrometry-based proteomics

1.4.3.1 Introduction

Mass spectrometry-based proteomics is a powerful and widely used technique in the field of biochemistry that has revolutionized the study of proteins and their functions. The last decade has been one of tremendous progress for the underlying technology of mass spectrometry. This has enabled comprehensive analysis of entire proteomes and has spurred rapid growth of the field.

Fortunately, the dominant technique for such proteome analysis is bottom-up proteomics, for which all proteins in a sample are first enzymatically digested, typically with the endopeptidase trypsin (Aebersold & Mann, 2003). Protein content is then inferred from the comprehensive sequencing and quantitation of all LC-MS detectable peptides. This means that the goals of bottom-up proteomics and immunopeptidomics analyses are largely aligned. Peptides in a sample are separated by liquid chromatography, ionized through electrospray ionization and detected in the mass spectrometer. While most computational tools have been developed for bottom-up proteomics, they have largely been successfully co-opted for immunopeptidomics. Immunopeptidomics remains the more challenging of the two applications. Tryptic digestion occurs predictably after arginine (R) or lysine (K) and in bottom-up proteomics this information is used to restrict the search space. Due to the lack of such restriction, the search space for immunopeptidomics is vastly inflated and analysis is far more computationally expensive and susceptible to false discovery. Also, the inherently diverse C-terminus of HLA binders often results in worse ionization and less informative ion fragmentation that makes MS detection more challenging (Faridi et al., 2018).

In the following I will introduce the proteomics techniques that I used in this thesis for the detection of peptides.

1.4.3.2 Reversed-phase liquid chromatography

Effective peptide separation is crucial to unlock the full potential of detection via mass spectrometry. Without effective chromatographic separation, sample complexity is usually overwhelming for even advanced mass spectrometers.

In reversed-phase liquid chromatography (RPLC), the nonpolar stationary phase is used for analyte retention and a polar mobile phase is used for elution. The most commonly used stationary phase in RPLC is based on different particulates that are bonded with hydrophobic alkyl chains, such as C18 or C8. These alkyl chains provide a nonpolar surface that nonpolar analytes interact with. The mobile phase is typically a mixture of water and an organic solvent such as acetonitrile. Peptides are first introduced to the stationary phase with a low-organic solvent to favor stationary interaction. Gradient elution, where the percentage of organic solvent in the mobile phase is gradually increased in what is termed an LC 'run', is then used to achieve peptide separation. Peptides with higher hydrophobicity tend to interact more strongly with the stationary phase and as a result elute later in the gradient. The gradient conditions, including the gradient slope, duration, and starting and ending concentrations of the organic solvent, need to be optimized to achieve the desired separation (Lenčo et al., 2022). An ion-pairing reagent, such as formic acid or trifluoroacetic acid, is added to form neutral pairs with charged peptides. This increases their hydrophobic character and enhances peptide retention and separation (Kota & Stolowitz, 2016).

1.4.3.3 Electrospray ionization

Electrospray ionization (ESI) is a soft ionization technique that generates gas-phase ions from analyte molecules in solution while largely preserving their structural integrity. The sample is delivered through an emitter, which is a capillary with a narrow tip. As liquid emerges from the tip, which is held at high voltage relative to a counter-electrode, a spray cone forms from which a fine spray of charged droplets is ejected. This is followed by solvent evaporation, which increases the charge density on droplet surfaces. Such highly charged droplets, due to electrostatic repulsion on the surface, break up into several smaller droplets. This reaches a critical point at which ions are ejected into gas phase (Ho et al., 2003; Niessen & Correa C, 2016). Peptides are typically ionized with the tip at a positive potential as they are easily protonated to form cations. They may acquire multiple positive charges, especially where this is favored by the presence of basic amino acids (Loo et al., 1989).

1.4.3.4 MS peptide detection

The mass spectrometer used for this thesis is a Thermo Scientific Orbitrap Exploris 480. It is a hybrid mass spectrometer that combines high performance quadrupole precursor selection with high-resolution accurate-mass detection in the Orbitrap (Makarov, 2000). Figure 5 shows the components that make up the instrument.

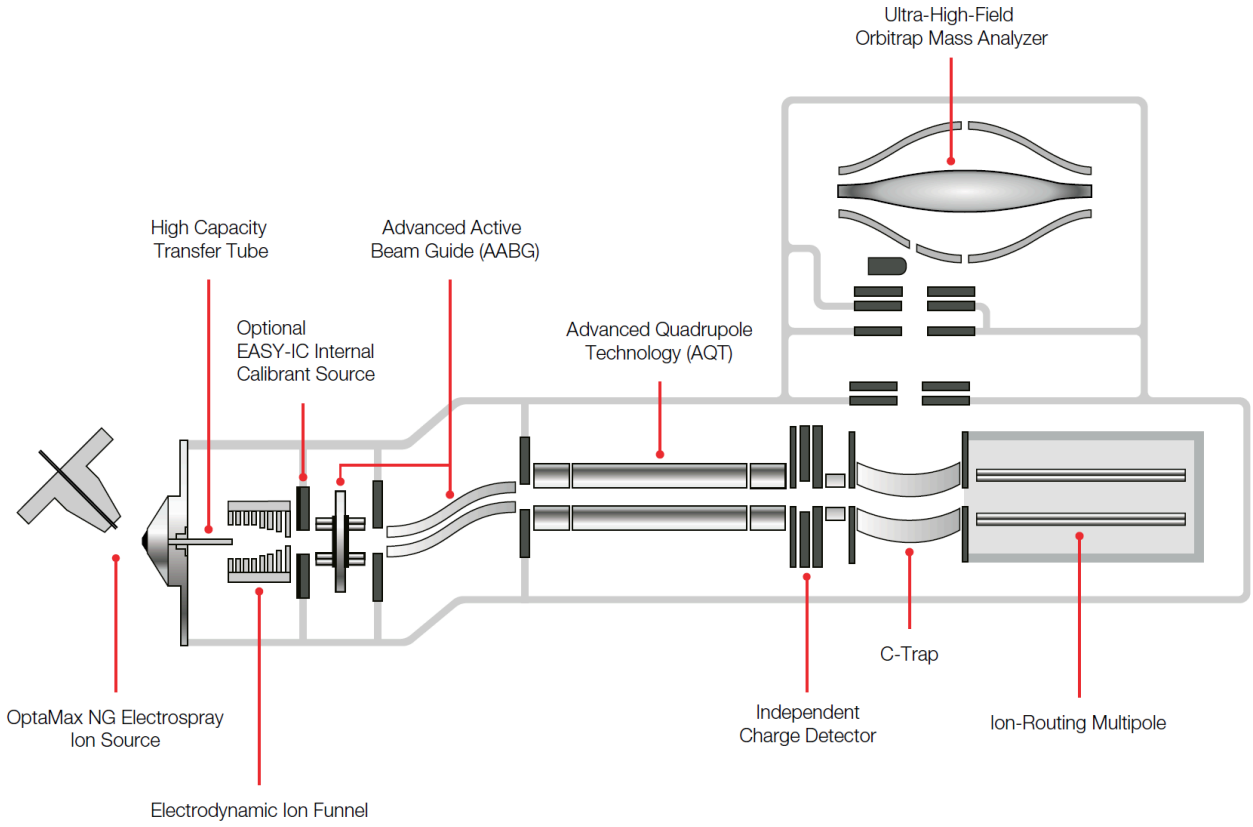


Figure 5. Schematic of the Orbitrap Exploris 480 MS used for this thesis. Ions are routed from the electrospray on the left side through the ion optics, which include the quadrupole. The quadrupole is a mass filter that allows for pre-selection of ions. Fragmentation is performed in the ion-routing multipole. Mass analysis is performed in the Orbitrap. Reprinted from product specification sheet with permission from Thermo Fisher.

The inside of the instrument is a vacuum chamber, to enable the precise and efficient routing of ions. The ions generated at the ESI source first encounter the instruments ion optics. They enter the instrument through the inlet ion transfer tube and the ion funnel, which is a stack of electrodes to which a radio frequency (RF) voltage is applied. Oscillating RF potentials can be used to confine and focus ions in gas phase. In this manner, the funnel radially focuses arriving ions. The focused ion beam is then guided through the advanced active beam guide (AABG), which is comprised of the injection filter (left) and the bent flatapole (right). The bent flatapole guides the ions through an arc whereby remnant neutral particles are cleared as they are not diverted (Thermo, 2021).

The quadrupole mass filter is a crucial part of the instrumentation that enables the pre-selection of ions. The Orbitrap Exploris uses a segmented quadrupole. Entrance and exit segments are RF-only segments that ensure that ions are focused whereas the middle section is the resolving quadrupole. This design is optimized for rectangular transmission window shapes, which is crucial to enable high efficiency ion selection at narrow isolation windows (Thermo, 2021).

The intensity of the permitted ions is then determined in the independent charge detector (ICD). This device increases the fidelity of the automatic gain control (AGC) that maintains an optimal quantity of ions for each scan. Next, ions pass through the curved linear trap (C-Trap) into the Ion-Routing Multipole (IRM). The IRM is filled with nitrogen. It traps the ions and optionally performs higher energy collisional dissociation (HCD). HCD is a variation of the low energy fragmentation method collision-induced dissociation (CID), which shows improved trapping of low-mass products (Olsen et al., 2007). In this technique, selected ions are accelerated towards the neutral collision gas which induces fragmentation. The collision energy can be adjusted by varying the offset voltage between

the C-Trap and the IRM. This value affects the outcome of the fragmentation such that different fragmentation results can be produced. After the trapping, the ions are passed back to the C-Trap which facilitates the simultaneous transfer of the ions into the Orbitrap (Shuken, 2023).

Mass analysis occurs at the Orbitrap, which consists of a spindle-shaped central electrode surrounded by an outer electrode that is split into bell-shaped halves. The basic principle is that for each scan, packets of ions are trapped in stable trajectories that combine rotation around the central electrode (radial position r) with oscillations along it (axial position z). Ions of the same mass-to-charge ratio (*m/z*) oscillate together along the z-axis while distributing evenly in the radial direction. The outer electrode detects the image current produced by the oscillations. Application of fast fourier transformation (FFT) to the image current then generates the frequencies of these z-axis oscillations, which correspond to the ion *m/z*. In this manner, each Orbitrap scan generates a spectrum of *m/z* values and corresponding intensities. Through acquisition over long transients (long scan time), spectra can be acquired at exceptionally high resolutions. The stated resolution usually reflects resolution at 200 *m/z* and scales inversely towards higher *m/z*. Due to this, measurement errors can best be understood as parts per million (ppm) deviations, which means that at higher *m/z*, higher mass errors are expected (Haag, 2016).

Due to the optional fragmentation step, two options for ion detection are available. In a so-called MS1 scan, the ions are left intact. This is typically done in a full MS1 scan, meaning that all or most ions are permitted through the quadrupole, generating an overview spectrum of all ions that are entering the instrument. MS2 scans are performed by pre-selecting ions (termed precursor ions or simply precursors) and fragmenting them. MS2 scans are the main tool for specific identification and sequencing of peptides.

Figure 6 shows the different ion that result from peptide backbone breaks through fragmentation. The accepted nomenclature was first introduced by Roepstorff and Fohlman (1984) and names the charged C-terminal fragment ions x, y or z and the charged N-terminal fragment ions a, b or c. In addition, the amino acid (AA) length of the fragment is given. Multiple terms for these fragment ions are in wide-spread use and can be used interchangeably, such as transition (from precursor to fragmentation product) or product ion. The neutral analyte molecule is named M and its protonated form is specified as $[M+H]^+$ or $[M+nH]^{n+}$ for n positive charges. This is typically the notation reserved for the precursor. The peptide amide bond is the most susceptible to fragmentation and produces the common y and b fragment ions (Tabb et al., 2003). b ions generally take on an oxazolone structure. This favors the subsequent loss of CO, producing an a ion (Johnson et al., 1988). y, b and a are the three dominant fragment ion types of interest produced by HCD fragmentation. Further ion-species derive from neutral losses that are common with HCD. Fragments containing AAs serine (S), threonine (T), asparagine (N) and glutamic acid (E) commonly show loss of water ($-H_2O$; -18 Da). AAs Asparagine (N), Glutamine (Q), Lysine (K) and Arginine (R) commonly show loss of ammonia ($-NH_3$, -17 Da) (Tabb et al., 2003). Also seen is the neutral loss of methanesulfenic acid ($-CH_3SOH$, -64 Da) from oxidized methionine (M[Oxi]) (Lagerwerf et al., 1996). When fragmentation products retain a charge from the precursor, they can be detected as the characteristic ions (e.g. $y_3^+-H_2O$). The combined detection of such fragments allows to deduce the AA sequence of the precursor peptide. This method for peptide sequencing is often hindered by detections with only incomplete sequence coverage. This happens because some backbone breaks are disfavored by collisional dissociation, which causes the corresponding ions to be undetectable, leaving a gap in the detected ion series. Therefore, further indications have to be considered, such as the characteristic relative intensity of the fragment ions or the LC retention time associated with the signal (Gallien & Domon, 2015; Toprak et al., 2014). In contrast to fragment masses, these are not available from theoretical calculations but are understood to be highly effective for peptide identification (Toprak et al., 2014; Searle et al.,

2018). Recently, deep learning approaches have made considerable progress in predicting these peptide characteristics with some accuracy but they can most reliably be determined through the acquisition of synthetic peptides (Gessulat et al., 2019; Bouwmeester et al., 2021; Gallien et al., 2015). Where a reference fragmentation pattern is known, the acquired signal intensities can be compared via different similarity measures. Toprak et al. (2014) compared several measures and established the highly discriminatory normalized spectral angle score (NSA) for peptide identification that I use here. It scores similarity between 0 and 1, where 1 indicates complete identity.

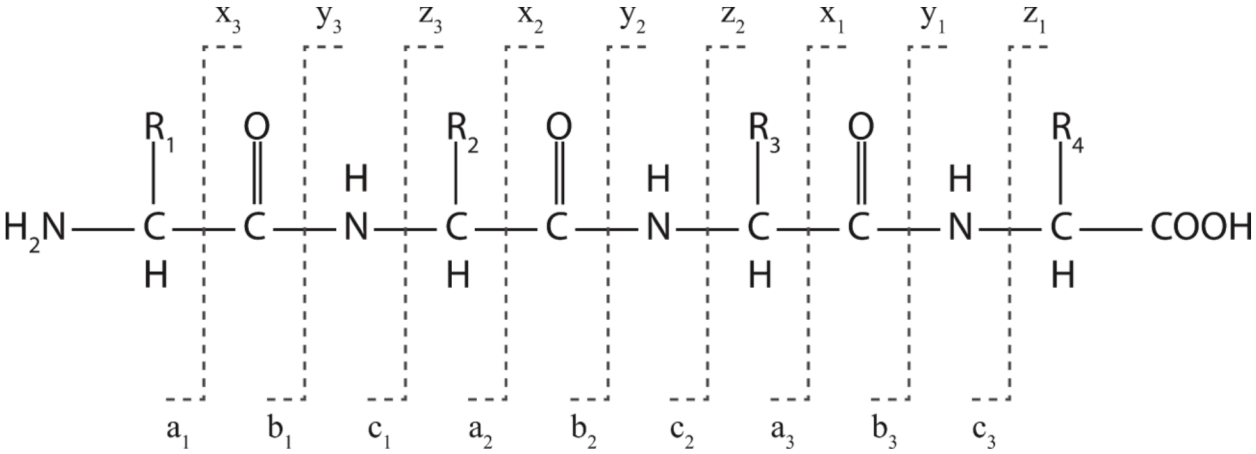


Figure 6. Peptide fragmentation products. All possible peptide backbone breaks are indicated. The accepted nomenclature labels fragment ions from the C-terminus x, y and z, and ions from N-terminus a, b and c. The letters indicate which type of bond was broken. A number specifies the length of the fragment ion in AAs. Created by H. Röst and M. Steiner licensed under CC BY-SA 3.0.

1.4.3.5 MS acquisition schemes

The hybrid quadrupole-Orbitrap configuration of the MS enables several different acquisition methods as illustrated in Figure 7. The two untargeted methods data-dependent acquisition (DDA) and data-independent acquisition (DIA) aim for complete analysis of the precursor ions available at any given time. For this, the DDA methods depends on ion detections in MS1. From MS1, potential precursors are identified and successively isolated and fragmented with individual scans. Due to time constraints, the method usually applies a TopN strategy to select the most promising precursors. Therefore, DDA has two requirements that limit sensitivity, as a precursor must be detectable in MS1 and further be selected for fragmentation. More recently, DIA has found widespread use as it effectively harnesses of the high resolution and speed of the Orbitrap. It aims to fragment all available precursors within broad *m/z* ranges of interest. To this end, several wide precursor windows are permitted and scanned in sequence. Spectra generated this way are typically hybrid spectra, that contain fragments derived from several precursors. For analysis, extracted ions from several spectra are projected on the retention time axis to create an extracted ion chromatogram (XIC). Signals originating from a single precursor cluster as a peak in the XIC and peptide identification is performed via analysis of the peaks. Sensitivity is limited due to the limited dynamic range of the detector, as dominant ions may quickly fill the Orbitrap to capacity, which causes low-intensity ions to fall below the limit of detection. MS1 is not strictly required for DIA as the acquisition scheme is data-independent but the additional precursor information is typically highly informative for peak identification (Hu et al., 2016). Another option is the targeted acquisition scheme termed parallel reaction monitoring (PRM). The name distinguishes this scheme from the targeted method used with previous generation instruments called single reaction monitoring (SRM), where a reaction refers to the transition of a precursor ion to a specific fragment ion. In SRM, transitions were acquired sequentially whereas in PRM, all ions generated from fragmentation can be acquired in one Orbitrap scan. PRM allows to specifically target any precursor with potentially long injection times and high

resolution for maximum sensitivity and mass accuracy, respectively (Hu et al., 2016). Target signal can readily be projected as XIC, which delivers more robust quantitation. MS1 is not required for PRM but is typically acquired for possible detection of the precursor and for diagnostics.

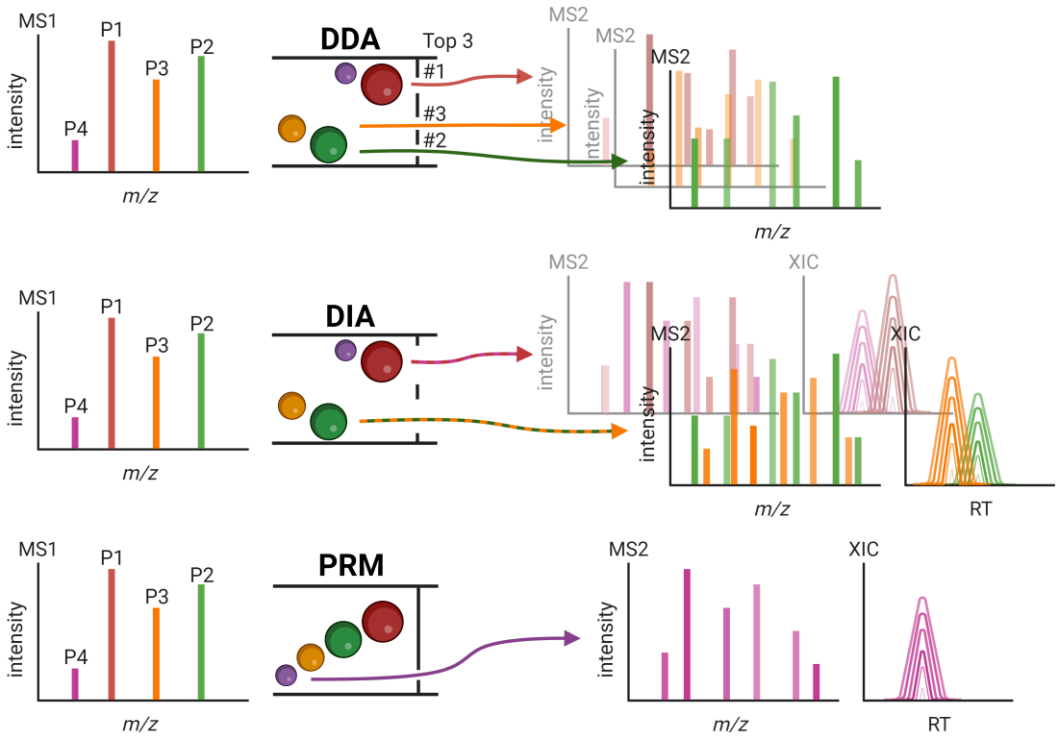


Figure 7. MS acquisition schemes used in this thesis. Two untargeted schemes, DDA and DIA, are illustrated. PRM is the targeted scheme. Generally, MS1 is acquired first, followed by acquisition of MS2, for which some form of precursor isolation is performed by the quadrupole (center). This happens fast enough so that ions detected in MS1 are usually still available for collection and isolation in the subsequent scans. For DDA, a top 3 strategy is shown. MS1 is non-obligatory for DIA and PRM. Adapted with permission from Dr. Jonas P. Becker.

1.4.4 Peptide HLA binding prediction

The distinct binding specificities of different HLA alleles result in a broad spectrum of peptides being presented across different individuals (Murphy et al., 2022). Due to the important role of peptide binding and presentation in the adaptive immune response and the clinical potential of identified epitopes, various tools for computational prediction of peptide HLA binding have been developed in the past decades (Rammensee et al., 1999; Nielsen et al., 2007; Bassani-Sternberg et al., 2017). Recently, the wealth of data generated by mass spectrometry-based immunopeptidomics and the use of deep learning techniques has driven rapid improvements in the performance of these tools (Jurtz et al., 2017; Reynisson et al., 2021). Pan-HLA-specific tools base predictions on the subset of the HLA protein sequence known to be in close proximity with the peptides, which has enabled binding predictions for HLA alleles of any known sequence (Nielsen et al., 2007).

However, the so-called eluted ligand (EL) data carry some limitations due to inherent biases of immunopeptidomics in its current form. An underrepresentation of cysteine-containing peptides is understood to stem from the susceptibility to oxidative modification (Ternette et al., 2023). Peptides that are too hydrophobic or too hydrophilic are systematically lost during the employed purification methods. Furthermore, due to common homologies in human proteins or domains within proteins, the frequency of certain motifs can be artificially enhanced in MS-data (Gfeller & Bassani-Sternberg, 2018).

By now, peptide binding predictions reportedly achieve about 90 % accuracy for common HLA alleles with performance expected to keep improving (Dersh et al., 2021). Prediction tools have become an invaluable tool for analysis of immunopeptidomics experiments and are being used in HLA epitope identification pipelines (Kovalchik et al., 2022; Coelho et al., 2020; Hundal et al., 2020). Still, the use of such tools alone fails to predict a significant amount of valid HLA binders (Bonsack et al., 2019). As an alternative, competition-based binding experiments using cell-bound HLA molecules and competing synthetic peptides can be performed to measure experimental IC50 values that approximate the true binding affinity (Kessler et al., 2004). These approaches can complement each other through the use of permissive thresholds for binding prediction and subsequent experimental validation. However, prediction of further factors such as proteasomal cleavage have not yet matured and currently, the detection via mass spectrometry-based immunopeptidomics remains the only available technique for direct proof of HLA presentation of peptides (Becker & Riemer, 2022).

1.5 Aims

1.5.1 Overall aim of the research group

The overall aim of the Division of Immunotherapy and Immunoprevention is the development of a therapeutic vaccine against HPV-induced malignancies and precursor lesions. The major focus is on the development of an epitope-specific vaccine that elicits highly specific immune responses against validated target epitopes. This vaccine should include a set of epitopes that combine for a majority population coverage of greater than 95%, such that the vaccine formulation would be widely compatible. The choice of epitopes for such vaccine development is best informed by the identification of truly HLA-presented HPV-derived T cell epitopes.

1.5.2 Generation of an HPV16 E6 and E7 epitome map

To describe suitable target epitopes to be included in the envisaged epitome map, it is necessary to ascertain the HLA binding of a given peptide, its HLA presentation and its immunogenicity. In preceding work, candidate HLA epitopes from HPV16 E6 and E7, covering a range of six HLA supertypes were determined experimentally. An expansive list of potential HLA binders was generated using multiple prediction tools with permissive thresholds. The candidates were subsequently narrowed down through experimental validation of the peptide:HLA binding affinities using synthetic peptides. With this study, a major prerequisite for the targeted MS detection of HPV-derived HLA binders was established by the research group (Bonsack et al., 2019).

MS detection of HPV-derived HLA-presented peptides through immunopeptidomics has been a goal of the group since its formation in 2010 which aimed to follow up on the first MS detection of an HPV-derived epitope in Riemer et al. (2010). Successful detection of eleven HPV-derived HLA-A*02:01-presented peptides in the cell line CaSki was a major milestone in this effort (Blatnik et al., 2018). These two studies found only non-cysteine-containing peptides and were limited to the HLA-A*02:01 allotype. No HPV16 peptides beyond the ones described in these studies had ever been detected by immunopeptidomics at the time this project was started.

In parallel to the immunopeptidomics experiments, a parallel project in the group tests all identified HLA-binding peptides for immunogenicity.

1.5.3 Aims of this thesis

The first aim of this thesis was the establishment of a sensitive immunopeptidomics technique that could effectively detect most HLA binders and would be scalable to hundreds of peptides. To this end, a novel state-of-the-art instrument was available.

The second aim of this thesis was to apply the newly established methodologies to assess a collection of HPV16-transformed cell lines for presentation of E6- and E7-derived peptides. To ensure the intended population coverage, the immunopeptidomics approach was to be expanded to cover further HLA alleles. As mentioned above, any detections for supertypes other than HLA-A2 or the detection of cysteine-containing peptides would constitute novel findings. With these experiments, the immunopeptidomics part of the epitome map experiments would be complete.

2 Materials and methods

2.1 Materials

2.1.1 General

2.1.1.1 Chemicals and reagents

Product	**Catalog number**	**Company**
Accutase	00-4555-56	Thermo Fisher Scientific, Waltham, MA, USA
Ammonium chloride (NH_4Cl)	P726.1	Roth, Karlsruhe
Ethanol (absolute)		Sigma-Aldrich, Taufkirchen
Ficoll-Paque™ PLUS	GE17-1440-03	Sigma-Aldrich, Taufkirchen
Hydrogen chloride (HCl)	30024.29	VWR International, Fontenay-sous-Bois, France
Isopropanol		Sigma-Aldrich, Taufkirchen
Potassium bicarbonate ($KHCO_3$)	X887.1	Roth, Karlsruhe
Potassium chloride (KCl)	6781.1	Roth, Karlsruhe
Potassium dihydrogen phosphate (KH_2PO_4)	3904.1	Roth, Karlsruhe
Sodium acetate ($C_2H_3NaO_2$)	6773.2	Roth, Karlsruhe
Sodium azide (NaN_3)	A1430,0100	AppliChem GmbH, Darmstadt
Sodium chloride (NaCl)	10428420	Thermo Fisher Scientific, Waltham, MA, USA
Sodium dodecyl sulfate (SDS)	05030-1L-F	Sigma-Aldrich, Taufkirchen
Sodium hydroxide (NaOH)	P031.2	Roth, Karlsruhe
Sodium phosphate dibasic dihydrate ($Na_2HPO_4 \cdot 2H_2O$)	12694947	Acros organics, Thermo Fisher Scientificm Geel, Belgium
Streptavidin-Alkaline Phosphatase	3310-10	Mabtech, Nacka Strand, Sweden
Tris(hydroxymethyl)amino-methane (TRIS)	167620010	Acros organics, Thermo Fisher Scientificm Geel, Belgium
Trypan blue stain (0.4 %)	T10282	Thermo Fisher Scientific, Waltham, MA, USA
Trypsin / EDTA (0.04 % / 0.03 %)	C-41000	PromoCell GmbH, Heidelberg
Tween20 (=Polysorbat 20)	Tween201	MP Biomedicals, Illkirch, France

2.1.1.2 Buffers and solutions

Name	Ingredients
10x phosphate buffered saline (PBS)	1.37 M NaCl 27 mM KCl 100 mM Na_2HPO_4 x 2(H_2O) 20 mM KH_2PO_4 pH 7.3 in ddH_2O
Sodium borate buffer	125 ml warm ddH_2O 5g borax anhydrous

2.1.1.3 Cell culture basal media and supplements

Product	Catalog number	Company
Adenine	A8626	Sigma-Aldrich, Taufkirchen
Cholera Toxin from Vibrio cholerae	C3012	Sigma-Aldrich, Taufkirchen
Dulbecco's Modified Eagle's Medium (DMEM) - high glucose	D5796	Sigma-Aldrich, Taufkirchen
Fetal calf serum (FCS)	10270	Thermo Fisher Scientific, Waltham, MA, USA
Ham's F-12 Nutrient Mix (F12)	21765029	Thermo Fisher Scientific, Waltham, MA, USA
HEPES	11560496	Life Technologies Europe BV, Bleiswijk, Netherlands
Human serum, type AB	H4522	Sigma-Aldrich, Taufkirchen
Hybri-Care Medium	46-X	ATCC, Manassas, VA, USA
Hydrocortisone	H0396	Sigma-Aldrich, Taufkirchen
Insulin from bovine pancreas	I6634	Sigma-Aldrich, Taufkirchen
Interleukin (IL)-1β, recombinant human	201-LB-100	Bio-techne, R&D Systems, Inc., Minneapolis, USA
L-glutamine	25030024	Thermo Fisher Scientific, Waltham, MA, USA
MEM Non-Essential Amino Acids Solution (100×)	11140050	Thermo Fisher Scientific, Waltham, MA, USA
Penicillin/Streptomycin (P/S) 10,000 U penicillin and 10 mg streptomycin per ml	P0781	Sigma-Aldrich, Taufkirchen
Phosphate buffered saline (PBS)	10010015	Life Technologies Europe BV, Bleiswijk, Netherlands

Recombinant Human Epidermal Growth Factor (EGF)	13247-051	Invitrogen, Carlsbad, CA, USA
RPMI-1640 liquid with L-Glutamine	E15-840	PAA Laboratories
Sodium bicarbonate	S5761	Sigma-Aldrich, Taufkirchen
Sodium pyruvate	25-000-CIR	Corning, Kaiserslautern, Germany

2.1.1.4 Cell culture media

Name	**Ingredients**
complete DMEM	DMEM 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) P/S
complete DMEM NEAA	DMEM 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) P/S 1 % (v/v) MEM NEAA
complete RPMI-H	RPMI-1640 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) HEPES 1 % (v/v) P/S
complete RPMI-P	RPMI-1640 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) sodium pyruvate (100 mM) 1 % (v/v) HEPES 1 % (v/v) P/S
ELISpot medium	RPMI 5 % (v/v) FCS 2 mM L-Glutamine 1× P/S 10 mM HEPES 0.1 mM 2-mercaptoethanol
F-medium	Ham‘s F-12 25 % (v/v) DMEM 5 % (v/v) FCS 1 % (v/v) P/S 0.4 mg / ml hydrocortisone 0.5 µg / ml cholera toxin 5 µg / ml insulin

	24.2 µg / ml adenine 0.1 µg / ml EGF
HB-122 medium	Hybri-Care Medium 1x 10 % (v/v) FCS 1.5 g / l sodium bicarbonate
Hb95 medium	RPMI-1640 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) P/S
hybridoma medium 1	DMEM 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) HEPES 1 % (v/v) P/S 1 % (v/v) sodium pyruvate (100 mM)
M-10 medium	Eagle‘s MEM 10 % (v/v) FCS 1 % (v/v) L-glutamine 1 % (v/v) P/S 1 % (v/v) MEM NEAA

2.1.1.5 Consumables

Product	**Company**
96-well Sample Collection Plate, 700 µL Round well	Waters, Milford, Ma, USA
Aluminium foil	CeDo GmbH, Mönchengladbach
Blood collection set (Safty-lok™)	BD, Franklin Lakes, NJ, USA
Blood collection tubes (Sodium Heparin, 170 I.U.)	BD, Plymouth, UK
Cell culture dish (100 mm × 20 mm)	TPP, Trasadingen, Switzerland
Cell culture flask (25 cm^2, 75 cm^2, 125 cm^2)	TPP, Trasadingen, Switzerland
Cell culture flask T-225 CytoOne	Starlab, Hamburg
Cell culture plate (12-, 24-, 48-well)	Corning, Corning, NY, USA
Cell culture plate (96-well), flat-bottom	BD, Franklin Lakes, NJ, USA
Cell culture plate (96-well), U-bottom	TPP, Trasadingen, Switzerland
Cell culture plate (96-well), V-bottom	Greiner Bio-One, Frieckenhausen

Cell culture plate “CytoOne” (6-well)	Starlab, Hamburg
Cell scraper	Sarstedt, Newton, NC, USA
CELLine Classic bioreactor flask CL 1000	Sigma-Aldrich, Taufkirchen
Cling film	CeDo GmbH, Mönchengladbach
Countess cell counting chamber slides	Invitrogen, Carlsbad, CS, USA
Cryogenic tubes Greiner Bio-One™ Cryo.s™	Thermo Fisher Scientific, Waltham, MA, USA
Cryogenic tubes Nalgene™	Thermo Fisher Scientific, Waltham, MA, USA
FACS tubes (5 ml Polystrene round-bottom tube)	BD, Franklin Lakes, NJ, USA
Filter microplate, 96-well, polypropylene, with 3 µm glass fiber and 10 µm polypropylene membranes, 2 ml/well, long drip	Agilent
Gloves	Microflex, Reno, NV, USA
Parafilm	Bemis, Neeah, WI, USA
PCR reaction tube CapStrips	Biozym Scientific GmbH, Oldendorf
PCR reaction tube SoftStrips	Biozym Scientific GmbH, Oldendorf
PCR reaction tubes, single	Biozym Scientific GmbH, Oldendorf
Pipette tips, with and without filter	Starlab, Hamburg
Poly-Prep Chromatography Columns	Bio-Rad, Hercules, CA, USA
Polypropylene Cap Mat Square Well for 96-well Plate	Waters, Milford, Ma, USA
Protein LoBind Tube 0.5 mL	Eppendorf, Hamburg
Protein LoBind Tube 1.5 mL	Eppendorf, Hamburg
Protein LoBind Tube 15 mL	Eppendorf, Hamburg
Protein LoBind Tube 2.0 mL	Eppendorf, Hamburg
Protein LoBind Tube 5.0 mL	Eppendorf, Hamburg
Reaction tubes (0.2 ml, 0.5 ml, 1.5 ml and 2 ml)	Starlab, Hamburg
Reaction tubes, black, flip cap (1.5 ml)	NeoLab, Heidelberg
Scalpel	Feather, Osaka, Japan
Sep-Pak tC18 96-well Plate, 100 mg Sorbent per Well, 37–55 µm	Waters, Milford, Ma, USA
Syringe (20 ml, 50 ml BD Plastipak Luer-Lok™)	BD, Drogheda, Ireland
Syringe filter (pore size 0.22 µm)	TPP, Trasadingen, Switzerland

Test tube (15 ml and 50 ml)	nerbe plus GmbH, Winsen/Luhe
TubeOne 1.5 ml Natural Flat Cap Microcentrifuge Tubes	Starlab, Hamburg
Vacuum Filter (bottle top, pore size 0.22 µm)	TPP, Trasadingen, Switzerland

2.1.1.6 Appliances

Equipment	**Product name**	**Company**
Agarose gel documentation	Gel Jet Imager 2006 with Printer P39D	Intas, Göttingen and Mitsubishi Electric, Tokio, Japan
Analytical balance	Entris	Sartorius AG, Göttingen
Automated cell counter	Countess 3	Thermo Fisher Scientific, Waltham, MA, USA
Cell freezing container	Nalgene Mr. Frosty Cryo 1°C Freezing Container	Thermo Fisher Scientific, Waltham, MA, USA
Centrifuge	Centrifuge 5417 R, 5430	Eppendorf, Hamburg
Centrifuge	Biofuge pico, fresco	Heraeus, Hanau
Centrifuge	Sunlab® Minizentrifuge SU1550	Labdiscount GmbH, Mannheim
Centrifuge	Heraeus Megafuge 16R	Thermo Fisher Scientific, Waltham, MA, USA
Centrifuge rotor for microcentrifuge	F45-30-11	Eppendorf, Hamburg
Centrifuge rotor for plates	M-20, 75003624	Thermo Fisher Scientific, Waltham, MA, USA
Centrifuge rotor for swinging buckets	TX-400, 75003629	Thermo Fisher Scientific, Waltham, MA, USA
Centrifuge rotor for ultracentrifuge	TF 12.5	Herolab, Wiesloch
Centrifuge rotor for vacuum concentrator	F45-48-11	Eppendorf, Hamburg
Electrophoresis chamber for agarose gels	Owl Easycast B2	Thermo Fisher Scientific, Waltham, MA, USA
Electrophoresis chamber for agarose gels	PerfectBlue™ gel system, Mini L	Peqlab, VWR International GmbH, Darmstadt
ELISpot plate reader	CTL-Immunospot® S6 Ultra-UV	CTL Europe, Bonn

Flow cytometer	FACS Canto II™	BD Biosciences, Franklin Lakes, NJ, USA
Flow cytometer	BD Accuri™ C6 with BD CSampler™ accessory kit	BD Biosciences, Franklin Lakes, NJ, USA
Glassware	Duran	Schott, Mainz
Glassware	Fischerbrand	Thermo Fisher Scientific, Waltham, MA, USA
Ice machine	FM 120 KE-50-HC	Hoshizaki, Tokio, Japan
Incubator (37 °C, 5 % CO_2, cell culture)	Heracell 150i	Thermo Fisher Scientific, Waltham, MA, USA
Incubator (37 °C, 5 % CO_2, cell culture)	C200	Labotec, Göttingen
Laminar flow hood	SterilGard® Class II laminar flow hood	The Baker Company, Sanford, USA
Laminar flow hood	Maxisafe 2020	Thermo Fisher Scientific, Waltham, MA, USA
Light microscope	Wilovert Standard 30 microscope	Hund Wetzlar, Wetzlar
Light microscope	Axiovert 25	Carl Zeiss Microscopy GmbH, Jena
Liquid nitrogen tank	Locator 8 plus	Barnstead/Thermolyne, Dubuque, IA, USA
Liquid nitrogen tank	ARPEGE110 NU	Cryopal, Bussy-Saint-Georges, France
Magnetic stirrer	RSM-01S	Phoenix Instrument GmbH, Garbsen
Magnetic stirrer, heatable	MR-Hel Standard	Heidolph Instruments, Schwabach
Microwave		Sharp, Osaka, Japan
Multichannel pipetting reservoir	Multi-channel pipettor Trifill reservoir	Roth, Karlsruhe
Nano Drop	ND-1000	Thermo Fisher Scientific, Waltham, MA, USA
Nano Drop	ND-8000	Thermo Fisher Scientific, Waltham, MA, USA
PCR cycler	Bioer Gene Touch Thermal Cycler	Biozym Scientific GmbH, Oldendorf
pH Meter	SevenCompact™ pH/Ionmeters S220 with pH Electrode InLab Ultra-Micro-ISM	Mettler Toledo, Glostrup, Denmark

Pipettes	2 µl, 20 µl, 200 µl, 1000 µl, 50 µl-multichannel and 300 µl-multichannel Finnpipette F2	Thermo Fisher Scientific, Waltham, MA, USA
Pipettes, glass		Hirschmann Labortechnik, Eberstadt
Pipetting controller	Pipetboy acu 2	Integra Biosciences, Biebertal
Pipetting device, 96-well format	Bel-Art™ SP Scienceware™ Vaccu-Pette	Bel-Art products, Wayne, USA
Positive pressure processor	Waters Positive Pressure-96 Processor	Waters, Milford, Ma, USA
Scale	Kern EG 4200-2NM	Kern & Sohn, Balingen
Surgical tweezer and scissors		Dimeda, Tuttlingen
Thermomixer	Thermomixer compact	Eppendorf, Hamburg
Tissue homogenizer	T 10 basic Ultra-Turrax	IKA, Staufen
Ultracentrifuge	UniCen HR	Herolab, Wiesloch
Vacuum concentrator	Concentrator plus	Eppendorf, Hamburg
Vacuum pump	LABOPORT N 86 KT.18	KNF, Freiburg
Vortexer	Vortex-Genie 2	Scientific Industries, Bohemia, USA
Water bath		GFL, Burgwedel

2.1.1.7 Cell lines

Authentication was performed with the commercial provider Multiplexion GmbH, Friedrichshafen, Germany via Single Nucleotide Polymorphism (SNP)-profiling. All cell lines were tested negative for mycoplasma contamination. HLA typing was performed by the DKMS Life Science Lab GmbH in Dresden (Germany).

Where applicable, cell lines are identified by Unique Research Resource Identifiers (RRID, Bandrowski and Martone (2016)).

Epitome map:

Name	Description	HPV-status	Culture medium	Reference	Source
866	human, cervical, adherent	HPV16 positive	F-medium	Brady et al. (2000)	Kindly provided by Peter L Stern, University of Manchester, Manchester, UK
879	human, cervical, adherent	HPV16 positive	F-medium	Brady et al. (2000)	Kindly provided by Peter L Stern, University of Manchester, Manchester, UK
C33A	human, cervical, adherent	negative	complete DMEM	RRID: CVCL_1094	Kindly provided by Felix Hoppe-Seyler, DKFZ
CaSki	human, cervical, adherent	HPV16 positive	complete DMEM	RRID: CVCL_1100	Kindly provided by Felix Hoppe-Seyler, DKFZ
DoTc2-4510	human, cervical, adherent	HPV16 positive	complete DMEM	RRID: CVCL_1181	ATCC, Manassas, VA, USA
Goerke	human, cervical, adherent	HPV16 positive	M10 medium	Brady et al. (2000)	Kindly provided by Andreas Kaufmann, Charité, Berlin
MRI-H-196	human, cervical, adherent	HPV16 positive	complete DMEM	RRID: CVCL_5721	Kindly provided by Elisabeth Schwarz, DKFZ
SCC090	human, HNSCC, adherent	HPV16 positive	M10 medium	RRID: CVCL_1899	Kindly provided by Susanne M Gollin, University of Pittsburgh, Pittsburgh, PA, USA
SCC152	human, HNSCC, adherent	HPV16 positive	M10 medium	RRID: CVCL_C058	Kindly provided by Susanne M Gollin, University of Pittsburgh, Pittsburgh, PA, USA

SCC154	human, HNSCC, adherent	HPV16 positive	M10 medium	RRID: CVCL_2230	Kindly provided by Susanne M Gollin, University of Pittsburgh, Pittsburgh, PA, USA
SiHa	human, cervical, adherent	HPV16 positive	complete DMEM	RRID: CVCL_0032	Kindly provided by Felix Hoppe-Seyler, DKFZ
SNU-1000	human, cervical, adherent	HPV16 positive	complete RPMI-H	RRID: CVCL_1000	Korean Cell Line Bank, Seoul, South Korea
SNU-1005	human, cervical, adherent	HPV16 positive	complete RPMI-H	KCLB NO 01005	Korean Cell Line Bank, Seoul, South Korea
SNU-1299	human, cervical, adherent	HPV16 positive	complete RPMI-H	RRID: CVCL_5021	Korean Cell Line Bank, Seoul, South Korea
SNU-17	human, cervical, adherent	HPV16 positive	complete RPMI-H	RRID: CVCL_5029	Korean Cell Line Bank, Seoul, South Korea
SNU-703	human, cervical, adherent	HPV16 positive	complete RPMI-H	RRID: CVCL_5085	Korean Cell Line Bank, Seoul, South Korea
SNU-902	human, cervical, adherent	HPV16 positive	complete RPMI-H	RRID: CVCL_5107	Korean Cell Line Bank, Seoul, South Korea
UD-SCC2	human, HNSCC, adherent	HPV16 positive	complete RPMI-P	RRID: CVCL_E325	Kindly provided by Thomas Hoffmann, Uniklinikum Essen, Essen
UM-SCC-104	human, HNSCC, adherent	HPV16 positive	complete DMEM NEAA	RRID: CVCL_7712	Kindly provided by Tom Carey, University of Michigan, Ann Arbor, MI, USA
UM-SCC-47	human, HNSCC, adherent	HPV16 positive	complete DMEM	RRID: CVCL_7759	Kindly provided by Tom Carey, University of Michigan, Ann Arbor, MI, USA
W12-20861	human, cervical, adherent	HPV16 positive	F-medium	(Jeon & Lambert, 1995)	Kindly provided by Paul Lambert, University of Wisconsin, Madison, WI, USA

Hybridoma:

Name	Description	Culture medium	Reference	Source
A11.1 M (HB-164)	mouse hybridoma	HB-122 medium	RRID: CVCL_J175	ATCC, Manassas, VA, USA
BB7.1 (HB-56)	mouse hybridoma	complete RPMI-P	RRID: CVCL_7245	ATCC, Manassas, VA, USA
BB7.2	mouse hybridoma	hybridoma medium 1	RRID: CVCL_7246	Hans-Georg Rammensee, University of Tübingen, Tübingen
GAP A3 (HB-122)	mouse hybridoma	hybridoma medium 1	RRID: CVCL_G652	ATCC, Manassas, VA, USA

2.1.2 HLA immunoprecipitation

Where applicable, reagents of higher purity were used for HLA immunoprecipitations (IPs), as specified below.

2.1.2.1 Chemicals and reagents

Product	Catalog number	Company
10× PBS	1058.1	Roth, Karlsruhe
Acetic acid (AA) 100 % for LC-MS	5330010050	Sigma-Aldrich, Taufkirchen
Acetonitrile (ACN) ULC/MS CC/SFC	1204101	Biosolve, Valkenswaard, Netherlands
Borax anhydrous	71997-100G	Sigma-Aldrich, Taufkirchen
cOmplete Mini Protease Inhibitor Cocktail (PIC)	11836153001	Roche, Mannheim
ddH_2O		self-generated
Dimethyl pimelimidate dihydrochloride (DMP)	80490-5G	Sigma-Aldrich, Taufkirchen
Dimethyl sulfoxide (DMSO) for molecular biology	D8418	Sigma-Aldrich, Taufkirchen
Ethanolamine	411000-500ML	Sigma-Aldrich, Taufkirchen
Ethylenediamine tetraacetate (EDTA)	1034	Gebru, Gaiberg
Formic Acid (FA) 99 % ULC/MS CC/SFC	069141A8	Biosolve, Valkenswaard, Netherlands
Gammabind Plus Sepharose	GE17-0886-01	Sigma-Aldrich, Taufkirchen
HEPES, 1 M (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	J61360.AP	Thermo Fisher Scientific, Waltham, MA, USA

Hydrogen Peroxide (H_2O_2) 30 %	LC-4458.4	neoFroxx, Einhausen
Iodoacetamide (IAA)	I1146-5G	Sigma-Aldrich, Taufkirchen
Methanol (MeOH) ULC/MS CC/SFC	13684102	Biosolve, Valkenswaard, Netherlands
Octyl β-D-glucopyranoside (OGP)	O8001-1G	Sigma-Aldrich, Taufkirchen
Phenylmethanesulfonylfluoride (PMSF)	6367.4	Roth, Karlsruhe
Pierce™ Trifluoroacetic Acid (TFA), Sequencing grade	28904	Thermo Fisher Scientific, Waltham, MA, USA
Recombinant Protein A-Sepharose 4B	101142	Thermo Fisher Scientific, Waltham, MA, USA
Sodium chloride (NaCl)	AM9759	Thermo Fisher Scientific, Waltham, MA, USA
Sodium deoxycholate	D6750-100G	Sigma-Aldrich, Taufkirchen
Tris(2-carboxyethyl)phosphine (TCEP)	HN95.1	Roth, Karlsruhe
tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl)	J22638-AP	Thermo Fisher Scientific, Waltham, MA, USA
Water (H_2O) ULC/MS CC/SFC	232141B1	Biosolve, Valkenswaard, Netherlands

2.1.2.2 Buffers and solutions

Name	**Ingredients**
IP Lysis buffer	0.25 % Sodium deoxycholate 1 % OGP 1 mM PMSF / 5 ml 1 mM EDTA / 5 ml 0,2 mM IAA / 5 ml 1 PIC / 5 ml in PBS
IP wash buffer A	150 mM NaCl 20 mM Tris-HCl pH 8 in H_2O
IP wash buffer B	400 mM NaCl 20 mM Tris-HCl pH 8 in H_2O
IP wash buffer C	20 mM Tris-HCl pH 8 in H_2O

2.1.2.3 Antibodies

Antigen	Clonality	Clone	Host	Catalog number	Company
pan-HLA class I	monocolonal	W6/32	mouse	311406	Biolegend,Amsterdam, Netherlands
HLA-A2	monocolonal	BB7.2	mouse	-	self-generated, hybridoma
HLA-A3	monocolonal	GAP A3	mouse	-	self-generated, hybridoma
HLA-A11/A24	monocolonal	A11.1 M	mouse	-	self-generated, hybridoma
HLA-B7	monocolonal	BB7.1	mouse	-	self-generated, hybridoma

2.1.3 LC-MS

Where applicable, reagents of highest purity were used for LC-MS experiments, as specified below.

2.1.3.1 Chemicals and reagents

Product	Catalog number	Company
Acetonitrile (ACN) Optima LC/MS Grade	A955-212	Thermo Fisher Scientific, Waltham, MA, USA
Dimethyl Sulfoxide (DMSO)	D8418	Sigma-Aldrich, Taufkirchen
Formic Acid (FA) 99 % ULC/MS CC/SFC	69141	Biosolve, Valkenswaard, Netherlands
Methanol (MeOH) Optima LC/MS Grade	A456-212	Thermo Fisher Scientific, Waltham, MA, USA
Trifluoroacetic Acid (TFA) ULC/MS CC/SFC	202341	Biosolve, Valkenswaard, Netherlands
Water (H_2O) Optima LC/MS Grade	W64	Thermo Fisher Scientific, Waltham, MA, USA

2.1.3.2 Buffers and solutions

Name	Ingredients
LC loading solvent A	99.9 % H_2O 0.1 % TFA
LC loading solvent B	99.9 % H_2O 0.1 % TFA
LC rear seal wash	84.9 % H_2O 15 % MeOH 0.1 % FA
LC solvent A	99.9 % H_2O 0.1 % FA
LC solvent B	99.9 % ACN 0.1 % FA
LC transfer solvent	99.9 % H_2O 0.1 % TFA
MS sample solvent	94.9 % H_2O 5 % ACN 0.1 % TFA

2.1.3.3 Consumables

Product	Catalog Number	Company
Acclaim PepMap 100 Cartrige, C18 100 Å 5 µm 300 µm × 5 mm	11362113	Thermo Fisher Scientific, Waltham, MA, USA
CoAnn Pulled Emitters 20 µm ID × 10 cm L × 360 µm OD	TIP36002010-10-5	MS Wil, Aarle-Rixtel, Netherlands
Fisherbrand 9 mm Short Thread Plastic Vial, Wide Opening, Flat Bottom	11707597	Thermo Fisher Scientific, Waltham, MA, USA
Fisherbrand 9 mm PP Short Thread Seal, Transparent, Center hole, Assembled septum	12353700	Thermo Fisher Scientific, Waltham, MA, USA
FS-115 Capillary, Fused Silica, ID 150 µm; OD 360 µm; 2 m	FS-115	Postnova Analytics, Landsberg
nanoEase M/Z Peptide BEH C18 Column, 130 Å, 1.7 µm, 75 µm × 200 mm	186008794	Waters, Milford, Ma, USA
PepMap Neo Trap Cartridge	174500	Thermo Fisher Scientific, Waltham, MA, USA
Protein LoBind Tube 0.5 mL	30108094	Eppendorf, Hamburg
Protein LoBind Tube 1.5 mL	30108116	Eppendorf, Hamburg

QuanRecovery with MaxPeak HPS 12 × 32 mm Screw Neck Vial, 300 µL	186009186	Waters, Milford, Ma, USA

2.1.3.4 Appliances

Equipment	Product name	Company
Autosampler	UltiMate WPS-3000	Thermo Fisher Scientific, Waltham, USA
Centrifuge	Centrifuge 5804	Eppendorf, Hamburg
Column to emitter link	Simple Link Uno	Fossiliontech, Madrid, Spain
HPLC System	UltiMate 3000	Thermo Fisher Scientific, Waltham, USA
Ion Source	Nanospray Flex	Thermo Fisher Scientific, Waltham, USA
Mass Spectrometer	Orbitrap Exploris 480	Thermo Fisher Scientific, Waltham, USA
Sonicator	Sonorex Super RK 31	BANDELIN, Berlin
Syringe 25 ul	Syringe 25 ul	Hamilton, Bonaduz, Switzerland
Syringe 500 µl	HPLC Autosampler Syringe 500 µl 365JLT41	Thermo Fisher Scientific, Waltham, USA
Vortexer	neoVortex® shaker	neoLab, Heidelberg

2.1.3.5 Kits

Name	Catalog number	Company
Pierce BSA Protein Digest, MS grade	88341	Thermo Fisher Scientific, Waltham, MA, USA
Pierce HeLa Protein Digest Standard	88328	Thermo Fisher Scientific, Waltham, MA, USA
Pierce Peptide Retention Time Calibration Mixture (PRTC)	88321	Thermo Fisher Scientific, Waltham, MA, USA
Retention Time Standardization Kit (JPTRT)	RTK-1-10pmol	JPT Peptide Technologies, Berlin, Germany

2.1.3.6 Stable isotope labeled peptides

Stable isotope-labeled (SIL) variants of all candidate peptides were ordered for custom synthesis at JPT Peptide Technologies GmbH, Berlin, Germany or SynPeptide Co., Ltd, Shanghai, China. Order specifications were minimum 0.1 mg dried crude synthesis product with one or two fully stable isotope labeled amino acids per peptide and no absolute quantification. To prevent excessive light peptide contamination, SIL labels were specified such that for nested peptide sequences, i.e. pairs such as PEPTIDE and PEPTIDER, no complete peptide sequence was present in unlabeled fashion as a subsection of a longer peptide. This means in the example case, PEPTIDER[+10] is invalid whereas PEPTI[+7]ER is valid.

Peptide	**Peptide**	**Peptide**	**Peptide**
AFRDL[+7]CIVY	AFRDLCIV[+6]YR	AGQAEPDR[+10]AHY	AHYNI[+7]VTF
AHYNIV[+6]TFCCK	AIVDKV[+6]PSV	APRTV[+6]ALTAL	AVCDKCL[+7]K
AVCDKCL[+7]KF	AVCDKCL[+7]KFY	AVCDKCL[+7]KFYS	AVILPPL[+7]SPYFK
AYVHMV[+6]THF	CIVYR[+10]DGNPY	CK[+8]QQLLRREVY	CLKFYSKI[+7]SEY
CPEEKQRHL[+7]	CQKPL[+7]CPEE	CTEL[+7]QTTIH	CV[+6]YCKQQLLR
CYSL[+7]YGTT	CYSL[+7]YGTTL	CYSL[+7]YGTTLE	CYSV[+6]YGTTL
CYSV[+6]YGTTLE	DFAFR[+10]DLCI	DIILECV[+6]YCK	DKKQR[+10]FHN
DKKQR[+10]FHNI	DLLIR[+10]CINCQK	DLLMGTLGI[+7]V	DLLMGTL[+7]GIVC
DLQPETTDL[+7]Y	DL[+7]YCYEQF	DPQER[+10]PRKL	DRAHYNI[+7]VTF
DTDHYFL[+7]RY	DTPTL[+7]HEY	EDLLMGTLGI[+7]V	EIILECV[+6]YCK
ELQTTI[+7]HDI	ELQTTIHEI[+7]	EPDRAHYNI[+7]V	ETTDL[+7]YCY
EVYDFAFR[+10]	EVYDFAFR[+10]DL	EYRHYCYSL[+7]	EYR[+10]HYCYSLY
EYR[+10]HYCYSV	EYR[+10]YYCYSLY	FAFRDLCI[+7]V	FAFRDL[+7]CIVY
FAFR[+10]DLCIVYR	FGPV[+6]NHEEL	FQDPQERPI[+7]	FQDPQERPIKL[+7]
FQDPQER[+10]PRK	FYSKI[+7]SEY	FYSKI[+7]SEYRH	FYSKI[+7]SEYRHY
GIVCPI[+7]CSQK	GQAEPDR[+10]AH	GQAEPDR[+10]AHY	GQAEPDR[+10]AHYN
GQVQKI[+7]VLY	GTLGIVCPI[+7]	GTTL[+7]EQQY	GTTL[+7]EQQYNK
GVCDKCL[+7]KFY	HDIILECV[+6]Y	HDIILECV[+6]YCK	HEIILECV[+6]YCK
HEIRLECV[+6]YCK	HGDTPTL[+7]HEY	HLDKKQR[+10]FH	HLDKKQR[+10]FHNI
HYCYSL[+7]YGTTL	HYNIV[+6]TFC	HYNIV[+6]TFCC	IADMGHL[+7]KY
IHDIILECV[+6]Y	IHEIILECV[+6]Y	IILECV[+6]YCK	ILECV[+6]YCK
IL[+7]ECVYCKQQL	IPRAALLPLL[+7]	IRTLEDL[+7]LMGT	ISEYR[+10]HYCY
ISEYR[+10]HYCYS	ISEYR[+10]YYCY	ISEYR[+10]YYCYS	IVCPI[+7]CSQK
IVYR[+10]DGNPY	IVYR[+10]DGNPYA	IVYR[+10]DGNPYAV	KAL[+7]INADEL
KCL[+7]KFYSKI	KFYSKI[+7]SEY	KFYSKI[+7]SEYR	KFYSKI[+7]SEYRH

KIADRFL[+7]LY	KISEYR[+10]HY	KI[+7]SEYRHYC	KISEYR[+10]HYCY
KISEYR[+10]YYCY	KISEYR[+10]YYCYS	KKQR[+10]FHNI	KKQR[+10]FHNIR
KLLNYAPL[+7]EK	KLPDLCTEL[+7]	KLPDL[+7]CTELQT	KLPQL[+7]CTEL
KLPQLCTEL[+7]QT	KPL[+7]CDLLI	K[+8]PLCDLLIRCI	KQQLLRREV[+6]Y
KQRFHNI[+7]RGR	KQRFHNI[+7]RGRW	KQRHL[+7]DKKQRF	KYPENFFLL[+7]
LCIVYR[+10]DGNPY	LCV[+6]QSTHVDI	LECVYCK[+8]QQL	LEQQYNK[+8]PL
LGIV[+6]CPICSQK	LIRCI[+7]NCQK	LIRCINCQK[+8]PL	LKFYSK[+8]ISEY
LLIRCI[+7]NCQK	LLMGTL[+7]GI	LLMGTLGI[+7]V	LLMGTL[+7]GIVC
LLMGTLGIV[+6]CP	LLRREV[+6]YDF	LLRREV[+6]YDFAF	LMGTLGIVCPI[+7]
LPQL[+7]CTEL	LPQL[+7]CTELQT	LQPETTDL[+7]Y	LQPETTDL[+7]YC
LQPETTDL[+7]YCY	LQTTI[+7]HDII	LQTTI[+7]HEII	LRREV[+6]YDFAF
LYGTTL[+7]EQQY	MFQDPQER[+10]PRK	MGTL[+7]GIVCPI	MLDL[+7]QPET
MLDL[+7]QPETT	NIRGR[+10]WTGRCM	NIV[+6]TFCCK	NPYAV[+6]CDKCLK
PDR[+10]AHYNIVTF	PYAV[+6]CDKCL	PYAV[+6]CDKCLK	PYAV[+6]CDKCLKF
QLLRREV[+6]Y	QLLRREV[+6]YDF	QQLLRREV[+6]Y	QQLLRREV[+6]YDF
QQYNK[+8]PLCDL	QYNK[+8]PLCDL	QYNK[+8]PLCDLL	QYNK[+8]PLCDLLI
RAHYNI[+7]VTF	RCINCQKPL[+7]CP	RCMSCCR[+10]SSR	REV[+6]YDFAF
RFHNI[+7]RGRW	RGR[+10]WTGRCM	RGR[+10]WTGRCMSC	RHYCYSL[+7]Y
RHYCYSV[+6]Y	RHYCYSV[+6]YGTT	RL[+7]CVQSTHV	RLECV[+6]YCK
RPR[+10]KLPQL	RPRKL[+7]PQLC	RPRK[+8]LPQLCT	RPR[+10]KLPQLCTE
RPTKL[+7]PQLCTE	RSSRTRRETQL[+7]	RTLEDL[+7]LMGT	RTLEDLL[+7]MGTL
RTRRETQL[+7]	RYYCYSL[+7]YGT	SEYR[+10]HYCY	SEYR[+10]HYCYSL[+7]
SEYR[+10]HYCYSLY	SEYR[+10]HYCYSVY	SEYR[+10]YYCYSL	SEYRYYCYSV[+6]
SKISEYR[+10]HY	SKISEYR[+10]HYCY	SKISEYR[+10]YYCY	SLYGTTL[+7]EQQY
SVYGTTL[+7]EQQY	TFCCK[+8]CDF	TFCCK[+8]CDFTL	TFCCKCDF[+10]TLR
TFCCK[+8]CDSTL	TI[+7]HDIILECV	TIHDIILECV[+6]Y	TIHDIILQCV[+6]Y
TI[+7]HEIILECV	TIHEIILECV[+6]Y	TIHEIRL[+7]ECV	TLEQQYNK[+8]
TL[+7]GIVCPI	TL[+7]GIVCPIC	TLHEYML[+7]DL	TLHEYML[+7]DLQP
TPTL[+7]HEYM	TPTL[+7]HEYML	TQMPDPKTF[+10]	TSAL[+7]PIIQK
TTDL[+7]YCYEQ	TTDL[+7]YCYEQL	TTDL[+7]YCYEQLS	TTI[+7]HDIILECV
TTIHEIIL[+7]ECV	TTIHEIRL[+7]ECV	TTL[+7]EQQYNK	VCDK[+8]CLK[+8]FYSK
VQSTHVDI[+7]	VYCK[+8]QQLL	VYCK[+8]QQLLR	VYCK[+8]QQLLRR
VYDFAFR[+10]DL	VYDFAFR[+10]DLC	VYDFAF[+10]RDLCI	VYGTTL[+7]EQ

VYGTTL[+7]EQQY	VYR[+10]DGNPY	VYR[+10]DGNPYAV	VYR[+10]DGNPYGV
WTGR[+10]CMSCCR	YAV[+6]CDKCLK	YAV[+6]CDKCLKF	YAV[+6]CDKCLKFY
YCYSL[+7]YGTTLE	YCYSV[+6]YGTTL	YDFAFR[+10]DLCI	YGTTL[+7]EQQYNK
YGVCDKCL[+7]KFY	YLL[+7]PAIVHI	YML[+7]DLQPE	YMLDL[+7]QPET
YMLDL[+7]QPETT	YMLDL[+7]QPETTD	YNI[+7]VTFCCK	YR[+10]HYCYSLY
YR[+10]HYCYSVY	YR[+10]YYCYSL	YRYYCYSV[+6]	YSKISEYR[+10]HY
YSL[+7]YGTTL	YSV[+6]YGTTL		

2.1.4 Software

The reproducibly defined dependencies, self-generated packages and the code for scheduling, evaluation, and diagnostics are freely accessible in a reproducible fashion in the source code repository at https://github.com/jonasfoe/ms_targeted_workflows.

2.1.4.1 General

Name	**Version**	**Company / source**
.NET SDK	6.0.20	Microsoft Corporation, Redmond, WA, USA
CTL ImmunoSpot Professional DC	5.1.36	CTL Europe, Bonn
EndNote	20.4	Thomas Reuter, Philadelphia, PA, USA
Excel	2019	Microsoft Corporation, Redmond, WA, USA
GIMP	2.10.24	GIMP Development Team, https://gimp.org
IEDB Analysis Tool Population Coverage	3.0.2	Bui et al. (2006)
Image Lab	6.0.1	Bio-Rad Laboratories, Hercules, CA, USA
Inkscape	1.3	Inkscape Project
Julia	1.9	Bezanson et al. (2017)
netMHCpan	4.1b	Reynisson et al. (2021)
PEAKS Studio X Pro	10.6	Bioinformatics Solutions Inc.
R	4.2	R Core Team (2023)
Rstudio	2023.09.0	Posit Software, Boston, MA, USA
Skyline	22.2.0.527	MacLean et al. (2010b)
Spectronaut	17.6	Biognosis, Schlieren, Switzerland

Thermo RawFileReader	5.0.88	Thermo Fisher Scientific, Waltham, MA, USA
Word	2019	Microsoft Corporation, Redmond, WA, USA
XCalibur Instrument software	2.0-2.0.122.16/2.0.182.25	Thermo Fisher Scientific, Waltham, MA, USA

2.1.4.2 R packages

Name	Version	Reference	Source
tidyverse	2.0.0	Wickham et al. (2019)	CRAN
conflicted	1.2.0	Wickham (2023)	CRAN
fs	1.6.2	Hester et al. (2023)	CRAN
ggpp	0.5.3	Aphalo (2023)	CRAN
ggseqlogo	0.1	Wagih (2017)	CRAN
here	1.0.1	Müller (2020)	CRAN
hrbrthemes	0.8.0	Rudis (2020)	CRAN
immunotation	1.7.0	Imkeller (2021)	github
IsoSpecR	2.1.3	Łącki et al. (2020)	CRAN
JuliaCall	0.17.5	Li (2019)	CRAN
markdown	1.7	Xie et al. (2023)	CRAN
MsRawAccess	0.1.2	-	self-generated
MSTargetedWorkflows	0.11.0	-	self-generated
patchwork	1.1.2	Pedersen (2022)	CRAN
processx	3.8.2	Csárdi and Chang (2023)	CRAN
quarto	1.2	Allaire (2022)	CRAN
readxl	1.4.3	Wickham and Bryan (2023)	CRAN
renv	1.0.2	Ushey and Wickham (2023)	CRAN
rmarkdown	2.23	Allaire et al. (2023)	CRAN
tictools	0.5.8	-	self-generated
UpSetR	1.4.0	Gehlenborg (2019)	CRAN
yaml	2.3.7	Garbett et al. (2023)	CRAN
eulerr	7.0.0	Larsson (2022)	CRAN
ggnewscale	0.4.9	Campitelli (2023)	CRAN

2.1.4.3 Julia packages

Name	Version	Reference	Source
CSV.jl	0.10.10	Quinn (2023)	Julia package registry
RCall.jl	0.13.15	Bates et al. (2023)	Julia package registry
Revise.jl	3.5.2	Holy (2023)	Julia package registry
Suppressor.jl	0.2.1	Castelló (2022)	Julia package registry
Underscores.jl	3.0.0	Foster (2021)	Julia package registry
MSUtils2.jl	0.1.0	-	self-generated
Peptidomics.jl	0.2.0	-	self-generated
DataFrames.jl	1.5.0	Harris et al. (2023)	Julia package registry

Data analysis and plotting was performed using R. For calculation of peptide precursor, fragment, and modification masses, I used the self-generated Julia package *Peptidomics*. *Peptidomics* is built with the Unimod database that specifies all relevant isotope masses, chemical peptide modifications and neutral losses (Creasy and Cottrell (2004), accessed: 2020-05-11).

2.2 Methods

If not indicated otherwise, everything described below was performed by myself.

2.2.1 Cell culture

Cell line cultivation was carried out by Rebecca Köhler, Alexandra Klevenz and Nika Vučković. Cells were maintained under sterile conditions in a humidified incubator at 37 °C and at 5 % CO_2 and 95 % relative humidity.

All cell lines for the epitome map were adherent cells. These were cultured in standard tissue culture flasks of appropriate size using respective media as indicated in the cell line section 2.1.1.7. When a cell confluence of 40–90 % was reached, cells were passaged. For passaging, the medium was removed and the cells were washed with PBS. Next, they were incubated with trypsin / EDTA solution until detachment. Upon detachment, fresh medium containing FCS was added for protease inhibition. The solubilized cells were collected in Falcon tubes, centrifuged at 1,300 rpm for 4 min and were resuspended in fresh medium. They were then reseeded in flasks at an appropriate ratio between 1:5 and 1:10 to ensure a minimum confluence for survival and proliferation.

For harvesting, cells were rinsed with PBS and detached with 5 ml Accutase per T-225 flask. They were washed three times with cold PBS with centrifugation steps at 300 ×g for 4 min at 4 °C. For counting, 10 µl of the cell suspension was mixed with 10 µl of trypan blue and cells were counted using the Countess automated cell counter. Aliquots of 50×10^6–100×10^6 cells were snap frozen for 30 sec in liquid nitrogen and stored as dry pellets at -20°C.

For production of antibodies, hybridoma cells were cultivated in CELLine CL 1000 bioreactors. Pre-cultures were cultured in standard tissue culture flasks of appropriate size using respective media as indicated in the cell line section 2.1.1.7. Growth was monitored daily by visual inspection and cell numbers and viability was determined by trypan blue staining and counting of the cells using the Countess automated cell counter. When pre-cultures contained at least $2x10^7$–$4x10^7$ viable cells, ideally in the log growth phase to prevent an initial lag phase, they were transferred to the

bioreactors. For this, the membrane was first equilibrated by the addition of 50 ml of culture medium to the medium compartment. Next, $2x10^7$–$4x10^7$ viable cells of the pre-culture were diluted in 16 ml of pre-warmed culture medium and added to the cell cultivation compartment. Finally, an additional 950 ml of culture medium were added to the medium compartment and the bioreactor was placed in the incubator. The growth of hybridoma cells was monitored every third day by determining cell numbers and viability. After the cell number reached a maximum of $4x10^8$ cells ($2.6x10^7$ cells / ml), the cells were harvested. To this end, the culture medium was removed from the medium compartment and cell suspension was removed from the cell cultivation compartment. $4x10^7$ viable cells in 15 ml fresh culture medium were put back to the cell cultivation compartment and 1 l of fresh culture medium was placed in the medium compartment. Typically, harvesting was performed weekly. The remaining cell suspension was cleared by centrifugation with 1,660 ×g for 15 min, filtered through a 0.45 µm syringe filter, and stored at -20 °C until usage for antibody purification.

2.2.2 Synthetic peptides

2.2.2.1 SIL peptide mixtures

Peptides were dissolved in DMSO to an assumed concentration of 2 nmol / µl, based on the 0.1 mg specification and the molecular weight of the unmodified peptide. They were combined into mixes of up to 50 peptides for LC-MS characterization. Mixtures were always either non-cysteine-containing or cysteine-containing, as cysteine-containing peptides required an extra alkylation step. To ensure robustness of DI-MS analysis, the mixtures were allotted such that no precursor of charges states +1 to +3 may end up in the same precursor isolation window as another precursor. A helper script to facilitate this is available in the github repository (see section 2.1.4). Mixes were dried as 10 pmol per peptide aliquots and stored at -80 °C. All synthetic peptide identities were confirmed via targeted mass spectrometry.

2.2.2.2 Peptide mixture alkylation

2 ng per peptide of vacuum dried peptide mixture was resolubilized in 20 µl 100 mM HEPES with 3 min sonication in a 0.5 µl protein LoBind microcentrifuge tube. 5 µl (500 pmol per peptide) of the sample was reduced by addition of 2 µl HEPES / 50 mM TCEP with incubation for 10 min at RT. Alkylation was performed by addition of 2 µl 400 mM IAA with incubation in the dark for 20 min at RT. Alkylation was quenched by addition of 2 µl 100 mM HEPES / 50 mM TCEP. The sample was acidified by adding 0.3 % TFA to a volume of 500 µl and desalted using a 100 mg sorbent well of a 96-well SepPak plate. The solid-phase extraction (SPE) protocol is similar to the one described in section 2.2.3.9 for the IP eluate. In brief, equilibration followed by washing twice, sample loading and again washing twice. The elution step was performed with 80 % ACN / 0.1 % TFA instead of with 28 % ACN / 0.1 % TFA, as there was no risk of coeluting proteins. The sample was split into 5 equal aliquots for 100 pmol equivalents.

2.2.3 HLA immunoprecipitation

This is the core wet-lab technique that I performed for this project. It encompasses cell lysis, immunoprecipitation of HLA class I:peptide complexes, acidic elution of peptides from HLA molecules, and solid-phase extraction of peptides from the eluate. After the initial establishment phase, all HLA immunoprecipitation and the related antibody production and beads preparation was carried out by a lab technician, Rebecca Köhler.

2.2.3.1 Sepharose beads suspension

Upon receiving commercial beads (Gammabind Plus Sepharose or Recombinant Protein A-Sepharose 4B), the product was distributed equally to eight 2 ml microcentrifuge tubes. The beads were centrifuged at 5500 rpm for 3 min at RT using a swing-out rotor and supernatant was removed. 1 ml

of 10 % acetic acid at pH 2 was added to each tube and the tubes were vortexed and beads were centrifuged. The beads were washed 5× with 1 ml PBS, vortexed and centrifuged. 0.02 % NaN_3 in PBS was added at a 1:1 volume ratio to the drained beads. The beads were vortexed, spun down and stored at 4 °C.

2.2.3.2 Antibody purification

The antibody was purified from hybridoma cell culture medium using Protein A Sepharose beads as described previously (Bassani-Sternberg, 2018). Polypropylene gravity-flow columns (10 ml Poly-Prep Chromatography Columns) were used. They were washed once with 10 ml of a 1 % SDS solution and 40 ml of ddH_2O before first use. Next, 4 ml of the beads suspension were loaded into the column and drained. The beads were washed by passing through 10 ml of 100 mM Tris-HCl (pH 8.0). The column tip was closed and up to 8 ml of cleared hybridoma cell culture supernatant was added. The column top was close and it was rotated for 30 min at RT. After 10 min, the hybridoma cell culture supernatant was allowed to drain. This was done twice to process a total of 16 ml of cleared hybridoma cell culture supernatant. Next, the columns were washed with 10 ml of 100 mM Tris-HCl (pH 8.0) and then with 10 ml of 20 mM Tris-HCl (pH 8.0). The antibodies were eluted with 6× 1 ml of 0.1 N acetic acid (pH 3.0) into tubes containing 300 µl of 1 M Tris-HCl (pH 8.0) to neutralize the solution. Antibody eluates were gently vortexed, pH was verified using pH test strips and the IgG concentration was determined using a NanoDrop 8000. Fractions with high antibody concentrations were pooled and purified antibodies were stored at -20 °C until further usage.

2.2.3.3 Antibody-beads crosslinking

For the use in HLA immunoprecipitation, the different monoclonal antibodies were covalently coupled to either Gammabind Plus Sepharose beads or Protein A Sepharose beads as described previously (Bassani-Sternberg, 2018). Where not indicated otherwise, anti-pan-HLA class I antibody W6/32 was used. Antibody was used at a ratio of 2.5 µg antibody per µl of beads suspension and crosslinked. Polypropylene gravity-flow columns (Poly-Prep Chromatography Columns) were used. They were washed once with 10 ml of a 1 % SDS solution and 40 ml of ddH_2O before first use. 4 ml of the beads suspension were loaded into the column and drained. The beads were washed by passing through 10 ml of 100 mM Tris-HCl (pH 8). For coupling, the column tip was closed and 10 mg of antibody was added per 4 ml of beads suspension. A 60 µl aliquot was stored for QC. The column top was closed and it was rotated for 30 min at RT. A 60 µl aliquot was stored for QC. The column was washed with 10 ml 0.2 M sodium borate buffer (pH 9). To prepare crosslinking, the tip was closed, 4 ml of 0.2 M sodium borate buffer (pH 9) was added and a 60 µl aliquot was stored for QC. Then, 31.2 mg DMP (final concentration: 40 mM) were added for crosslinking and the column was rotated for 30 min at RT. A 60 µl aliquot was stored for QC. Under a fume hood, the column was washed with 5 ml of 0.2 M ethanolamine to initiate quenching. 4 ml of 0.2 M ethanolamine was added and the column was rotated for 2 h at RT. The column was washed with 10 ml PBS / 0.02 % NaN_3. 2 ml PBS / 0.02 % NaN_3 was added for a 1:1 beads suspension and suspension was spread to 2× 2 ml microcentrifuge tubes and stored at 4 °C.

2.2.3.4 Tissue homogenization and lysis

All tubes and solutions were kept on ice during the homogenization. The sample tube of the tissue was weighed without and later with sample to deduce tissue sample weight. The frozen sample was placed in a 2 ml Protein LoBind tube with 2 ml of freshly prepared cold IP lysis buffer and briefly rested for up to 2 min. If tissue was not fully submerged, more IP lysis buffer was added as necessary. Homogenization was performed with the T 10 basic Ultra-Turrax high speed setting for 5 s. Further 10 ml of lysis buffer were added per 1 g of sample and the lysis commenced for 1 h on ice with gentle

agitation every 10 min. Afterwards, the lysate was centrifuged with 40,000 ×g for 60 min at 4 °C. A 30 µl aliquot was stored for QC.

2.2.3.5 Cell pellet lysis

Frozen cell pellets were lysed by resuspending with 1 ml freshly prepared cold lysis buffer per 50×10^6 cells for 1 h at 4°C and vortexed intermittently. The lysate was centrifuged with 40,000 ×g for 30 min at 4 °C in protein LoBind tubes. A 30 µl aliquot was stored for QC.

2.2.3.6 Batch IP

The amount of beads used was determined according to the number of cells at 170 µl beads suspension per 100×10^6 cells. The IP was performed in protein LoBind tubes of appropriate size up to 15 ml. The following spin down steps with IP beads were performed for 3 min at 4 °C with a swing-out rotor either at 5500 rpm for microcentrifuge tubes or at 3500 rpm for falcon tubes. For optional pre-clearing, uncoupled beads were washed 2× with 10 ml PBS, spun down and supernatant was removed. The fresh cell lysate was added. The sample was incubated for 1 h at 4 °C on the rotating wheel and spun down afterwards. A 30 µl aliquot of the supernatant was stored for QC. Antibody-crosslinked beads were washed 2× with 10 ml PBS, spun down and supernatant removed. Supernatant from pre-clearing or, if no-preclearing was performed, cell lysate was added to the beads. The sample was incubated for 4 h at 4 °C on the rotating wheel and spun down afterwards. A 30 µl aliquot of the supernatant was stored for QC. The supernatant was removed. The IP beads were washed 3× with IP wash buffer A, 3× with IP wash buffer B and 3× with IP wash buffer C. The supernatant was removed after each wash step. For elution, peptides were dissociated by addition of 0.3 % TFA at 3× the volume of the dry beads. The sample was incubated for 20 min at RT in shaking incubator and spun down afterwards. The supernatant was either transferred to a fresh LoBind tube or directly transferred to solid-phase extraction.

2.2.3.7 96-well plate-based IP

The Waters Positive Pressure-96 Processor was used to perform the different steps on the 96-well plate filter plate (3 µm glass fiber, 10 µm membrane, Agilent). Up to an equivalent of 100×10^6 cells of IP eluate was processed in a single well. Larger samples were split equally into multiple wells. The tumor tissue IP eluate was split into five wells. Each plate well was prepared by passing through 1 ml 100 % ACN, 1 ml of 80 % ACN, 0.1 % TFA and 1 ml of 0.1 % TFA. Equilibration was done by passing through 2× 1 ml of 0.1 M Tris-HCl (pH 8). 150 µl beads suspension was used per well. For optional pre-clearing, a separate plate was used that was prepared the same way and loaded with uncoupled beads. 1 ml of 0.1 M Tris-HCl (pH 8) was passed through and the beads were conditioned by passing through 400 µl IP lysis buffer. The cell lysate was loaded by gravity flow at 4 °C. Flow through was collected in the IP plate and a 30 µl aliquot was stored for QC. If no pre-clearing was performed, cell lysate was directly loaded onto the IP wells. Again, by gravity flow at 4 °C. The IP was washed by passing through 8× 1 ml wash IP buffer A, 8× 1 ml IP wash buffer B, 8× 1 ml IP wash buffer A and 4× 1 ml IP wash buffer C. For IP elution, 2× 500 µl of 0.3 % TFA were added, with incubation for 3 min each and slow elution at maximum applied positive pressure of 1.5 psi. The eluate was collected in a 96-well sample collection plate. The eluate was either transferred to a fresh protein LoBind tube or directly transferred to solid-phase extraction.

2.2.3.8 Alkylation

Alkylation was an optional step between IP and SPE. 1 ml of IP eluate was reduced with 100 µl HEPES / 50 mM TCEP with incubation for 10 min at RT. Alkylation was performed with 100 µl of 400 mM IAA with incubation in the dark for 20 min at RT. Alkylation was quenched with 100 µl of 1 M HEPES / 50 mM TCEP with incubation for 3 min at RT. The sample was acidified again with 60 µl of 10 % TFA.

2.2.3.9 Solid-phase extraction

Solid-phase extraction (SPE) was performed with 100 mg sorbent wells of a Sep-Pak tC18 96-well plate. Up to an equivalent of 100×10^6 cells of IP eluate was processed in a single well. Larger samples were split equally into multiple wells. Solutions were either pulled through the wells with a vacuum pump or pushed through with the Waters Positive Pressure-96 Processor. The plate was prepared by equilibrating each well with 1 ml of 80 % ACN / 0.1 % TFA and washing 2× with 1 ml of 0.1 % TFA. The IP eluate was loaded. The wells were washed with 1 ml of 0.1 % TFA. The bottom of the plate was cleaned with 100 % MeOH and a 96-well sample collection plate was placed below. Elution was performed with 2× 400 µl of 28 % ACN / 0.1 % TFA. The eluate was transferred into microcentrifuge tubes, dried in the vacuum concentrator at 30 °C and stored at -20 °C.

2.2.3.10 IP Oxidation

A performic acid solution was prepared fresh on the day: 5 µl of 30 % H_2O_2 and 45 µl of 10 % formic acid were combined in a glass vial and let rest for 5 min. The resulting performic acid solution was diluted 1:200 with H_2O.

The oxidation step was performed on the 96-well SepPak plate as an optional step after sample loading. After the sample loading step, wells were washed 1× with 1 ml of 0.1 % TFA. The oxidation was performed by brief application of 1 ml of the performic acid solution. The SPE procedure was continued as usual with 2× post-sample-load wash.

2.2.4 LC-MS

2.2.4.1 Synthetic peptide resolubilization

To resolubilize synthetic peptide mixtures I added 20 µl of 5 % ACN / 0.1 % TFA to dried mixtures of 10 pmol per peptide, sonicated for 3 min and stored these aliquots at -20 °C. For reference acquisition, I used several dilutions in sequence to achieve appropriate signal for all peptides. The samples were diluted in QuanRecovery screw neck vials. The highest load per injection was set to 150 fmol per peptide with two 1:25 dilutions following. Retention time standard peptides were spiked in at 50 fmol PRTC kit per LC-MS injection and 5 fmol JPTRT kit per LC-MS injection.

2.2.4.2 IP sample resolubilization

I resolubilized dried IP eluate in 5–20 µl LC sample solvent. For targeted MS, retention time standard peptides were spiked in at 50 fmol PRTC kit per LC-MS injection and 5 fmol JPTRT kit per LC-MS injection. The sample was sonicated for 3 min, briefly spun down, transferred to QuanRecovery screw neck vials and spun down again.

2.2.4.3 Titration of SIL peptide spike-in mixture

The SIL peptides for assay-internal confirmation of 26 peptide detections were prepared as a mixture in high concentrations. I created a master mixture of nominally 100 pmol / µl per synthetic peptide. I injected 2 µl of a 1:1000 dilution of this mixture with targeted LC-MS acquisition. According to each precursors' peak area in MS2, the master mixture was adjusted by adding additional stock for weak peptides. After two rounds of adjustments, the master mixture was finalized. The 1:1000 dilution to

the final concentration was performed shortly before spike-in to IP samples to reduce uneven loss of the different precursors, primarily due to different peptide hydrophobicity.

2.2.4.4 Automatic sample injection

The UltiMate WPS-3000 autosampler was set to use different batches of transfer solvents when switching from reference acquisition of synthetic peptides to immunopeptidomics samples, to avert cross-contamination. The samples were stored in the autosampler at 5 °C in 300 µl screw neck vials with a septum cover. A 20 µl sample loop was used.

2.2.4.5 LC gradients

The LC gradients on the UltiMate 3000 HPLC system progressed from 99 % or 98 % LC Solvent A (0 % organic) to 80 % LC Solvent B (100 % organic). The gradients used are: Grd4 starting at 1 % B and progressing through 1 % B at 0.5 min, 2.5 % B at 13 min, 28.6 % B at 100 min, 38.7 % B at 101.4 min, 80 % B at 104 min, 80 % B at 109 min, 1 % B at 109.1 min. The trap was in line with the column from 3 min to 110 min. OptiGrd3 starting at 2 % B and progressing through 2 % B at 0.5 min, 6 % B at 5 min, 28.5 % B at 80 min, 48 % B at 82 min, 80 % B at 84 min, 80 % B at 89 min, 2 % B at 89.1 min. The trap was in line with the column from 3 min to 90 min. Grd4_fast starting at 1 % B and progressing through 1 % B at 0.5 min, 2.5 % B at 8.8 min, 28.6 % B at 66.8 min, 38.7 % B at 67.8 min, 80 % B at 69.5 min, 80 % B at 74.5 min, 1 % B at 74.6 min. The trap was in line with the column from 3 min to 75.5 min. Grd_DDA starting at 2 % B and progressing through 2 % B at 0.5 min, 6 % B at 5 min, 28.5 % B at 40 min, 48 % B at 41 min, 80 % B at 42 min, 80 % B at 47 min, 2 % B at 47.1 min. The trap was in line with the column from 3 min to 40 min. A flow rate of 300 nl / min was used on the analytical column and electrospray. LC wash runs ran a total of 50 minutes with three fast gradients from 1 % B to 80 % B on the analytical column and three fast gradients from 0 % B to 80 % B on the trap column.

2.2.4.6 Electrospray ionization

Electrospray ionization was driven with an applied potential of 2500 V. A CoAnn pulled emitter with 20 µm inner diameter (ID) to 10 µm ID at the tip was used.

2.2.4.7 Untargeted LC-MS DDA acquisition for metallothionein-2

The LC gradient was Grd_DDA. Expected LC peak width was set to 14 s. APD was enabled. Lock mass was $[(Si(CH_3)_2O)_6+H]^+$ (445.120025 *m/z*) from ambient air in scan-to-scan mode. RF Lens was set to 50 %. The full MS1 scan was set to 120k resolution, 300–1650 *m/z*, 300 % AGC target, 35 ms max. injection time. Precursor selection was using peptide MIPS, 1×10^2 min. intensity, charge-states 1–3, dynamic exclusion for 20 s at 6 ppm tolerance, excluding isotopes. MS2 was cycled for 10 scans. Full MS2 scan was set to 1.2 *m/z* isolation window, 28 % NCE, 30k resolution, automatic scan range, 200 % AGC target, 150 ms max. injection time.

2.2.4.8 Untargeted LC-MS DDA acquisition for peptide alkylation test

The LC gradient was OptiGrd3. Expected LC peak width was set to 16 s. APD was enabled. Lock mass was $[(Si(CH_3)_2O)_6+H]^+$ (445.120025 *m/z*) from ambient air in scan-to-scan mode. RF Lens was set to 50 %. The full MS1 scan was set to 120k resolution, 300–1650 *m/z*, 300 % AGC target, 35 ms max. injection time. Precursor selection was using peptide MIPS, charge-states 1–4, dynamic exclusion for 20 s at 6 ppm tolerance, excluding isotopes. Dependent MS2 scans were cycled for 3 seconds. Full MS2 scan was set to 1.2 *m/z* isolation window, 27 % NCE, 15k resolution, automatic scan range, 200 % AGC target, 120 ms max. injection time.

2.2.4.9 Untargeted LC-MS DIA acquisition for HLA IPs

The LC gradient was Grd4. Expected LC peak width was set to 16 s. APD was enabled. Lock mass was $[(Si(CH_3)_2O)_6+H]^+$ (445.120025 *m/z*) from ambient air in timed mode for the first 3 min. RF Lens was set to 50 %. The full MS1 scan was set to 120k resolution, 300–1650 *m/z*, 300 % AGC target, 35 ms max. injection time. Independent MS2 scans used 44 dynamic precursor windows from 299.5 *m/z* to 1650.5 *m/z* with 1 *m/z* overlaps. Full MS2 scan was set to 30 % NCE, 30k resolution, scan range from 200 *m/z* first mass, 3000 % AGC target, automatic max. injection time.

2.2.4.10 General targeted LC-MS acquisition settings

Settings for targeted experiments are as stated here if not specified differently. The LC gradient was OptiGrd3 or Grd4. Expected LC peak width was set to 16 s. APD was enabled. Lock mass was $[(Si(CH_3)_2O)_6+H]^+$ (445.120025 *m/z*) from ambient air in timed mode for the first 3 min. RF Lens was set to 50 %. Retention time standard MS2 was acquired with scheduled acquisition with 30 % NCE, 15k resolution, automatic scan range, 50 % AGC target, 40 ms max. injection time, all scheduled target cycling. The dynamic retention time (RT) feature was set to use the PRTC kit with 6 min RT window for the first peptide and 2 min RT windows for others. Precursor isolation was set to minimum vendor recommendation: 0.4 *m/z* for precursor ≤ 400 *m/z*, 0.7 *m/z* for ≤ 700 *m/z*, 1 *m/z* for ≤ 1000 *m/z*, 1.5 *m/z* for ≤ 1500 *m/z*, 2.0 *m/z* for ≤ 200 *m/z*, 3 *m/z* for ≤ 2500 *m/z*.

2.2.4.11 MS scan parameter tuning LC-MS acquisition

The LC gradient was Grd4_fast. The full MS1 scan was set to 120k resolution, 360–803 *m/z*, 300 % AGC target, 25 ms max. injection time. The dynamic RT feature was off. Target peptide RT schedule was defined such that no targets overlap. Base settings for target peptide MS2 were scheduled acquisition with 30 % NCE, 120k resolution, automatic scan range, 1000 % AGC target, 25 ms max. injection time, scheduled target cycling for 2.1 s. Only parameters to be tuned were adjusted where needed.

2.2.4.12 SIL reference targeted LC-MS acquisition

The full MS1 scan was set to 60k resolution, 220–1450 *m/z*, 300 % AGC target, 25 ms max. injection time. Target peptide MS2 was acquired with scheduled acquisition with custom NCE, 60k resolution, automatic scan range, 1000 % AGC target, 7 point dynamic max. injection time, scheduled target cycling for 2.6 s. Precursor charge-states 1–4 were all included in the acquisition until the best precursor were selected.

2.2.4.13 IP targeted LC-MS acquisition

The full MS1 scan was set to 60k resolution, 220–1450 *m/z*, 300 % AGC target, 25 ms max. injection time. Where spiked-in, SIL peptide MS2 was acquired with scheduled acquisition with custom NCE, 60k resolution, automatic scan range, 500 % AGC target, 350 ms max. injection time, all scheduled target cycling. Target peptide MS2 was acquired with scheduled acquisition with custom NCE, 60k–480k resolution, automatic scan range, 1000 % AGC target, 5–7 point dynamic max. injection time, scheduled target cycling for 2.6–3.1 s.

Depending on the targeted HLA supertypes, sets of commonly presented allotypic peptides were targeted as positive controls Table 1. These were helpfully suggested by Stefan Stevanovic. For supertype A02, a set of negative controls derived from mouse genes were also included.

Table 1. Allotypic peptides acquired as controls.

Supertype	Peptide #1	Peptide #2
A01	DTDHYFLRY (sp|Q969N2|PIGT_HUMAN)	IADMGHLKY (sp|P12004|PCNA_HUMAN)
A02	YLLPAIVHI (sp|P17844|DDX5_HUMAN, sp|Q92841|DDX17_HUMAN)	AIVDKVPSV (sp|Q9Y678|COPG1_HUMAN)
A03	KIADRFLLY (sp|P61968|LMO4_HUMAN)	KLLNYAPLEK (sp|P62875|RPAB5_HUMAN)
A11	TSALPIIQK (sp|Q99541|PLIN2_HUMAN)	AVILPPLSPYFK (sp|O95167|NDUA3_HUMAN)
A24	AYVHMVTHF (sp|P55061|BI1_HUMAN)	KYPENFFLL (sp|P36873|PP1G_HUMAN, sp|P62136|PP1A_HUMAN, sp|P62140|PP1B_HUMAN)
B07	APRTVALTAL (sp|P04440|DPB1_HUMAN)	IPRAALLPLL (sp|Q92743|HTRA1_HUMAN)
B15	TQMPDPKTF (sp|Q9Y5V0|ZN706_HUMAN)	GQVQKIVLY (sp|Q00610|CLH1_HUMAN)
A02 mouse derived neg. Ctrl	FGPVNHEEL (sp|P46414|CDN1B_MOUSE)	KALINADEL (sp|P16546|SPTN1_MOUSE)

2.2.5 Direct infusion MS (DI-MS)

Dried mixtures of synthetic peptides at 10 pmol per peptide were resolubilized in 5 µl of 50 % ACN / 4% FA for a concentration of 5 pmol / ul. The sample was sonicated for 3 min and spun down. A 150 µm inner diameter fused silica capillary with a length of ~40 cm was connected on one end to a 25 µl syringe, with the other end open. The capillary was filled with 5 % ACN. Next, 5 µl sample were drawn into the capillary. The emitter (CoAnn, ID 20 µm, tip ID 10 µm) was connected through the Simple Link Uno at the sample end. At the syringe end, a 500 µl syringe was connected. The syringe contained water and enclosed an air bubble to allow to set a consistent pressure. The 250 µl air bubble was compressed to 50 µl to set a pressure and start flow at the emitter. The spray was driven and applied potential of 2500 V and the MS acquisition method was started once strong signal was observed in MS1. Expected LC peak width was set to 30 s. APD was enabled. Lock mass was $[(Si(CH_3)_2O)_6+H]^+$ (445.120025 *m/z*) from ambient air in scan-to-scan mode. RF Lens was set to 50 %. Sets of 5,000 MS2 scans were specified with precursor target tables. For quality control, an intermittent full MS1 scan was used at the start and after every completed set of MS2 scans. MS1 was set to 120k resolution, 220–2000 *m/z*, 300 % AGC target, 100 ms max. injection time. MS2 was set to 45k resolution, automatic scan range, standard AGC target, 180 ms max. injection time. For each synthetic peptide the mixture, precursor charge states +1 to +4 were targeted successively. Methionine oxidation was set as variable modification. Cysteine alkylation was set as fixed modification. For each precursor, NCE ranging from 4 % to 42 % in steps of 2 % were scanned in 5 iterations so that 5 datapoints were acquired per NCE value. Each iteration used a randomized order with a fixed seed. The resulting pseudorandom pattern is the same for all precursors. Some

optimization data that were acquired early on in the project spanned slightly different NCE ranges and acquired only 4 datapoints per NCE value. The method runtime was set to 300 min and was typically ended manually upon reaching the end of the scheduled scans which was signaled with a continuously looping dummy scan.

2.2.6 Data analysis

2.2.6.1 Supertype population coverage calculation

Supertype population coverage was calculated using the offline version of the IEDB Population Coverage Tool. The tool was repurposed for supertype coverage by specifying a complete set of supertype associated alleles per epitope input. For single supertype coverage, a single mock epitope input was used. For combinations of supertypes, one mock epitope was specified per supertype. HLA supertype allele assignments were as specified in Sidney et al. (2008) using the restriction "exact match(es) in the B and F pockets".

2.2.6.2 Human proteome search for candidate peptides

The candidate peptides were searched in an extensive human UniProtKB fasta (proteome:up000005640 OR organism_id:9606, canonical and isoform, accessed: 2023-10-10) using regular expressions (regex) in R. The regex was ambiguous for MS-indistinguishable amino acids I and L ([IL]). The only two matches were from trEMBL with tr|Q8V9K8|Q8V9K8_HUMAN, tr|Q9Y4Y4|Q9Y4Y4_HUMAN which are unreviewed and denoted to relate to protein E6.

2.2.6.3 MS data and metadata extraction

Rawfiles were accessed using a custom .NET command line tool 'rawfile_cli_base64' that accesses the MS data files (.raw) using the Thermo RawFileReader library. The tool is used for extraction of extracted ion chromatograms (XICs), spectra, scan metadata and charge-state resolved total ion currents (TICs). For charge-state resolved TIC, the tool iterates over every centroided datapoint in every full MS1 scan to read out the charge state annotations and accumulates per scan the total intensity values resolved by charge state. Data is returned through stdout and was accessed through R.

2.2.6.4 Normalized spectral contrast angle

For comparison of acquired peptide signal to SIL reference signal, the normalized spectral contrast angle (NSA) as defined by Toprak et al. (2014) was used. This is also the default approach termed 'dotp' in the Skyline software. The reference signal was derived from XIC of synthetic peptides (typically SIL) which was extracted with Skyline. Fragment criteria for the reference library were all fragment lengths of types y, b, a, and p with up to two neutral losses and 2 *m/z* exclusion around precursor *m/z*. Allowed neutral losses were loss of water (D, E, S, T; -18 Da), ammonia loss (K, N, Q, R; -17 Da) and loss of CH_4SO (-64 Da) loss from oxidized methionine. The reference is indexed by ion type (e.g. y_3^+, a_5^{++}-18) and not *m/z*, to facilitate heavy vs. light intensity comparison. For comparison of XICs, peak area was used as intensity. Similarly, for comparison of spectra, the extracted fragment intensities were used and unidentified background signal was not considered. The resulting two vectors, one for reference and one for the acquired signal, were sorted equally, according to the fragment rank in the reference and missing fragments were filled with 0 intensity. Fragments labeled non-quantitative (n.q.) were excluded from both vectors.

The vectors were both normalized according to Equation 1, where VO is the output vector and VI is the input vector. The similarity score NSA is then calculated for the two normalized vectors V1 and V2 according to Equation 2.

Equation 1. Vector normalization for NSA scoring.

$$VO_i = \frac{\sqrt{VI_i}}{\sqrt{\sum_{j=1}^{n} VI_j}}$$

Equation 2. NSA calculation.

$$\mathrm{NSA} = 1 - \frac{2\arccos(\mathrm{V}_1 \cdot \mathrm{V}_2)}{\pi}$$

2.2.6.5 MS parameter optimization

Targeted experiments were analyzed with the software Skyline, which was used for data extraction and peak curation. Top 5 reference fragments were enabled. Fragment criteria were fragment *m/z* > precursor *m/z* using fragments of types y, b, a with up to two neutral losses and 2 *m/z* exclusion around precursor *m/z*. Allowed neutral losses were loss of water (D, E, S, T; -18 Da), ammonia loss (K, N, Q, R; -17 Da) and loss of CH_4SO (-64 Da) from oxidized methionine. MS2 XICs were extracted in centroided mode with 3 ppm tolerance, except for resolution tests, where 7 ppm tolerance was used. Where transitions were affected by interference, they were marked non-quantitative (n.q.) for exclusion from NSA calculation and a further reference transition was enabled in its stead. To count the amount of transitions detected, top 50 reference fragments ≥ 3 amino acids were enabled in a separate assessment. Data was exported as .csv tables for downstream analysis in R. Transition counts were calculated by iterating from the top transition down to further transitions until the first point when < 90 % of transitions were detected. Reverting to the last detected transition then indicated the count.

2.2.6.6 Collision energy optimization

Data was extracted and analyzed using R with the *MsRawAccess* package. For each precursor, possible fragments and their *m/z* were generated using the Julia package *Peptidomics*. For fragmentation, y, b, and a ions and the precursor were used. 2 *m/z* exclusion window around precursor *m/z*. Allowed neutral losses were up to two of loss of water (D, E, S, T; -18 Da), ammonia loss (K, N, Q, R; -17 Da) and loss of CH_4SO (-64 Da) loss from oxidized methionine. Charge states were from +1 to the precursor charge state. MS2 data were extracted in centroid mode with 9 ppm tolerance. For selection of the optimal normalized collision energy (NCE), the target was to maximized one of the 4th–6th most intense transitions. Furthermore, at least 6 transitions must be detectable at the optimal NCE. Quantification of the optimization benefit was performed with an unpaired one-sided t test comparing the 5th most intense fragment at the determined optimal NCE against the 5th most intense fragment at 30 % NCE. For quality assurance, precursors in the overall efficacy evaluation are limited to precursors that were determined to be relevant in LC-MS, meaning either dominant or with comparable intensity to the dominant precursor. This assessment was made independently from the optimization as it was routinely performed to decide which precursors to target in the targeted LC-MS experiments.

2.2.6.7 HLA MAC decoding, supertype assignment and binding prediction

Multiple allele codes (MAC) represent groups of HLA alleles and are useful when the HLA typing is ambiguous and does not allow to narrow down one single allele from a list of alleles. This can occur for example where HLA typing is not based on full sequencing of an HLA gene, usually because the employed technique is designed to resolve only differences in the peptide binding domains (Marsh

et al., 2010). MAC are registered centrally with the US National Marrow Donor Program. The website[1] offers a service to decode the MAC, which returns a list of sometimes dozens of HLA alleles that match the HLA typing results.

HLA alleles specified using MAC codes were expanded to all associated alleles using the R package *immunotation*, which provides access to the MAC Service of the National Marrow Donor Program. HLA allele prevalence was calculated using the all world output of the IEDB Population Coverage Tool. HLAs without prevalence data were discarded. Supertypes assigned at this stage were unambiguous in all cases, meaning no two HLA alleles derived from a MAC code in this way were found to belong to distinct supertypes. Peptide HLA binding predictions were performed with netMHCpan for all supported alleles. Unsupported alleles were discarded. Finally, for each peptide and HLA allele, a single Rank EL (%) binding prediction was selected by picking the weighted median value, where weights are the HLA prevalences.

2.2.6.8 HeLa digest untargeted DIA LC-MS

Data-independent acquisition (DIA) experiments were analyzed with the search engine Spectronaut directDIA. Settings were automatic tolerances, Trypsin / P digest mode with a maximum of two missed cleavages, and a max of 2 post-translational modifications (PTMs) per peptide. 1 % peptide-spectrum match (PSM) false discovery rate (FDR) and 1 % peptide FDR was set. Carbamidomethyl (C) was set as a fixed modification. Acetylation (N-term) and Oxidation (M) were set as variable modifications. Where the oxidation step was diagnosed, the modification was set as variable. The database was Uniprot Human reference proteome (accessed: 2021-10-21). Data was exported as .csv tables for downstream analysis with R.

2.2.6.9 SIL reference targeted LC-MS

Targeted reference acquisitions were analyzed with the software Skyline, which was used for data extraction and peak curation. Initial peak selection was performed with fragments of ≥ 3 amino acids in length of types y, b with up to one neutral loss and 2 *m/z* exclusion around precursor *m/z*. Allowed neutral losses were loss of water (D, E, S, T; -18 Da), ammonia loss (K, N, Q, R; -17 Da) and loss of CH_4SO (-64 Da) loss from oxidized methionine. MS2 XICs were extracted in centroided parallel reaction monitoring (PRM) mode with 4–8 ppm tolerance. Following the initial peak picking, further transitions were enabled by inclusion of all fragments of ≥ 1 amino acid in length of types y, b, a, and additional precursor-derived fragments (type p) with up to two neutral losses. The now available extensive list of fragments was curated to exclude fragments not adhering to the peak shape. For each peptide, the dominant precursor charge-states were noted for later targeting in IPs. Data were exported as spectral library (.blib) and .csv tables for downstream analysis in R.

2.2.6.10 IP untargeted DDA LC-MS

Data-dependent acquisition (DDA) experiments were analyzed with the database search engine PEAKS Studio. Settings were parent mass tolerance of 10.0 ppm, fragment mass tolerance of 0.02 Da, unspecific digest mode and a maximum of 3 PTMs per peptide. 1 % peptide FDR was set. Carbamidomethyl (C), Oxidation (M) and Acetylation (N-term) were set as variable modifications. The databases were Uniprot Human and HPV16 reference proteomes (accessed: 2020-04-22). Data was exported as .csv tables for downstream analysis with R.

[1] https://bioinformatics.bethematchclinical.org/hla-resources/allele-codes/ (accessed: 2023-10-17)

2.2.6.11 IP untargeted DIA LC-MS

DIA experiments were analyzed with the search engine Spectronaut directDIA. Settings were automatic tolerances, unspecific digest mode and a maximum of 2 PTMs per peptide. 1 % PSM FDR and 1 % peptide FDR was set. Carbamidomethyl (C) was set as a fixed modification. Acetylation (N-term) was set as variable modification. Oxidation (M) was by default set as fixed modification. Where the oxidation step was diagnosed, the modification was set as variable. The databases were Uniprot Human and HPV16 reference proteomes (accessed: 2021-10-21), including a manual addition of known HPV16 E6 and E6 mutations. Data was exported as .csv tables for downstream analysis with R.

HLA supertypes were assigned according to Sidney et al. (2008). HLA allele prevalence was calculated using the all world output of the IEDB Analysis Tool Population Coverage. HLA binding predictions (Rank EL (%)) were performed with netMHCpan. Peptides were considered binders for HLA alleles where Rank EL (%) was ≤ 2 %. HLA binder population density for violin plots was calculated after log10 transformation using a gaussian kernel with bandwidth 0.1 and spanning values from 1×10^{-4} to 1×10^{2}.

2.2.6.12 IP targeted LC-MS

Targeted experiments were analyzed with the software Skyline, which was used for data extraction and peak curation. Initial peak selection was performed with top 5 reference fragments enabled. Fragment criteria were ≥ 3 amino acids in length of types y, b, a with up to two neutral losses and 2 *m/z* exclusion around precursor *m/z*. Allowed neutral losses were loss of water (D, E, S, T; -18 Da), ammonia loss (K, N, Q, R; -17 Da) and loss of CH_4SO (-64 Da) loss from oxidized methionine. XICs were extracted for MS2 in centroided PRM mode with 4–8 ppm tolerance. Following the initial selection, further transitions were enabled, including top 25 reference fragments, ≥ 1 amino acid and additional precursor-derived fragments (type p). The now available extensive list of fragments was curated to exclude low-specificity fragments (< 3 amino acids or precursor-derived) that were affected by interference (as determined by non-adherence to the peak shape). Further fragments not affected by interference, were marked non-quantitative (n.q.). The list of fragments was then trimmed back to the rank at which most fragments are detected. Peaks scoring with an NSA value ≥ 0.85, based on at least 5 specific fragments (≥ 3 amino acids, types y, b, a), were indicated as detected. Data were exported as .csv tables for downstream analysis in R.

2.2.7 Patient tumor genotyping

The tumor sample I received was acquired as part of the CAPTURE/CATCH prospective precision oncology program (Hlevnjak et al., 2021). Whole genome sequencing had already been performed at the DKFZ Division of Molecular Genetics. In brief, DNA and RNA from tumor sample was isolated according to standardized workflows using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen). DNA concentrations were determined using fluorometric Quant-iT dsDNA BR Assay Kit (Thermo Fisher Scientific). The quality of DNA was tested using an automated electrophoresis system (genomic DNA ScreenTape for TapeStation System, Agilent). DNA libraries of fresh-frozen samples were prepared for whole-genome sequencing (WGS) using the TruSeq Nano DNA Kit (Illumina) and were sequenced paired-end on a NovaSeq 6000 platform (Illumina) yielding paired-end (PE) 151 bp long reads with 332.47 Gigabases. HPV16 read extraction from the raw WGS sequencing reads was performed using Kraken2 (Wood et al. (2019), v2.09, confidence 0.1) with a database built on the human genome (hg19) and the NCBI nonredundant nucleotide database (nt) (January 2020). Reads assigned by Kraken2 to taxonomy id 333760 were extracted from the fastq files based on their read id.

To determine single nucleotide polymorphisms, I realigned the 4498 HPV16 reads (each 151 nucleotides) using the NCBI basic local alignment search tool (BLAST) service (accessed: 2023-

10-07) (Sayers et al., 2022). The subject sequence was set to GenBank K02718.1, the algorithm to megablast, and the maximum display number to 5,000 alignments.

2.2.8 ELISpot test for donor PBMC reactivity against synthetic peptide

The preliminary immunogenicity data generated by enzyme-linked immunospot (ELISpot) are not results of this thesis and were performed by Dr. Maria Bonsack and Kathrin Wellach.

Buffy coat preparations of the blood from anonymous healthy female donors were obtained from DRK Blutspendedienst Mannheim. Material of donors of at least 40 years of age was used to increase the probability of prior exposure to HPV. HLA typing was performed by the DKMS Life Science Lab GmbH in Dresden (Germany). Peripheral blood mononuclear cell (PBMC) isolation was performed using a standard density gradient procedure with Ficoll-Paque PLUS in Leucosep tubes (Greiner Bio-One). The isolated PBMCs were subsequently washed with PBS, and residual erythrocytes were lysed by 5 min incubation with ACK lysis buffer (150 mM NH_4Cl, 10 mM $KHCO_3$, and 0.1 mM EDTA). PBMCs were washed, suspended in human serum with 10 % DMSO, and cryopreserved in the gas phase of liquid nitrogen.

PBMCs were thawed and one T cell culture per antigen stimulation was set up with 1–2×10^6 cells / well in 24-well plates in T cell medium (RPMI-1640 with 10 % human serum, 2 mM L-glutamine, 10 mM HEPES, and 1 % P/S) supplemented with 10 ng / ml recombinant human Interleukin 7 (rhIL-7; R&D Systems). Each T cell line was stimulated with 10 mg / ml of either one of the HPV16 E6- and E7-derived synthetic peptides or 10 mg / ml of HIV-, CMV- or EBV-derived synthetic control peptides, or 1 mg / ml per peptide of the positive control CEF peptide pool. On day 3 and day 7, cultures were fed with 20 U / ml of recombinant human Interleukin 2 (rhIL-2; PeproTech), with a half medium change at day 7. At day 12, 1–2×10^5 cells / well were transferred to sterile Multiscreen-HA membrane plates (Millipore) coated with 1:500 anti-human IFNγ (clone 1-D1K; Mabtech) in PBS. Cell lines were re-stimulated with respective peptide or CEF peptide pool (# PA-CEF-001, PanaTecs) (four replicates). Concanavalin A (Sigma Aldrich, 2 mg / ml) and DMSO (Sigma Aldrich, 1 ml / ml) (two replicates each) were used as positive and negative controls. After 24 h incubation, cells were discarded and plates were washed with PBS and developed with 1:1000 biotinylated anti-human IFNγ (clone 7 B6-1-Biotin; Mabtech) in PBS, 1:2000 Streptavidin-Alkaline Phosphatase solution (Mabtech) in PBS, and filtered substrate (NBT/BCIP; Millipore). Plates were washed with ELISpot wash buffer (PBS with 0.05 % Tween 20) between steps. Spots were counted with an automated ImmunoSpot reader (CTL-Immunospot S6 Ultra-UV, CTL, Bonn, Germany). For each well, spot forming units (SFU) per 10^6 cells and the stimulation index (SI) (spot count divided by mean spot count in DMSO-treated control wells) were calculated. Peptides were defined as immunogenic, if the detected response induced SFU ≥ 100 per 10^6 cells and SI ≥ 2. Additionally, if the SFU of the DMSO background control was < 5, responses with SFU ≥ 50 per 10^6 cells and SI ≥ 4 were defined as positive.

3 Results

3.1 Optimization of the HLA immunoprecipitation protocol (Wet-lab)

3.1.1 Establishment of the HLA immunoprecipitation protocols

The HLA immunoprecipitation (IP) techniques used in this thesis were established over the course of this project in collaborative efforts with two postdoctoral fellows in the group, first with Dr. Mogjib Salek for a sensitive and robust batch IP technique, and later with Dr. Jonas Becker for the implementation of the high-throughput 96-well plate-based protocol. The resulting protocols are adaptations of the most influential published protocols in the field (Bassani-Sternberg et al., 2015; Chong et al., 2018; Purcell et al., 2019). My contributions in this process were the implementation of the targeted acquisition method to measure the outcome in terms of HPV16-derived peptide detections, the development of quality control analyses after liquid chromatography-mass spectrometry (LC-MS) acquisition and the establishment of chemical peptide modifications to facilitate optimal detection of peptides containing methionine and cysteine. The high-throughput protocol showed similar performance to the batch IP protocol and was therefore used interchangeably. Whether the batch or high-throughput protocol was used is indicated throughout this thesis and in Appendix Table 1.

The main cell line used for the establishment and refinement of the method in our lab is the widely used CaSki cell line. It is derived from a cervical cancer metastasis and contains an integrated HPV16 genome. It is an adherent cell line that is well suited for cell culture and reliably expresses HLA and the proteins E6 and E7 (Choo et al., 1994). A further cell line, C33A, was used as a faster growing alternative that was also derived from a cervical squamous cell carcinoma, but is HPV16 negative and therefore does not present HPV16-derived peptides (negative control cell line).

3.1.2 Charge state-resolved total ion current

One marked property of the common immunoprecipitation protocols is that the eluted HLA class I binding peptides, while chemically diverse, have a limited length of typically less than 12 amino acids and therefore ionize at the electrospray with charge states between +1 and +4 at the most. This property creates a signal that is immediately distinguishable from the common proteinaceous contaminants and polymeric detergents, which may acquire many more charges. On high resolution mass spectrometers, these charge states can reliably be determined for most of the detected ions. Specifically, the Orbitrap Exploris MS utilizes an advanced peak determination (APD) algorithm to extract this information by observing isotopic clusters of ions and the spacing of the contained isotopes on the mass-to-charge (*m/z*) axis (Hebert et al., 2018). To harness these data, which were not accessible otherwise, I developed a command line utility that uses the low-level Thermo RawFileReader .NET library for deep access to the experiment data (.raw files). This utility iterates through every full MS1 scan and all ion signals, reading out the charge state annotations and accumulating total intensity values resolved by charge state.

The extracted data enables the creation of charge state-resolved total ion current (TIC) chromatograms. TICs are a graphical representation of all the ions detected by a mass spectrometer over a period of time, which is in this case the complete LC gradient. In Figure 8, an example is shown where two exemplary IPs show similar TICs (left panels) but resolving by charge state (right panels) reveals drastically different character of the signal. A large proportion of the ions produced by the IP in the upper row (labeled ‘contaminated’) are of very high charge states. The additional signal colored in dark red is of undetermined charge state, likely a consequence of the resulting degradation in spectrum quality. In contrast, the IP in the lower row (labeled ‘optimal’) shows a vast majority of

signal to be of charge states +2–3, which matches well with the properties of the target analytes, i.e. HLA binders. Co-eluting non-target ions are known to impact MS sensitivity through ion suppression and through adversely affecting the dynamic range on the MS1 level as well as through co-isolation in MS2 (Kaufmann et al., 2010; Domon & Gallien, 2015). The charge state-resolved diagnostic plots have proven much more effective in revealing contamination issues than standard TIC chromatograms. I made these quality control plots available from R and via a graphical user interface and they have been a staple tool for quality control for our research group and over the course of this thesis work.

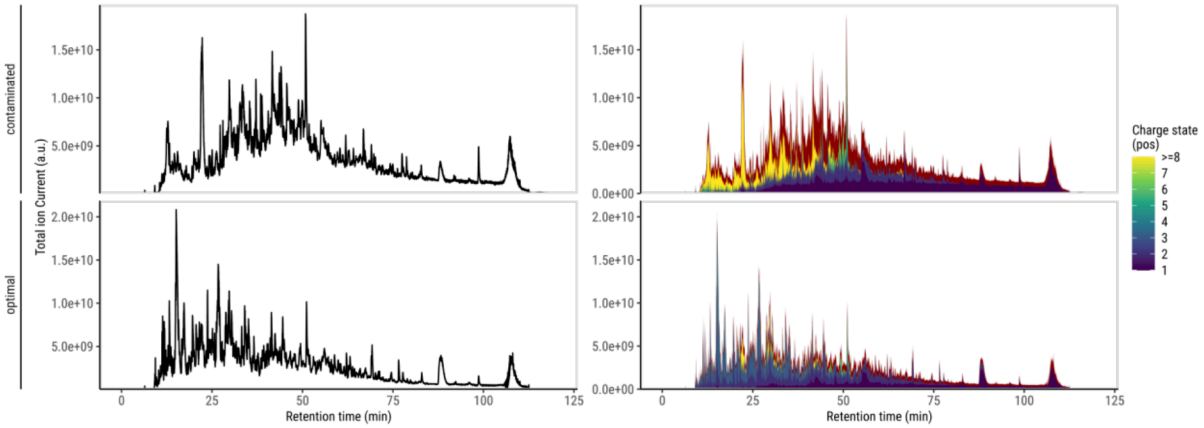


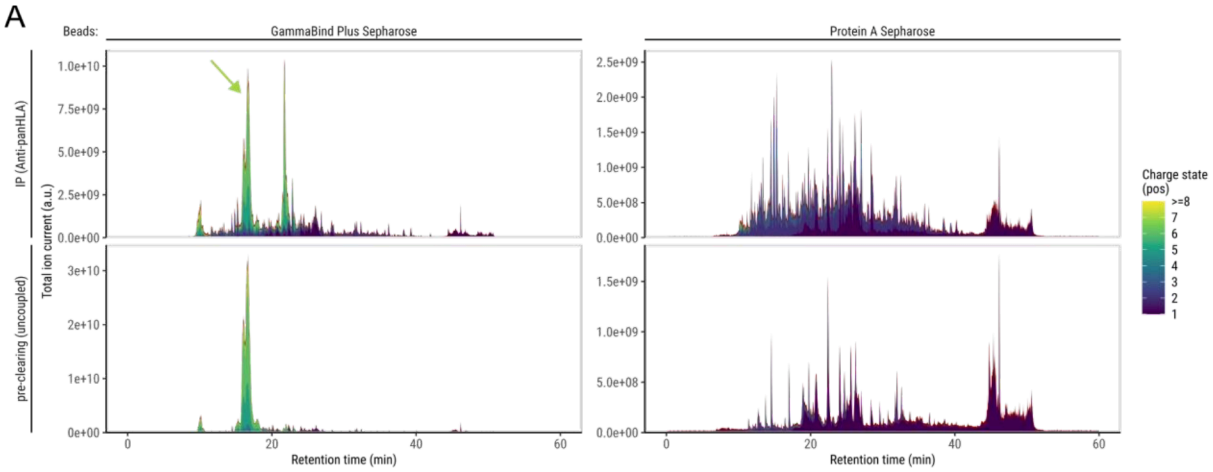
Figure 8. Charge state-resolved total ion current chromatograms. Two different immunopeptidomics experiments are shown (top vs. bottom). The left panels show commonly used total ion current (TIC) chromatograms. The right panels show the same chromatograms resolved by charge state of the acquired ions. Charge states 1–4 can be considered viable peptide signal. Dark red signal is of undetermined charge state.

3.1.3 Metallothionein-2 contamination

The use of charge-state resolved TIC chromatograms revealed a distinct contamination that appeared in almost all immunopeptidomics experiments early on in the project. Exemplarily, the contamination was present in IP experiments of the cell line C33A, where a cluster of mainly +6 charged ions appeared dominant early in the gradient (Figure 9A, green arrow). A signal of similar nature consisting of multiple ion species of 5.93 kDa to 6.25 kDa in mass appeared with most cell line-based immunopeptidomics experiments. This was clearly an undesirable species of ions and in testing for the source of the contamination, I found that the signal can already be detected in eluates from a pre-clearing step. This optional step in the IP protocol, which is commonly used when tissue sample is processed, is performed with bare Sepharose beads, i.e. beads that have not been coupled to the HLA-specific antibodies. Typically, this step is used to deplete unspecific antibodies (pre-clearing). Antibodies are not expected in this sample so this was purely exploratory. In principle, the step could serve to deplete this contamination too, but the depletion is incomplete. The pre-clearing step (bottom left panel) does not resolve the contamination in the IP step of the same experiment (top left panel). Using beads only without cell lysate does not produce the signal (data not shown). I found that the use of Protein A Sepharose beads instead of the Gammabind Plus Sepharose beads resolves this contamination completely (right panels). To understand the nature of the contaminant, I worked with Dr. Mogjib Salek who performed a tryptic digestion of an IP that was prepared with Gammabind Plus Sepharose beads and I performed untargeted LC-MS analysis of the sample. Database search, limited to tryptic peptide identification, identified only the protein Metallothionein-2 (MT2_HUMAN) with major coverage (Figure 9B). This is a small cytosolic protein of canonical mass 6042 kDa, fitting well with the signature of the contamination. Other proteins were only detected with minor coverage, likely due to a subset of HLA-presented peptides that resemble the products of

a tryptic digestion. It was therefore determined that the contamination results from co-precipitation of the protein Metallothionein-2 with Gammabind Plus Sepharose beads. This cytosolic protein is likely present in high abundance in cell lysate. The contamination was resolved by using a different Sepharose bead product.

A



B

	Accession	-10lgP	Coverage	#Peptides	Avg. Mass
	Proteins				
1	P02795|MT2_HUMAN	166.21	100%	7	6042
2	P61769|B2MG_HUMAN	88.84	19%	2	13715
3	P62861|RS30_HUMAN	51.95	19%	1	6648
4	P63173|RL38_HUMAN	52.70	14%	1	8218
5	P62805|H4_HUMAN	58.00	13%	1	11367
6	Q96FQ6|S10AG_HUMAN	55.56	11%	1	11801
7	P28799|GRN_HUMAN	69.95	9%	2	63545

Figure 9. GammaBind Plus Sepharose beads elute Metallothionein-2. (A) Results of two C33A HLA IPs using different Sepharose beads and including a pre-clearing step. Charge state-resolved TIC chromatograms reveal a dominant contamination of majority charge state +6 at ~6 kDa (green arrow). (B) Protein identifications from a tryptic digestion of a GammaBind Plus HLA IP eluate. Panel B adapted from (Salek & Förster et al., 2023). Experiment in panel B was performed by Dr. Mogjib Salek. Analysis by me.

3.1.4 Methionine oxidation

The presence of the amino acid methionine in many of the target peptides posed an additional challenge for the detection via LC-MS. Methionine is readily oxidizable to methionine sulfoxide and peptides containing methionine may therefore be detectable in one of two states with very distinct properties (Figure 10A). The oxidized form typically elutes several minutes earlier and has a different precusor mass (+16 Da) and fragmentation pattern. In initial targeted IPs, I measured various ratios of oxidized vs. reduced forms of methionine-containing peptides in the LC-MS. I concluded that likely sources of this variance are oxidation during the IP processing and LC separation. Consequently, a sensitive targeted method would have to target both forms for each peptide at the cost of additional instrument cycles and, in case of a 50% oxidation rate, may only receive half of the ions each. An effective way to tackle this is to force oxidation of methionine, which is by now also described for immunopeptidomics elsewhere (Pollock et al., 2021).

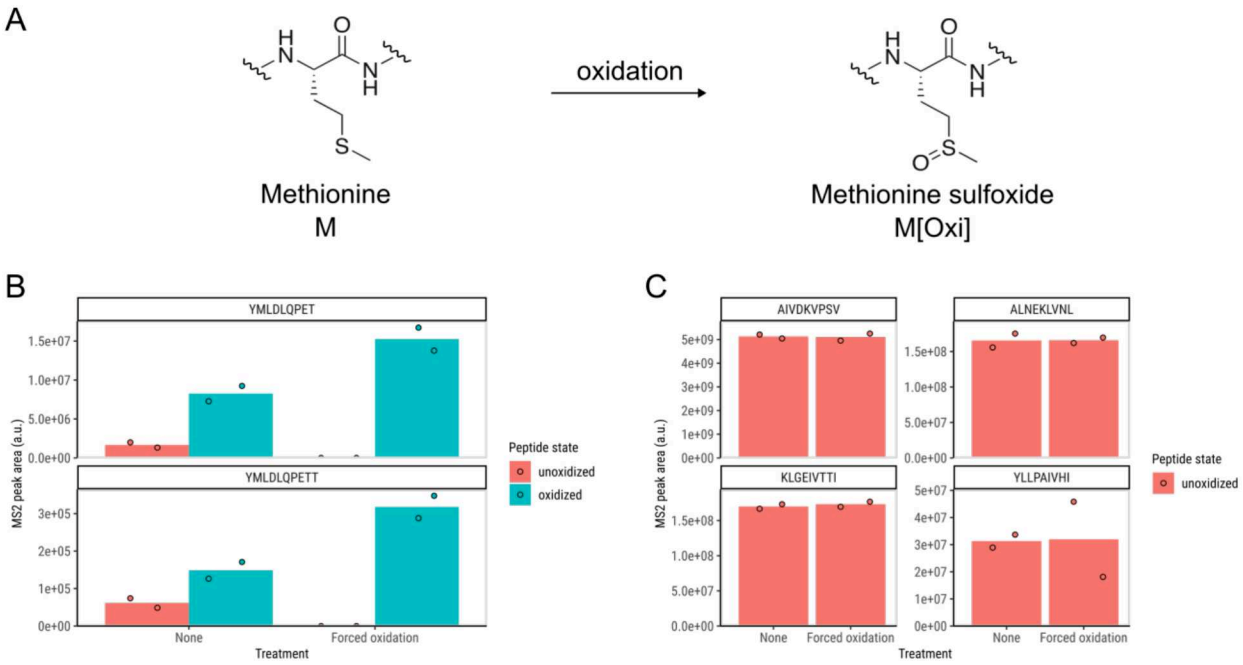


Figure 10. Oxidation of methionine-containing peptides in an HLA IP eluate. (A) Chemical structure of methionine and methionine sulfoxide produced through oxidation. (B) Quantification of signal for HPV16-derived peptides containing methionine. HLA IP samples either received oxidation treatment or no treatment. Peptides were acquired both in the oxidized and the unoxidized form. (C) Quantification of signal for known HLA peptides containing no methionine. (B–C) two technical replicates per condition.

I tested a treatment of the peptides with an oxidation solution of weakly concentrated performic acid that is used to rapidly oxidize methionine and which can be briefly applied during the solid phase extraction step. Figure 10B shows the measured MS2 intensities of the two detectable methionine containing target peptides from a CaSki HLA IP, either with a standard IP protocol or with the extra oxidation step. For this experiment, the eluate of a large-scale batch IP with a load of 400M[2] cells was split into four equal aliquots for two samples per condition. Without treatment, a substantial amount of the peptide is detectable in the reduced form. By applying the oxidation step, the peptides were fully converted to the oxidized form and the signal was boosted considerably. Interestingly, in this example, the consolidated signal after treatment is for both peptides more than the sum of what was measured without treatment. The cause for this was not determined but it is likely that the reduced and oxidized MS2 signals do not appear in true proportion as the two species may ionize with different efficiency and may fragment more or less suitably for quantitation. Control measurements of known endogenous HLA binders showed no detrimental effects of the treatment on peptides not containing methionine (Figure 10C). This oxidation step was therefore applied to most IPs in the project, except for some early IPs and where there were no methionine containing peptides targeted. Whether the step was applied in any IP is specified in Appendix Table 1.

[2] For cell counts, the 'M' suffix is used to denote millions.

3.1.5 Cysteine alkylation

The HPV16 genes E6 and E7 are rich in cysteine and therefore, of the 242 candidate peptides, 148 contain cysteine. The detection of cysteine-containing peptides is particularly challenging for MS experiments due to the propensity of the thiol side chain to undergo oxidative modifications. The resulting dispersion of peptides into various forms does not allow for sensitive detection in targeted experiments. For this reason, the preceding study did not target any cysteine-containing peptides (Blatnik et al., 2018).

One of the established solutions in the proteomics field is the reduction of the sample, exposing the free thiols of the cysteines, and subsequent alkylation using reagents such as iodoacetamide (IAA) (Figure 11A). Alkylation of cysteines has also been implemented for immunopeptidomics before. In our group, Dr. Jonas Becker established a reduction/alkylation protocol for IP samples using tris(2-carboxyethyl)phosphine (TCEP) for reduction, IAA for alkylation with a subsequent quenching by TCEP, all buffered in HEPES. It is applied to the IP eluate prior to solid phase extraction. I implemented a corresponding protocol for alkylation of the synthetic peptide mixtures, which allowed me to develop the targeted assay. With these prerequisites established, the application of the alkylation step for the immunopeptidomics experiments allowed for detection of cysteine-containing HPV16-derived peptides.

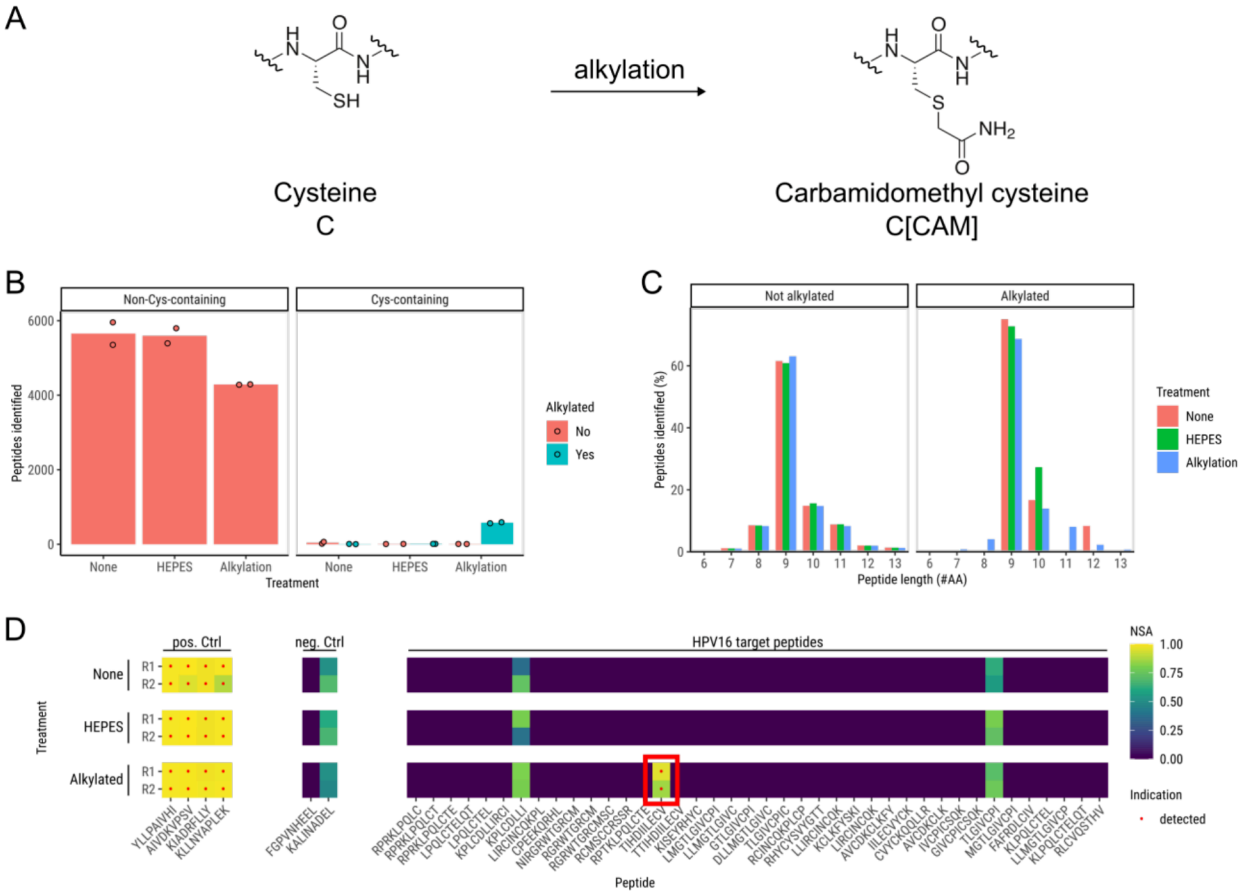


Figure 11. Alkylation of cysteine-containing peptides in an HLA IP eluate. (A) Chemical structure of cysteine and carbamidomethyl cysteine produced through alkylation. (B–C) Results of untargeted DDA acquisition of treated HLA IP eluate. Equal aliquots of a 300M cell load CaSki HLA IP. (B) Peptide identification counts separated by non-cysteine-containing and cysteine-containing peptides. (C) Peptide length distributions of the detected peptides. The right panel reflects the small populations of alkylated peptides in (B). (D) Results of targeted acquisition of the samples. NSA ≥ 0.85 is the main criterion for detection (red dot). Red box marks the novel detection of a cysteine containing peptide after alkylation.

To test for detection of cystine-containing HLA binders, I set up an experiment based on a single large-scale batch IP with 300M[3] cell load which was split for testing into six equal aliquots (50M load equivalent). Two aliquots received no treatment. Two aliquots were treated with the same steps and solvents as the alkylation protocol, but without the reactants TCEP and IAA. This was done with the objective to judge whether potential losses are due to sample handling or due to the induced reactions. The last two aliquots received the full alkylation treatment. I performed targeted and untargeted data-dependent acquisition (DDA) LC-MS analysis with each sample using 56 % (28M) and 44 % (22M), respectively. Figure 11B–C show the results of the untargeted acquisitions. The overall most peptide detections were achieved using no treatment (Figure 11B). The HEPES treatment did not meaningfully affect the result. The alkylation treatment resulted in an overall reduction of peptide identifications but the right panel for cysteine-containing peptides shows that it enabled the detection of cysteine-containing peptides in the alkylated form. None of the other samples showed an appreciable amount of cysteine-containing peptide detections in any form. Figure 11C shows that the alkylated peptides resulting from the alkylation treatment (right panel, blue) follow the same characteristic size distribution for HLA binders as the non-alkylated peptides. This is a strong indication of their validity. Figure 11D shows the results of the targeted experiments in a heatmap format, where high confidence detections (indicated by NSA ≥ 0.85) are marked with a red dot. The alkylation treatment enabled the novel detection of a single cysteine-containing HLA binder TIHDIILECV (red box). Taken together, these results confirmed that the alkylation protocol allows for the detection of cysteine-containing peptides. Due to this, I decided to include cysteine-containing peptides as viable targets for this thesis. To account for the observed losses incurred by alkylation, I split the initial rounds of epitome map experiments into a set targeting non-cysteine-containing peptides and a distinct set of alkylated HLA IPs targeting cysteine-containing peptides.

3.2 Optimization of peptide LC-MS detection

MS detection of peptides using the newly acquired hybrid quadrupole-Orbitrap mass spectrometer is based on drastically different principles than the method of the preceding study that used a time-consuming and labor-intensive MS3 technique on a triple quadrupole linear ion trap (QTRAP) mass spectrometer. Therefore, I re-evaluated and optimized most aspects of peptide LC-MS acquisition with the aim of establishing a method for targeted acquisition that achieves optimal sensitivity and is scalable for the detection of hundreds of peptides.

3.2.1 Detection diagnostics

The high-resolution accurate-mass Orbitrap mass spectrometer used for this thesis resolves ions of extremely similar *m/z* and determines the ion *m/z*-ratio very precisely. The tolerance on the *m/z* axis that is used to identify an ion of interest is typically below 8 parts per million (ppm). This allowed to extract trace ion signal even from the most complex of spectra and project them as extracted ion chromatograms (XICs). The classic way that peptides are identified is through the interpretation of a single spectrum but the comparison of XIC peaks is a potent technique for parallel reaction monitoring (PRM) and data-independent acquisition (DIA). Both these approaches are contrasted in Figure 12.

[3] For cell counts, the 'M' suffix is used to denote millions.

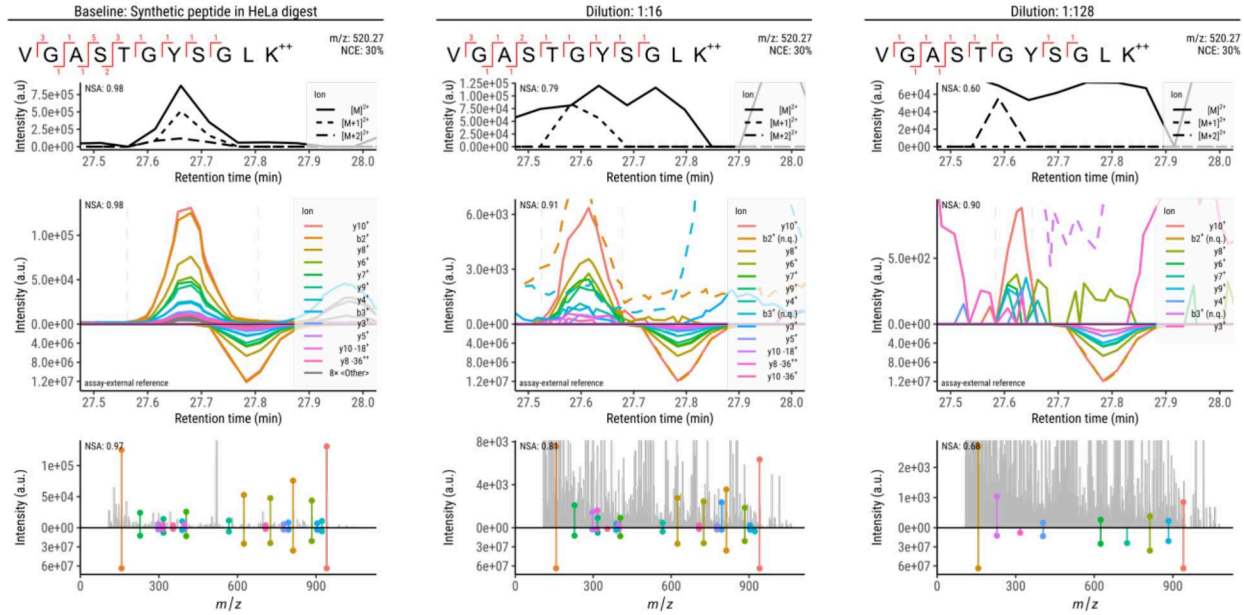


Figure 12. Peptide detection diagnostics. Synthetic peptide detection in a complex sample. Three different concentrations of the peptide in the same complex matrix are shown in the three columns. Diagnostics from top to bottom are: Precursor sequence and charge state with detected fragmentation sites marked in red. Precursor *m/z* and normalized collision energy (NCE) are annotated. The upper chromatogram shows precursor isotopic envelope acquired in MS1 and the normalized spectral contrast angle (NSA) quantifying the similarity to the theoretical distribution. The colored chromatogram is the MS2 extracted ion chromatogram. The extracted ions, including further ions <Other>, are the known fragmentation products according to the synthetic reference acquisition. The assay-external reference signal is mirrored below for comparison. The NSA quantifies the similarity of the fragmentation pattern to the reference. The spectrum at the bottom shows the best spectrum that was acquired, compared to the reference spectrum. Spectrum y-axes are zoomed to the most intense characteristic ion (here: b_2^+). The NSA quantifies the similarity of the identified spectrum signal to the reference.

I used a dilution of a known synthetic peptide in a constant amount of a complex peptide matrix (tryptic digest of HeLa cells), to determine which techniques are suitable to enable sensitive detection with high confidence. A representative example is shown to illustrate the effect of decreasing analyte concentration (Figure 12). The left panel represents a clear detection due to strong target signal. At the top, the sequence coverage achieved by the detection is marked along the precursor sequence. The red marks indicate the location of the peptide backbone breaks that were observed through the resulting y-, b- or a-type fragments. A red number specifies how many different fragments were observed for each mark. Good sequence coverage increases confidence that the observed signal does not stem from a peptide of the same or similar amino acid composition with a sequence permutation. Below, a full MS1 chromatogram was extracted for the precursor ion. In this case, the three dominant isotopes can be seen and appear with characteristic hierarchy. This is quantified with the normalized spectral contrast angle (NSA) comparing the observed peak areas to the theoretical isotopic distribution of the precursor (termed isotopic envelope). Below, in the MS2 extracted ion chromatogram a large set of different characteristic ions show. The relative intensities of the ions can similarly be quantified and scored via NSA to compare them to a known reference acquisition, in this case scoring near the optimum of 1. The reference acquisition is included as the bottom panel of the mirror plot. This configuration allows to gauge the similarity of the fragmentation to the reference and in addition shows a comparison of retention time and peak shape. These two additional factors, based on the LC separation, are not as precise as the MS-derived data but serve an important role in the determination of detection validity. In the bottom panel, the best acquired spectrum is shown. It allows for a clear visual correlation of the extracted signal to the reference. The dominant peaks of the acquired spectrum clearly match the reference, with the exception of the precursor window at ca. 520 *m/z* and some undetermined ions in the low *m/z* range.

The precursor window is susceptible to contamination from off-target ion species that co-isolate with the precursor but don't fragment due to a different chemical composition. Low *m/z* ions may or may not originate from the precursor of interest but are generally not as characteristic as ions that are composed of multiple amino acids. Due to this they are often noise-affected and are less informative about the precursor sequence.

The dilution series reveals several general trends. In the 1:16 dilution (Figure 12) of the analyte, I reduced the number of extracted ions, as lower ranked ions did not produce discernible signal. The MS1 signal does not produce a viable peak anymore. For the MS2 extracted ion chromatogram (XIC), extraction at 3 ppm tolerance doesn't entirely prevent remnant off-target signal that stems from the complex matrix from disturbing the characteristic peak. In this case, the b_2^+ ion was clearly eluting before and after the peak of interest and is therefore not quantitative. Manually marking ions clearly affected by this as non-quantitative (n.q., dashed) ions allows to exclude them from the NSA calculation and thereby to focus on the most informative signal for producing the NSA score. Surprisingly, the spectrum at this point is already dominated by off-target ions. Visual comparison of the spectra is only enabled by highlighting the ions of interest and there is context missing to understand why the b_2^+ ion is stronger than the reference suggests. This results in a degraded NSA score. In the 1:128 dilution (Figure 12), even fewer fragments are of use, negatively affecting the sequence coverage. Still, remnant signal in the MS2 XIC appears to be highly characteristic, with all 9 remaining fragments appearing in one of the spectra that comprise the peak. The spectrum is even less informative than before. This led me to conclude that the MS2 XIC along with the NSA score is by far the most informative evidence for peptide detection on the Orbitrap MS. Additionally keeping track of sequence coverage can help ensure that the extracted signal is characteristic of the peptide.

I found these observations to be highly relevant for the detection of the HPV16 peptides. Figure 13 shows that this dynamic for low intensity detections also manifests with HLA binders in immunopeptidomics experiments. The peptide E7/11–19 is clearly detectable in the cell line CaSki but is barely detectable in SNU-703. Peptides frequently presented first with very low level of evidence that could typically be improved upon through more focused acquisition and replication. To indicate successful detections consistently, I decided on a minimum threshold for the NSA score of 0.85, where scoring includes at least the top 5 reference ions larger than two amino acids in length.

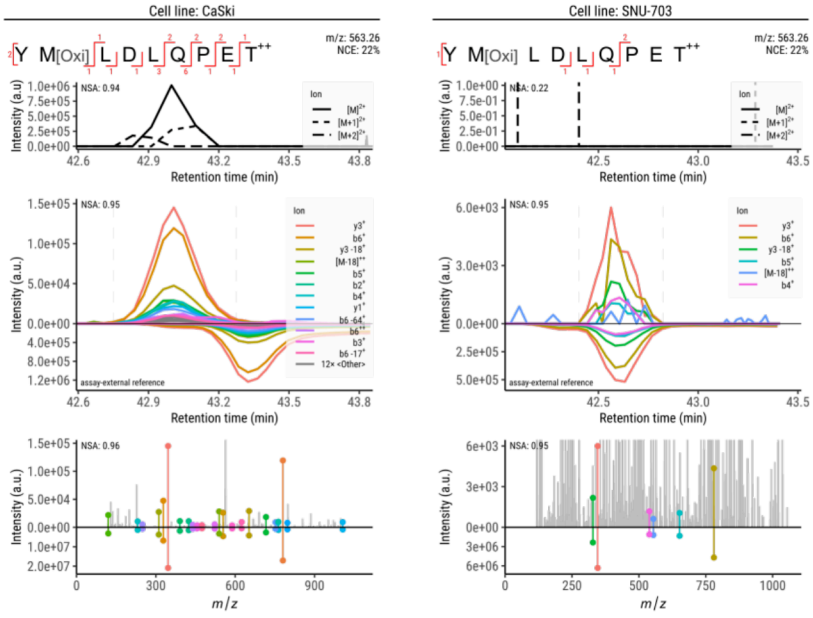


Figure 13. HLA-presented peptide identification at limit of detection. Detection diagnostics as explained before. The same peptide as detected in two different cell lines.

3.2.2 Retention time shifts

Many of the initial HPV16-derived peptide detections showed a considerable discrepancy on the retention time (RT) axis between the target signal and the separately acquired (assay-external) reference (Figure 14A). While some fluctuations are expected on low-flow nano-LC, the observed discrepancies on the order of minutes were problematic. Proving exact co-elution of the target with the reference is an important addition to the fragmentation data, as HLA binders tend to fragment in unbalanced fashion, allowing only for limited sequence coverage. A sequence permutation of the amino acids that are not covered typically produces a considerable retention time shift and thus exact co-elution can serve to alleviate remaining uncertainties due to limited sequence coverage.

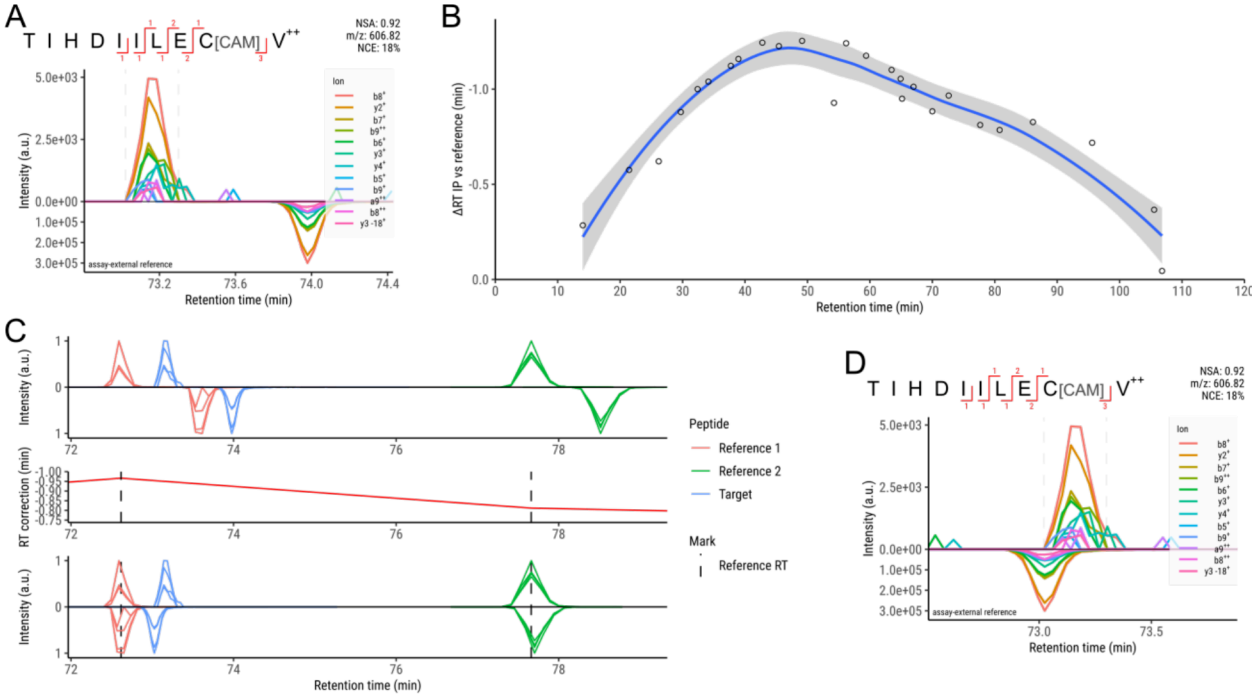


Figure 14. HLA IP retention time shifts and alignment using standard peptides. (A) MS2 XIC produced from an HLA IP with mirrored reference XIC. The reference XIC is shown with RT as-is. (B) Retention time shift of spiked-in reference peptides observed with an HLA IP, compared to an LC control acquisition immediately before. Negative delta indicates early elution in HLA IP. The x-axis is that of the HLA IP. The method for the blue trendline is locally weighted scatterplot smoothing (Loess) in R with formula y ~ x. (C) Scheme for RT correction of (A) using linear interpolation between adjacent reference peptides (red, green). Peak maximum intensities are normalized to 1. Above the x-axis are peaks from the HLA IP. Below x-axis are peaks from the reference acquisition. Top panel is the unaligned case. Bottom panel is the realigned case. (D) MS2 XIC from (A) with realigned mirrored reference XIC.

To better understand the issue, I used synthetic reference peptides that I spiked into the IP samples and into basic negative control samples that were run between the IPs. These synthetic peptides are unrelated to the candidate peptides and the sequences of the peptides do not occur in any natural protein. They are specifically designed to cover most of the retention time range (Zolg et al., 2017). Comparing retention times of these peptides between control and the experiment in Figure 14A revealed a systematic shift towards early elution mid-gradient (Figure 14B). This was a common occurrence in the epitome map experiments. I therefore used the reference peptides to correct the reference chromatogram retention times. The technique is shown in Figure 14C. The top panel shows the mismatch of the two LC-MS runs in Figure 14A, with the IP on top and reference run mirrored below the x-axis. Reference peptides adjacent to the target peptides indicate the local RT shift and all RT values of the reference are adjusted according to a linear interpolation between the two (middle

panel). In the bottom panel, the corrected result is shown. Figure 14D shows the improved alignment. Consequently, I used this technique for all targeted immunopeptidomics experiments where assay-external references are shown.

I observed RT shifts with different magnitudes between different IPs and noticed that it was likely related to the load on the liquid chromatography system. To determine whether this is intrinsic to the system and if there are negative effects, I ran a series of tryptic Hela digest acquisitions with different loads (in ng of peptide). A low-load reference acquisition with 50 ng and three further runs with up to 800 ng were acquired in sequence (Figure 15A, top to bottom). Comparing retention times for the 71,811 peptides detected in all samples shows the same trend that was observed for the IPs, where peptides elute early mid-gradient. However, the effect is confirmed to be transient, as the LC reset back to the initial condition for the 50 ng run after high-load 800 ng run (Figure 15A, very bottom). Notably, the shifts at any timepoint come with considerable variance between peptides. I observed that this even leads to elution-order inversions, which have been described in literature before (Gupta et al., 2019). These non-linear effects indicate that the approach I used for correction of retention time shifts is limited in power and retention time mismatches are still to be expected.

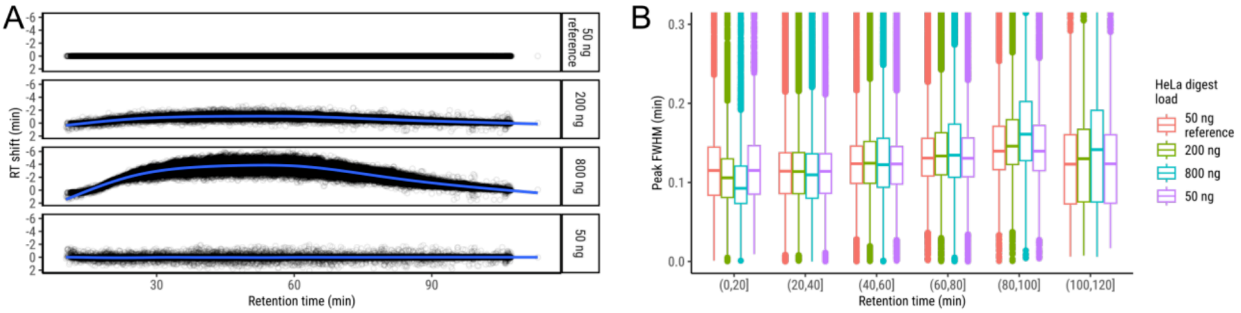


Figure 15. Systematic RT shifts as a function of peptide load on the LC system. (A) RT shift observed for peptides in untargeted acquisition of different amounts of tryptic HeLa digest. Only peptides detected in all runs are shown. The top run serves as the reference for the RT shift. Sequential runs from top to bottom. The method for the blue trendline is a generalized additive model (GAM) in R with formula y ~ s(x, bs = 'cs'). (B) Peak full width at half maximum (FWHM) observed in the runs in (A).

I also compared the full width at half maximum (FWHM) of the peptide peaks and found only small effects on the sharpness of the peaks (Figure 15B). I therefore determined that the retention time shifts are an intrinsic load-related property of the LC setup that I cannot fully correct for within the scope of this project.

3.2.3 MS acquisition scheduling

For scheduling of the target acquisition during the LC-MS run, the retention time shifts created an additional challenge. For each peptide, a reference retention time is known and the MS instrument can be scheduled to focus on this peptide over the expected duration of the elution. While the peak elution typically spans less than 20 seconds, uncertainties on the retention time axis necessitate much larger windows. Figure 16A shows how for discrete peptide retention times (red dashes), this leads to simultaneous targeting of multiple peptides. Depending on the RT window, the number of targets having to be cycled at any timepoint can increase considerably.

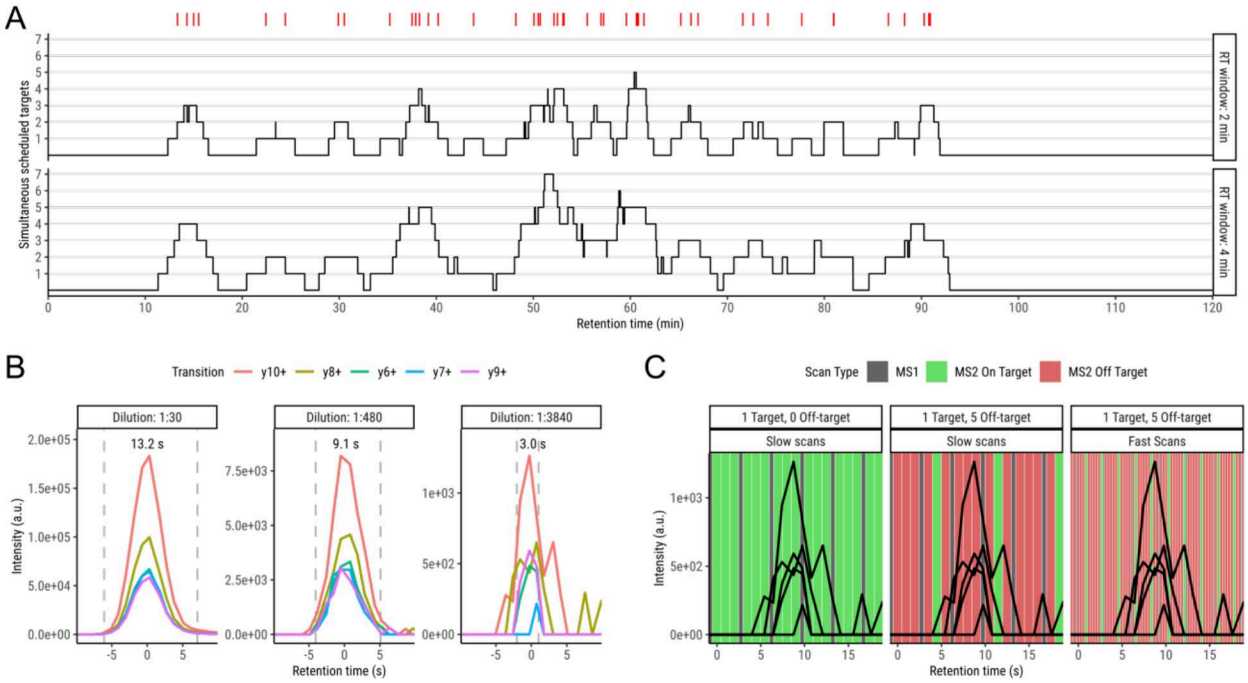


Figure 16. MS acquisition scheduling concerns. (A) The effect of the per-precursor retention time window size on the amount of simultaneously circled MS2 targets. Red marks indicate discrete target peptide expected retention times. (B) The effect of precursor intensity on the RT window for successful acquisition. Numbers indicate the width of the RT window between the dashed gray lines. (C) A low-intensity XIC overlays a hypothetical scan configuration with different amounts of simultaneously targeted precursors and scanning speeds.

This clashes with the objective to target all peptide with maximum achievable sensitivity. As shown for the detection of a synthetic peptide in Figure 16B, even when the typical peak width is on the order of tens of seconds, the timespan for successful detection of any peptide may be as low as 3 seconds when the intensity is low. As instrument settings allow for precursor accumulation for a long as 1000 ms per target, this can lead to excessive cycle times when cycling more than one target at a time. This is shown schematically in Figure 16C, where a simulated succession of MS1, on-target MS2 and off-target MS2 scans are overlaid with a low-intensity MS2 peak. When scheduling a single target, all MS2 scans are on-target and hit several times near the peak maximum. With 5 off-target scans between each on-target scan, it becomes likely to miss the peak maximum entirely. In such cases, faster scanning (less accumulation, lower sensitivity) was used in a trade-off to allow for better peak coverage.

To keep the simultaneous targeting low, I activated the dynamic retention time correction feature of the MS instrument (Gallien et al., 2014). This feature works with synthetic retention time standard peptides that are spiked into every IP and corrects schedule windows on-the-fly during the LC-MS run. The feature has limited precision and caused jumps in scheduling but allowed me to typically keep RT windows below 2.5 minutes.

I also made sure to only target, for each cell line, the appropriate peptides, based on matching the HLA allotypes of the cells to the HLA supertype association of the target peptides and on matching the peptide sequences to the exact HPV variant lineage present in each cell line (as previously determined in the group).

3.2.4 LC gradient

The choice of LC gradient (i.e. the gradual shift from low 1–2 % organic to 80 % organic solvent that enables the reversed-phase liquid chromatography) was also informed by the need to target few peptides simultaneously. The initial gradient was devised by Dr. Mogjib Salek for immunopeptidomics experiments (Figure 17A, OptiGrd3) and is a long gradient at 100 minutes method runtime, starting with 2 % organic solvent and quickly progressing to 5 % at 6 minutes. This is the gradient used in the initial part of the project. I developed a second gradient (Grd4) in collaboration with Dr. Jonas Becker to account for an observed clustering of peptide detections around the 10 minute mark in his untargeted experiments. I observed a similar clustering of target peptides for my targets (Figure 17B, marked *), when listing all targets at once. Peptides at this RT were not properly resolved, likely attributable to the quick progression to 5 % organic solvent of OptiGrd3 that is applied to the LC column 5–10 minutes earlier. The new 120 min gradient starts at 1% organic is flattened in the early part. The major clustering of early peptides was thereby resolved. Which gradient was used for any of the IPs is listed in Appendix Table 1.

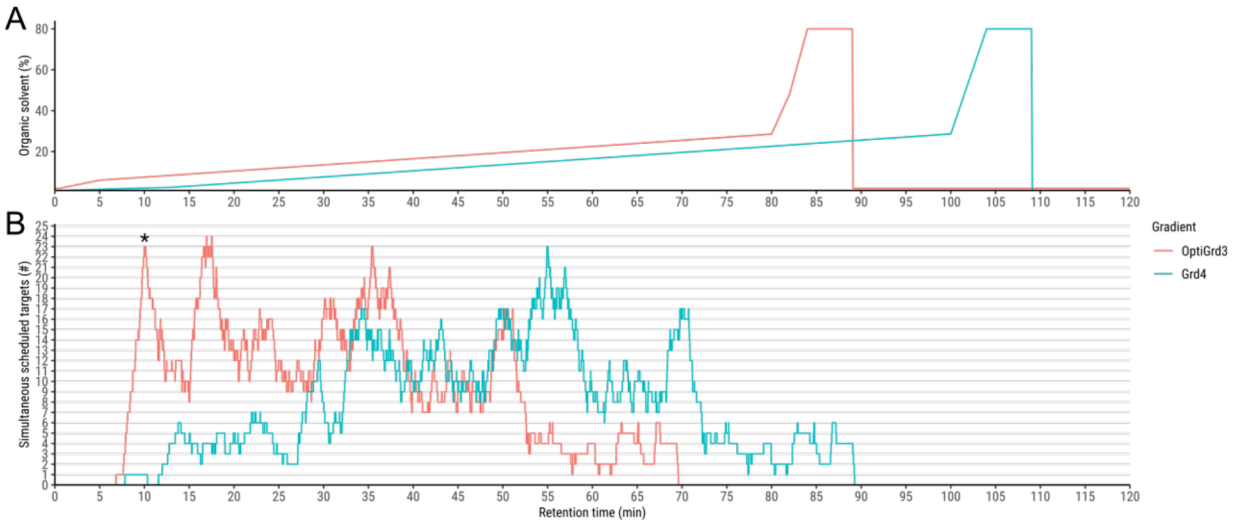


Figure 17. The two main LC gradients used in this thesis. (A) The LC gradient progression from low to high organic solvent percentage. (B) The distribution of all targeted peptides with a 2.5 min RT window. The black asterisk indicates a cluster of improperly resolved peptides.

3.2.5 MS scan parameter tuning

3.2.5.1 Rationale

The Orbitrap MS instrument has several key acquisition parameters, which may affect sensitivity of MS2 scans. The general trend for the instrument is that the more time is allocated for each scan, the better the sensitivity (Bourmaud et al., 2016). However, at the top end, this trend has not been validated and several caveats of over-tuning the parameters are described in literature. Details are provided in the following sections on the parameters Orbitrap resolution, maximum MS2 injection time and precursor isolation window.

Over the course of this thesis, I noticed that detections were often at the very limit of detection and that I would have to perform replicate experiments where the focus is on validation of as few as one peptide. This would allow me to spend the full instrument time on a single precursor. The particular challenges with any detection can be multifaceted, due to issues such as interfering off-target ions or limited intensity. For this reason, I set up an experiment with 24 target precursors, so that general trends can be observed. I selected a subset of peptides from the synthetic PROCAL peptide mix, which is designed to be distributed evenly across the retention time range (Zolg et al., 2017). Figure 18 shows all these peptide and the non-overlapping retention time windows. I created high quality spectral libraries, allowing me to compare clearly how many MS2 fragmentation product ions (transitions) are detected and what the NSA is. In three test series, I tested detection sensitivity for each of these peptides with different parameter settings by successively diluting them in a complex background mixture (HeLa tryptic digest). While this background is likely still less complex than an immunopeptidomics sample, it provides an abundance of coeluting and interfering ions that make analyte detection challenging. The evaluation of each parameters' effect was then performed in similar fashion and is described in the following sections.

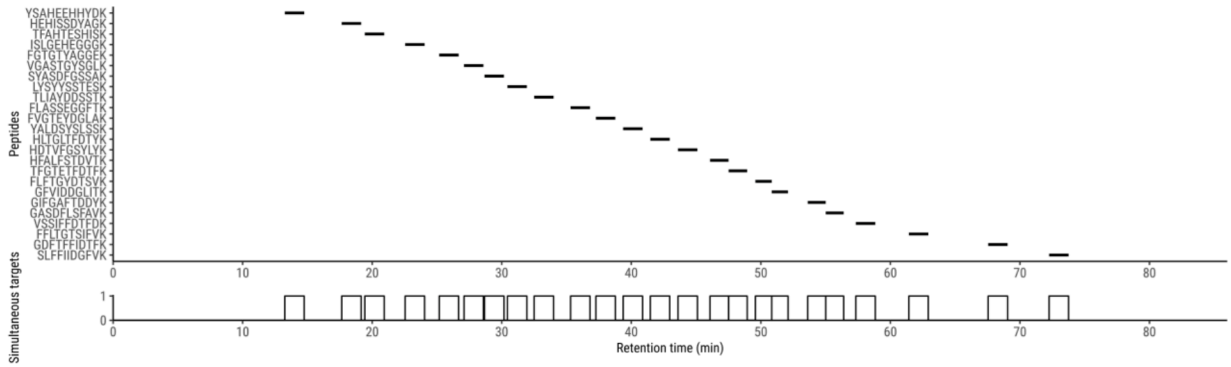


Figure 18. Retention time windows of selected synthetic peptides of the JPTRP standard. The selected peptides are listed and retention time windows of up to two minutes are shown. All windows are non-overlapping.

3.2.5.2 Orbitrap resolution

The main effect of the resolution parameter is described as the increase in resolving power on the *m/z* axis. This is generally desirable, as it allows to distinguish ions with very similar *m/z* values and also improves the accuracy of the determined *m/z* value. The parameter affects how long the ion signals are acquired in the Orbitrap detector. There is a clear positive relationship between transient time and sensitivity, driven by a linear increase in background noise versus a quadratic increase in ion signal. In contrast, potential issues with long transient times are described, where similar ions may coalesce into one population or where ion packets decay due to collisions in gas phase, which can degrade *m/z* and intensity readouts (Hecht et al., 2019).

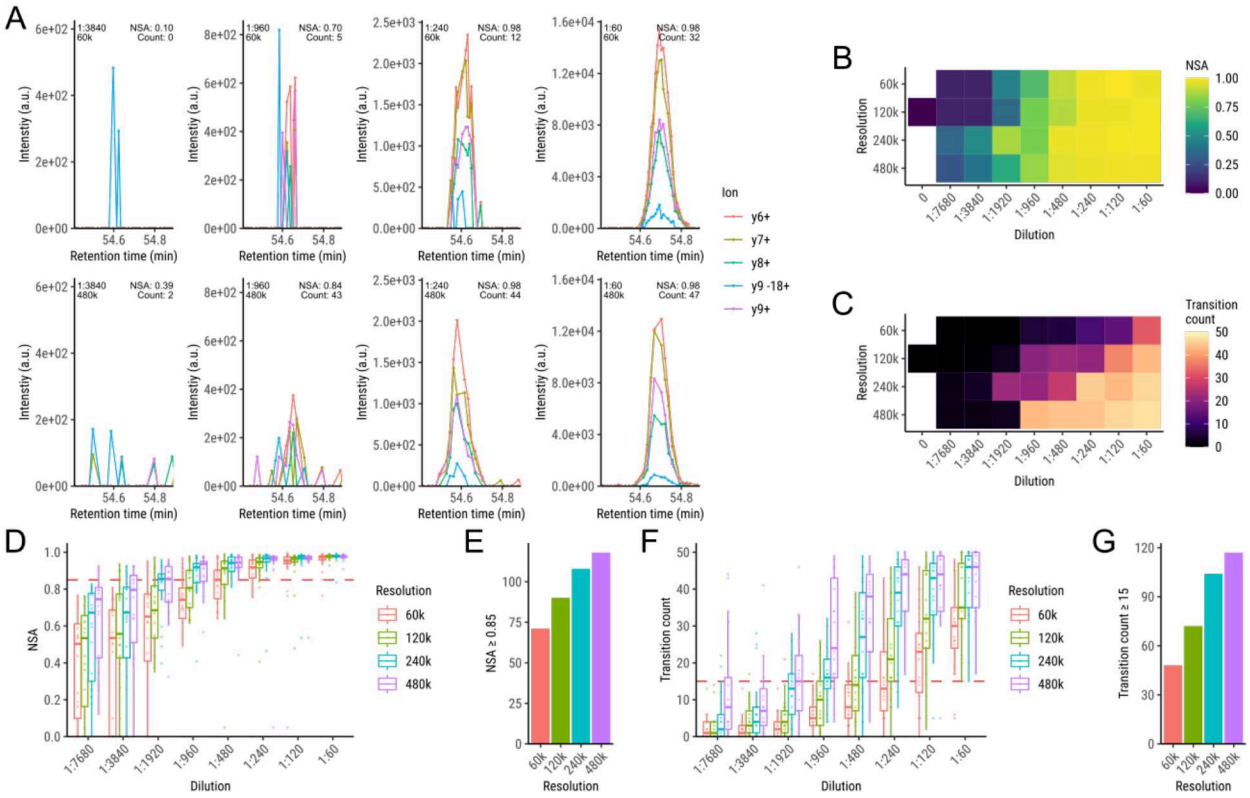


Figure 19. Test series on the effect of MS resolution on detection sensitivity. The synthetic peptide kit is diluted against a consistent complex matrix of tryptic HeLa digest. (A) Exemplary XIC for the precursor GASDFLSFAVK^{++}. Annotated at the top left of each panel is the dilution and MS resolution used. Annotated at the top right are the NSA and total count of detected transitions. (B) Full panel of NSA results for the precursor in (A). (C) Full panel of transition counts for the precursor in (A). (D) The distribution of NSA that were measured for all precursors at all dilutions and resolutions. The dashed red line indicated the threshold of NSA ≥ 0.85. (E) Summary of all events in (D) that passed the threshold. (F) The transition counts that were measured for all precursors at all dilutions and resolutions. The dashed red line indicates the threshold of ≥ 15. (G) Summary of all events in (F) that passed the threshold.

In this series of LC-MS runs, I tested four different resolutions, 60k, 120k, 240k and 480k, which is the maximum possible setting. I tested exponential dilutions of 1:60 to 1:8000 of the peptide mix and one control with no peptides. The wide range of dilutions is owed to fact that the different precursors show intensities spanning several orders of magnitude, which is common in mass spectrometry and especially in immunopeptidomics. The series was started with more diluted peptides, progressing towards less diluted peptides. In Figure 19A, XICs are shown for an example precursor GASDFLSFAVK^{++} for resolutions 60k and 480k and at dilutions 1:4000, 1:960, 1:240 and 1:60. The NSA comparing the five transitions shown to the reference transitions is annotated. Also annotated is the total count of reference transitions observed within the peak boundaries (data not shown). The count allows for some missed reference transitions, enforcing a 90% hit rate. At a dilution of 1:4000, no valid signal is present and it appears that with 480k resolution, more undesirable off-target signal is produced. At 1:960 dilution, the signal with 60k resolution does not reflect a typical peak shape and is limited by a low NSA of 0.70 and a transition count of 5. With 480k resolution, a peak shape can be recognized and the NSA reaches 0.84 with an extremely high transition count of 43. At lower dilutions, detections become immediately apparent. Figure 19B and C show the results for this precursor in terms of NSA and transition count for all conditions. For the NSA, a weak trend of better results with higher resolutions can be seen. For the transition count, a very strong trend is apparent with higher resolutions achieving higher counts. Figure 19D shows the NSA metric for all viable

precursors in the experiment in all conditions. The dashed line indicates the same minimum NSA threshold of 0.85 for a valid detection that is used throughout this thesis. Peptides cross the threshold at different dilutions for several reasons: The concentrations of the synthetic peptides are not absolutely determined, the ionization efficiencies may be drastically different, and most importantly, the varying effectiveness of the parameter adjustment. The count of all detections that passed the threshold, shows the best result with 480k resolution (Figure 19E). A similar evaluation is performed for the transition count, where the threshold was set to 15 transitions (Figure 19F–G). Here, 480k resolution also performed best. Due to this, high resolutions were set whenever the resulting instrument acquisition cycle time was not excessive.

3.2.5.3 Maximum MS2 injection time

The maximum MS2 injection time (max. IT) determines how long precursor ions are accumulated before fragmentation and subsequent acquisition in the Orbitrap. In principle, the response should be linear, such that accumulation for twice as long allows for twice as much sensitivity, up to the maximum allowed value of 1000 ms. The parameter is specified as a maximum limit because the instrument limits injection already before the time has elapsed in case too many ions accumulate. The drawback is a lower frequency of scans which allows for fewer samples (Bourmaud et al., 2016).

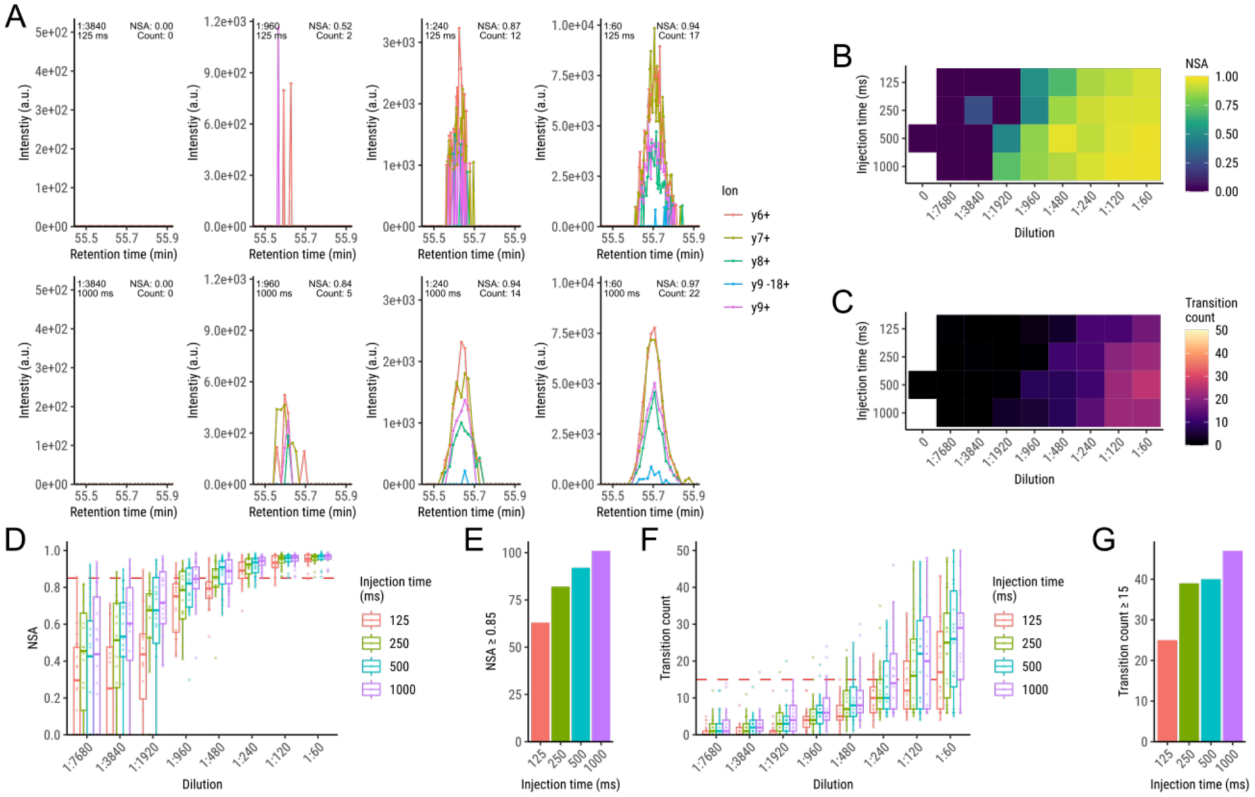


Figure 20. Test series on the effect of maximum MS2 injection time on detection sensitivity. The synthetic peptide kit is diluted against a consistent complex matrix of tryptic HeLa digest. (A) Exemplary XIC for the precursor GASDFLSFAVK^{++}. Annotated at the top left of each panel is the dilution and max. IT used. Annotated at the top right are the NSA and total count of detected transitions. (B) Full panel of NSA results for the precursor in (A). (C) Full panel of transition counts for the precursor in (A). (D) The distribution of NSA that were measured for all precursors at all dilutions and max. ITs. The dashed red line indicated the threshold of NSA ≥ 0.85. (E) Summary of all events in (D) that passed the threshold. (F) The transition counts that were measured for all precursors at all dilutions and max. ITs. The dashed red line indicates the threshold of ≥ 15. (G) Summary of all events in (F) that passed the threshold.

In this series of LC-MS runs, I tested four different max. ITs, 125 ms, 250 ms, 500 ms and 1000 ms, which is the maximum possible setting. I tested exponential dilutions of 1:60 to 1:8000 of the peptide mix and one control with no peptides. As above, the series was started with more diluted peptides, progressing towards less diluted peptides. In Figure 20A, XICs are shown for an example precursor GASDFLSFAVK^{++} for injection times 125 ms and 1000 ms and at dilutions 1:4000, 1:960, 1:240 and 1:60. As above, the NSA comparing the five transitions shown to the reference transitions is annotated. Also annotated is the total count of reference transitions observed within the peak boundaries (data not shown). The count allows for some missed reference transitions, enforcing a 90% hit rate. At 1:4000, no valid signal is present. At 1:960, only the 1000 ms injection time example shows appreciable signal, albeit still below the threshold. Detections at 1:240 and above show a clear difference in quality. Figure 20B and C show the results for this precursor in terms of NSA and transition count for all conditions. For the NSA, a trend of better results with longer injection times is appreciable. For the transition count, the effect does not seem strong with 500 ms injection time performing best. Figure 20D shows the NSA metric for all viable precursors in the experiment in all conditions. The dashed line indicates the same minimum NSA threshold of 0.85 for a valid detection that is used throughout this study. As above, peptides cross the threshold at different dilutions for several reasons: The concentrations of the synthetic peptides are not absolutely determined, the ionization efficiencies may be drastically different, and most importantly, the varying effectiveness of the parameter adjustment. The count of all detections that passed the threshold, shows the best result with 1000 ms injection time (Figure 20E). A similar evaluation is performed for the transition count, where the threshold was set to 15 transitions (Figure 20F–G). Here, 1000 ms injection time also performed best. Due to this, a maximum injection time of 1000 ms was used for MS2 scans.

3.2.5.4 Precursor isolation window

The precursor isolation window determines the *m/z* range that the quadrupole permits in the precursor accumulation step. The adjustment of the parameter comes with concrete tradeoffs. Narrow isolation windows are typically used to exclude interfering off-target ions that are commonly present in close proximity to the target precursor in highly complex samples (Gallien et al., 2014). However, this comes at a cost as the permissive window is typically not perfectly sharp. When isolation is very narrow, this can cause loss even in the center of the isolation window. Furthermore, peptides have several isotopes beyond the monoisotopic peak available. Non-monoisotopic precursor isotopes can produce a fraction of monoisotopic transitions and thereby contribute signal, which for small peptides is predominantly monoisotopic. To test the effect of the isolation window parameter, I tested values 5 *m/z*, 2.5 *m/z*, 1.8 *m/z* and a narrow *m/z* setting that implements the lower limit recommendations of the instrument vendor which depend on the central *m/z* (e.g.: 0.7 *m/z* window for ≤ 700 *m/z* precursor, see methods). As the effectiveness of narrow precursor isolation is likely highly dependent on the presence of interfering ions, optimally a test would be performed with highly complex immunopeptidomes. A tryptic HeLa matrix is used to approach the complexity of an immunopeptidome but will likely not elicit the full benefit of narrow isolation due to fewer off-target ion species. Nevertheless, the test is crucial to confirm whether very narrow precursor isolation may hamper sensitivity.

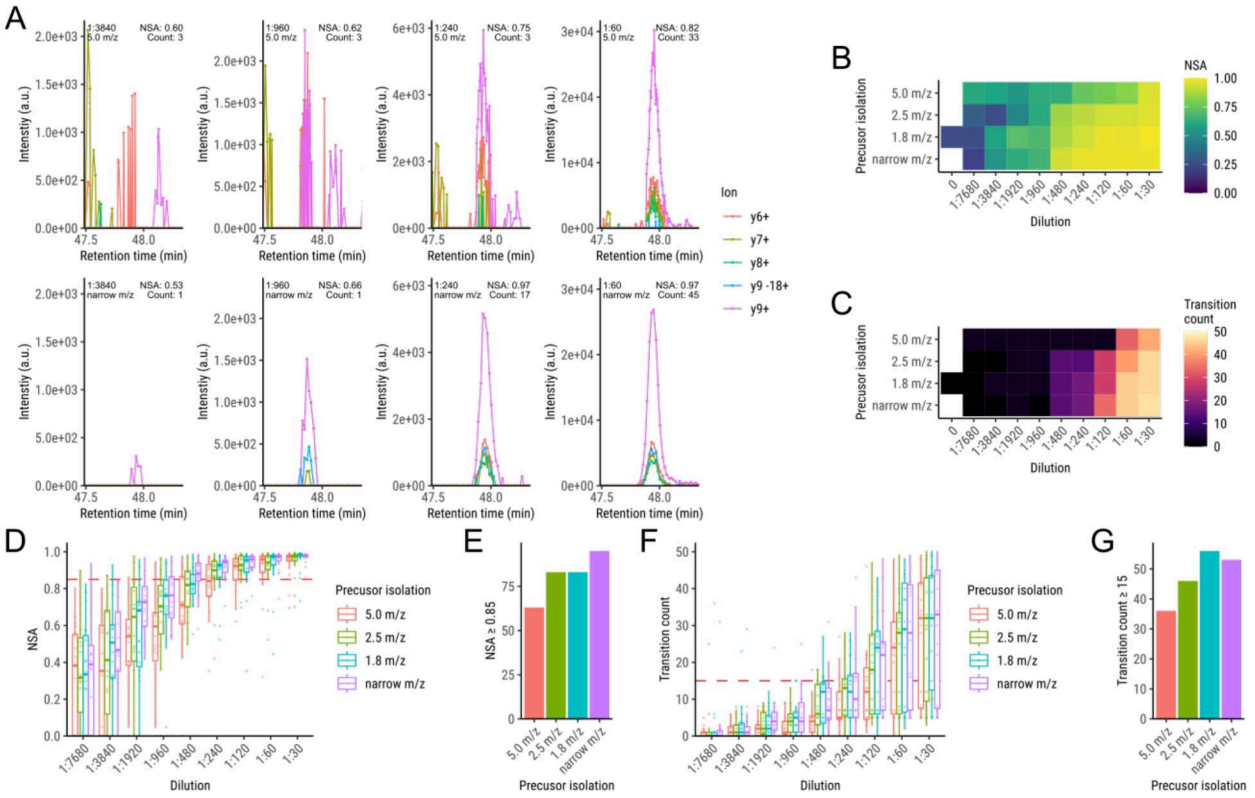


Figure 21. Test series on the effect of the precursor isolation window on detection sensitivity. The synthetic peptide kit is diluted against a consistent complex matrix of tryptic HeLa digest. (A) Exemplary XIC for the precursor TFGTETFDTFK^{++}. Annotated at the top left of each panel is the dilution and isolation window used. Annotated at the top right are the NSA and total count of detected transitions. (B) Full panel of NSA results for the precursor in (A). (C) Full panel of transition counts for the precursor in (A). (D) The distribution of NSA that were measured for all precursors at all dilutions and isolation windows. The dashed red line indicated the threshold of NSA ≥ 0.85. (E) Summary of all events in (D) that passed the threshold. (F) The transition counts that were measured for all precursors at all dilutions and isolation windows. The dashed red line indicates the threshold of ≥ 15. (G) Summary of all events in (F) that passed the threshold.

In this series of LC-MS runs, I tested four different precursor isolation windows in 5.0 *m/z*, 2.5 *m/z*, 1.8 *m/z* and the narrow *m/z*, which is the minimum recommended setting. I tested exponential dilutions of 1:30 to 1:8000 of the peptide mix and one control with no peptides. As above, the series was started with more diluted peptides, progressing towards less diluted peptides. In Figure 21A, XICs are shown for an example precursor TFGTETFDTFK^{++} for precursor windows 5.0 *m/z* and narrow *m/z* and at dilutions 1:4000, 1:960, 1:240 and 1:60. As above, the NSA comparing the five transitions shown to the reference transitions is annotated. Also annotated is the total count of reference transitions observed within the peak boundaries (data not shown). The count allows for some missed reference transitions, enforcing a 90% hit rate. At 1:4000, off-target signal is seen with 5.0 *m/z* isolation vs. a trace of the dominant ion for narrow *m/z* isolation. At 1:960, the narrow *m/z* isolation shows a distinct peak shape but already the second most intense ion is missing. At 1:240, the peptide is clearly detected using narrow *m/z* isolation whereas 5.0 *m/z* isolation still doesn't achieve a high NSA at 1:60 dilution. Figure 21B and C show the results for this precursor in terms of NSA and transition count for all conditions. For both the NSA and transition count, a trend of better results with narrower precursor isolation is appreciable with narrow *m/z* isolation performing best. Figure 21D shows the NSA metric for all viable precursors in the experiment in all conditions. The dashed line indicates the same minimum NSA threshold of 0.85 for a valid detection that is used throughout this study. As above, peptides cross the threshold at different dilutions for several

reasons: The concentrations of the synthetic peptides are not absolutely determined, the ionization efficiencies may be drastically different, and most importantly, the varying effectiveness of the parameter adjustment. The count of all detections that passed the threshold, shows the best result with narrow *m/z* precursor isolation (Figure 21E). A similar evaluation is performed for the transition count, where the threshold was set to 15 transitions (Figure 21F–G). Here, superior performance is achieved for 1.8 *m/z* isolation. These results confirmed very narrow precursor isolation windows as a viable strategy, which benefits peptide detection in a complex background matrix. The dynamic narrow *m/z* isolation window was therefore used for targeted immunopeptidomics.

3.2.6 Collision energy optimization

The collision energy that is applied for the fragmentation of the precursor ion is a further parameter that can be modified on the mass spectrometer. This setting differs in nature from the other parameters as the effect is highly specific to each precursor. Collision energy optimization on a per transition basis, where transition refers to an MS2 product ion, has long been a common procedure with single reaction monitoring (SRM) targeted mass spectrometry methods (Lange et al., 2008; MacLean et al., 2010a; Sherwood et al., 2011). With these methods, each precursor to fragment transition through collisional dissociation has to be acquired individually and as such benefits from a specific optimal collision energy. The parallel reaction monitoring (PRM) method of the Orbitrap mass spectrometer used here, collects all fragmentation products in one scan and as such does not allow for the same flexibility in optimization of collision energies. Furthermore, the instrument vendor has implemented a collision energy normalization feature that adjusts collision energies based on precursor *m/z* and ion charge state with a basic formula (Thermo, 2012). Collision energies derived like this can be tuned as a percentage setting, termed normalized collision energy (NCE). As this method works well in general, PRM is commonly performed without any targeted optimization of the parameter (Révész et al., 2023; Gallien & Domon, 2015).

For the epitome map, I decided to determine optimal collision energies for all targeted precursors. This would also potentially provide insight into whether this is a worthwhile step to include in future studies. To facilitate this, my colleague Dr. Mogjib Salek set up a direct infusion (DI) setup, which delivers synthetic peptides to electrospray ionization without prior LC separation. This allows to acquire MS2 data for any precursor in a mixture of typically 10 to 100 peptides under consistent conditions for the extended timespans needed to scan through a range of collision energies and with high resolution scans. I found an NCE range of 2–42 % covers the full range of interest for peptide fragmentation. I then developed comprehensive methods for acquisition and evaluation of these data as no tools for this purpose were available. For extraction of ion traces, I generated a list of all possible precursor ions, y, b and a fragments, and further allowed for up to two neutral losses. Figure 22A shows an overview of the data that is generated for each precursor, in this case for the example of DLQPETTDL[+7]Y++, for which optimization was highly effective. Error bars indicating standard deviation are available due to 5x acquisition at each NCE percentage. The data shows an expected trend where very low collision energy causes almost no fragmentation such that the intact precursor ([M+2H]2+) presents the most intense signal. Progressing towards higher energies, a peak of especially informative transitions with a length of ≥ 3 amino acids appears, spanning the range from ca. 10–20 % NCE. Smaller less informative transitions of less than 3 amino acids in length that start to appear at a similar point seem to be favored by higher collision energies. To accommodate the previously determined focus on achieving high NSA in peptide detections, I decided to select an optimum NCE such that the intensity at 5 transitions into the spectrum is maximized (Figure 22B). This means that NCE values producing heavily imbalanced spectra with only few very intense fragments may be disregarded in favor of more balanced spectra that provide better signal for more

transitions. In the example, this optimum was determined at 14 % (green mark) as opposed to the common default NCE of 30 %.

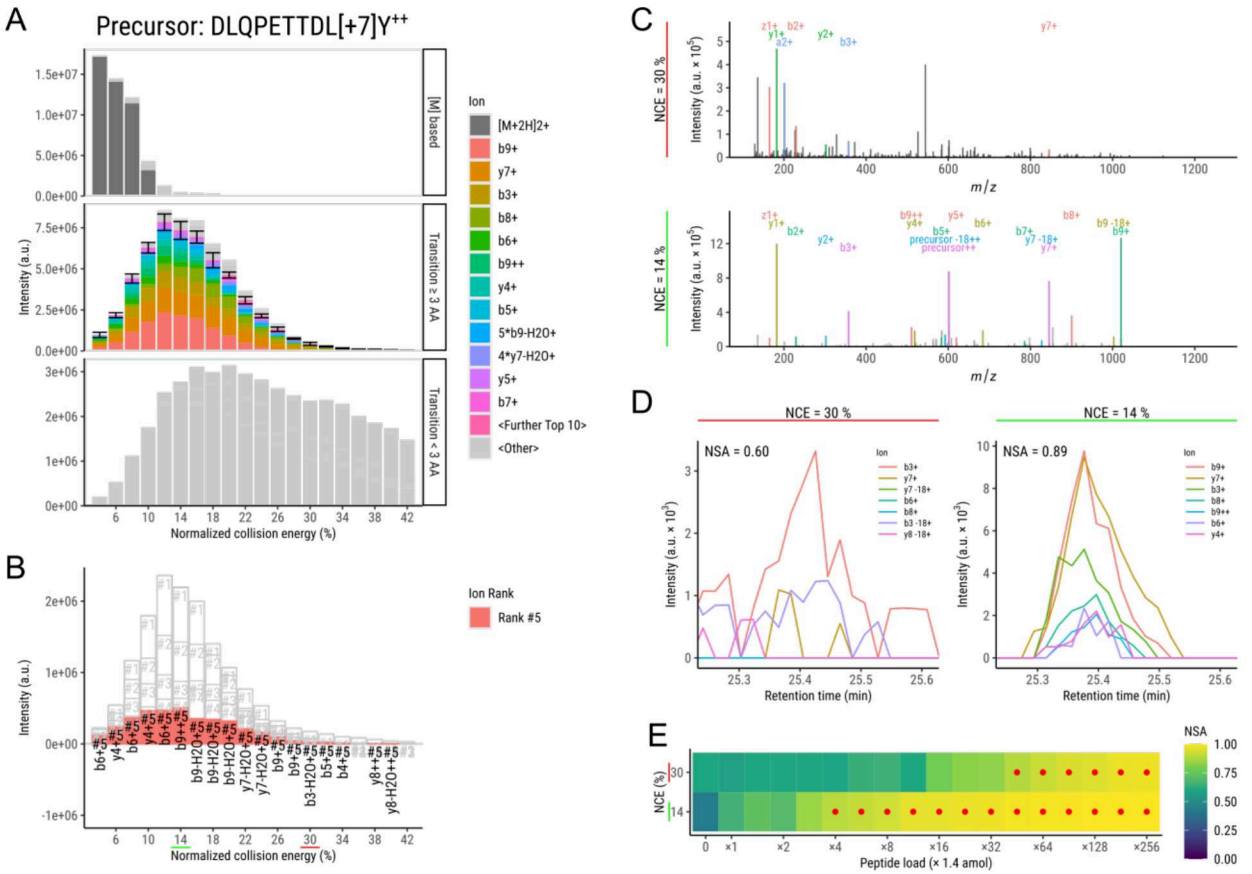


Figure 22. Collision energy optimization for target peptides. Stacked quantification of MS2 signal acquired via DI-MS for peptide DLQPETTDL[+7]Y at precursor charge state +2 over a range of normalized collision energies (NCEs). For each NCE, five spectra were acquired and mean ion intensities are shown. The main focus is on highly specific fragments of length ≥ 3 AA and the most dominant fragments are colored. Ions of lower specificity are separated as precursor ([M])-based and of length < 3 AA. Confidence intervals are the standard deviation of the sum of the top 10 ions. (B) Overlaid (not stacked) mean intensity of the top 5 ions at each NCE. The intensity of the rank #5 ion was used to pick the optimum. The default NCE of 30 % is marked in red and the optimized value for the precursor of 14 % is marked in green. (C) MS2 spectra at default and optimized NCE for the precursor in (A) acquired in an LC-MS PRM run. Both spectra were acquired close to the peak maximum (< 1 s apart). All ions that could be matched with 7 ppm tolerance are annotated. (D) Extracted fragment ion chromatograms from a single LC-MS run alternating between the two NCE settings. 5.5 amol of the peptide in (A) was loaded with 50 ng HeLa tryptic digest to produce a signal close to the limit of detection. Transitions of length ≥ 3 AA are shown. (E) NSA results comparing acquired peaks to a reference library in a dilution series down to 1.4 amol of the peptide in (A). A background matrix of 50 ng HeLa tryptic digest was used for all runs. High confidence detections with an NSA ≥ 0.85 are marked with a red dot. Adapted from (Salek & Förster et al., 2023). Authored by me.

To confirm whether this strong effect is true with the more specific signal available in a full LC-MS run, I performed a targeted experiment with high amounts of synthetic peptide. In this run, scanning was alternated between the two relevant NCE values. The best spectrum for each value is shown in Figure 22C. At the 30% NCE value only a fraction of the ions present were annotated, indicating that fragmentation may have produced less common products such as internal fragments that I did not target. The main observation of consequence though, is that the overall intensity of the spectrum is severely diminished (note y-axis scale) and a significant fraction are clearly less informative small ions. The benefit of the optimization also materializes with XICs produced from injection of a low

amount of the peptides (ca. 5.52 amol) in a matrix of HeLa digest (50 ng) (Figure 22D). In Figure 22E, the full titration series of the synthetic peptide is shown and based on the 0.85 NSA threshold for detection, an approximately 11-fold higher concentration of the peptide was needed for successful detection without the optimization (14% NCE: ×4; 30% NCE: ×45.25).

With the > 1000 precursors acquired this way, I observed various trends. Figure 23A shows a precursor where very high collision energies were required for productive fragmentation. Figure 23B shows a precursor where the spectrum is dominated by two extremely intense ions (y_7^{++}, y_8^{++}) around 20 % NCE. A more balanced spectrum can be achieved with 28 % NCE. Therefore, the selected NCE percentage should reflect the desired properties of the signal. Few dominant transitions may enable improved quantification while a more balanced top 5 ions allow for more effective NSA scoring.

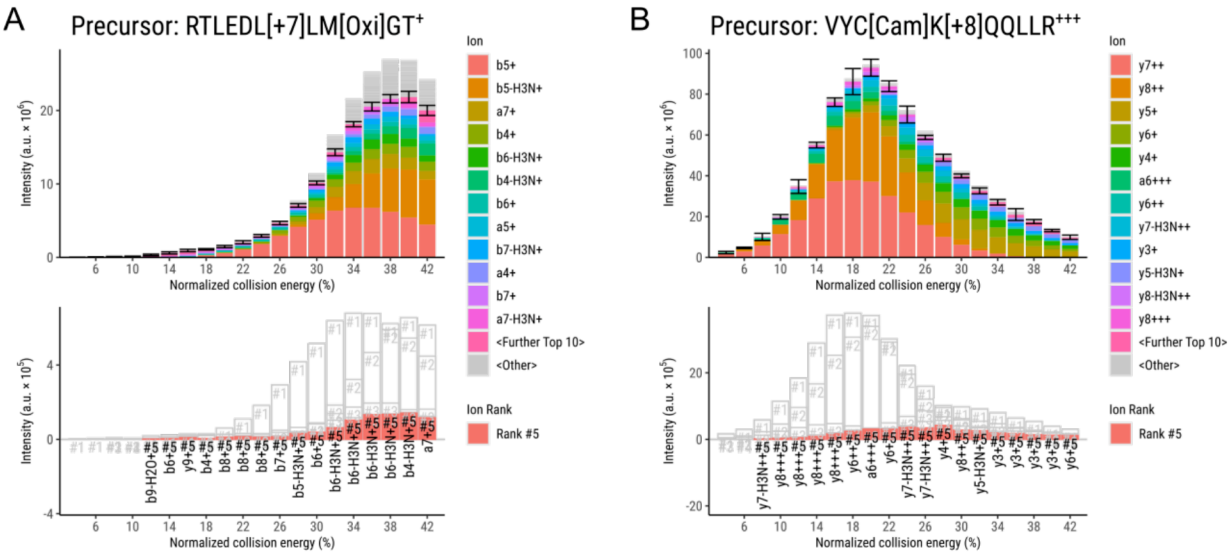


Figure 23. Further examples of NCE optimization. NCE data of the main ions of interest (fragments ≥ 3 AA in length) and the matching top intensity quantification as presented before.

As the optimization was performed with all peptides targeted in this thesis and also further sets of synthetic peptides that I acquired for related studies, I was able to conduct an evaluation of the benefit of the optimization step with a set of 1343 precursors from 796 peptides (Figure 24). For each precursor, I determined the optimal collision energy according to the criteria stated earlier and compared it to the default value of 30 % (Figure 24A). For comparison, the intensity of the fifth most intense transition (≥ 3 AA) was used. With five datapoints each available for the default and optimized NCE, I calculated the benefit with an unpaired one-sided t-test. Figure 24B presents the results of this comparison for all relevant precursors. The evaluation indicates a benefit of at least 50 % (red vertical dashed line) at a p-value of more than 0.5 (red horizontal dashed line) for 38.8 % of peptides that I classified as being of tryptic character (C-terminal AA: R or K) and for 63.1 % of all other putative HLA binders. I therefore concluded that the optimization provides a general benefit in signal and is especially effective for immunopeptidomics. The criteria used for the selection of the optimum evolved over the course of this project and due to this the collision energies used for the acquisition of target peptides changed in some instances. To account for this, I annotated the exact collision energy used with each XIC that displays an epitope detection.

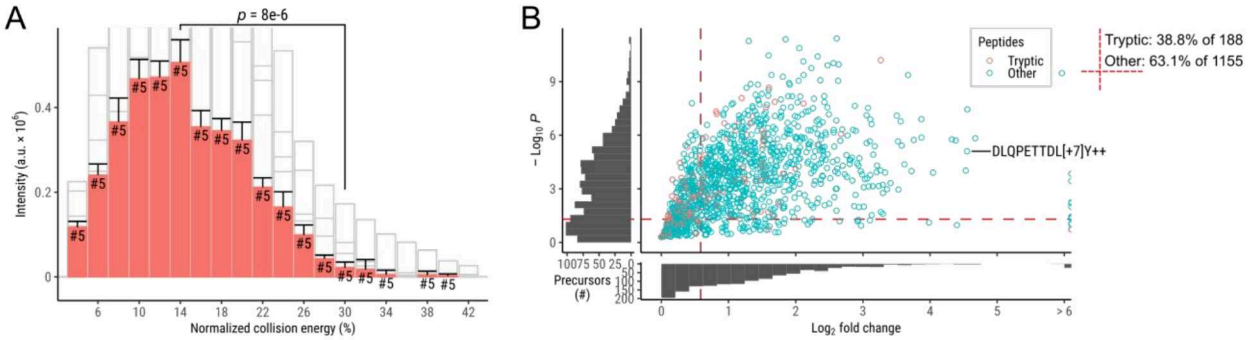


Figure 24. General evaluation of the NCE optimization benefit. (A) The rank #5 ion intensity quantification for Figure 22B. The rank #5 ion is shown in red. Further top ions 1–4 are drawn in the background. Confidence intervals are standard deviation based on 5 datapoints per NCE. The comparison an unpaired one-sided Welch's t-test with 4.5 degrees of freedom. (B) Optimization results for all relevant precursors. Compared is the signal at a standard NCE of 30% to the optimal value for each precursor. Histograms for the x- and y-axis show the overall distribution of points (precursors) in the plot. Precursors in the top-right corner, delimited with dashed red lines, show a signal gain ≥ 50% at a p-value < 0.05. The data point corresponding to the peptide analyzed in Figure 22 is indicated. The percentage of precursors appearing in the top-right section and an overall count are indicated to the right of the panel. Panel B adapted from (Salek & Förster et al., 2023). Authored by me.

3.2.7 Stable isotope labelling concerns

Stable isotope-labeled (SIL) synthetic peptides were essential for this project as they allowed to work with the target peptides at high concentration during establishment of the method. SIL refers to the incorporation of stable heavy isotopes, such as carbon-13 (13C) or nitrogen-15 (15N) (Gevaert et al., 2008). For synthetic peptides, this was accomplished via synthesis with one or more heavy amino acids, where ^{12}C and ^{14}N are completely replaced. For example, a mass difference of +8 Da results from labelling of lysine ($C_6H_{14}N_2O_2$) due to the replacement of 6 ^{12}C and 2 ^{14}N. The MS instrument can then exploit the physical difference in mass to cleanly separate the two species for detection. Without the isotopic labels, trace-level cross contamination is essentially unavoidable in a regular LC-MS sample delivery setting, where the same needle, capillaries and column are used to pick up sample and achieve separation. SIL peptides are also commonly spiked directly into experiments to produce an assay-internal "heavy" reference signal, to which the signal acquired for the "light" cognate can be compared. This is because 13C 15N SIL peptides have identical chromatographic properties to the unlabeled variants and elute at the same retention time (Brun et al., 2007).

In working with Dr. Mogjib Salek on the establishment of the spike-in method, we observed spurious detection of light peptide with synthetic SIL peptide mixtures. Figure 25 shows how for a negative control acquisition, where no signal in the light precursor channel was expected, several peptides showed specific signal for the light precursor. We found this to be a systematic issue with commercially available SIL labeled peptides and we published diagnostics and conclusions on the source of this phenomenon in a shared first authorship publication (Salek & Förster et al., 2022).

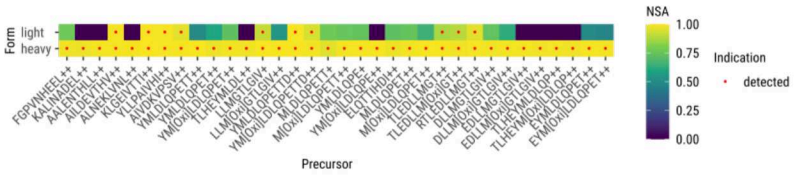


Figure 25. Light contamination with synthetic SIL peptides. Negative control acquisition of unlabeled 'light' precursors in an experiment with spiked-in 'heavy' synthetic SIL peptides shows contaminations indicated by red dots (NSA ≥ 0.85) present for the light forms.

Due to this, I concluded that the spike-in of SIL peptides for reference was impractical for the epitome map, as it would necessitate careful titration of the hundreds of target peptides down safe levels. This is especially challenging as the peptides' intensities often responded non-linearly to the titration. In addition, acquisition of assay-internal reference consumes crucial instrument time. In contrast, I had implemented high quality comparison of target signal with assay-external reference peaks as shown in sections 3.2.1–3.2.2.

Upon finalizing an initial round of the epitome map IPs, I found that there are several limitations inherent to low-intensity detections on the LC-MS system at hand. First, HLA binders often fragment in a way that does not allow for major sequence coverage at low intensities. Second, considerable separation power of the nano-LC system was not being exploited to its full potential, as the alignment of assay-external reference chromatograms is fuzzy. Even though I had included a small set of negative control peptides with supertype A02 IPs, which were never detected, the overall low frequency of peptide detections with hundreds of targets meant that this did not provide a firm indication on whether false discovery is probable with the detection criteria I defined in section 3.2.1. Therefore, I decided to validate each peptide with an additional SIL spike-in experiment which would provide an as yet unattainable binary outcome of whether the target coelutes exactly with the reference or not.

To this end, I titrated the limited set of target peptides in mixtures at levels where each peptide was detectable but did not produce signal for the light precursor (Figure 26A). Spiking this mixture into IP eluates allowed for validation of exact coelution of the target peak with the assay-internal reference (Figure 26B). The same spike-in was performed for the HPV16 negative control cell line and showed no specific signal (Figure 26C). This control was performed for each SIL spike-in experiment (labeled: "+ SIL").

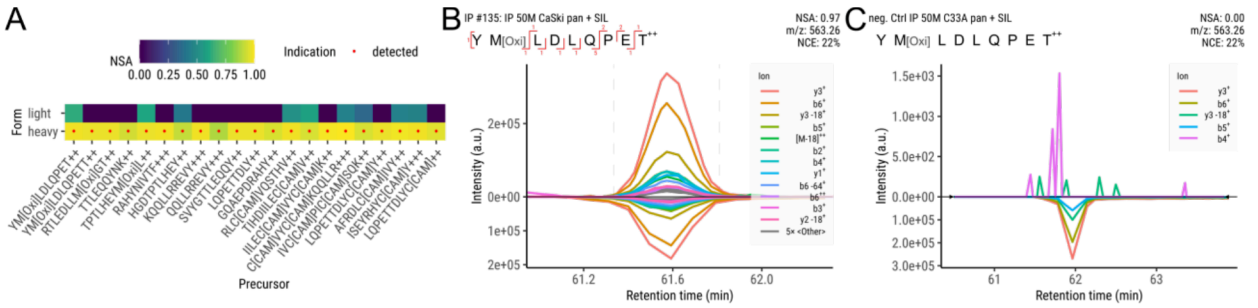


Figure 26. Calibration of SIL spike-in for detected peptides. (A) Titration of each peptide separately before injection as a mixture allowed for acquisition of all candidate assay-internal heavy references without detectable light contamination. (B–C) Results of SIL spike-in into HLA IPs. (B) The assay-internal reference XIC is mirrored below the x-axis without the RT offsets common with assay-external reference. (C) The negative control with cell line C33A does not show specific target signal.

The use of SIL peptides as references comes with further challenges. A subset of the fragments for any SIL precursor appear with offset *m/z* compared to the light precursor. This is corrected for in software, where fragment intensities are compared according to their identity as opposed to the *m/z* at which they appear. E.g. a y_3^+ fragment of the heavy precursor is compared to the y_3^+ fragment of the light precursor, even if y_3^+ for the heavy precursor has a +8 Da offset. Despite this, I encountered several edge cases that required manual curation of the XIC comparison. While y and b fragments represent an even peptide bond break, a-type fragments cut off a C_{13} if the adjacent amino acid is heavy (fully labeled). The software Skyline, which is crucial for the workflow as it allows interactive signal extraction and peak curation, does not support this distinction and applied only the C_{12} offset for the fragment. A further edge-case that I encountered was where a precursor had a combination of C-terminal and N-terminal ions with the same chemical composition. These were typically a y- and

a b-type ion, often with a loss of water, such that e.g. y6-18 would have the same mass as b5. Such ions cannot be distinguished by the MS and are observed with combined intensity. If a SIL amino acid only affects one of these ions, then the fragments become disjointed again. This issue can manifest such that the SIL reference has the combined fragments and the light precursor does not or vice versa. Again, Skyline was not able to account for this and calculated low NSA scores. Manual inspection and peak curation were required in order not to miss valid detections.

3.2.8 Acquisition sequence and controls

Due to the projects aim to detect peptides at trace level in the LC-MS experiments, it was crucial to ensure that any target signal truly originates from presentation at the cell surface. During establishment of the experimental procedures, two major sources for contaminations became apparent. First, the occasional use of concentrated synthetic peptides for unrelated purposes in our lab, such as for immunogenicity assessments, can potentially lead to trace-level introduction into the samples. Due to the sensitivity of the instrument, this can lead to spurious detection. While this was only clearly determined on one occasion, and measures were taken to prevent this, I decided nevertheless to include paired negative control experiments with each IP. For this purpose, I used the cell line C33A, which is HPV-negative. I confirmed that none of the target peptide sequences are present, beyond what is associated with E6 and E7, in the human proteome (Swiss-Prot), including isotypes and further translated nucleotide sequence data (TrEMBL). Due to this, the cell line C33A is not expected to present any of the target peptides. The second kind of contamination is due to the properties of the LC system, where for any series of experiments, the analytes pass through the same needle, capillaries and analytical column. In my experiments this led to the possibility that a preceding IP would introduce strong signal for a peptide, thereby contaminating the LC system in the following runs. Signal derived from SIL peptide light contamination due to reference acquisition in preceding runs can present similarly. Such contamination can be controlled for by acquisition of an LC control run prior to each sample. For this control run, I included the same RT reference peptides that were spiked into each IP run, stabilized by a low complexity bovine serum albumin (BSA) digest. The combination of the LC control and the IP negative control runs meant that for each targeted HLA IP acquisition, 4 LC-MS runs were performed in sequence, using the same LC-MS method (targets, gradient, settings). The sequence consists of an LC control 1, a C33A negative control IP, an LC control 2 and the main epitome map HLA IP as can be seen exemplarily in Figure 27A. Before the targeted runs, LC wash runs were used to clear the analytical column. These runs inject only blank solvents and apply multiple swift gradients. In some cases, where the targeted peptides were highly similar for multiple HLA IPs, a single negative C33A control IP was used to cover all, with LC control runs performed as usual.

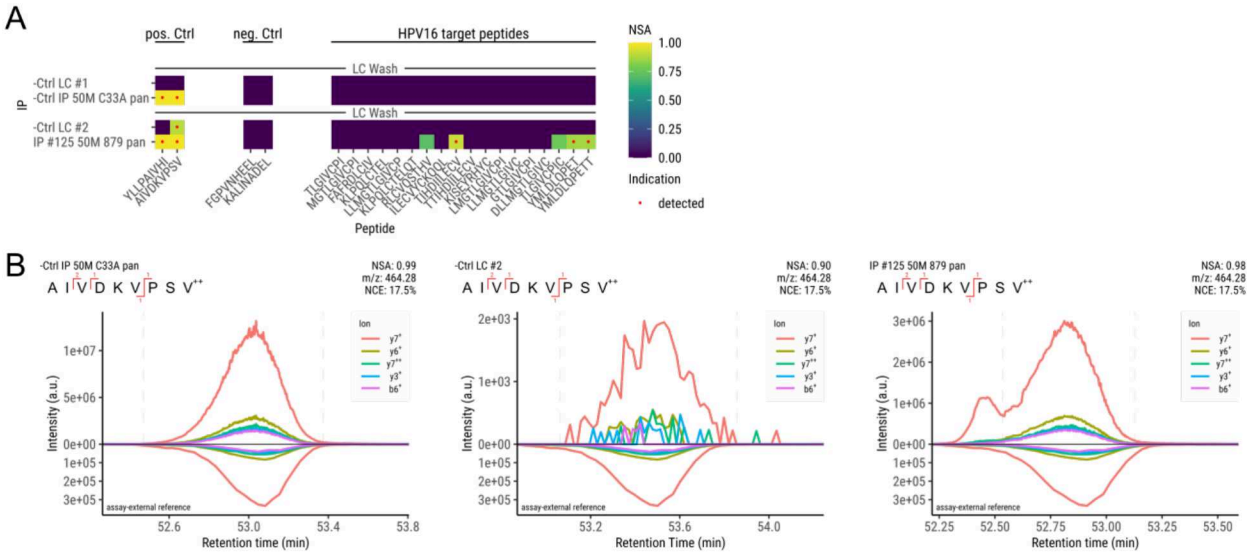


Figure 27. Targeted HLA IP acquisition sequence and controls. (A) The complete results of the sequence of LC-MS runs that control one HLA IP experiment. Positive control, negative control and target peptides are acquired equally with all runs in the sequence. The sequence of runs progresses from top to bottom rows. Peptides are sorted in no particular order. (B) XICs for positive control peptide AIVDKVPSV acquired in (A).

In the best case, and in practice in most experiments, the three controls were entirely free of specific target peptide signal. No target peptide was detected in any of the C33A negative controls, except for cases where the LC control already indicated contamination of the LC system. Two peptides most commonly showed contamination already with the LC controls (RAHYNIVTF, LLMGTLI). I could not precisely determine the source of this irregular occurrence but as the instrument was used for other projects, too, I decided to focus on a well implemented control scheme to ensure valid detections. In the case of the example in Figure 27A the controls are negative for almost all peptides. A set of positive and negative control peptides are shown in addition to the main targets. These peptides are what has been described as allotypic, which means they are typically dominant in the immunopeptidome and can be detected in a majority of cells that express the respective HLA allotype (Hillen & Stevanovic, 2006). The positive control peptide AIVDKVPSV, derived from the protein Coatomer subunit gamma-1, was strongly detectable already in C33A. This led to detection also in the following LC control #2. Usually, in cases where the controls already indicated a detection, the successful detection of the peptide in the main HLA IP was rejected. The example is a case where rejection was not necessary, as the intensity of the detection in the HLA IP (#125) was several orders of magnitude stronger than in the LC control run (Figure 27B).

Further positive and negative control peptides, which I included for targeted detection whenever applicable, are listed in the methods section. Negative controls based on mouse peptides were used with supertype A02 and were never detected. Positive control detections were unreliable, likely due to the differences in HLA affinity and cell line gene expression and therefore did not prove helpful to determine IP success. The consistent application of the control scheme ensured that no analyte had been introduced during the experimental procedures.

With this newly designed methodology for sensitive and validated LC-MS detection of peptides that is scalable for the targeting of hundreds of peptides, the basis for the targeted immunopeptidomics experiments of the HPV16 epitome map was set.

3.3 HLA supertype population coverage

The overarching aim of the establishment of an HPV16 E6 and E7 epitome map defines a scope of six HLA supertypes that are to be included. These would combine for a majority population coverage of greater than 95%, such that a putative vaccine formulation would be widely compatible. The basis for this goal was derived from a study by Reche and Reinherz (2007) where it was calculated that, according to the supertype classifications available at the time, such coverage could be achieved by combining the supertypes A02, A03, A24, B07 and B15. The experimental binding tests that are the source of the candidate peptides for my work are based on this specification and have been performed using an extended set of supertype representative HLA alleles, namely A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, B*07:02 and B*15:01 (Bonsack et al., 2019). The use of cell lines for this study required me to assess how to best map the alleles available in each cell line to the binders identified with supertype representative alleles. To this end, I used a comprehensive assignment of HLA alleles to supertypes published by Sidney et al. (2008). This source also refined previous supertype classifications such that the supertype B15 is separated into supertypes A01 and B62. To adhere to the labelling used in our research group, I will herein refer to this supertype B62 as B15. The study also combines supertypes A3 and A11 into a supertype A03 herein referred to from now on as A03_11. As the peptide:HLA experimental binding data was compiled more granularly, with the representative types A*03:01 and A*11:01, I decided to keep separate all A*11 alleles in a supertype A11, where it pertains to peptides targeted in the experiments.

The set of supertypes targeted In this thesis is therefore A01, A02, A03_11, A24, B7, B15; matching what is named in Sidney et al. (2008) as A01, A02, A03, A24, B7, B62. The available data on phenotypic frequency of HLA alleles has increased considerably since the target for 95% population coverage was formulated in 2007 and due to this, I was able to assemble a new detailed assessment of the achievable coverage. The immune epitope database (IEDB) analysis tool for epitope population coverage can be used to calculate supertype population coverages (Bui et al., 2006). It takes into account allelic and haplotype frequencies and specifies results stratified by ethnicity or as complete world population coverage. The calculations of the tool are based on data extracted from the allele frequency net database (AFND) in 2013 and list ethnicities according to the database's specification (Gonzalez-Galarza et al., 2020). Table 2 presents six prevalent ethnicities and the world population coverage that can be achieved with each supertype or with a full combination of the six supertypes. The calculated supertype coverage for the combined set is 99.6% and as such exceeds the initial goal.

Table 2. HLA supertype population and all world coverage.

	Ethnicity						
Supertype	Caucasoid	Oriental	Mixed	Amerindian	Black	Asian	**World**
A01	39.2	14.6	42.4	5.4	33.1	27.5	**31.5**
A02	50.7	46.2	46.5	77.4	42.1	38.6	**46.9**
A03_11	48.0	60.8	49.5	32.2	51.4	59.7	**51.6**
A24	22.5	40.2	20.9	43.4	24.8	29.1	**29.3**
B07	49.8	35.4	41.5	32.6	53.1	43.2	**43.7**
B15	13.6	42.5	11.7	9.4	5.0	26.9	**19.8**
Combined	99.7	99.9	99.6	100.0	99.1	99.5	**99.6**

With this new approach, I decided to review whether the supertype combination that was derived from literature can still be considered optimal. The collection does not include the HLA C locus and

favors the HLA A locus. This reflects the reported low expression and low contribution of HLA C alleles to the immunopeptidome (Zemmour & Parham, 1992; Apps et al., 2015). To investigate whether this is also reproducible from my data, I retrospectively quantified the contribution of the HLA A, B and C alleles as determined in my untargeted experiments described in section 3.4.2 (Figure 28A). Indeed, the HLA C alleles contributed by far the least to the immunopeptidomes of the cell lines assessed in this thesis. In contrast, the HLA A alleles had the biggest contribution. I also found HLA A supertypes to be the most potent to achieve large population coverage. I calculated the maximum coverage for all permutations of supertypes and found that several combinations of supertypes can allow for very similar population coverages with the top combinations being a combination of six HLA A supertypes (Figure 28B, red) and a combination with supertype B44 instead of B15 (Figure 28B, green) with 99.8 % and 99.7 % population coverage, respectively. While these differences are not consequential, I calculated population coverage for all permutations of missing supertypes from these three different sets. This calculation gives insight on the robustness of the approach in case suitable epitopes cannot be established for the full set. I determined that the set including only HLA A supertypes is less robust, likely due to the large population coverage being the result of low redundancy. The set of supertypes including B44 was indicated to be the most robust, but the set used for this study (Figure 28B, blue) is confirmed to be very close to optimal for the development of a globally applicable vaccine.

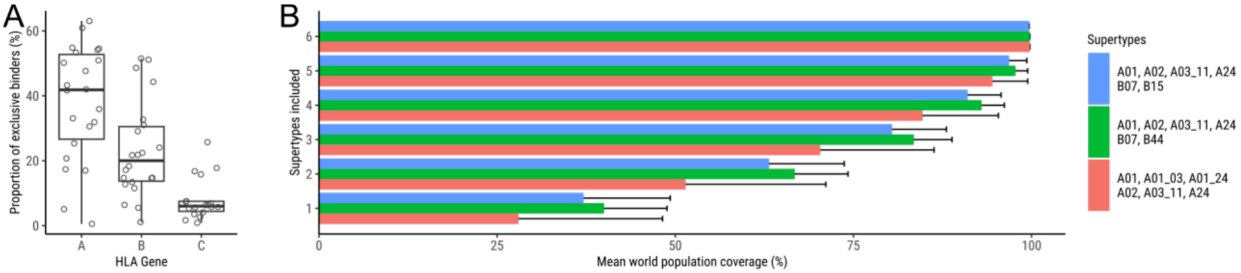


Figure 28. HLA supertype combination metrics. (A) Contribution of HLA loci to the immunopeptidomes acquired in this thesis. Plotted are all untargeted experiments and the relative size of the exclusive binder peptide populations for each HLA locus. (B) Robustness of world population coverage. Shown are, for each combination of supertypes, all permutations where one or multiple supertypes are left out. Missing supertypes reduce the population coverage. Confidence intervals show the standard deviation.

3.4 Immunopeptidomics experiments for the HPV16 Epitome map

3.4.1 Outline

In this thesis, I tested a total of 20 HPV16-transformed cell lines for the presentation of HPV16-derived peptides (Table 3). Initially, only cell lines derived from cervical cancer (CxCa) were used. To increase the availability of HLA supertypes A01 and A24, I decided to include further HPV16-positive cell lines derived from head and neck squamous cell carcinoma (HNSCC). As can be seen from Table 3, this resulted in a good representation of the targeted HLA supertypes across the cell line panel. Each cell line has a unique HLA peptide surface presentation profile and therefore provides a unique opportunity to detect the target peptides. With respect to the epitome map, the HLA supertypes of the cell lines' HLA alleles and the single amino acid variants (SAVs) of the proteins E6 and E7, are of primary concern. The E6 and E7 SAVs appear as patterns that allow the deduction of the HPV16 variant lineage and possibly the sublineage (Kaba et al., 2023). Together, HLA supertypes and SAVs indicate, for each cell line, which of the candidate peptides may be presented.

Table 3. Cell lines targeted for the epitome map. The E6 single amino acid variant (SAV) positions are based on the first start codon that encodes the 158 amino acid variant of the protein. HPV16 lineages and sublineages were determined according to characteristic SAVs of E6 and E7.

Cell line	Disease	HPV16 variant lineage (sublineage)	E6 SAVs	E7 SAVs	Relevant supertypes	IPs (targeted)
CaSki	CxCa	A	R17T, L90V	-	A02, A03, B07	8
MRI-H-196	CxCa	A	L90V	-	A02, A03, B07	4
W12-20861	CxCa	A	L90V	-	A11, A24, B07, B15	9
SNU-17	CxCa	A (A4)	D32E	N29S	A02	3
SNU-703	CxCa	A (A4)	D32E	N29S	A02, B07	2
SNU-902	CxCa	A (A4)	D32E	N29S, S63F	A01, A11, B07	6
SNU-1299	CxCa	D	Q21H, H85Y, L90V	-	A02, A03	7
SNU-1000	CxCa	A (A4)	D32E, I34R	N29S	A02, A11	4
SNU-1005	CxCa	A (A4)	D32E	N29S	A24, B15	7
866	CxCa	B	R17I, Q21D, H85Y	N29S	A11, A24, B07, B15	4
879	CxCa	A	-	-	A02, A03	4
Goerke	CxCa	A	-	-	A01	5
DoTc2-4510	CxCa	D	Q21H, H85Y, L90V	-	A03, B07	2
SiHa	CxCa	A	L90V, E120D	L28F	A24	5
SCC090	HNSCC	A	R17T, L90V	-	A02, A03, B07	3
SCC152	HNSCC	A	R17T, L90V	-	A02, B07	5
SCC154	HNSCC	D	Q21H, H85Y, L90V	-	A01	2
UD-SCC2	HNSCC	A	-	H51N	A01	2
UM-SCC-47	HNSCC	B	R17I, Q21D, E36Q, A68G, H85Y	N29S	A01, A24	2
UM-SCC-104	HNSCC	A	L90V	-	A02, A24, B07	4

3.4.2 Untargeted immunopeptidomics LC-MS acquisition

3.4.2.1 Rationale

For any cell line I selected candidate epitopes according to the experimental binding data for the present HLA that were determined by HLA genotyping. This selection comes with two caveats. First, it is unknown whether the desired HLA allele is actually expressed and functional. More importantly, there are typically several HLA allotypes present, for which experimental binding data is not available and which may offer alternative hypotheses for the HLA-restriction of a detected peptide. I therefore decided that untargeted LC-MS acquisition experiments could add valuable context for the epitome map by generating a comprehensive dataset on what kind of peptides are presented by each cell line. Untargeted experiments also offer a chance to detect HPV16-derived peptides which are not amongst the list of candidates for the targeted experiments. The list of candidates is limited in scope and there could be further peptides not covered for a variety of reasons, such as false negative binding predictions or incompatibility with the representative HLA allele of the binding tests.

I decided to use a data-independent acquisition (DIA) method for immunopeptidomics established by Dr. Jonas Becker in our lab. The goal was to perform one experiment per cell line. This was done in quick succession as the method does not need to be tuned to the same degree as the targeted experiments and sensitivity was not a primary concern.

3.4.2.2 Cell line immunopeptidome analysis

I evaluated the DIA experiments with the Spectronaut Direct-DIA peptide search engine, which typically yielded several thousand peptide identifications per cell line. As an example, the results of an untargeted acquisition of 50M[4] CaSki cells are being shown in Figure 29.

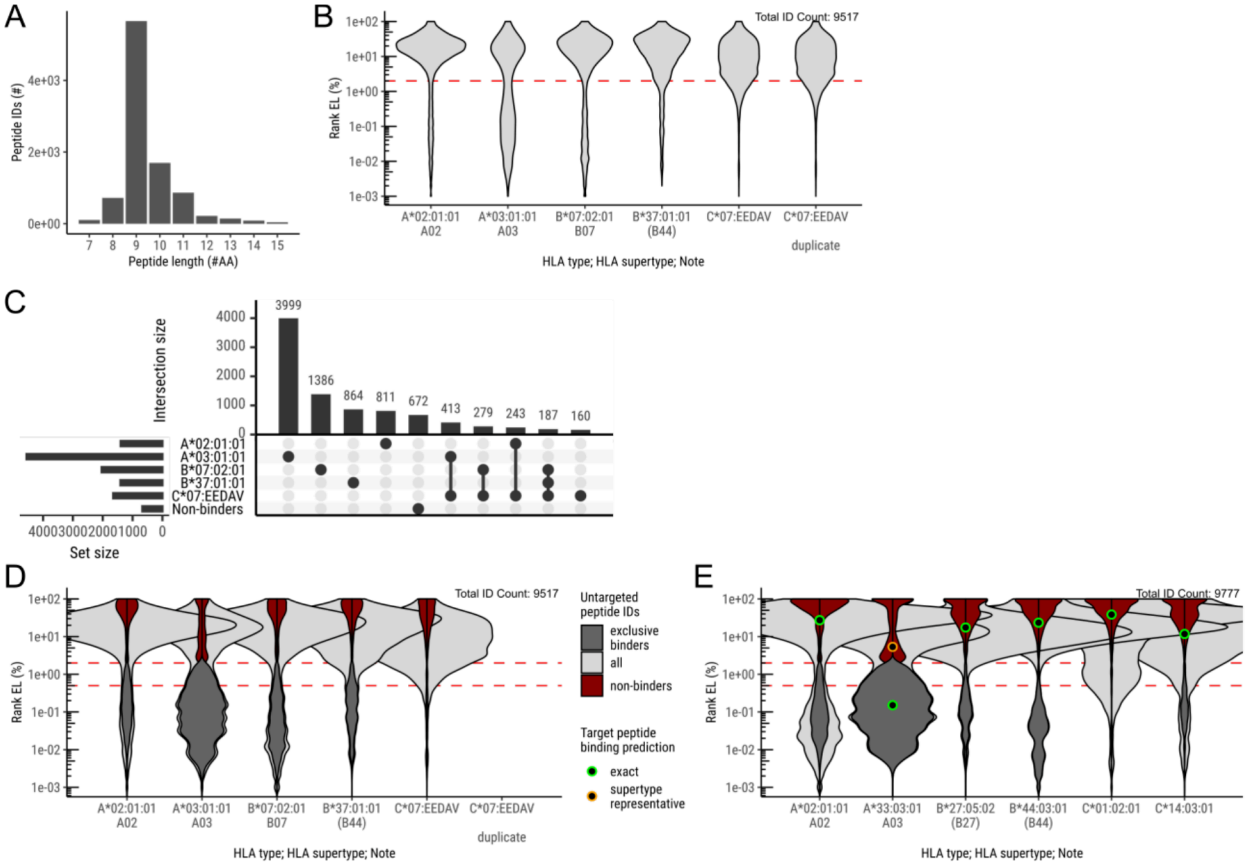


Figure 29. Evaluation of untargeted immunopeptidomics experiments. (A–D) Results for cell line CaSki. (A) Peptide length distribution of all detected peptides. (B) The eluted ligand (EL) predictions for each peptide to each of the HLA allotypes of the cell line. Lower ranks indicate better binding prediction. The dashed red line indicates the recommended top 2 % rank cutoff for weak HLA binders. The x-axis specifies the HLA-allele and the matching supertype, if any. Supertypes not targeted in this thesis are marked with parentheses. (C) Upset plot for all predicted HLA-allele:peptide matches according to the weak binder cutoff. (D) Enhanced violin plot with focus on predicted binder populations. Exclusive binders (according to weak binder cutoff) and non-binders are highlighted. The two dashed red lines indicate 2 % and 0.5 % thresholds for weak and strong binders, respectively. (E) Results for cell line SNU-1299 detection of peptide MFQDPQERPR. The target peptide Rank EL predictions are overlaid as green and orange dots.

The detected peptides are confined almost exclusively to lengths of 8 to 11 amino acids, matching well the expected distribution for peptides eluted during an HLA class I IP (Figure 29A). All peptides are therefore assumed to be potential HLA binders and I performed NetMHCpan 4.1 binding predictions for all peptides with all known HLA alleles of the cell line (Figure 29B). The plot displays the eluted ligand score rank percentage (Rank EL (%)) which is currently the state-of-the-art score for assigning the HLA allele of origin for peptides detected by mass spectrometry (IEDB, 2023b; Kovalchik et al., 2022). On the x-axis of this plot are aligned the six HLA alleles that are present in the cell line according to the HLA typing. Most HLA alleles directly belonged to an HLA supertype, as shown with a second x-axis label. A third line of the x-axis label is reserved for notes, here specifying

[4] For cell counts, the 'M' suffix is used to denote millions.

where alleles are duplicates. For some alleles, instead of the classic HLA nomenclature, a multiple allele code (MAC) was specified in the form of a 5-letter code after the allele group. In the example, this is the case for the C*07 allele. These are not uniquely assigned to one known allele but rather link to a group of alleles due to ambiguity of the typing result (see methods section 2.2.6.7). Due to this, I calculated rank scores as a median rank weighted by population coverage, as specified in the methods section. The distribution of the generated Rank EL scores is such, that typically the bulk of peptides fall into the non-binder range of high rank percentages. The authors of NetMHCpan recommend a threshold of ≤ 2 % Rank EL for classification as weak binders (as opposed to predicted non-binders above the threshold), which is indicated by the dashed red line. With peptides classified like this, one can quantify how many peptides derive from any of the HLA alleles (Figure 29C). In the case of the CaSki cell line, the identified peptides are dominated by the A*03:01 allele. Most of these A*03:01 binders are exclusive binders, meaning they are not also classified as binders for any of the other alleles. In this panel, the four leftmost columns are such exclusive binder populations. HLA alleles A*02, B*07 and B*37 each get assigned similar amounts of peptides. A remainder of fewer than 1000 peptides have not been assigned to an HLA allele and are classified as non-binders of unknown origin (bottom row). Notably, the C-allele only has a very small set of exclusive binders (160, rightmost column) and most are shared with the other alleles. The binding predictors are known to systematically underperform for C-alleles and results involving these alleles may therefore not reflect well the true contributions (Reynisson et al., 2021). As the threshold-based grouping loses information on the rank distribution, I decided to focus on tuning the violin plot to present concisely the data of interest (Figure 29D). To this end, I allowed overlap of the distributions in the upper section to focus on the relevant section below the 2 % Rank EL threshold. All shown distributions are zoomed by the same factor on the x-axis, so that the areas remain in true proportion. Both 2 % and 0.5 % thresholds, indicating cutoffs for weak and strong binder classifications, respectively, are marked with dashed red lines. In addition, the non-binders are highlighted in dark red and exclusive binders (according to the weak threshold) are highlighted with darker gray. Duplicate distributions such as the second C*07 allele of the example are not shown. Large populations of non-binders typically indicate low quality HLA IPs but can also indicate alleles where binding thresholds are too strict. Figure 29E shows that for any peptide of interest, the binding predictions can be overlaid to provide context on whether the peptide falls into a major population of binders and what the predictions are for the other HLA alleles. As the experimental binding tests were performed on specific HLA alleles (deemed supertype representative) that may show discrepancies compared to the exact HLA allele in the cell line, the supertype representative prediction is shown separately on the same axis.

I compiled this additional context for all peptides detected in the epitome map project.

3.4.2.3 Testing antibody specificity

Antibodies binding the HLA alleles of interest more specifically were available for this thesis. The available antibodies beside the main anti-pan-HLA class I antibody (clone W6/32) are antibodies termed anti-A02 (A*02 alleles, clone BB7.2), anti-A03 (A*03 alleles, clone GAP A3), anti-B07 (B*07 alleles, clone BB7.1) and anti-A11A24 (A*11 and A*24 alleles, clone A11.1 M). Initially, I routinely performed separate IPs with these antibodies, where appropriate according to the cell line's HLA allotypes. These targeted experiments, however, did not seem to provide a clear benefit compared to pan-HLA class I IPs. I therefore compared a set of 50M cell CaSki IPs using the four antibodies applicable for the cell line. Figure 30A shows the binding score distributions for the peptides detected in these IPs. It can be clearly determined that each of the antibodies also pulls down some of the other HLA allotypes and that the specificity is therefore limited. The A03:01 allele is clearly still present in the anti-A02 and anti-B07 IPs and seems to be only slightly enriched in the anti-A03 IP. I compared the total amount of peptide IDs by assigning each peptide to the best scoring allele (at minimum weak binder) or classifying it as non-binder. I found that the alternative antibodies produced overall much lower yields from the IPs (Figure 30B). While this is expected due to the aim of excluding 5 of 6 HLA-alleles with each of the antibodies, the resulting peptide identification counts for the targeted HLA-alleles are also lower than with anti-pan-HLA class I for anti-A03 and anti-B07. Only the anti-A02 antibody provides a clear benefit.

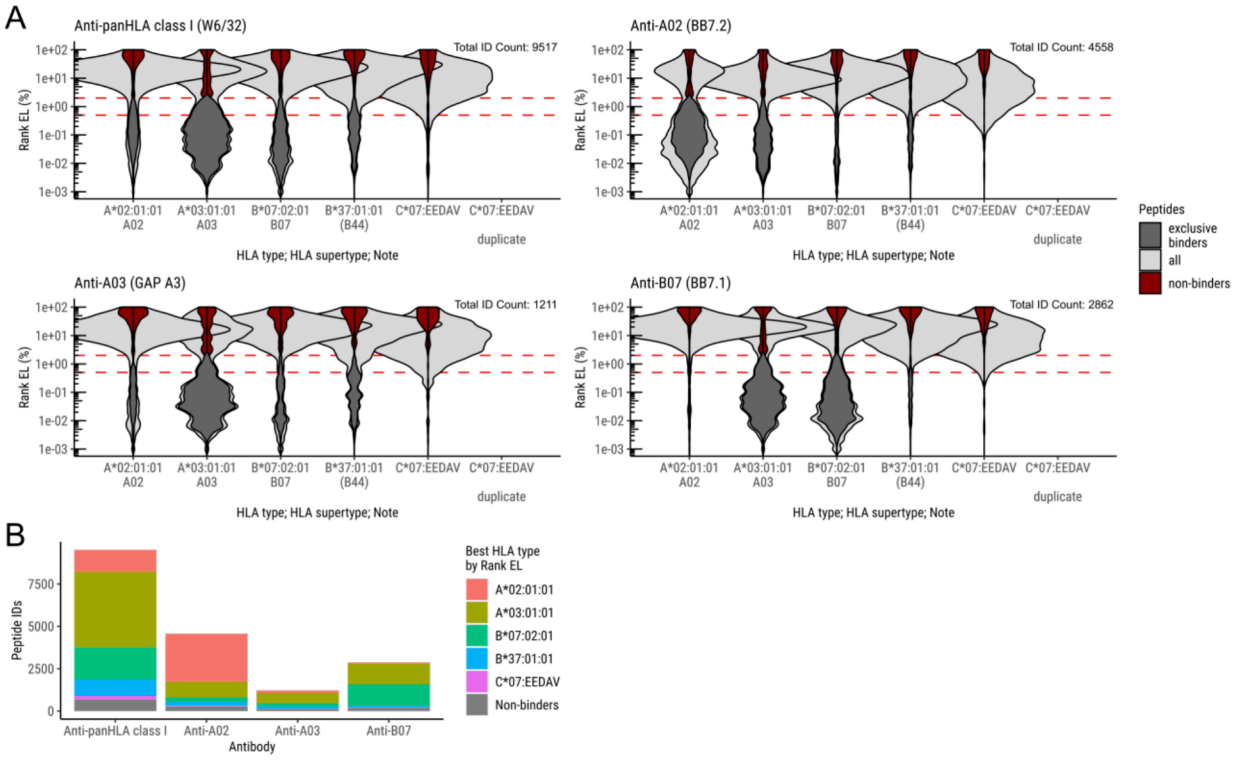


Figure 30. Effect of HLA IP antibody selection on detected peptides with cell line CaSki. Data is from 50M cell load CaSki IPs using different antibodies as indicated. (A) Predicted binder populations resulting from the use of antibodies as indicated. Antibody clone is specified in parentheses. (B) Total counts of identified binders with different antibodies. Peptides are assigned to the best-predicted HLA-allele.

I performed a similar analysis with the anti-A11A24 antibody using the cell line SNU-1005 (Figure 31). The result was again an enrichment of binders for the desired HLA-allele with only limited specificity and at the cost of the overall identification count.

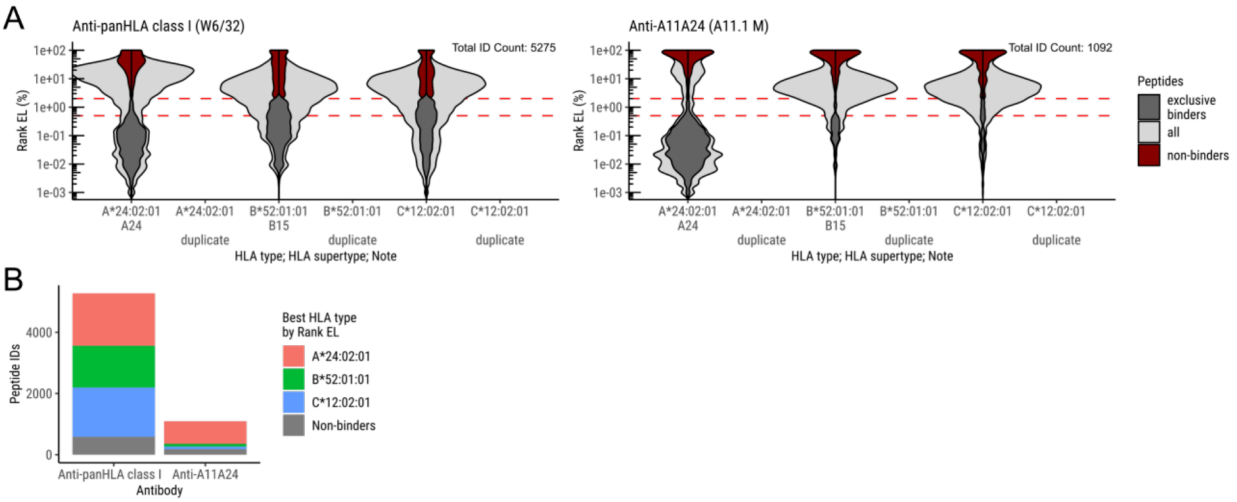


Figure 31. Effect of HLA IP antibody selection on detected peptides with cell line SNU-1005. Data is from 50M cell load SNU-1005 IPs using different antibodies as indicated. (A) Predicted binder populations resulting from the use of antibodies as indicated. Antibody clone is specified in parentheses. (B) Total counts of identified binders with different antibodies. Peptides are assigned to the best-predicted HLA-allele.

I concluded that the specific antibodies are not well suited to provide confidence in the origin of peptide detections and typically reduce the IP yields considerably. Due to this, I stopped routine use of alternative antibodies and only used them in cases where a specific problem could be addressed. This can for example be the case where a possible peptide detection is strongly affected by dominant coeluting ions, which may be peptides originating from an undesired HLA allele.

3.4.2.4 HPV16 peptide detections by untargeted immunopeptidomics

The search engine for the DIA experiment was set up to search for the complete HPV16 proteome with all sequence variants, regardless of any cell line specifics. Across all experiments, this led to a small set of HPV16 associated detections (Table 4). Four of the six detected peptides were also candidates for the targeted experiments. Untargeted detections are susceptible to false discovery and so each of these detections was aimed to be confirmed in targeted experiments. Indeed, only five of the six peptides were confirmed. These results will be presented with the detected peptides in section 3.4.4.

Table 4. Peptide detections from untargeted HLA IP acquisitions.

Peptide	Cell lines	Candidate	Targeted confirmation
YMLDLQPET	CaSki, 866, SNU-1299	Yes	Yes
GQAEPDRAHY	866, W12-20861	Yes	Yes
IILECVYCK	W12-20861	Yes	Yes
RAHYNIVTF	W12-20861	Yes	Yes
AEPDRAHY	866, W12-20861	No	Disproven
MFQDPQERPR	SNU-1299	No	Yes

3.4.3 Targeted HPV16 peptide detection by cell line

Targeted experiments were performed with all 20 cell lines. For each cell line, I confined the list of candidate HLA binders being targeted to those, which have been experimentally confirmed to bind to at least one of the HLA supertypes present (more precisely, to bind to an HLA allele representative for the supertype). A further criterion was that each peptide also matches the respective E6 or E7 protein sequences of the cell line.

In the following sections, the results of the targeted LC-MS experiments are presented for a set of exemplary cell lines, with all other cell line results being presented in Appendix B.

3.4.3.1 CaSki

The cell line CaSki has been the main cell line for all establishment and optimization of the IP experiments in our lab. The cell line is derived from a cervical carcinoma and is HPV16-transformed. It has a long history of use for analysis of the HPV16 immunopeptidome due to being naturally transformed and having robust surface expression of the HLA-A*02:01 allele. It has the E6 sequence mutations R17T and L90V. In full agreement with the HLA-typing, untargeted analysis confirmed supertype A02, A03, B07 and B44 surface presentation, with negligible contribution from the C alleles (Figure 32A). Over the course of this project, dozens of IPs were performed with this cell line. I summarized the IPs of note and peptides targeted in NSA heatmaps (Figure 32B). The IPs listed are the subset of IPs needed to satisfy the requirement of at least 2 replicate detections per cell line and one stable isotope-labeled (SIL) validation per peptide. IP IDs marked with # are given for reference with Appendix Table 1. The peptides are listed on the x-axis. The major separation is whether the peptides are cysteine-containing, where all non-cysteine containing peptides are listed first and then cysteine-containing peptides second. This is to reflect the fact that I typically performed one complete scan for cysteine-containing candidates and one complete scan for non-cysteine containing candidates. Further sorting is by HLA supertypes, as indicated above the heatmap. Some peptides are viable targets for more than one supertype. Last, peptides are arranged by gene and position, with E6 before E7. For CaSki, the combination of 3 HLA supertypes allowed for targeting for 75 peptides, with 31 targets for A02, 19 for A03 and 26 for B07. This resulted in the detection of 5 peptides, with 3 detections for A02 and 1 detection each for A03 and B07. Thus, while A03 was dominating the untargeted detections, this was not reflected in the targeted peptide detections.

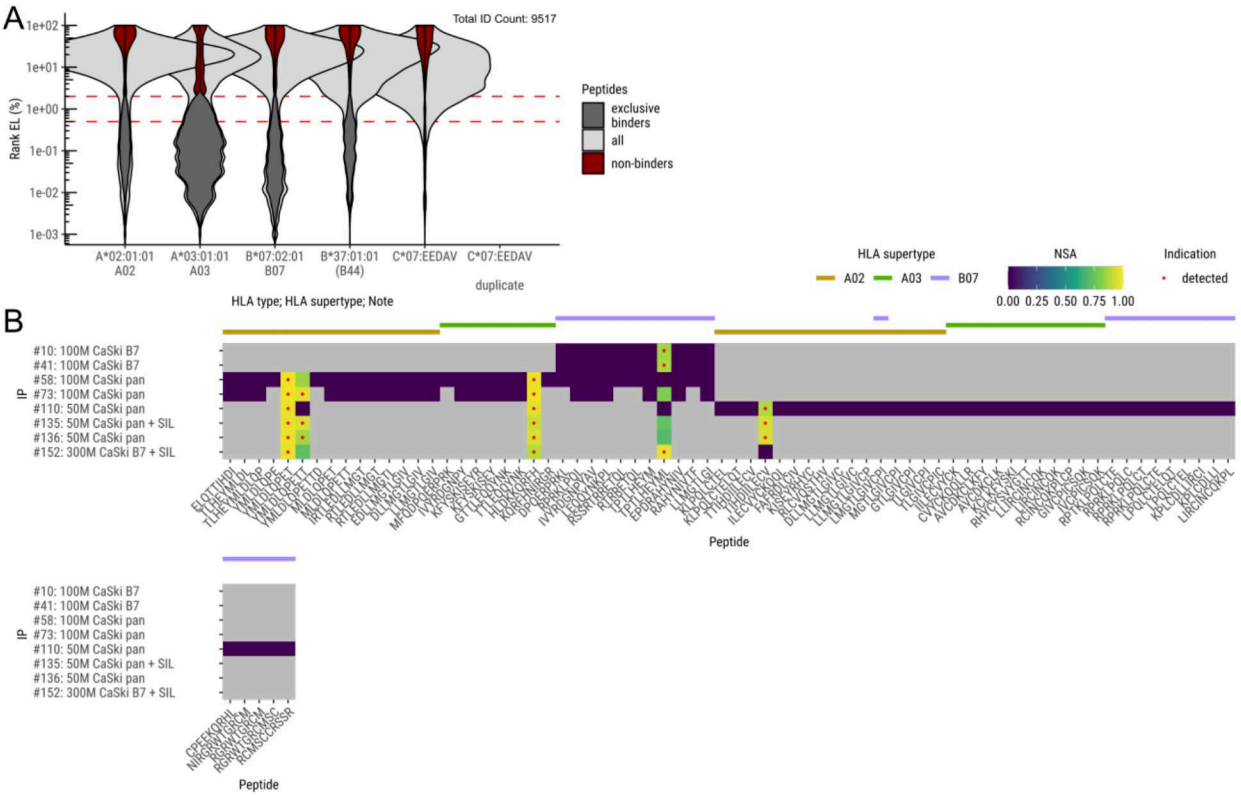


Figure 32. Cell line CaSki result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load ('M' denotes millions), cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

3.4.3.2 W12-20861

The cell line W12-20861 is notable, as it led by far to the most detections. In the untargeted experiment, a total of 15731 peptides were identified, which is the highest amount out of all the cell lines measured (Figure 33A). The cell line also has HLA alleles matching four different supertypes, which were all targeted in this thesis and all of which showed substantial amounts of exclusive binders in the untargeted experiment. Interestingly, the distribution of predicted non-binders shows a bimodal distribution for the A*11 allele, where a subset clusters close to the 2 % Rank EL threshold, suggesting that it is too strict in this case (red asterisk). The important B*15 allele showed a significant fraction of non-exclusive binders. To understand whether this is an artifact or whether this indicates that B*15 detections may be unspecific, I analyzed which allele is the best binder for this population of non-exclusive binders (light gray, below 0.5% threshold). In Figure 33B, I collected this information for the B*15 allele and additionally for the C alleles. B*15 is for almost all of its non-exclusive binders the best candidate and I conclude from this that detections from the population of predicted B*15 binders are highly specific. In contrast, the C alleles both have considerable amounts of predicted binders which favor other alleles. Taken together, the C alleles again show binding predictions of low quality but cannot be dismissed entirely, as at least C*15 seems to have a clear contribution to the peptides detected. The cell line has the E6 sequence variant L90V and in combination with the four supertypes this led to a large target list of 151 targets, with 36 A11, 51 A24, 25 B07 and 60 B15 targets (Figure 33C). Of these, 8 B15 binders, 2 A11 binders and one ambiguous A24, B15 and B07 binder were detected. IP #172 additionally targeted the peptide

AEPDRAHY (the very last listed peptide), which was rejected despite high NSA as explained in section 3.4.4.26.

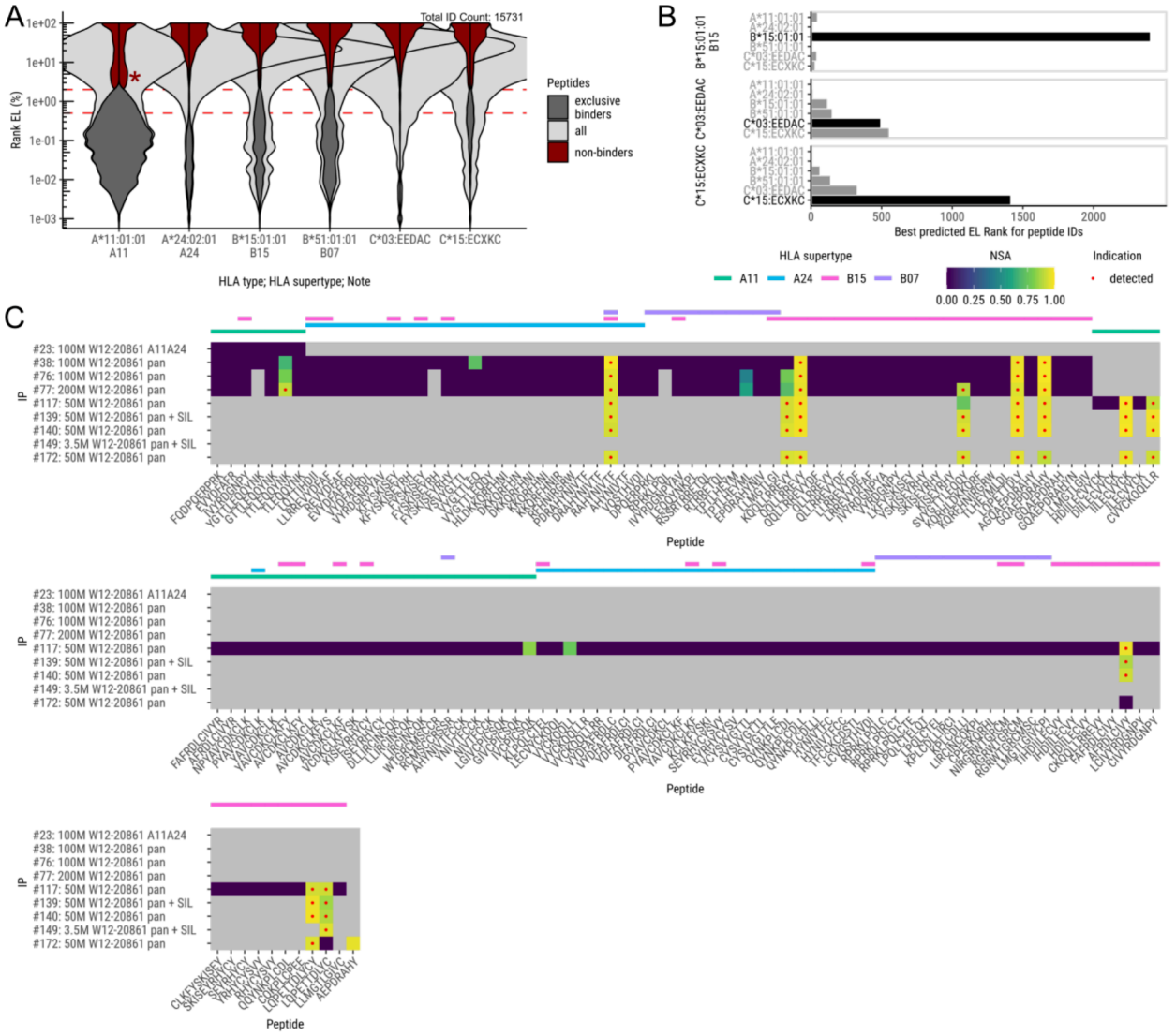


Figure 33. Cell line W12-20861 result compilation. (A) Untargeted cell line immunopeptidome evaluation. Red asterisk marks a population of non-binders that score just below the binder threshold for A*11 (B) The distribution of the best predicted HLA association for HLA allotypes showing considerable non-exclusive binder populations. Each panel presents one HLA allotype, lists it as the main hypothesis (black bar) and additionally displays the number of binders for which other allotype achieve better scores (gray bars). Only predicted strong binders are included. (C) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

3.4.3.3 SCC090 and SCC152

The cell line SCC090 is a HNSCC cell line naturally transformed by HPV16. It has the E6 sequence mutations R17T and L90V. In full agreement with the HLA-typing, untargeted analysis confirmed supertype A02, A03, B07 and B44 surface presentation, with minor contribution from the C alleles (Figure 34A). The resulting target list consists of 69 targets, with 31 A02, 18 A03 and 21 B07 targets (Figure 34B). Of these, 3 B15 binders, 2 A03 binders were detected. LLMGTLGI, a B07 binder, was rejected due to also appearing in the negative controls (data not shown). IVCPICSQK, an A03 binder, was only very weakly detected and not further followed up on due to successful A03-derived detection in another cell line.

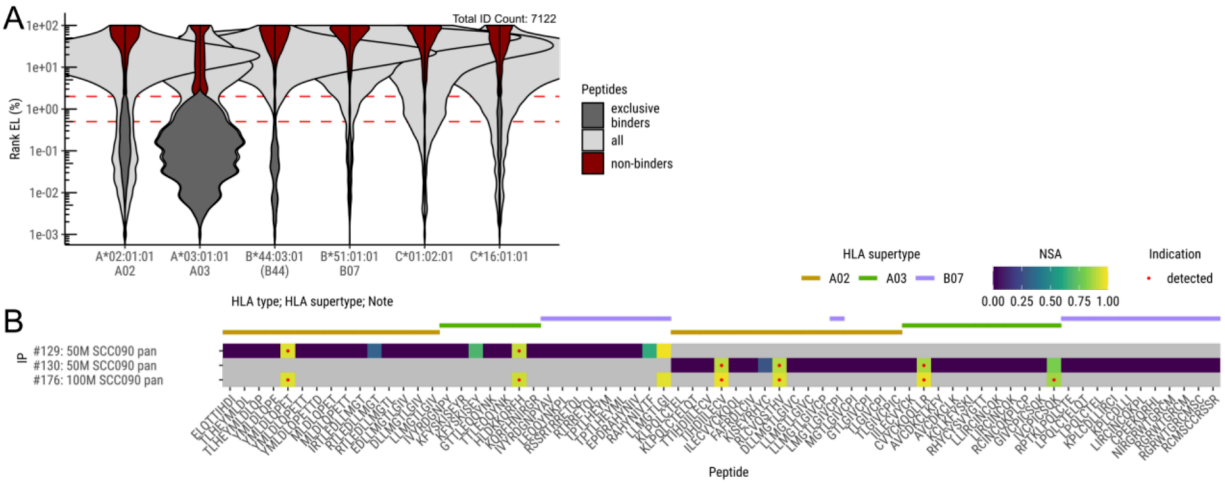


Figure 34. Cell line SCC090 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the peptide to HLA supertype associations from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

The cell line SCC152 originates from the same patient as SCC090, the source being a metastasis of the same cancer. Due to this, I assumed the same HLA types, even though HLA-typing did not list A*03:01 and B*44. In the untargeted analysis, it became clear that indeed, A*03, B*44 and likely also C*16 did not contribute to HLA peptide presentation anymore (Figure 35A). This may indicate partial or complete loss of a copy of chromosome 6, on which all the HLA genes are located. The two predicted binding distributions for A*03 and B*44 show a complete lack of peptides below the threshold. In contrast, a considerable amount of peptides are still considered binding candidates for C*16. This again shows a tendency for binding predictions with C alleles to be highly unreliable. Computing predictions with only the three remaining HLA alleles still covers the detected binders almost completely (Figure 35B). The targeted experiments were set to target A03 still and the resulting target list is the same as for SCC090. The same 3 A02 binders were detected. In agreement with the loss of the A*03 allele, no trace of the A03 binders was detectable.

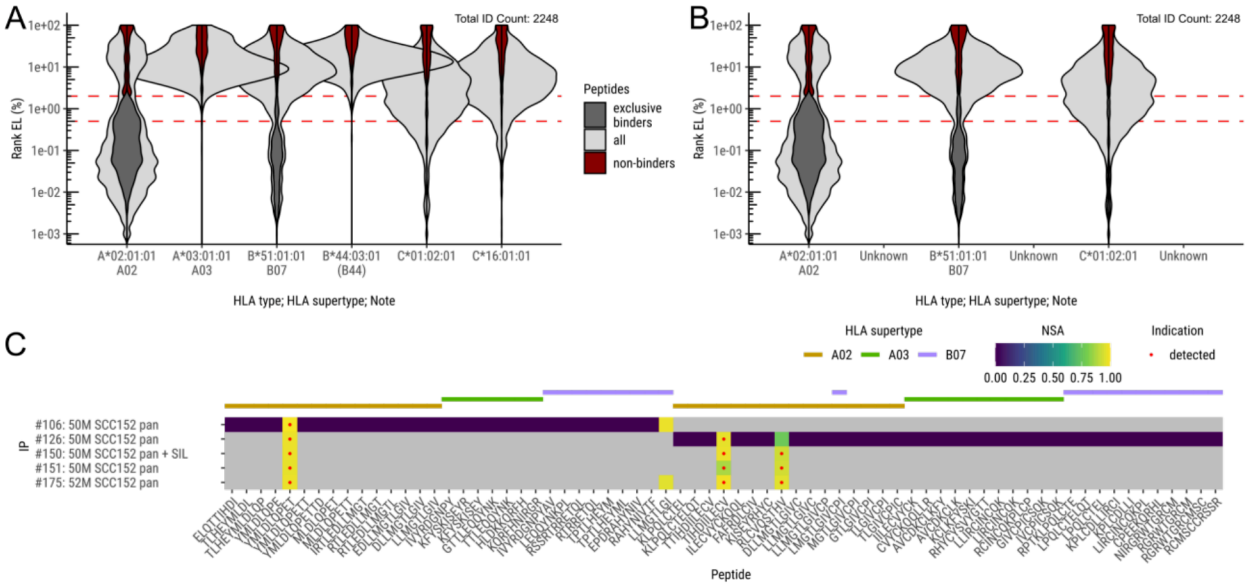


Figure 35. Cell line SCC152 result compilation. (A) Untargeted cell line immunopeptidome evaluation using the complete HLA panel from the patient. (B) Untargeted cell line immunopeptidome evaluation using the revised HLA-typing results. (C) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the peptide to HLA supertype associations from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

3.4.4 Targeted detections by peptide

In the following sections, the targeted peptide detection evidence is presented and contextualized. First, peptides E7/11–19, E7/11–20, E7/49–57, and E7/5–13 are shown as they are well suited to present the strategy and the challenges of the peptide detections. Further peptides are ordered according to the protein of origin, with E6 shown first, and then by sequence position. The two peptides that were originally found only through untargeted acquisitions are presented last.

3.4.4.1 E7/11–19: YMLDLQPET

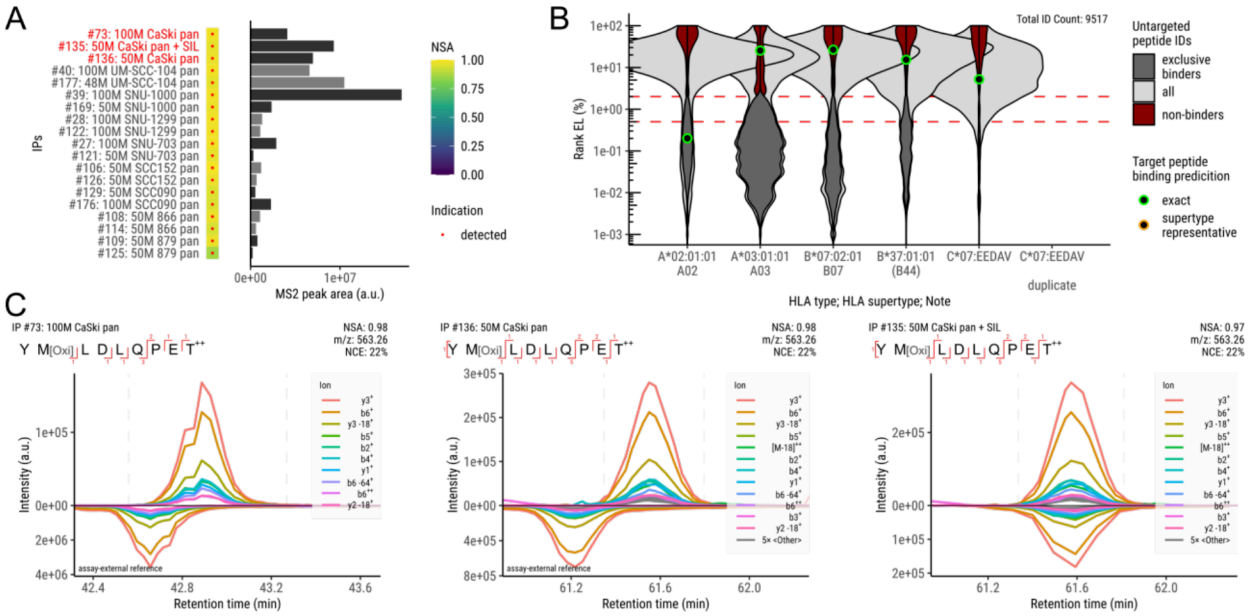


Figure 36. E7/11–19 YMLDLQPET targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load ('M' denotes millions), cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the CaSki immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide YMLDLQPET is at position 11–19 of the E7 sequence, which is a region that has not been found to be affected by any of the known variant mutations. It was targeted as a supertype A02 binder (exp. BA: 01.40±0.94; ± indicates standard deviation) and was the most readily detectable epitope, being detected in nine cell lines as shown in Figure 36A. This figure indicates for which cell lines peptide signal was observed, with the aim of two replicate detections per cell line to validate detections. The IPs for which extracted ion chromatograms (XICs) are shown below are marked in red. In addition, a quantification of the peak area is shown for each detection. These quantifications have very large variance by design, as the experiments were set up exclusively for maximum sensitivity. Nevertheless, they enable to assess the order of magnitude of the signal. Usually, low intensity coincides with detections that are difficult to reproduce. Next, the question of whether the detection is specifically attributable to the HLA supertype of interest is answered by overlaying the predicted binding affinity on the detected HLA binder populations of the untargeted experiments (Figure 36B). For the CaSki cell line, the A*02:01 allele was the only predicted binding allele and the detection matches well with the bulk of A*02:01 binders. Further, XICs from three replicates are shown below (Figure 36C). They produce very high sequence coverage, with IP #135 being a SIL spike-in experiment that confirms the exact coelution with the reference. Therefore, the peptide was determined to be presented by the supertype A02.

E7/11–19 represents the first MS-detected HPV16-derived peptide, restricted by the same HLA allotype as described in this thesis (Riemer et al., 2010). It has since also been detected in HPV16-associated cancers (Keskin et al., 2011; Peng et al., 2023). It is a promising vaccine candidate that has shown efficacy in several studies as described in the introduction. Multiple T cell receptors (TCRs) against this epitope were isolated from tumor infiltrating lymphocytes (TILs) from HPV16-associated cancers (Jin et al., 2018; Santegoets et al., 2022).

3.4.4.2 E7/11–20: YMLDLQPETT

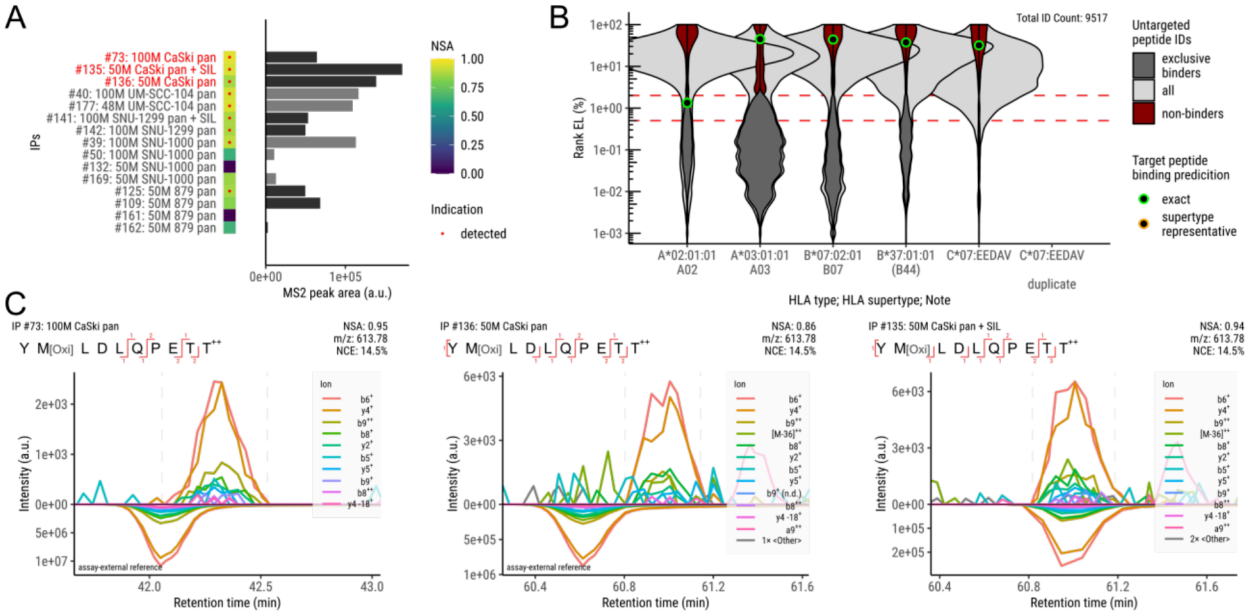


Figure 37. E7/11–20 YMLDLQPETT targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the CaSki immunopeptidome. (C) XICs for peptide detections indicated in (A).The peptide YMLDLQPETT is at position 11–20 of the canonical E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A02 binder (exp. BA: 02.19±2.71) and was detected in three cell lines s (CaSki, UM-SCC-104 and SNU-1299), with 2 more cell lines showing characteristic signal that was too weak to confirm conclusively (SNU-1000 and 879, Figure 37A). In the CaSki cell line, the A*02:01 allele was the only predicted binder (Figure 37B). In contrast to E7/11–19 it is predicted to be a weak binder, which also matches with the much lower observed peak areas. The detection matches the tail end of the A*02:01 binders, confirming the specificity of the detection as a supertype A02 detection. XICs from three replicates are shown below (Figure 37C). The two detections without spike-in were at low intensity and with limited sequence coverage. Nevertheless, SIL spike-in experiment (right XIC) confirms the exact coelution with the reference. Therefore, the peptide was determined to be presented by the supertype A02.

The peptide has been targeted for MS detection in prior studies but was either undetectable (Riemer et al., 2010) or too weak to confirm with certainty (Blatnik et al., 2018). The peptide:HLA-A*02 combination was previously found to induce IFNγ release in peripheral blood mononuclear cells (PBMCs) from healthy donors (Ressing et al., 1995; Wentworth et al., 1996; Blatnik et al., 2018). In addition, a TCR has been isolated from PBMCs from CxCa patients and showed specific killing of HPV16-transformed cell lines (Youde et al., 2000).

3.4.4.3 E7/49–57: RAHYNIVTF

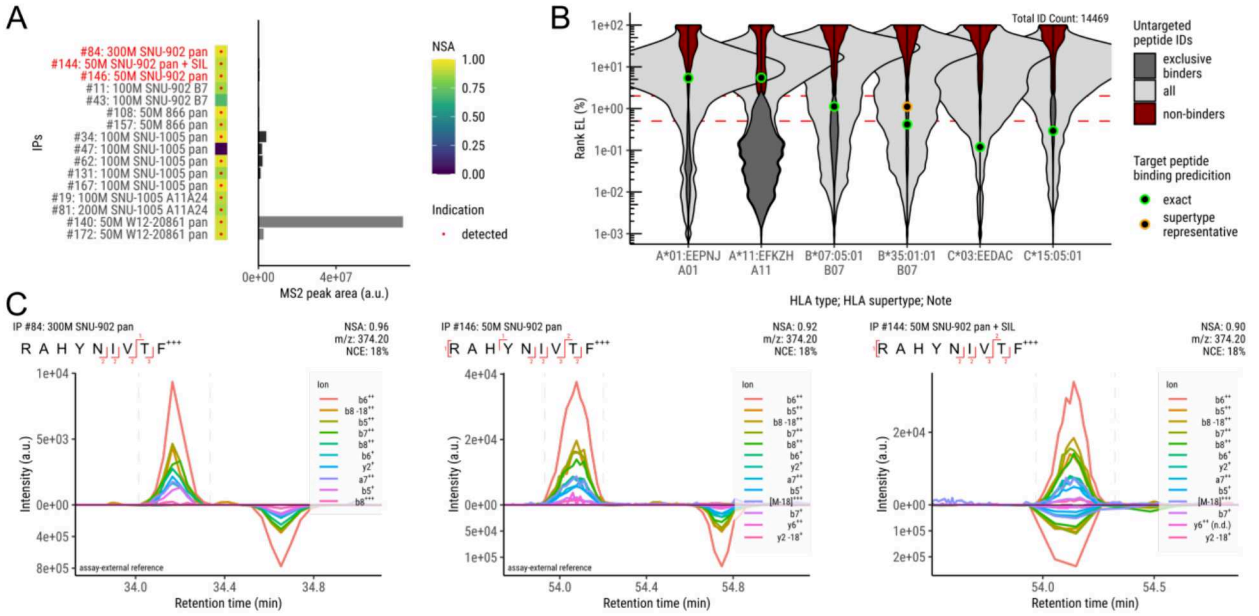


Figure 38. E7/49–57 RAHYNIVTF targeted detections for supertype B07. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the SNU-902 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide RAHYNIVTF is at position 49–57 of the E7 sequence. The position is affected by the known variant mutation H51N, which this peptide does not possess. It was targeted as a supertype A24 (exp. BA: 0.15±0.07), B07 (exp. BA: 04.65±2.28) and B15 (exp BA: 15.72±15.79) binder and it was detected in four cell lines (Figure 38A). In the SNU-902 cell line, the peptide was targeted for the B07 supertype. It is predicted as a binder to both B alleles, both of which belong to the B07 supertype (Figure 38B). Unexpectedly, the peptide is predicted as a strong binder for both present C alleles but as noted before (Sections 3.4.2.2 and 3.4.3.2), these predictions are of low quality. There is no experimental binding data to confirm these C allele predictions. XICs from three replicates are shown below (Figure 38C). The two detections without spike-in show very limited sequence coverage. Nevertheless, SIL spike-in experiment (right XIC) confirms the exact coelution with the reference.

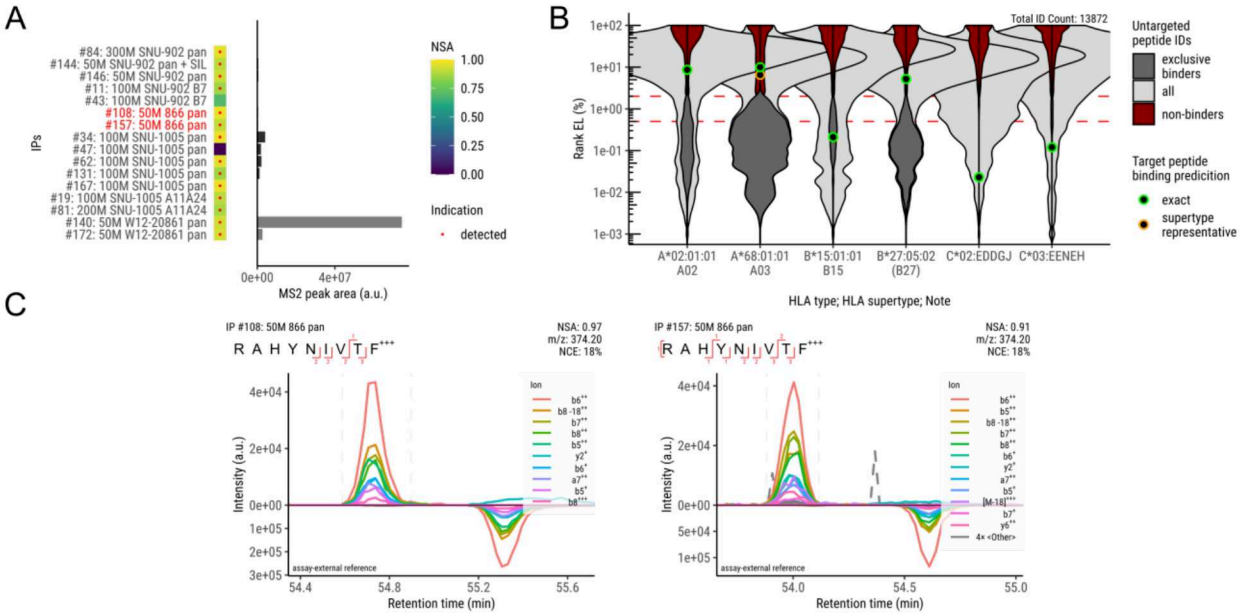


Figure 39. E7/49–57 RAHYNIVTF targeted detections for supertype B15. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the 866 immunopeptidome. (C) XICs for peptide detections indicated in (A).

Detection of the peptide specifically attributable to the B15 supertype was attempted with the cell line 866 (Figure 39A). Here, it is predicted as a strong binder to the B*15:01 allele (Figure 39B). There are similar strong predictions for the C alleles, too, similarly to the observations with the cell line W12 20861 (Section 3.4.3.2). The B*15 allele is a strong contributor in cell line 866 and the predictions for the C alleles cannot be confirmed. Of the A* and B* alleles, the B*15 allele is the only candidate for presentation of the peptide. XICs from two replicates are shown below (Figure 39C).

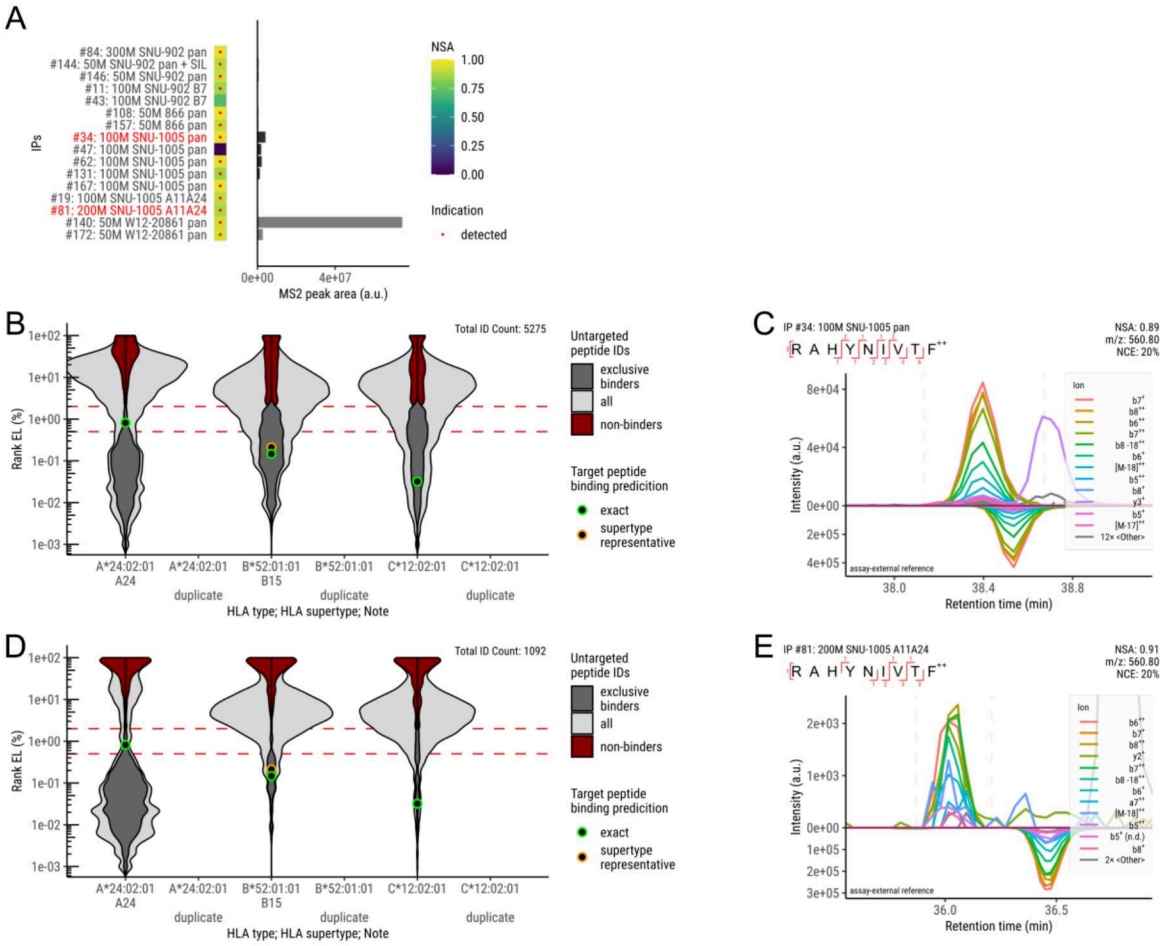


Figure 40. E7/49–57 RAHYNIVTF targeted detections for supertype A24. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B and D) and which results are shown in (C and E). (B) Peptide binding predictions overlaid on the untargeted analysis of the SNU-1005 immunopeptidome isolated with anti-pan-HLA class I. (C) XIC for peptide detection using anti-pan-HLA class I. (D) Peptide binding predictions overlaid on the untargeted analysis of the SNU-1005 immunopeptidome isolated with anti-HLA-A11A24. (E) XIC for peptide detection using anti-A11A24.

Detection of the peptide specifically attributable to the A24 supertype was attempted with the SNU-1005 cell line (Figure 40A). Here, it was targeted as a supertype A24 (exp. BA: 0.15±0.07) and B15 (exp BA: 15.72±15.79) binder (Figure 40B). In addition, the cell line shows a major contribution by the C*12 allele, which may also present the peptide. An XIC produced with an anti-pan-HLA class I IP shows strong signal (Figure 40C). Due to the contribution of the other HLA alleles, more specific confirmation was attempted with the use of the anti-A11A24 antibody. The untargeted evaluations show that the antibody did not achieve the desired specificity (Figure 40D). The XIC produced with anti-A11A24 shows much weaker signal (Figure 40E). It is therefore likely that the peptide as detected with the anti-pan-HLA class I IP was presented by other HLA alleles and that this contribution was only partially removed by using the alternative antibody.

Taken together, the peptide was determined to be presented by the supertypes B07 and B15, whereas it was not conclusively determined that the peptide is presented by supertype A24. The peptide:HLA-B*07 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018). The peptide:HLA-B*15 combination has not previously been found through MS detection or immunological tests. The peptide is very well known from murine studies, as it is immunodominant in context of the murine MHC allotype H-2 Db. It is a potent vaccine target

in all murine HPV16 tumor models in a mouse strain C57BL/6 background (Feltkamp et al., 1993). Even though being a binder to three HLA allotypes and shown here to be presented by at least two of them, it has not yet found the same use in humans.

3.4.4.4 E7/5–13: TPTLHEYML

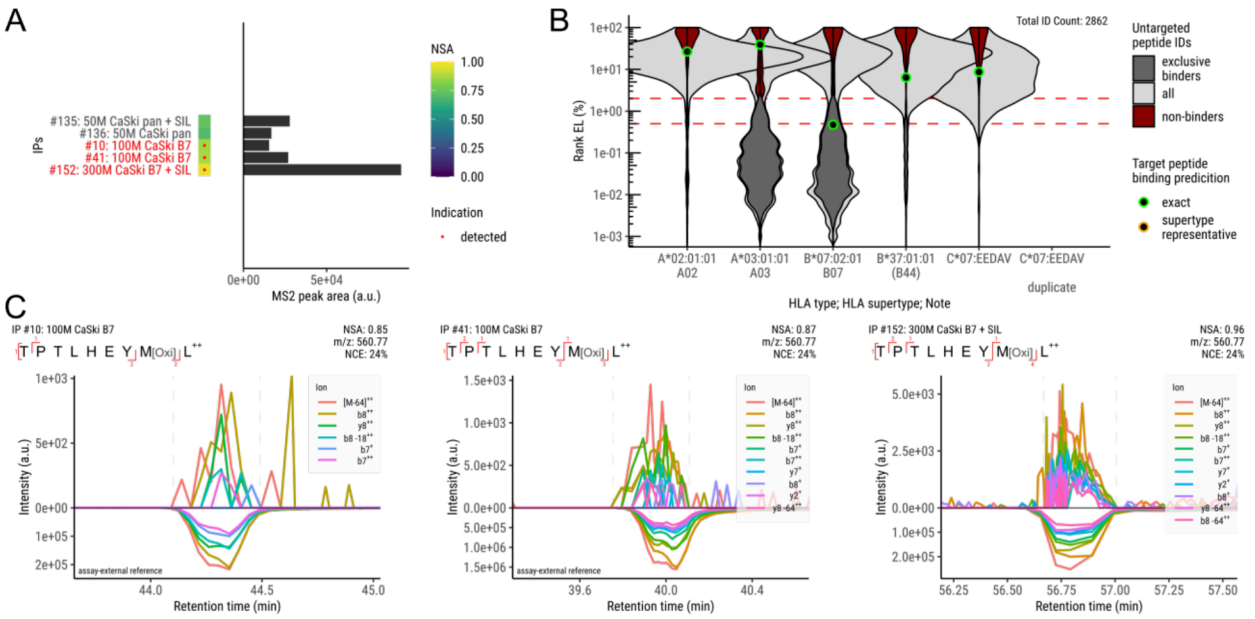


Figure 41. E7/5–13 TPTLHEYML targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the CaSki immunopeptidome isolated with anti-HLA-B7 IP. (C) XICs for peptide detections indicated in (A).

The peptide TPTLHEYML is at position 5–13 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B07 (exp. BA: 04.50±3.30) binder and was detected in the cell line CaSki (Figure 41A). In this cell line, it is predicted as an exclusive strong binder to the B*07:02 allele, confirming the specificity of the detection as a supertype B07 detection (Figure 41B). XICs from three replicates are shown below (Figure 41C). As signal from IPs using the anti-pan-HLA class I antibody was strongly distorted by off-target interfering ions, the anti-B7 antibody was used to achieve more specific signal. The three detections using anti-B7 antibody were at very low intensity and have low sequence coverage. Even though the SIL spike-in experiment showed a distorted peak due to off-target interference, it succeeded in acquiring highly specific signal, exactly coeluting with the reference. Therefore, the peptide was determined to be presented by the supertype B07.

The peptide:HLA-B*07 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018).

3.4.4.5 E6/15–22: RPRKLPQL

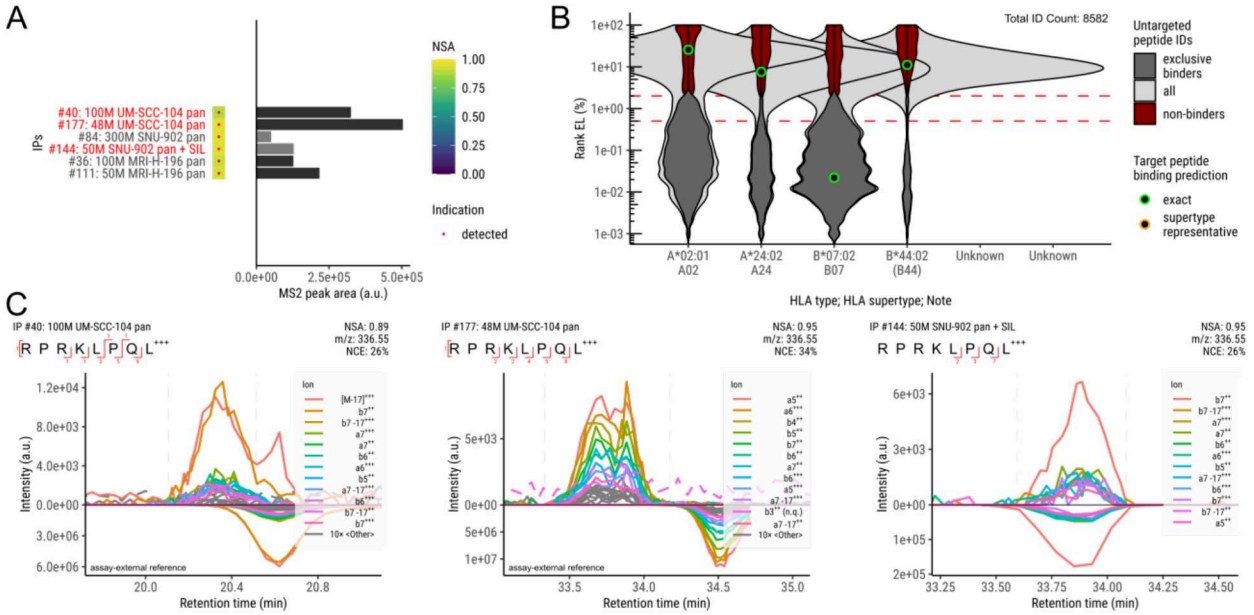


Figure 42. E6/15–22 RPRKLPQL targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the UM-SCC-104 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide RPRKLPQL is at position 15–22 of the E6 sequence. The position is affected by the known variant mutations R17I and R17T, which this peptide does not possess. It was targeted as a supertype B07 (exp. BA: 0.20±0.06) binder and was detected in three cell lines (Figure 42A). In the cell line UM-SCC-104, it is predicted as an exclusive binder to the B*07:02 allele, confirming the specificity of the detection as a supertype B07 detection (Figure 42B). XICs from three replicates are shown below (Figure 42C). Detections were at moderate intensity and with good sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A03.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.6 E6/29–38: TIHDIILECV

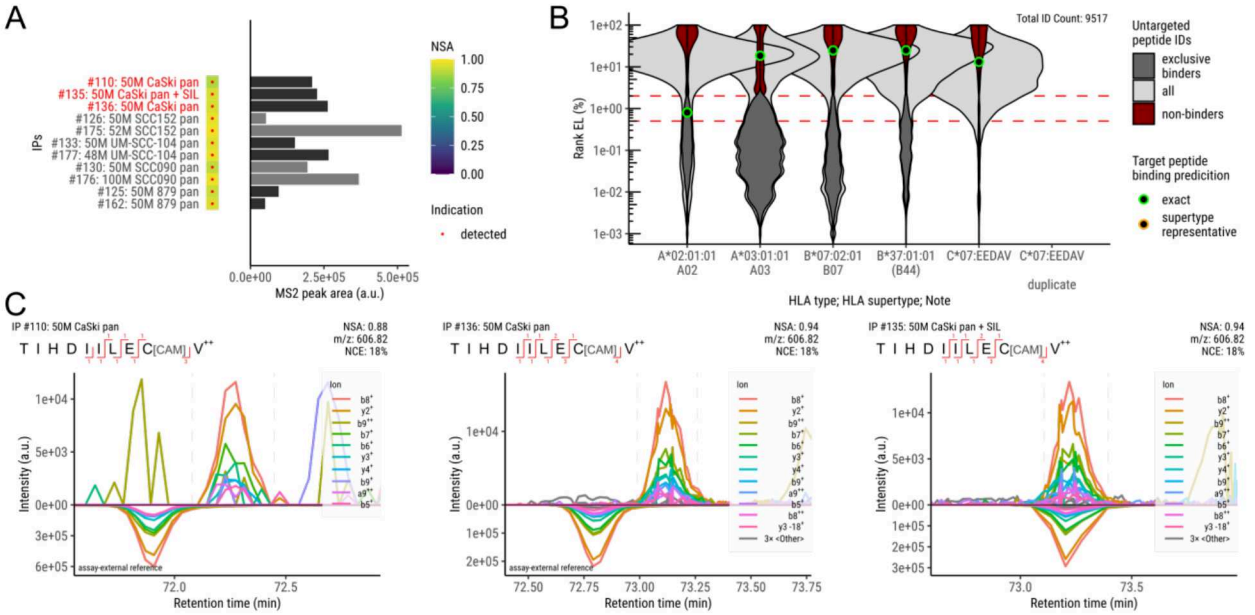


Figure 43. E6/29–38 TIHDIILECV targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the CaSki immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide TIHDIILECV is at position 29–38 of the E6 sequence. The position is affected by the known variant mutations D32E, I34R and E36Q, which this peptide does not possess. It was targeted as a supertype A02 (exp. BA: 03.69±2.17) binder and was detected in five cell lines (Figure 43A). In the cell line CaSki, it is predicted as an exclusive binder to the A*02:01 allele, confirming the specificity of the detection as a supertype A02 detection(Figure 43B). XICs from three replicates are shown below (Figure 43C). Detections were at low intensity with missing sequence coverage towards the N-terminus. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A02.

The peptide:HLA-A*02 combination has previously been targeted for MS detection in preceding work but the acquired signal was too weak to confirm with certainty (Riemer et al., 2010). It was found to induce IFNγ release in TILs from HPV-associated cancers, leading to the identification of specific TCRs (Santegoets et al., 2022). In addition, a TCR has been isolated from TILs and showed specific killing of HPV16-transformed cell lines, which led to a clinical trial to test its therapeutic use in patients (Draper et al., 2015).

3.4.4.7 E6/33–41: IILECVYCK

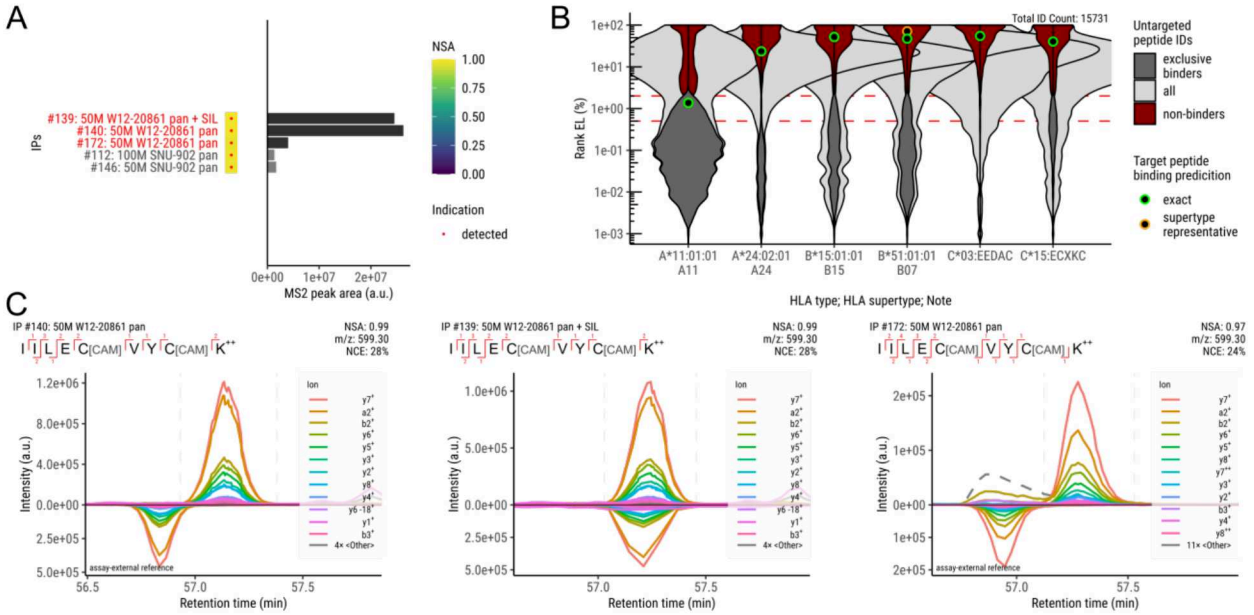


Figure 44. E6/33–41 IILECVYCK targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide IILECVYCK is at position 33–41 of the E6 sequence. The position is affected by the known variant mutations I34R and E36Q, which this peptide does not possess. It was targeted as a supertype A03 (exp. BA: 11.02±3.27) and A11 (04.98±1.33) binder and was detected in two cell lines (Figure 44A). In the cell line W12-20861, the peptide was targeted for the A11 supertype (Figure 44B). It is predicted as an exclusive binder to the A*11:01 allele, confirming the specificity of the detection as a supertype A11 detection. XICs from three replicates are shown below (Figure 44C). Detections were at high intensity with complete sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A11.

T cells binding the peptide:HLA-A*11 complex have previously been detected in PBMCs from patients with cervical lesions (Jiang et al., 2022). In addition, the related peptide:HLA-A*03 combination was found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018).

3.4.4.8 E6/37–46: CVYCKQQLLR

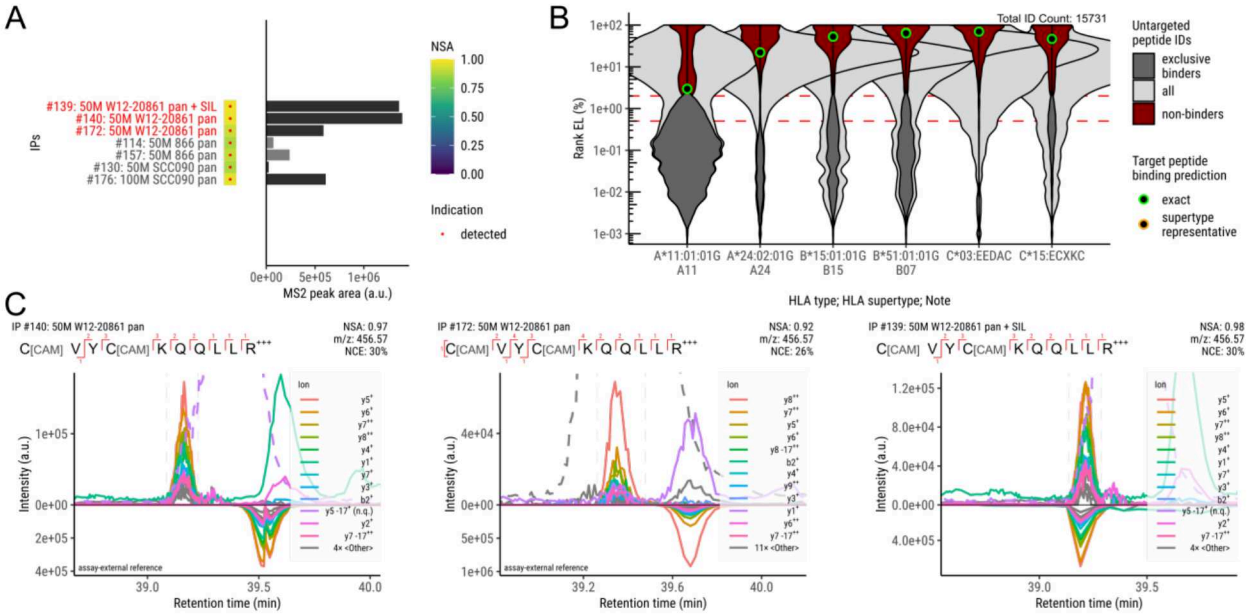


Figure 45. E6/37–46 CVYCKQQLLR targeted detections for supertype A11. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide CVYCKQQLLR is at position 37–46 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A03 (exp. BA: 36.33±19.78) and A11 (36.49±6.55) binder and was detected in three cell lines (Figure 45A). In the cell line W12-20861, the peptide was targeted for the A11 supertype (Figure 45B). It does not pass the threshold of a predicted binder for the present HLA alleles owing to the threshold for the A*11:01 being too strict, as was noted in section 3.4.3.2. Due to the experimentally verified binding, the detection can be specifically attributed to the A11*01 allele. XICs from three replicates are shown below (Figure 45C). Detections were at moderate intensity with complete sequence coverage. The SIL spike-in experiment confirmed exact coelution.

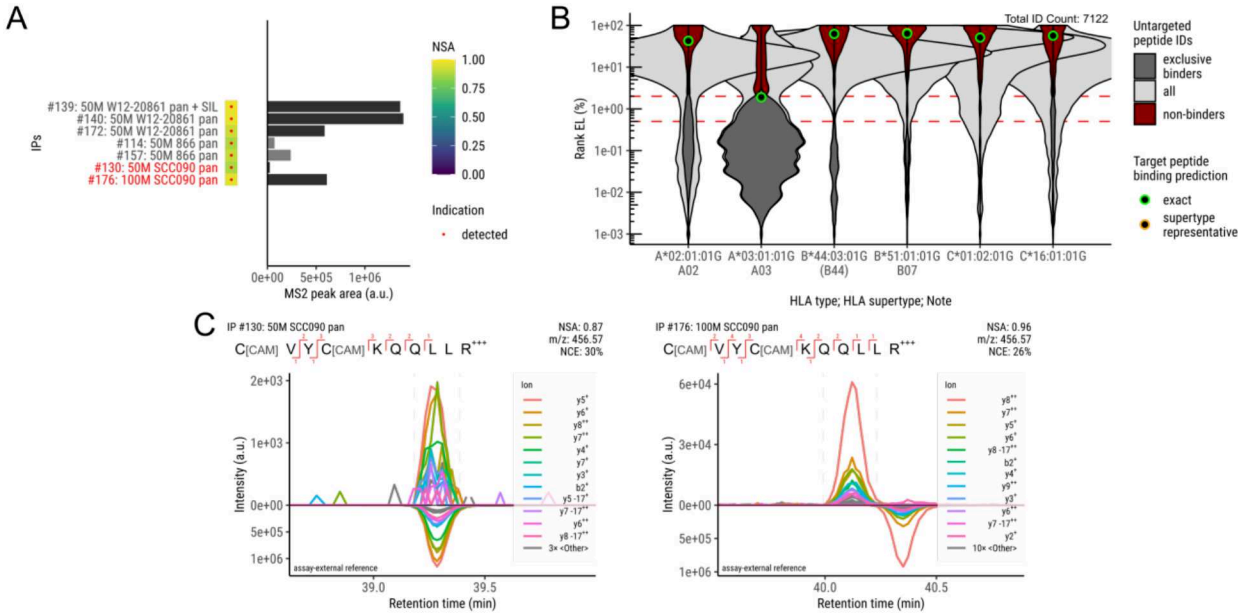


Figure 46. E6/37–46 CVYCKQQLLR targeted detections for supertype A03. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the SCC090 immunopeptidome. (C) XICs for peptide detections indicated in (A).

Detection of the peptide specifically attributable to the A03 supertype was enabled by the SCC090 cell line (Figure 46A). Here, it is predicted as an exclusive binder to the A*03:01 allele, confirming the specificity of the detection as a supertype A03 detection (Figure 46B). XICs from two replicates are shown below (Figure 46C). Detections differed in intensity by over an order of magnitude, with the second detection achieving complete sequence coverage.

Taken together, the peptide was determined to be presented by the supertypes A11 and A03. The peptide:HLA-A*03 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018). The peptide:HLA-A*11 combination has not previously been found through MS detection or immunological tests.

3.4.4.9 E6/38–45: VYCKQQLL

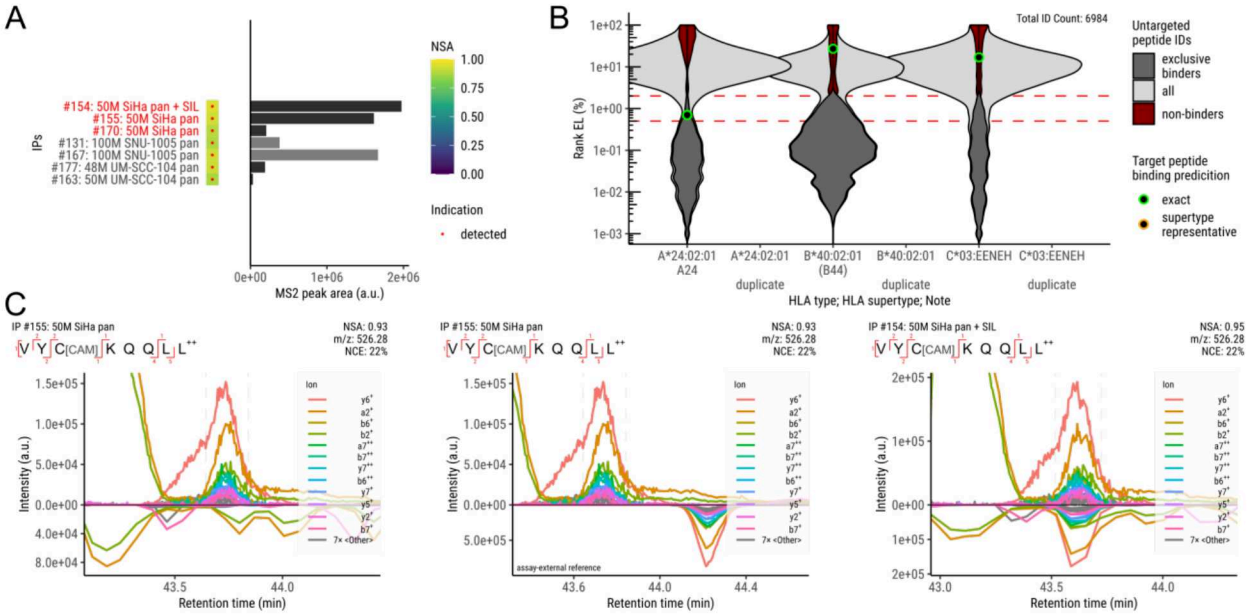


Figure 47. E6/38–45 VYCKQQLL targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the SiHa immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide VYCKQQLL is at position 38–45 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A24 (exp. BA: 03.51±1.64) binder and was detected in three cell lines (Figure 47A). In the cell line SiHa, it is predicted as an exclusive binder to the A*24:02 allele, confirming the specificity of the detection as a supertype A24 detection (Figure 47B). XICs from three replicates are shown below (Figure 47C). Detections were at moderate intensity with high sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A24.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.10 E6/41–50: KQQLLRREVY

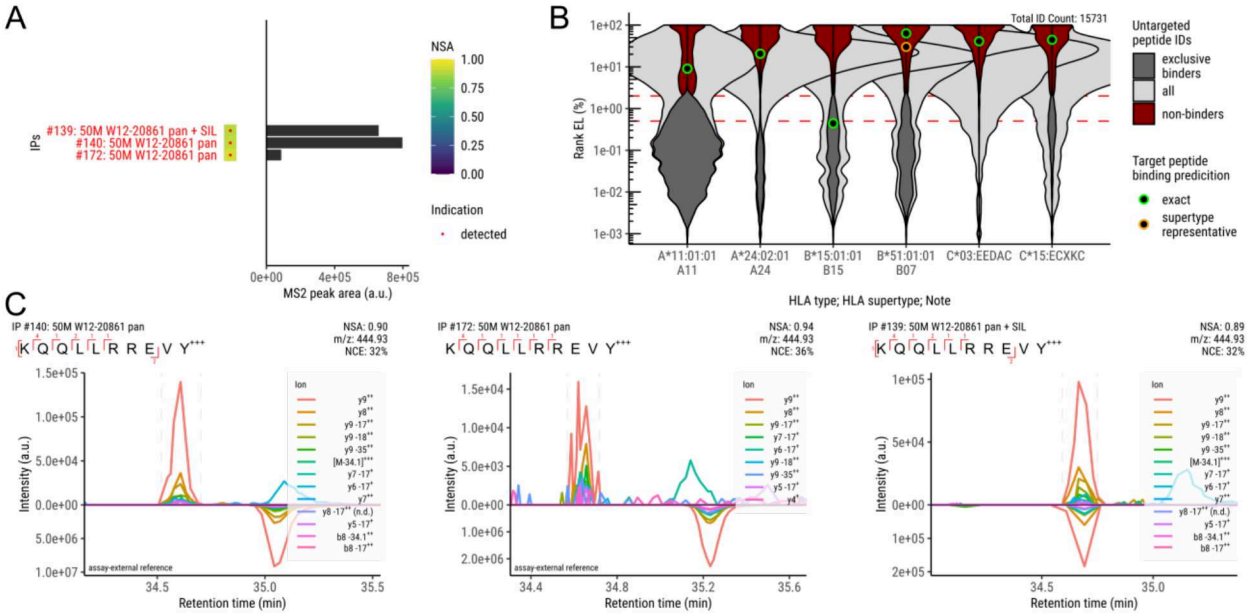


Figure 48. E6/41–50 KQQLLRREVY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide KQQLLRREVY is at position 41–50 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B15 (exp. BA: 05.80±2.97) binder and was detected in one cell line, W12-20861 (Figure 48A). In this cell line, it is predicted as an exclusive binder to the B*15:01 allele, confirming the specificity of the detection as a supertype B15 detection (Figure 48B). XICs from three replicates are shown below (Figure 48C). Initial detection at low intensity and with good sequence coverage showed a strong discrepancy on the retention time axis. The second replicate satisfied the criteria for a valid detection but was overall of lower quality. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.11 E6/42–50: QQLLRREVY

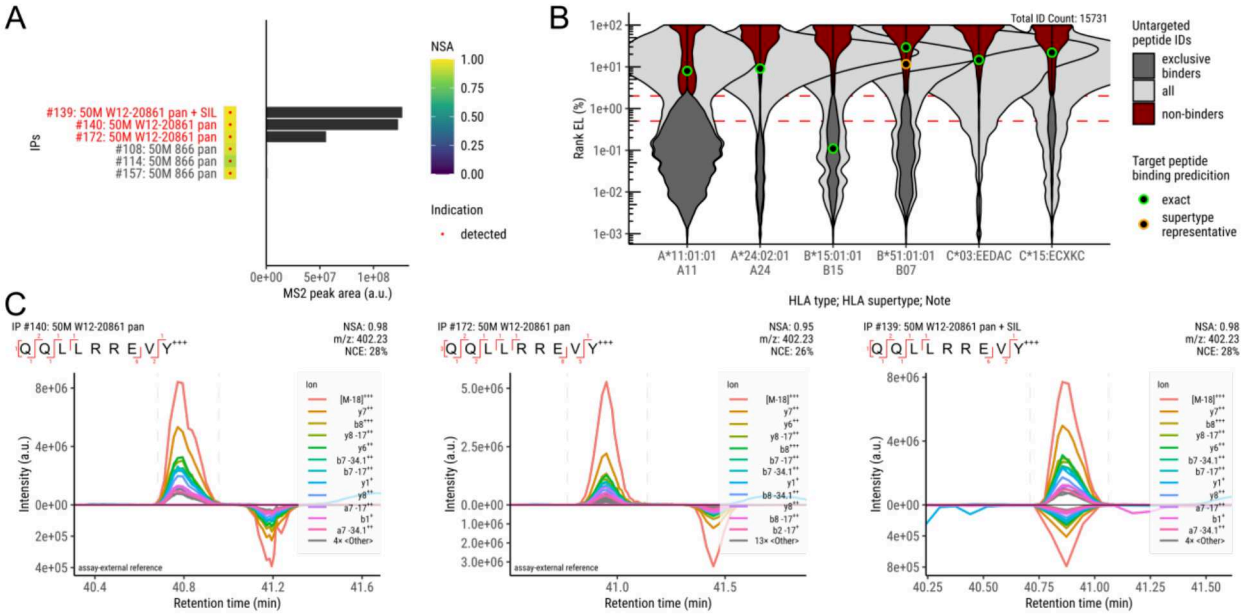


Figure 49. E6/42–50 QQLLRREVY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide QQLLRREVY is at position 42–50 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B15 (exp. BA: 08.12±3.38) binder and was detected in two cell lines (Figure 49A). Between the two cell lines, there is a large discrepancy in the intensity of the peptide, such that no bar is visible for 866. In the cell line W12-20861, the peptide is predicted as an exclusive binder to the B*15:01 allele, confirming the specificity of the detection as a supertype B15 detection (Figure 49B). XICs from three replicates are shown below (Figure 49C). Detections were at moderate intensity. The sequence coverage has a gap of four amino acids and the reference comparison showed a strong discrepancy on the retention time axis. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.12 E6/53–61: AFRDLCIVY

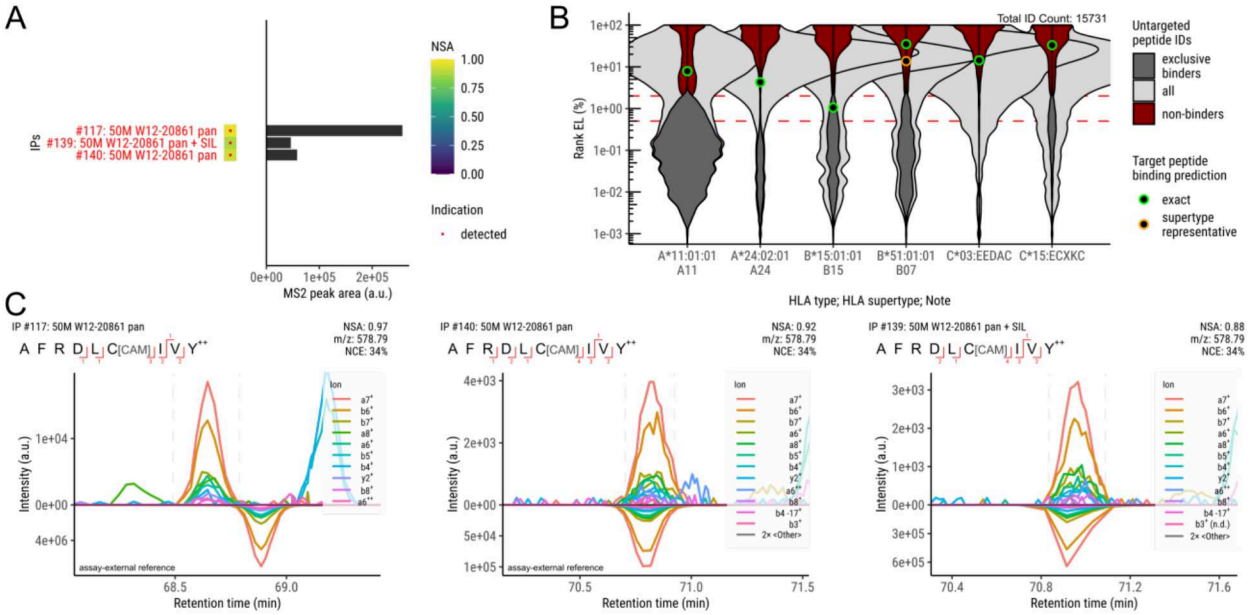


Figure 50. E6/53–61 AFRDLCIVY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide AFRDLCIVY is at position 53–61 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B15 (exp. BA: 06.13±2.15) binder and was detected in one cell line, W12-20861 (Figure 50A). In this cell line it is predicted as an exclusive binder to the B*15:01 allele, confirming the specificity of the detection as a supertype B15 detection (Figure 50B). XICs from three replicates are shown below (Figure 50C). Detections were at moderate intensity with a gap in sequence coverage towards the N-terminus. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide:HLA combination. However, the peptide in association with other HLA-allotypes B*07:02 and C*07:02 was previously found to induce IFNγ release in TILs from HPV-associated cancers, leading to the identification of a specific TCR (Santegoets et al., 2022).

3.4.4.13 E6/80–88: ISEYRHYCY

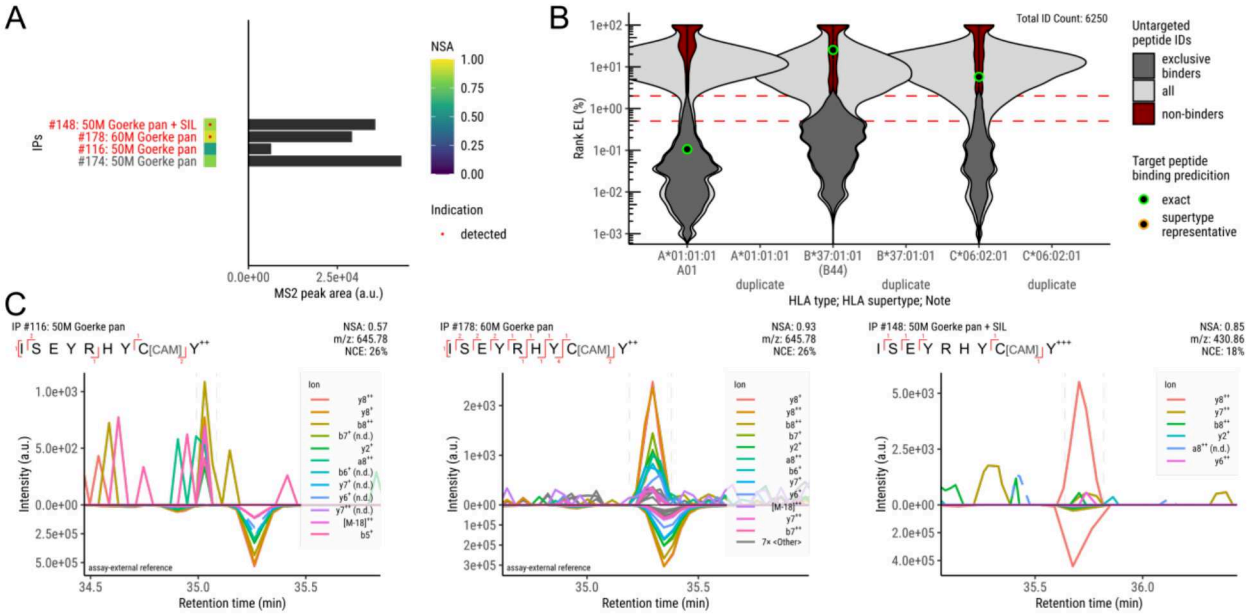


Figure 51. E6/80–88 ISEYRHYCY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the Goerke immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide ISEYRHYCY is at position 80–88 of the E6 sequence. The position is affected by the known variant mutation H85Y, which this peptide does not possess. It was targeted as a supertype A01 (exp. BA: 01.20±0.96) and A11 (exp. BA: 62.36±12.84) binder. It was only detected as an A01 binder in one cell line, Goerke, with limited reproducibility due to low intensity (Figure 51A). In this cell line, it is predicted as a strong and exclusive binder to the A*01:01 allele, confirming the specificity of the detection as a supertype A01 detection (Figure 51B). XICs from three replicates are shown below (Figure 51C). The peptide was initially targeted for the +2 and +3 charge states, with +3 being the primary choice due to slightly higher MS2 intensity in the synthetic tests. The initial trace-level signal did only show for the +2 charge state, with +3 showing no signal (not shown). The signal also did not satisfy the criteria for a valid detection. In a follow-up experiment strong evidence was gathered, again with the +2 charge state. The SIL spike-in experiment was performed with the +3 charge state, due to the strong y_6^{++} ion. It also showed trace-level signal with exact coelution at NSA ≥ 0.85. Therefore, the peptide was determined to be presented by the supertype A01.

The peptide:HLA-A*01 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018).

3.4.4.14 E6 L90V/89–99: SVYGTTLEQQY

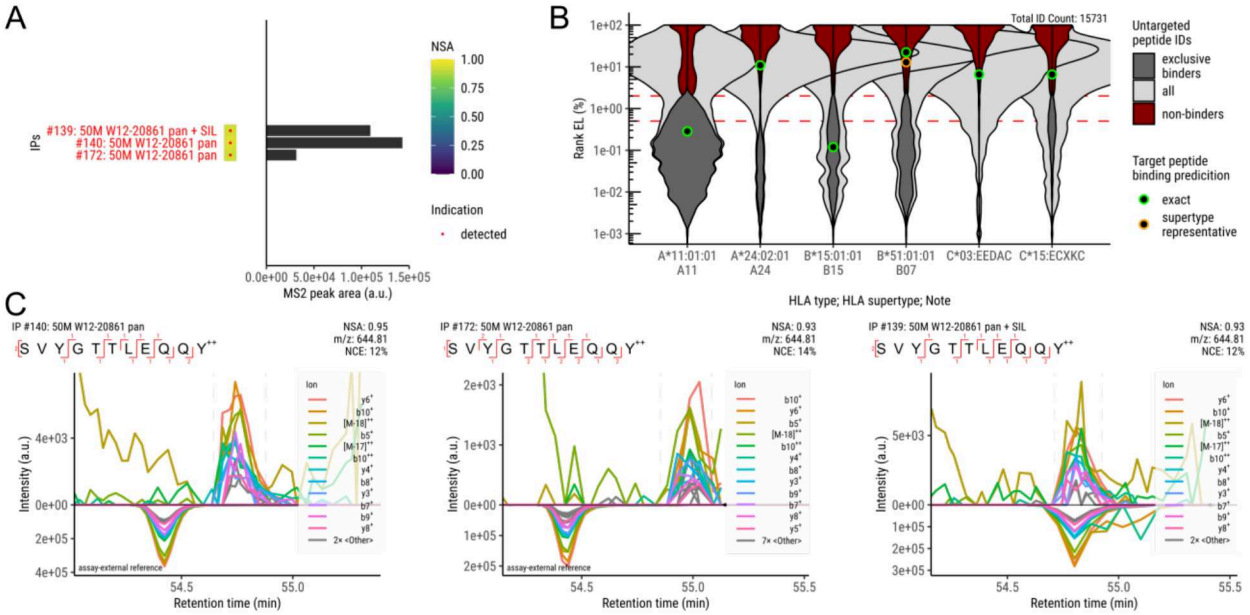


Figure 52. E6/89–99 SVYGTTLEQQY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide SVYGTTLEQQY is at position 89–99 of the E6 sequence. The position is affected by the known variant mutation L90V, which this peptide possesses. It was targeted as a supertype B15 (exp. BA: 07.51±2.18) binder and was detected in one cell line, W12-20861 (Figure 52A). In this cell line, it is predicted as a strong binder to the B*15:01 allele, but also to the A*11:01 allele (Figure 52B). Binding to A*11:01 was rejected in an experimental binding experiment. This confirms the detection being specific to the B15 supertype. XICs from three replicates are shown below (Figure 52C). Detections were at moderate intensity and with good sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.15 E6/93–101: TTLEQQYNK

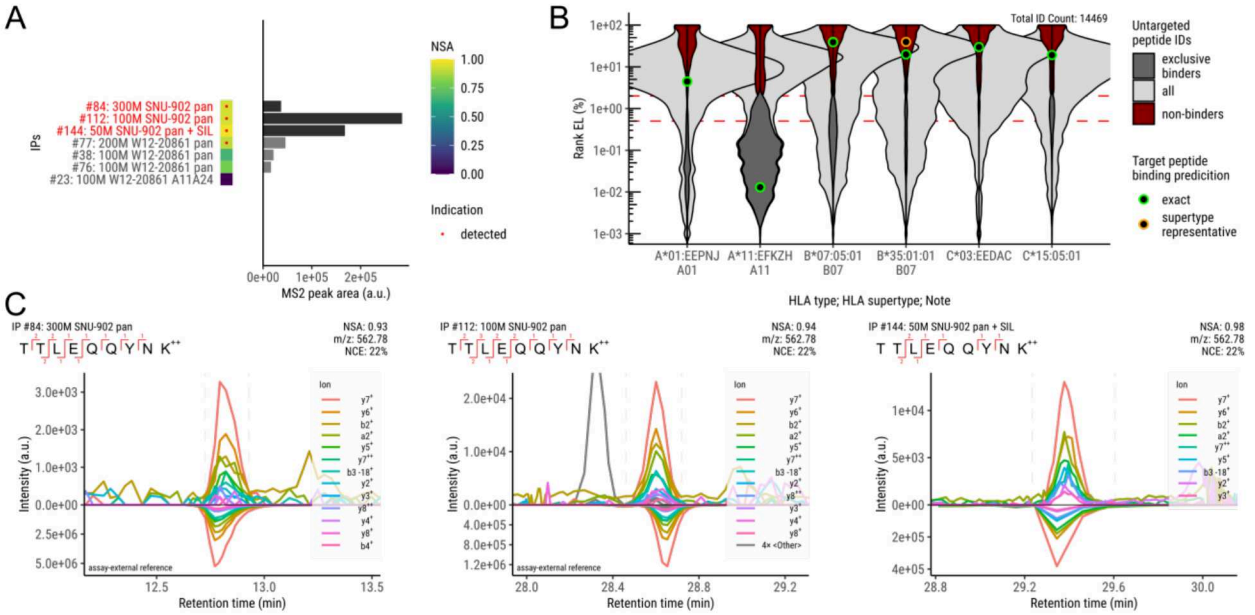


Figure 53. E6/93–101 TTLEQQYNK targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the SNU-902 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide TTLEQQYNK is at position 93–101 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A03 (exp. BA: 15.55±5.05) and A11 (04.71±1.74) binder and was detected in the cell line SNU-902, with cell line W12-20861 showing characteristic signal that was too weak to confirm conclusively (Figure 53A). In the cell line SNU-902, the peptide was targeted for the A11 supertype (Figure 53B). It is predicted as an exclusive strong binder to the group of alleles indicated by the MAC A*11:EFKZH, confirming the specificity of the detection as a supertype A11 detection. XICs from three replicates are shown below (Figure 53C). Detections were at moderate intensity and with good sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A11.

Initially, only the related peptide:HLA-A*03 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018). Later, the peptide:HLA-A*11 combination was found to induce IFNγ release in PBMCs from healthy donors, leading to the isolation of a specific TCR, which enabled specific killing of HPV16-transformed cell lines (Xiong et al., 2022). A chimeric TCR has also been produced (Wang et al., 2021).

3.4.4.16 E6/125–133: HLDKKQRFH

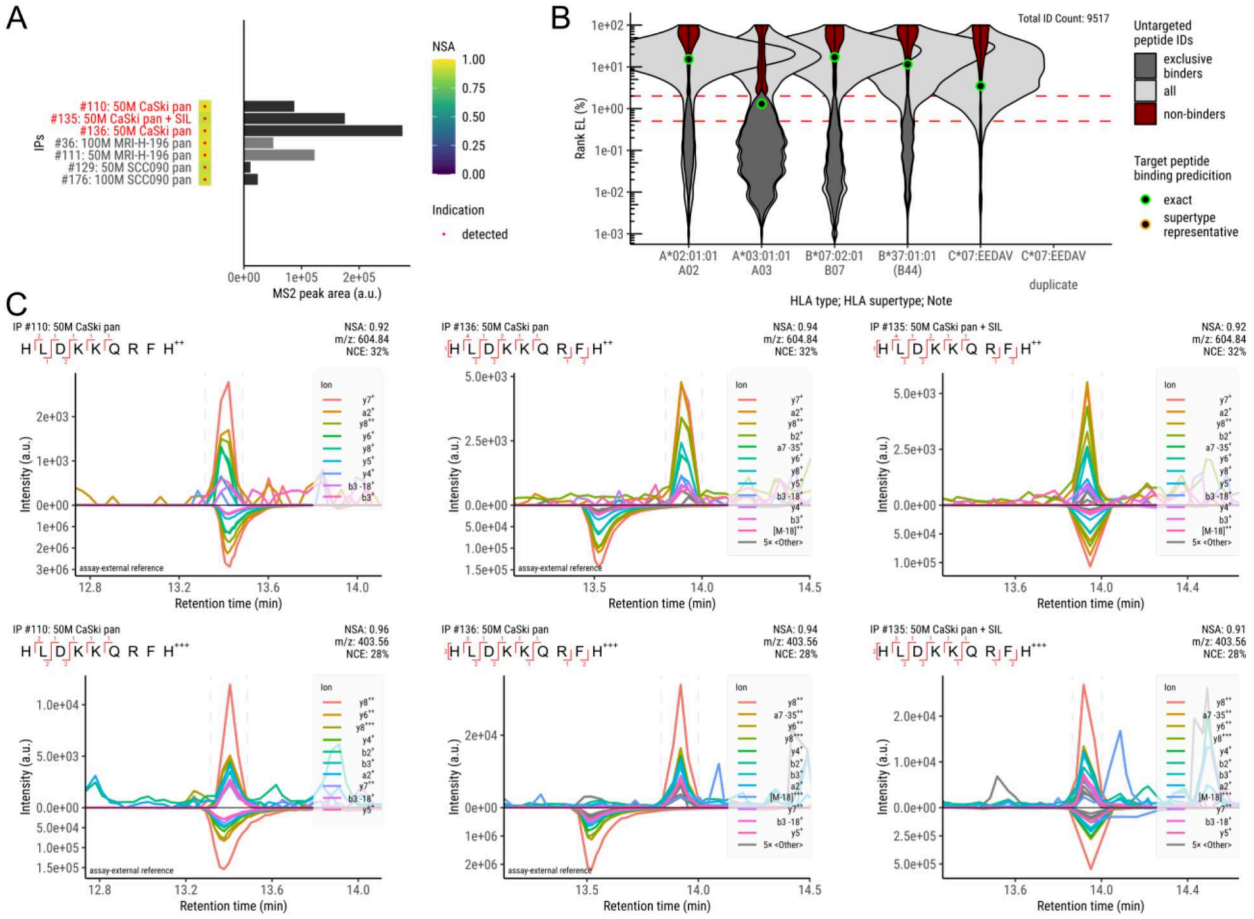


Figure 54. E6/125–133 HLDKKQRFH targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the CaSki immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide HLDKKQRFH is at position 125–133 of the E6 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A03 (exp. BA: 29.58±13.63) binder and was detected in three cell lines (Figure 54A). In the cell line CaSki, it is predicted as an exclusive binder to the A*03:01 allele, confirming the specificity of the detection as a supertype A03 detection (Figure 54B). XICs from three replicates are shown below (Figure 54C). The peptide was detectable with charge states +2 and +3 at similar intensities and with good sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A03.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.17 E7/2–11: HGDTPTLHEY

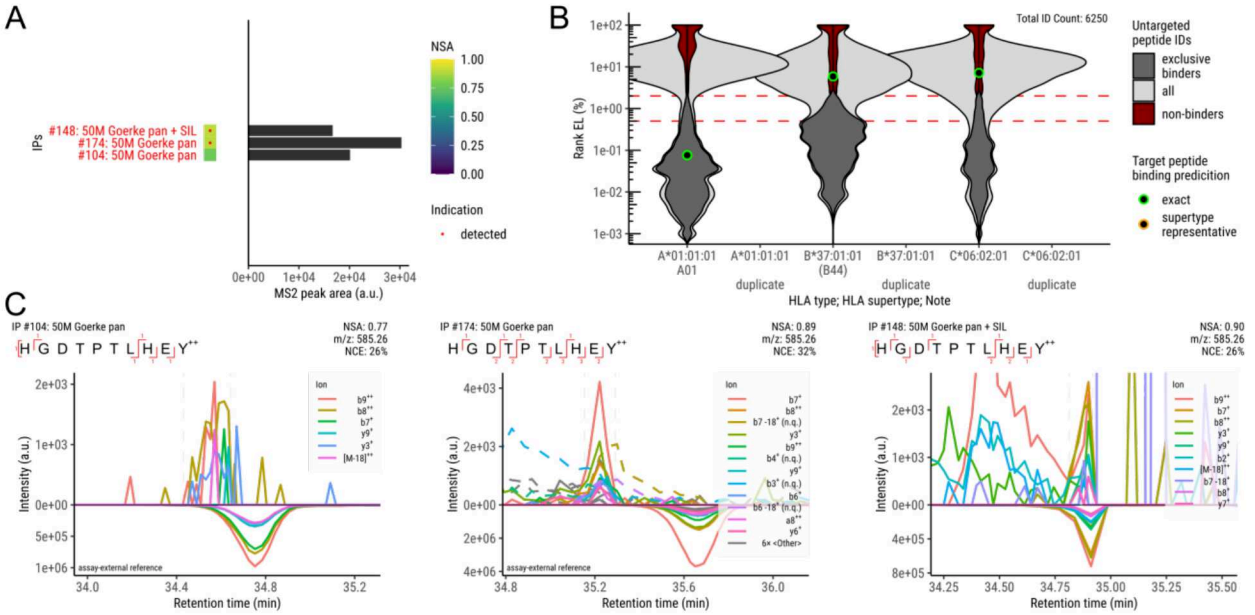


Figure 55. E7/2–11 HGDTPTLHEY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the Goerke immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide HGDTPTLHEY is at position 2–11 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A01 (exp. BA: 01.20±0.96) binder and was detected in one cell line, Goerke, with limited reproducibility due to low intensity (Figure 55A). In this cell line, it is predicted as a strong and exclusive binder to the A*01:01 allele, confirming the specificity of the detection as a supertype A01 detection (Figure 55B). XICs from three replicates are shown below (Figure 55C). Intensities were extremely low, so that the initial trace evidence of the peptide did not satisfy the criteria for a clear detection. The characteristic signal led to follow-up experiment showing stronger evidence somewhat perturbed by coeluting ions. The SIL spike-in experiment showed even stronger perturbation but confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A01.

The peptide:HLA-A*01 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018).

3.4.4.18 E7/15–23: LQPETTDLY

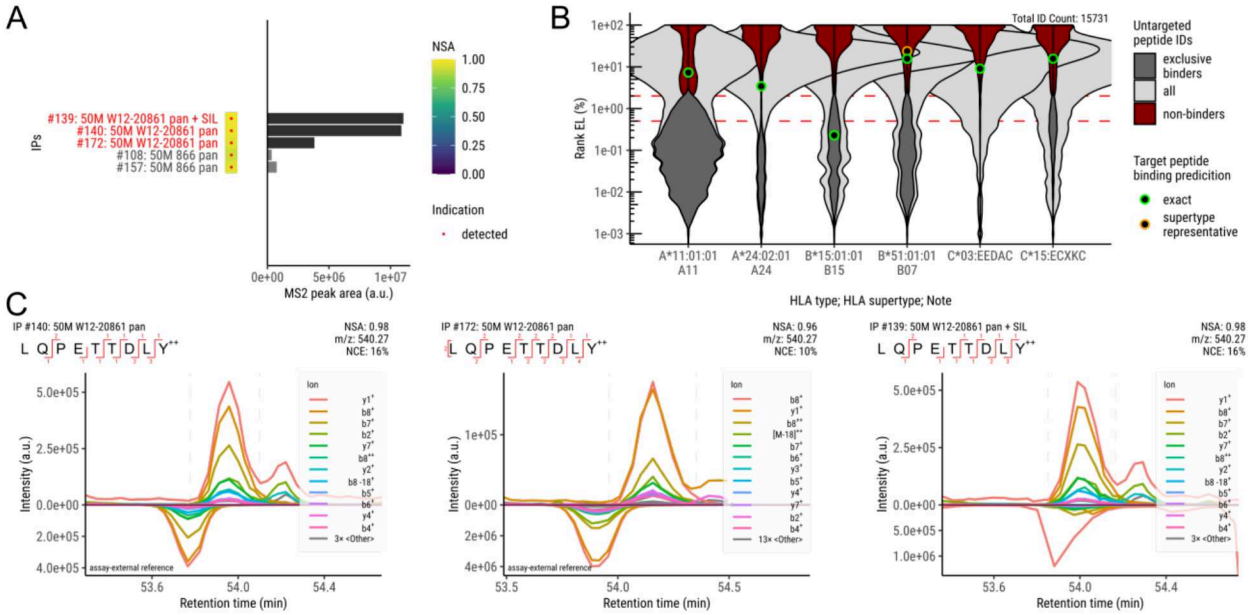


Figure 56. E7/15–23 LQPETTDLY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide LQPETTDLY is at position 15–23 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B15 (exp. BA: 04.68±1.28) binder and was detected in two cell lines (Figure 56A). In the cell line W12-20861, it is predicted as an exclusive strong binder to the B*15:01 allele, confirming the specificity of the detection as a supertype B15 detection (Figure 56B). XICs from three replicates are shown below (Figure 56C). Detections were at moderate intensity and with good sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

The peptide:HLA-B*15 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018).

3.4.4.19 E7/15–24: LQPETTDLYC

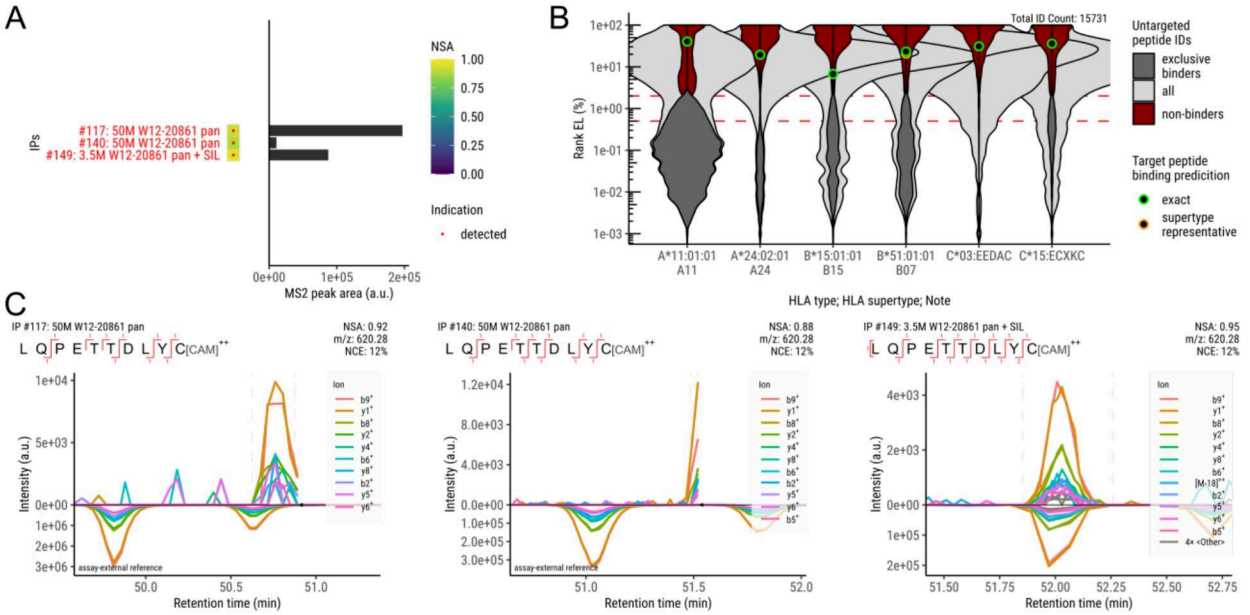


Figure 57. E7/15–24 LQPETTDLYC targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide LQPETTDLYC is at position 15–24 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B15 (exp. BA: 38.45±10.85) binder and was detected in one cell line, W12-20861 (Figure 57A). In this cell line, it does not pass the threshold of a predicted binder for any of the present HLA (Figure 57B). Due to the experimentally verified binding, the detection can be specifically attributed to the B*15:01 allele. XICs from three replicates are shown below (Figure 57C). Detections were at moderate intensity with high sequence coverage but the reference comparison showed a strong discrepancy on the retention time axis. This also led to a near-miss of the peak in IP #140, as the targeting of the precursor was scheduled to stop. Due to this, the peak was cut off half way but shows specific signal. The SIL reference showed two distinct peaks which likely reflect disparate conformations during the LC separation, of which the first one is dominant and was selected for targeting. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.20 E7/15–25: LQPETTDLYCY

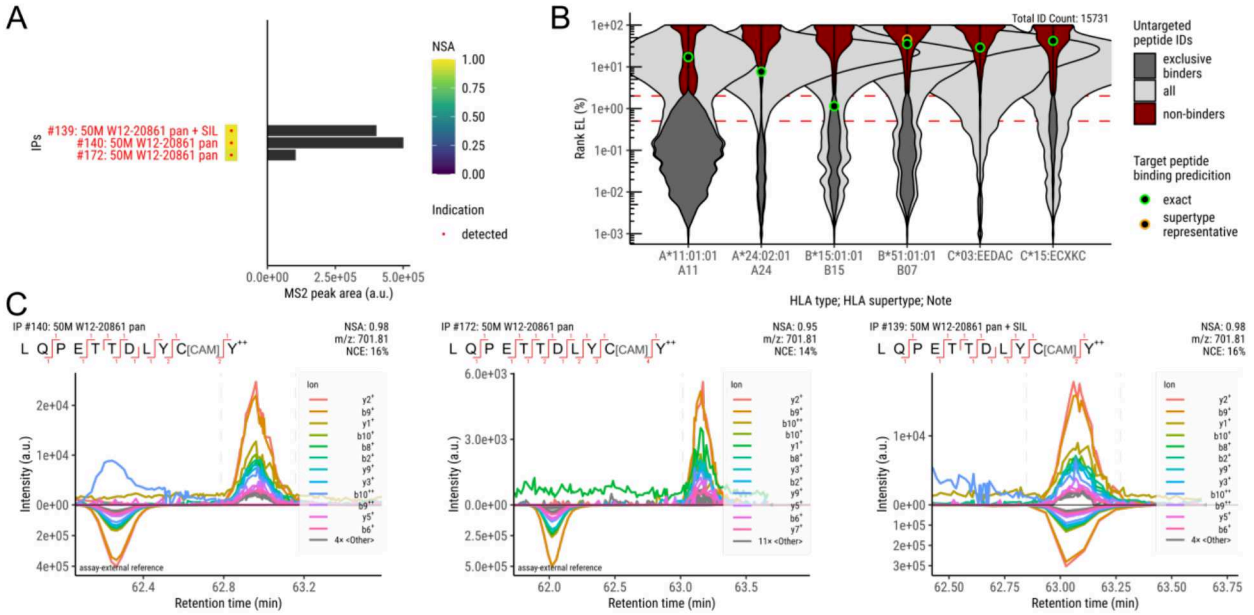


Figure 58. E7/15–25 LQPETTDLYCY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide LQPETTDLYCY is at position 15–25 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype B15 (exp. BA: 07.30±2.00) binder and was detected in one cell line, W12-20861 (Figure 58A). In this cell line it is predicted as an exclusive binder to the B*15:01 allele, confirming the specificity of the detection as a supertype B15 detection (Figure 58B). XICs from three replicates are shown below (Figure 58C). Detections were at moderate intensity with high sequence coverage but the reference comparison showed a strong discrepancy on the retention time axis. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.21 E7/43–52: GQAEPDRAHY

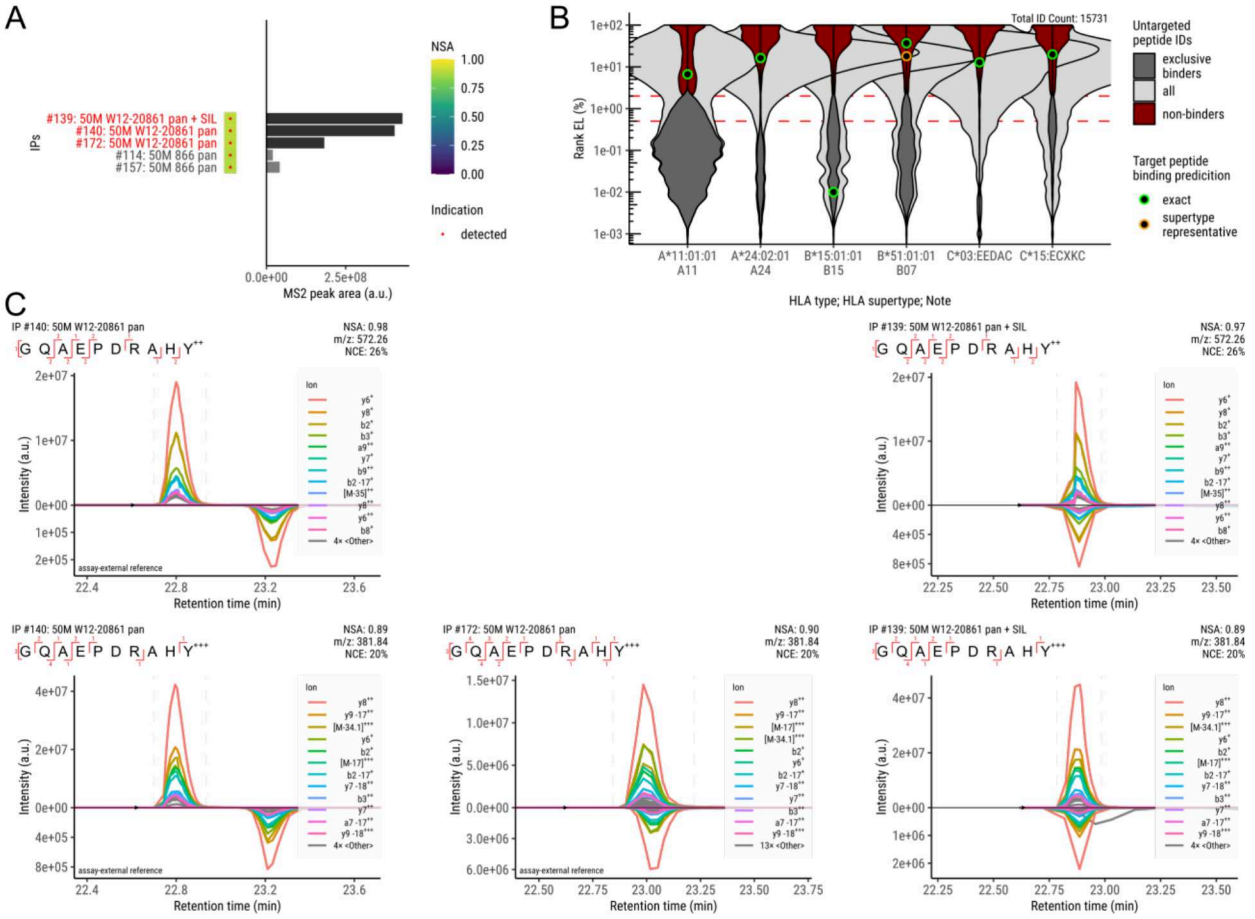


Figure 59. E7/43–52 GQAEPDRAHY targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the W12-20861 immunopeptidome. (C) XICs for peptide detections indicated in (A). In IPs #140 and #139, the peptide was detectable with charge states +2 and +3, simultaneously.

The peptide GQAEPDRAHY is at position 43–52 of the E7 sequence. The position is affected by the known variant mutation H51N, which this peptide does not possess. It was targeted as a supertype B15 (exp. BA: 04.07±1.50) binder and was detected in two cell lines (Figure 59A). In the cell line W12-20861, it is predicted as an exclusive strong binder to the B*15:01 allele, confirming the specificity of the detection as a supertype B15 detection (Figure 59B). XICs from three replicates are shown below (Figure 59C). Detections were at high intensity. The peptide was detectable with charge states +2 and +3 at similar intensities. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype B15.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.22 E7/66–74: RLCVQSTHV

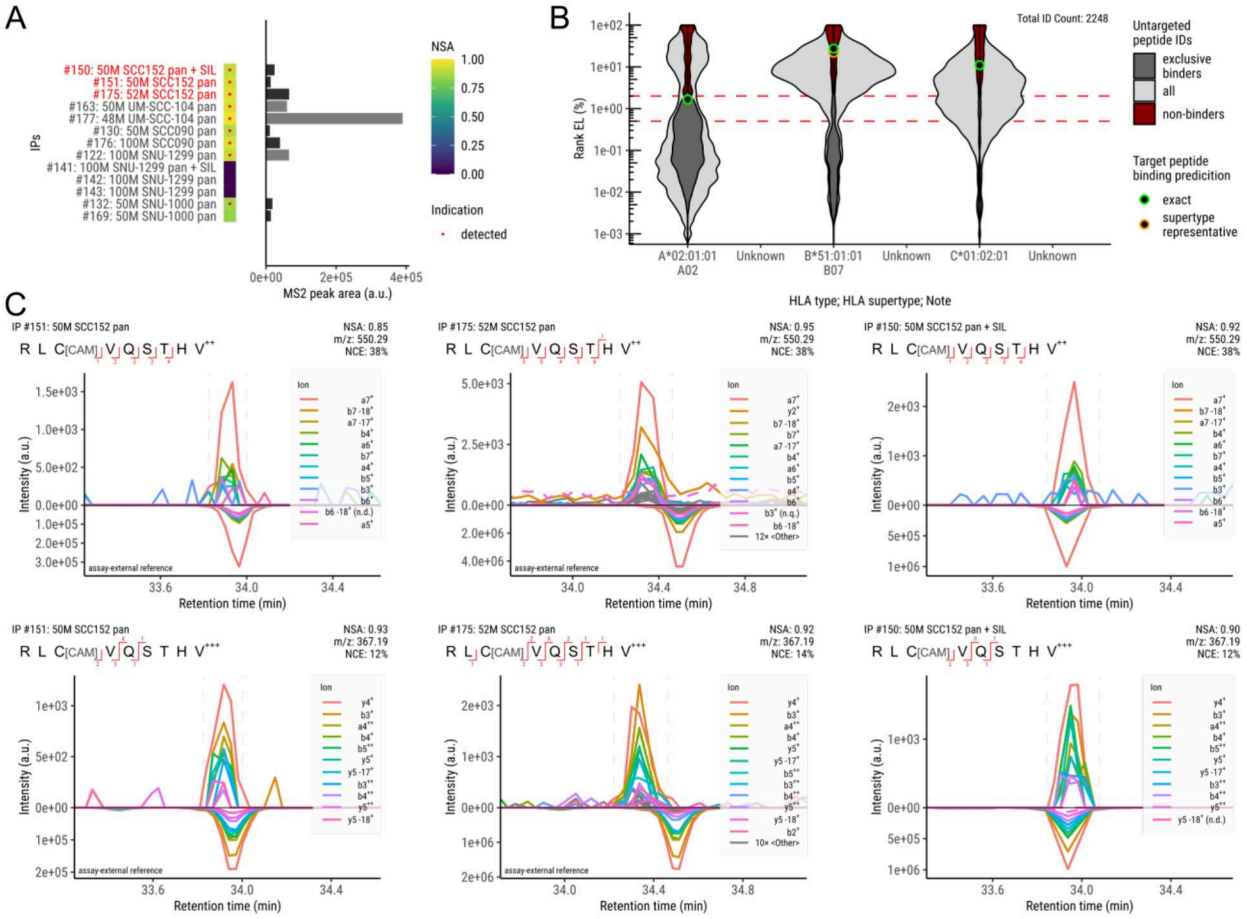


Figure 60. E7/66–74 RLCVQSTHV targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the SCC152 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide RLCVQSTHV is at position 66–74 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A02 (exp. BA: 37.18±10.65) binder and was detected in three cell lines (SCC152, UM-SCC-104, SCC090), with two more cell lines showing characteristic signal that was too weak to confirm conclusively (SNU-1299, SNU-1000, Figure 60A). In the cell line SCC152, it is predicted as an exclusive binder to the A*02:01 allele, confirming the specificity of the detection as a supertype A02 detection (Figure 60B). XICs from three replicates are shown below (Figure 60C). The peptide was detectable with charge states +2 and +3 at similar intensities and with moderate sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A02.

There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.23 E7/77–86: RTLEDLLMGT

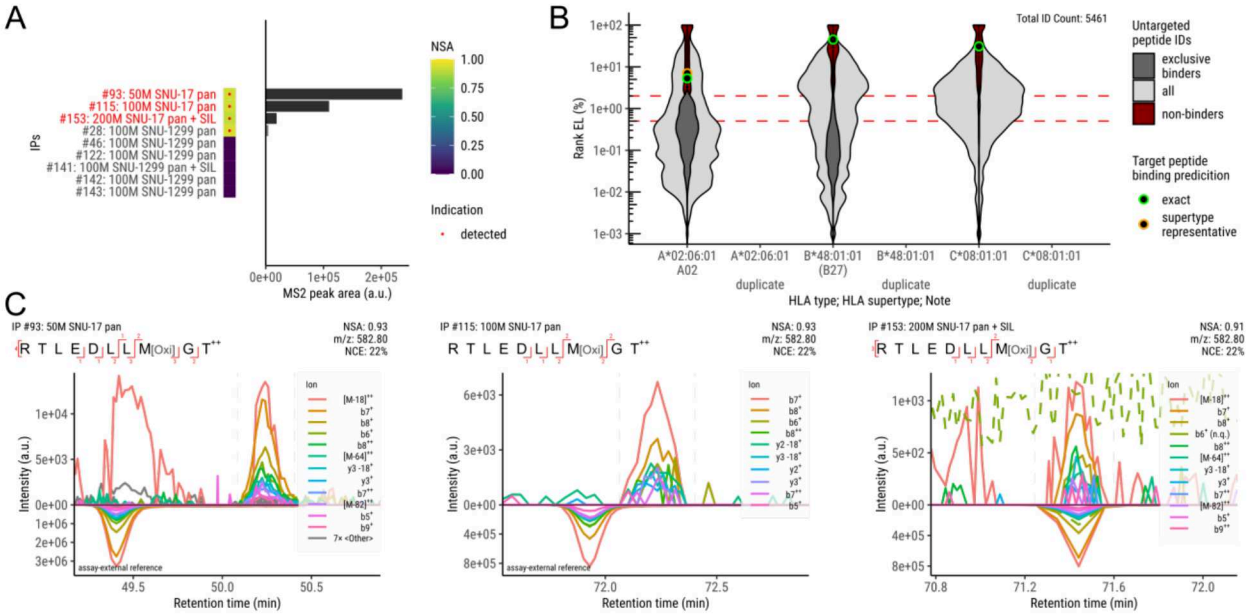


Figure 61. E7/77–86 RTLEDLLMGT targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the SNU-17 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide RTLEDLLMGT is at position 77–86 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A02 (exp. BA: 72.72±17.36) binder. In contrast to the preceding study, it was not detected again in the cell line CaSki. Instead, it was readily detectable in the cell line SNU-17 and seen only irreproducibly in SNU-1299, where it was only observed once at very low intensity (Figure 61A). In the SNU-17 cell line, the peptide was targeted for the A02 supertype (Figure 61B). The cell line has the A*02:06 allele. In comparison to the reference allele A*02:01, which is also the allele present in CaSki, it is a predicted to be a moderately better binder (green vs. orange dot). Nevertheless, it is not predicted to be a binder to any of the present HLA alleles and the detection is therefore clearly attributable due to the experimentally confirmed binding with the A02 supertype representative allele. XICs from three replicates are shown below (Figure 61C). Initial detection at low intensity and with moderate sequence coverage showed a strong discrepancy on the retention time axis. The SIL spike-in experiment showed overall weaker signal but confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A02.

This peptide was MS-detected in the preceding study, based on the same peptide:HLA-A*02 combination (Blatnik et al., 2018). This combination was also previously found to induce IFNγ release and cytolysis in PBMCs from healthy donors (Tsang et al., 2017; Blatnik et al., 2018).

3.4.4.24 E7/89–97: IVCPICSQK

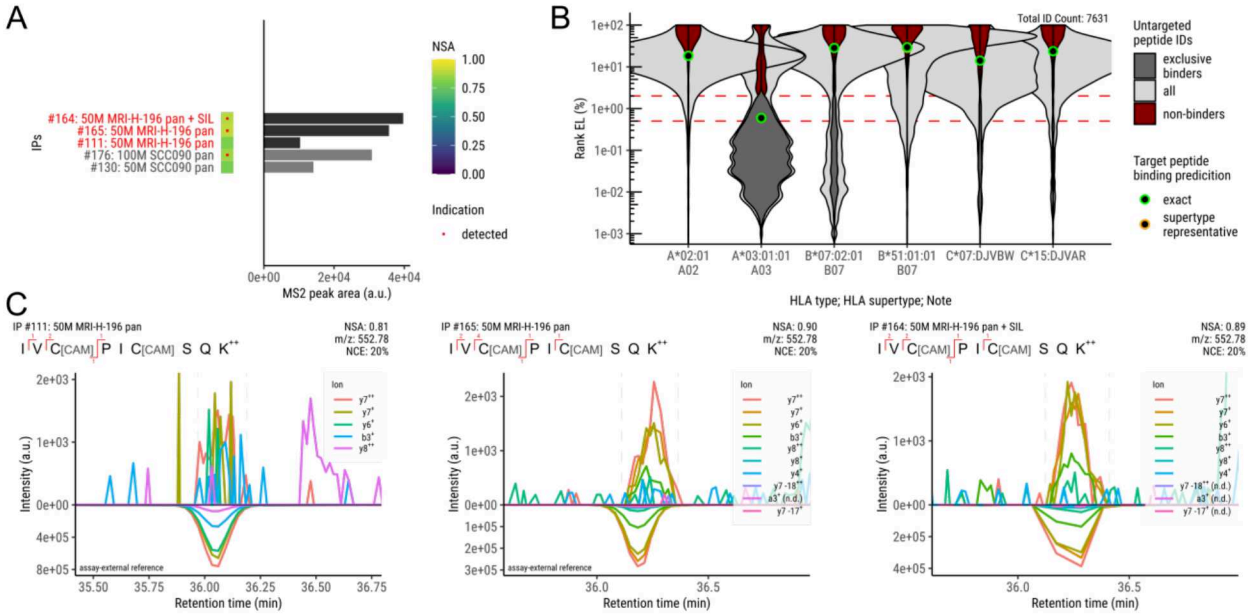


Figure 62. E7/89–97 IVCPICSQK targeted detections. (A) All relevant epitome map IPs where traces of the peptide were detected. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which cell line is shown in (B) and which results are shown in (C). (B) Peptide binding predictions overlaid on the untargeted analysis of the MRI-H-196 immunopeptidome. (C) XICs for peptide detections indicated in (A).

The peptide IVCPICSQK is at position 89–97 of the E7 sequence. The position has not been found to be affected by any known variant mutations. It was targeted as a supertype A03 (exp. BA: 04.06±1.38) and A11 (exp. BA: 03.16±0.78) binder and was detected in cell line MRI-H-196, with cell line SCC090 showing characteristic signal that was too weak to confirm conclusively (Figure 62A). In the cell line MRI-H-196, the peptide was targeted for the A03 supertype (Figure 62B). It is predicted as an exclusive binder to the A*03:01 allele, confirming the specificity of the detection as a supertype A03 detection. XICs from three replicates are shown below (Figure 62C). Detections were at low intensity with low sequence coverage. The SIL spike-in experiment confirmed exact coelution. Therefore, the peptide was determined to be presented by the supertype A03.

The peptide:HLA-A*03 combination has not previously been found through MS detection or immunological tests. However, the related peptide:HLA-A*11 combination was previously found to induce IFNγ release in PBMCs from HNSCC patients (Krishna et al., 2018). This combination was also previously found to induce IFNγ release in PBMCs from healthy donors, leading to the isolation of a specific TCR, which enabled specific killing of HPV16-transformed cell lines (Xiong et al., 2022).

3.4.4.25 E6/8–17: MFQDPQERPR — Confirmation of untargeted detection

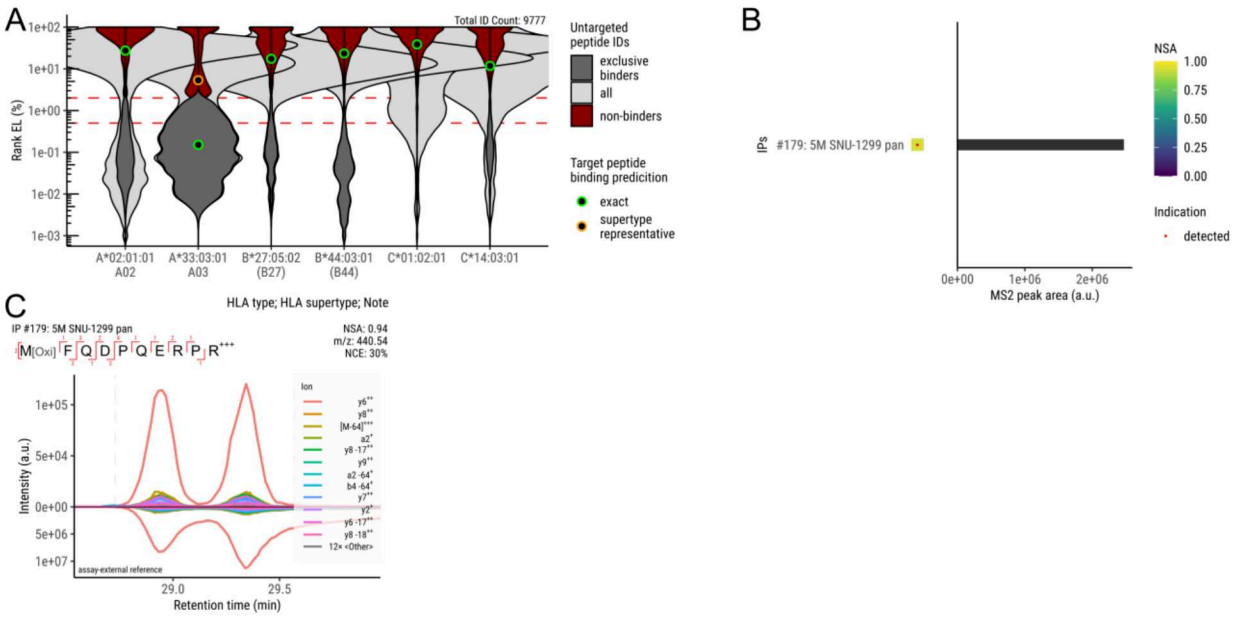


Figure 63. E6/8–17 MFQDPQERPR untargeted detection and targeted validation. (A) The untargeted experiment with cell line SNU-1299 that resulted in the identification of the peptide by the search engine. (B) The epitome map IP where the untargeted detection was validated in a targeted acquisition. Specification for the IP contains: ID, cell load, cell line, antibody, SIL spike-in. (C) XIC for the targeted peptide detection.

The peptide MFQDPQERPR is at position 8–17 of the E6 sequence. The position is affected by the known variant mutations R17I and R17T, which this peptide does not possess. It was initially detected in the untargeted acquisition of cell line SNU-1299 at a reported q value of 2×10^{-14} and at a retention time of 29.05 min. It is a predicted strong binder for the allele A*33:03, which belongs to supertype A03 (Figure 63A). Due to the fact that it is not predicted to bind to the A03 supertype representative allele (A*03:01), there is no experimental binding data available for this peptide. A targeted IP was performed with the small remaining aliquot of the same IP sample to confirm the detection through comparison to the synthetic SIL reference (Figure 63B–C). Due to the strong signal, detection was readily confirmed at high intensity and with full sequence coverage. Therefore, the peptide was determined to be presented by the supertype A03.

However, it was not determined whether the peptide is presented by the more common supertype representative allele A*03:01 (world population coverage: 16.8 % vs. 4.8 % for A*33:03). There are no previous findings through MS detection or immunological tests for this peptide.

3.4.4.26 E7/45–52: AEPDRAHY — Rejection of untargeted detection

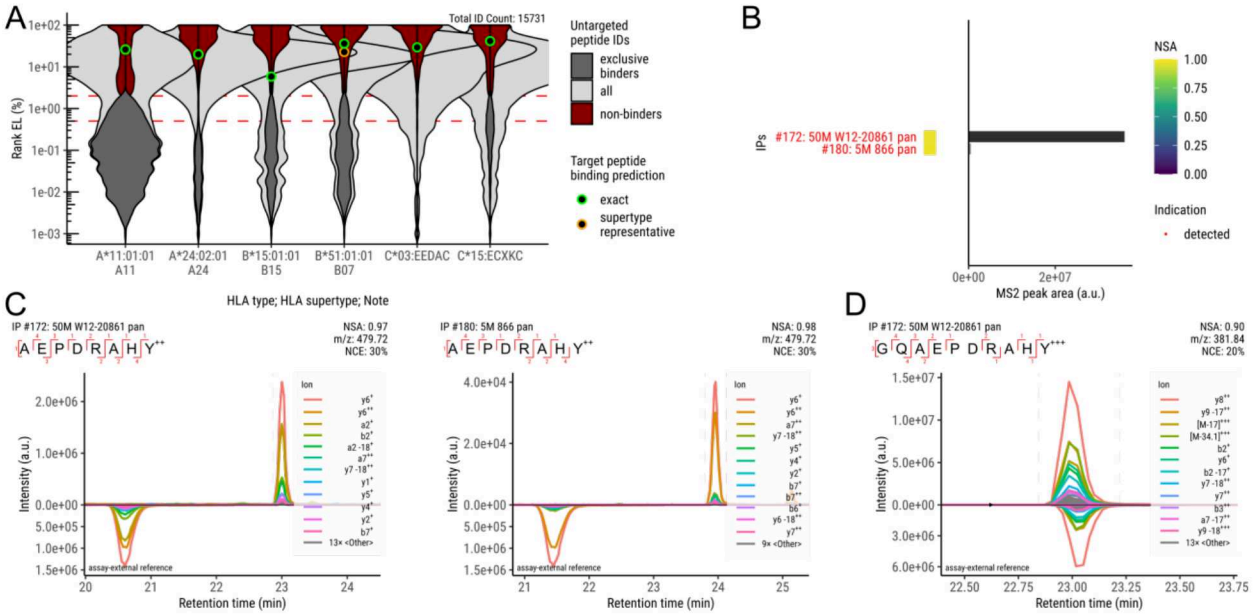


Figure 64. E7/45–52 AEPDRAHY untargeted detection and targeted rejection. (A) The untargeted experiment with cell line W12-20861 that resulted in the identification of the peptide by the search engine. (B) The epitome map IPs where the untargeted detection was validated in a targeted acquisition. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. Changes in bar shade indicate next cell line. Red text indicates which results are shown in (C). (C) XICs for the targeted peptide acquisitions. (D) XIC for related peptide GQAEPDRAHY.

The peptide AEPDRAHY is at position 45–52 of the E7 sequence. The position is affected by the known variant mutation H51N, which this peptide does not possess. It was initially detected in the untargeted acquisitions of cell lines W12-20861 and 866 at reported q values of 6.2e-8 and 1.5e-30 and at retention times of 23.0 min and 23.9 min, respectively. It is not predicted to bind to any of the HLA alleles in the cell line W12-20861 (Figure 64A), with the same being true for cell line 866 (data not shown). Experimental binding experiments showed that this peptide does not bind to B*15:01. Confirmation of the detection through a targeted IP, enabling comparison to the synthetic SIL reference, was performed with both cell lines (Figure 64B). The cell line 866 targeted experiment was performed with a small aliquot of the untargeted IP sample. I detected highly specific MS signal for the peptide in both replicates, with full sequence coverage (Figure 64C). On the LC-axis, however, a strong discrepancy was observed, beyond the more common shifts of up to one minute. Upon inspection, I found that the LC-MS peak in IP #172 coelutes exactly (at retention time 23.0 min) with the LC-MS peak for the longer peptide GQAEPDRAHY, which is ca. three orders of magnitude stronger (Figure 64D). Therefore, I concluded that the untargeted detection was the result of in-source fragmentation, which has been described as one of the pitfalls peptide identification in immunopeptidomics (Fritsche et al., 2021). This is a phenomenon where an analyte will undergo LC separation as one species but is then partially fragmented already during the electrospray ionization. Due to this, the MS may truly acquire specific signal for a peptide, which is not in this form present in the sample. Therefore, the untargeted detection was rejected in the targeted confirmation experiments.

3.4.5 Summary

In total, 25 peptides have been detected. They are summarized in Table 5. Only two of the peptides had previously been detected by MS, namely E7/11–19 (Riemer et al., 2010) and E7/77–86 (Blatnik et al., 2018). Two further peptides, namely E6/29–38 (Riemer et al., 2010) and E7/11–20 (Blatnik et al., 2018) were previously noted only as plausible MS detections due to weak evidence. Of the 27 peptide:HLA combinations, 12 had previously been found to be immunogenic. Using the common combination of supertypes A03 and A11 to A03_A11, this rises to 13 previous findings. Two peptides, namely E6/37–46 and E7/49–57 were detected with two peptide:HLA combinations.

Table 5. Summary of MS-detected peptides and associated literature. * indicates MS detection noted only as plausible due to weak evidence. † indicates that the described peptide:HLA combination is an allotype matching only the the larger A03_A11 combined supertype instead of the respective A03 or A11 classification of the detection.

Peptide	HLA supertype	Cell lines	Previously MS-detected	Previously described
E6/8–17 MFQDPQERPR	A03	SNU-1299	-	
E6/15–22 RPRKLPQL	B07	MRI-H-196, UM-SCC-104	-	
E6/29–38 TIHDIILECV	A02	CaSki2020, 879, SCC090, SCC152, UM-SCC-104	*Riemer et al. (2010)	Ressing et al. (1995) Wentworth et al. (1996) Draper et al. (2015) Santegoets et al. (2022) (many more)
E6/33–41 IILECVYCK	A11	W12-20861, SNU-902	-	Jiang et al. (2022)_
E6/37–46 CVYCKQQLLR	A03	866, SCC090	-	†Krishna et al. (2018)
	A11	W12-20861	-	Krishna et al. (2018)
E6/38–45 VYCKQQLL	A24	SiHa, SNU-1005, UM-SCC-104	-	
E6/41–50 KQQLLRREVY	B15	W12-20861	-	
E6/42–50 QQLLRREVY	B15	W12-20861, 866	-	
E6/53–61 AFRDLCIVY	B15	W12-20861	-	
E6/80–88 ISEYRHYCY	A01	Goerke	-	Krishna et al. (2018)
E6 L90V/89–99 SVYGTTLEQQY	B15	W12-20861	-	
E6/93–101 TTLEQQYNK	A11	SNU-902	-	Wang et al. (2021) Xiong et al. (2022)

E6/125–133 HLDKKQRFH	A03	CaSki2020, MRI-H-196, SCC090	-	
E7/2–11 HGDTPTLHEY	A01	Goerke	-	Krishna et al. (2018)
E7/5–13 TPTLHEYML	B07	CaSki2020	-	Krishna et al. (2018)
E7/11–19 YMLDLQPET	A02	CaSki2020, 866, 879, SNU-703, SNU-1000, SNU-1299, SCC090, SCC152, UM-SCC-104	Riemer et al. (2010) Keskin et al. (2011) Blatnik et al. (2018) Peng et al. (2023)	Ressing et al. (1995) Wentworth et al. (1996) Tsang et al. (2017) Blatnik et al. (2018) Jin et al. (2018) Santegoets et al. (2022) (many more)
E7/11–20 YMLDLQPETT	A02	CaSki2020, UM-SCC-104	*Blatnik et al. (2018)	Ressing et al. (1995) Wentworth et al. (1996) Blatnik et al. (2018) Youde et al. (2000) Santegoets et al. (2022) (many more)
E7/15–23 LQPETTDLY	B15	W12-20861, 866	-	Krishna et al. (2018)
E7/15–24 LQPETTDLYC	B15	W12-20861	-	
E7/15–25 LQPETTDLYCY	B15	W12-20861	-	
E7/43–52 GQAEPDRAHY	B15	W12-20861	-	
E7/49–57 RAHYNIVTF	B07	W12-20861, SNU-902	-	Krishna et al. (2018)
	B15	W12-20861, 866, SNU-1005	-	
E7/66–74 RLCVQSTHV	A02	SCC090, SCC152	-	
E7/77–86 RTLEDLLMGT	A02	SNU-17	Blatnik et al. (2018)	Tsang et al. (2017) Blatnik et al. (2018)
E7/89–97 IVCPICSQK	A03	MRI-H-196	-	†Krishna et al. (2018) †Jiang et al. (2022) †Xiong et al. (2022)

In targeting a total of 239 peptides in 20 cell lines, I completed the initial scope of this thesis. In 16 of the 20 cell lines, at least one peptide was detected, leaving four cell lines, where peptides were detected.

Table 6 summarizes the outcome with regard to the peptides targeted and detected. In preceding experiments, 242 experimentally verified HLA binders for HPV16 oncogenes E6 and E7 were identified. All of these were initial candidates. A majority of these binders are cysteine-containing. Of the extensive collection of HPV16-positive cell lines of the research group, I found 20 cell lines to be suitable for inclusion this study. For each cell line, I assigned candidate peptides by matching HLA types and sequence variants of E6 and E7 against the total list of candidates. In total, 239 HLA binders were matched with a cell line using these criteria. Only three peptides had to be excluded. Two supertype A24 binders including the sequence mutation E7 S63F were not targeted as no cell line matched A24 and E7 S63F (binders: E7 S63F/56–63, E7 S63F/56–65). The B15 binder E6/83–91, based on the canonical E6 sequence, was not targetable due to the prevalence of the H85Y and L90V variants in the cell lines. All candidates were synthesized as SIL variants and I characterized them in DI-MS and LC-MS, so that reference peaks were available. As a result, I targeted all 239 cell line-matched peptides in at least one IP. This led to the LC-MS detection of 24 HLA binders and one additional binder that was not derived from the experimental binder list but was detected in the untargeted experiments. The success-rate for non-cysteine-containing peptides at 15 % was higher than for cysteine-containing peptides at 7 %, resulting in a total success rate of 10 % of candidate HLA binders being validated as HLA-presented peptides. The validated peptide HLA-restrictions cover all six of the complete set of supertypes A01, A02, A03_A11, A24, B07 and B15, for which the cumulative population coverage was determined in section 3.3.

Table 6. Summary of the epitome map outcome. In addition to the total, results for cysteine-containing and non-cysteine-containing peptides are listed separately. The peptide that stems from untargeted experiments and was not originally considered a binder is not counted for percentage calculations (indicated '+1').

Peptides: Candidate verified binders	Total: 242	
	No Cys: 94	Cys: 148
Cell lines	20	
Peptides: matching cell line HLA types & genotypes	Total: 239	
	No Cys: 94	Cys: 145
Peptides: MS-detected	Total: 24 (10 %) + 1	
	No Cys: 14 (15 %) + 1	Cys: 10 (7 %)
Supertypes covered	6 of 6	
Peptides: detected & reactive in ELISpot with healthy donor PBMCs	Total: 14 (58 %)	
	No Cys: 10 (71 %)	Cys: 4 (40 %)

To further contextualize the results, I here include preliminary results on the immunogenicity of all candidate peptides. These were not produced by me and derive from the parallel project in the research group that is part of the epitome map and is led by Dr. Maria Bonsack and Kathrin Wellach. The peptides were tested for their capacity to elicit immune reactions in HLA-matched healthy donor peripheral blood mononuclear cells (PBMCs) using enzyme-linked immunospot (ELISpot) assays. Donors were all female and at least 40 years old to increase the likelihood of a past encounter with HPV16. Positive results in an ELISpot assay, where reactivity against synthetic peptide is tested, are considered an indicator of epitope immunogenicity. The dataset targets the same set of supertypes as were used in this thesis. The total number of positive immunogenic peptides in the dataset is 119, which corresponds to a 49 % rate of observed immunogenicity for the 242 binders. Of the 24 MS-detected peptides that have been tested, immunogenicity is indicated for 14. The fraction of immunogenic peptides is higher for non-cysteine-containing peptides at 71 % than for cysteine-containing peptides at 40 %. Therefore, the cysteine-containing majority of candidate peptides showed a lower success rate for LC-MS detection and in preliminary results also a lower immunogenicity.

To understand how the detected peptides are distributed along the sequences of the proteins E6 and E7, I generated heatmaps for both protein sequences that show the clustering of experimentally verified HLA binders, MS-detected binders and immunogenic binders at each sequence position. Figure 65 shows the sequence for the protein E6. The verified HLA binders show a strong cluster roughly spanning positions 81–88. Remarkably, this abundance of binders did not translate to MS detections. This cluster of peptides is also replicated in the immunogenicity results. A cluster for MS-detected peptides is present spanning positions 37–46 with up to four peptides covering a position. It is striking that this cluster does not stem from a hotspot of HLA binders and does not produce a cluster of immunogenic peptides. Overall, there appears to be an exceptional correlation between HLA binders and immunogenic peptides, whereas MS-detected peptides cluster differently. Deactivating mutations from literature, that are commonly used to ensure safety of vaccine formulations utilizing the whole protein,, are indicated with red marks. The epitope E6/53–61 spans two such positions and is therefore incompatible with vaccine formulations including F54R or L57G (Li et al., 2017).

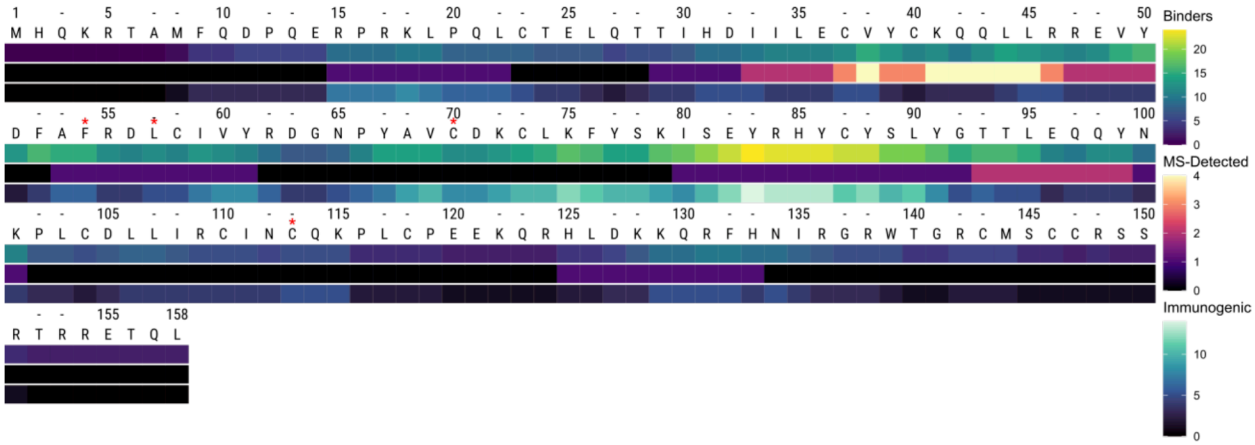


Figure 65. Distribution of experimentally verified binders, MS-detected peptides and immunogenic peptides along the sequence of HPV16 E6. For each position, a count of how many peptides cover that position is shown. Peptides with identical start and end positions but distinct sequence mutations together only contributed one count. 182 binder peptides in total. Only the 12 MS-detected peptides and 81 immunogenic peptides that are a subset of the binders are included. Displayed is the canonical protein sequence from the first start codon that encodes the 158 AA variant of the protein. Red * mark locations that have been used for inactivating mutations in vaccine formulations. Data for row 'Immunogenic' was produced by Dr. Maria Bonsack and Kathrin Wellach.

The analysis of the three peptide classes in Figure 65 was a highly condensed approach that omitted the HLA-restriction of the peptide detections. To understand if the HLA compatibilty of certain sequence regions had an outsized influence on the detection results, I performed a further analysis showing the HLA binders and the resulting MS detections with respect to their HLA-restriction (Figure 66). The cluster of MS-detected peptides from positions 37–47 consists of 7 HLA:peptide combinations for supertypes A02, A03 A11, A24 and B15. Furthermore, binders for A01 were available but not detected. Taken together, detections of these binders derive from 10 cell lines, indicating that this may be a common hotspot for HLA presentation. The availabilty of binders for 6 supertypes was, however, not unusual, as binders are well-distributed along the sequence. *E.g.* the cluster of binders spanning positions 81–88 consist of binders for all of six supertypes A01, A02, A03, A11, A24, B07, B15 but only resulted in one MS detection.

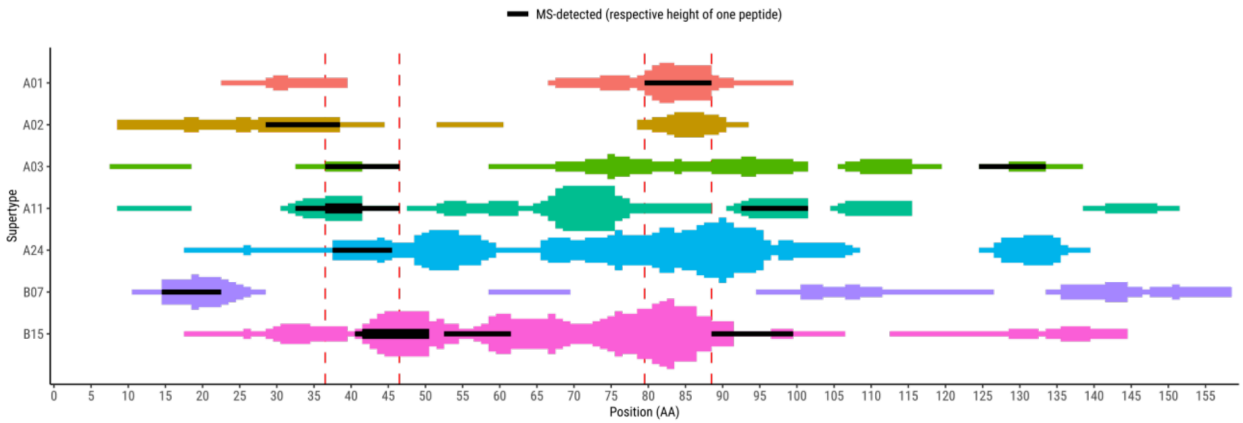


Figure 66. Pile-up plot of the overlapping peptide sequences for the candidate peptides and MS-detected peptides of E6. All candidate peptides are shown in color. Where multiple peptide sequences overlap, the bars grow incrementally wider. MS-detected peptides that are a subset of the binders are overlaid in black. The cluster for MS-detected peptides (37–47) and the cluster for HLA binders (81–88) are indicated with dashed red lines.

Figure 67 shows the sequence for protein E7. The verified binders show three major clusters with again an exceptional correlaton of HLA binders and immunogenic peptides. The main cluster of interest roughly spans positions 11–23. It shows for all three classifications. There is a weak indication for a second cluster around position 50 but it only represents two overlapping MS-detected peptides. A major cluster of HLA binders around position 88 is not reflected in MS detections. As above, deactivating mutations from literature, that are commonly used to ensure safety of vaccine formulations utilizing the whole protein, are indicated with red marks. Three MS-detected peptides overlap the region 23–25, where three inactivating mutations are situated and may therefore be incompatible with vaccine formuations including Y23G, C24G or Y25G. A further artificial mutation at C91G is incompatible with E7/89–97 (Smith et al., 2012; Gan et al., 2014; Li et al., 2017). Overall, for protein E7, the MS-detected peptides more closely reflect the HLA binders and immunogenicity results than for protein E6.

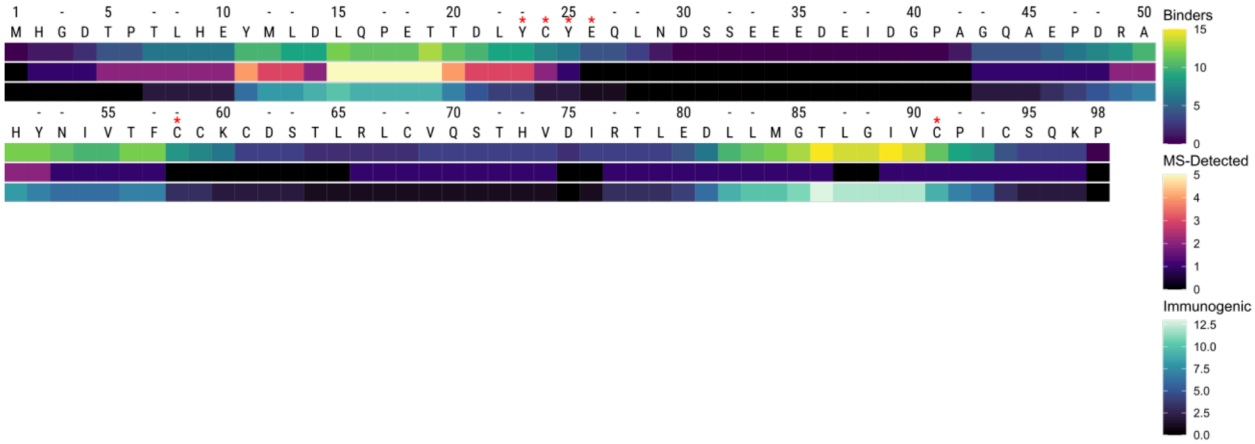


Figure 67. Distribution of experimentally verified binders, MS-detected peptides and immunogenic peptides along the sequence of HPV16 E7. For each position, a count of how many peptides cover that position is shown. Sequence mutations were not considered. Peptides with identical start and end positions but distinct sequence mutations together only contributed one count. 60 binder peptides in total. All 12 MS-detected peptides and only the 38 immunogenic peptides that are a subset of the binders are included. Red * mark locations that have been used for inactivating mutations in vaccine formulations. Data for row 'Immunogenic' was produced by Dr. Maria Bonsack and Kathrin Wellach.

As above, I performed the analysis showing the HLA binders and the resulting MS detections with respect to their HLA-restriction (Figure 68). The cluster of MS-detected peptides from positions 11–23 consists of 7 HLA:peptide combinations for supertypes A01, A02, B07 and B15. Furthermore, binders for A24 were available but not detected. Taken together, detections of these binders derive from 11 cell lines, indicating that this may be a common hotspot for HLA presentation. Here too, no unusual compatibility with multiple HLA types was found at this cluster. *E.g.* the cluster of binders around position 88 consist of binders for five supertypes.

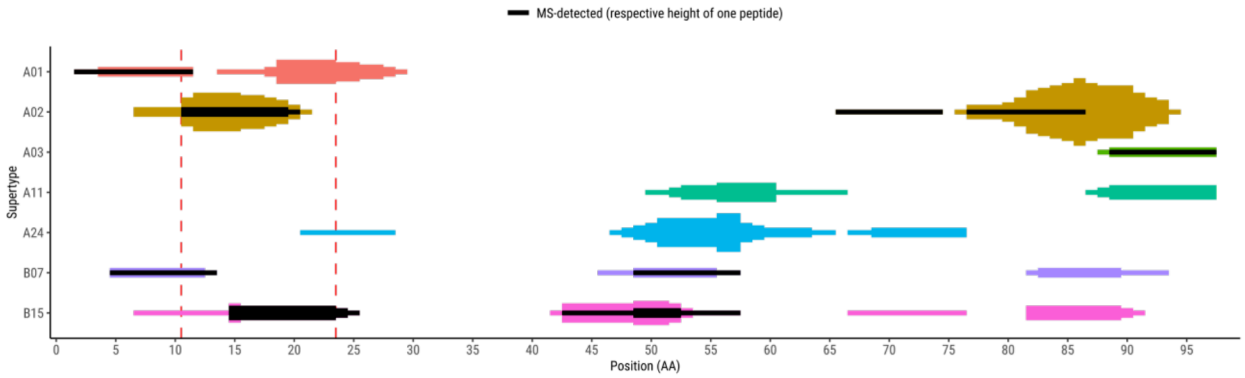


Figure 68. Pile-up plot of the overlapping peptide sequences for the candidate peptides and MS-detected peptides of E7. All candidate peptides are shown in color. Where multiple peptide sequences overlap, the bars grow incrementally wider. MS-detected peptides are overlaid in black. The cluster for MS-detected peptides (11–23) is indicated with dashed red lines.

As I decided to include HNSCC cell lines in an effort to expand the cell line panel and thereby potentially detect more peptides, I determined for the detected peptides whether they were detected in CxCa cell lines, in HNSCC cell lines, or in both. Figure 69 shows the Venn diagrams with the source associations of the cell lines and the detected peptides. All detected peptides, except for the A02 binder E7/66–74 RLCVQSTHV were detected at least once from a CxCa source. The proportion of CxCa to HNSCC for both, the number of cell lines and the number of detected peptides, is very similar, which suggests that the main factor for the different detection counts was the amount of cell lines used.



Figure 69. Result CxCa versus HNSCC association. (A) The type of the cell lines used for the experiments. Distinct by definition. (B) The source types associated with each of the detected peptides. A detected peptide was either detected only in CxCa cells, only in HNSCC cells or in cells of both types.

3.5 Targeted HPV16 epitope detection in patient tumor sample

The detection successes with cell lines posed the question of whether the developed epitope detection workflow is sufficiently sensitive to detect HPV16-derived HLA-presented peptides also in tumor samples. The procurement of a suitable patient tumor tissue sample from the NCT Heidelberg allowed me to perform this test. The sample is from a recurrent neuroendocrine tumor of the cervix.

Whole genome sequencing (WGS) had been performed at the DKFZ, which confirmed the presence of HPV16 and was used to determine HLA types (Supertypes A01, A02, B07). To test which HPV16 E6 and E7 sequence variants are present, I used 4498 WGS reads that were assigned to the complete HPV16 genome. I aligned these to the HPV16 reference sequence using BLAST. Figure 70 presents the alignment results for genes E6 (panel A) and E7 (panel B). The BLAST pile-up view of the aligned reads shows single nucleotide polymorphisms (SNPs) in red. Both gene sequences are fully covered by the aligned reads. I identified a single SNP T350G for E6, which corresponds to the amino acid substitution L90V. The E7 sequence was determined to be the canonical sequence.

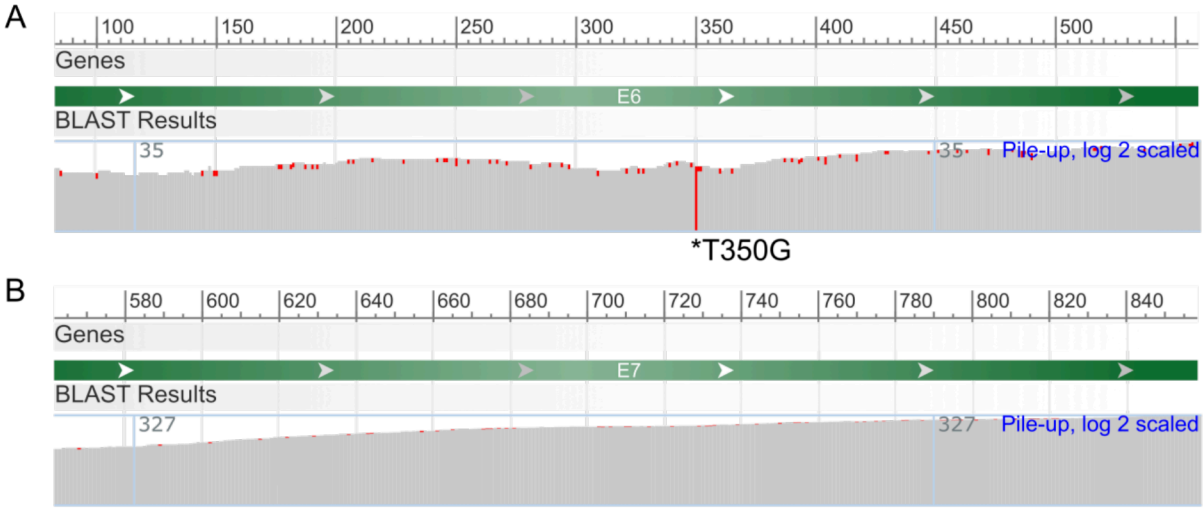


Figure 70. Blast results for patient WGS reads that were assigned to HPV16. Below the respective genes (green bars), a pile-up view of the aligned reads is shown. SNPs of the reads are marked in red. (A) Section spanning the full E6 gene from base pairs (bps) 83 to 559. The asterisk marks the SNP T350G of the genome that corresponds to E6 L90V (B) Section spanning the full E7 gene from bps 562 to 858.

The tissue sample had a weight of 1031.3 mg and the tumor cell content was specified to be 30 % alive cells and 60 % dead cells, with further 10 % of cells being stroma. For the IP, the tissue was homogenized and pre-cleared. Patient sample immunopeptidomics experiments are typically single shot experiments and to account for this here, I allotted a small aliquot (10 %) for untargeted assessment and two aliquots for technical replicates of targeted acquisition (2× 45 %).

The untargeted acquisition resulted in the detection of 1506 peptides with a low proportion of non-binders, indicating a successful IP experiment (Figure 71A). In full agreement with the HLA-typing, the analysis confirmed supertype A01, A02 and B07 surface presentation, with minor contribution from the C alleles. The resulting target list consists of 70 targets, with 15 A01, 31 A02 and 24 B07 targets (Figure 71B). In addition, two control peptides (as described in section 2.2.4.13) are indicated with '+' and '-'. The control peptides showed the expected outcomes. Replicate R1 indicated no valid peptide detection. LLMGTLGI, a B07 binder, was rejected due to also appearing in the negative controls. YMLDLQPET (E7/11–19), a A02 binder, only showed trace level evidence with an NSA below the imposed threshold for a valid detection (Figure 71C). For the second replicate, I reduced the target list drastically to enable acquisition of the most plausible targets with maximum

sensitivity. This approach produced an XIC for YMLDLQPET with much better evidence. Better temporal resolution and maximum resolution scans produced a high NSA of 0.93. Full sequence coverage is achieved, likely due to the combined effect of long injection time and maximum resolution (480k) acquisition. Therefore, the A02 binder E7/11–19 YMLDLQPET was determined to be presented in the patient tumor tissue.

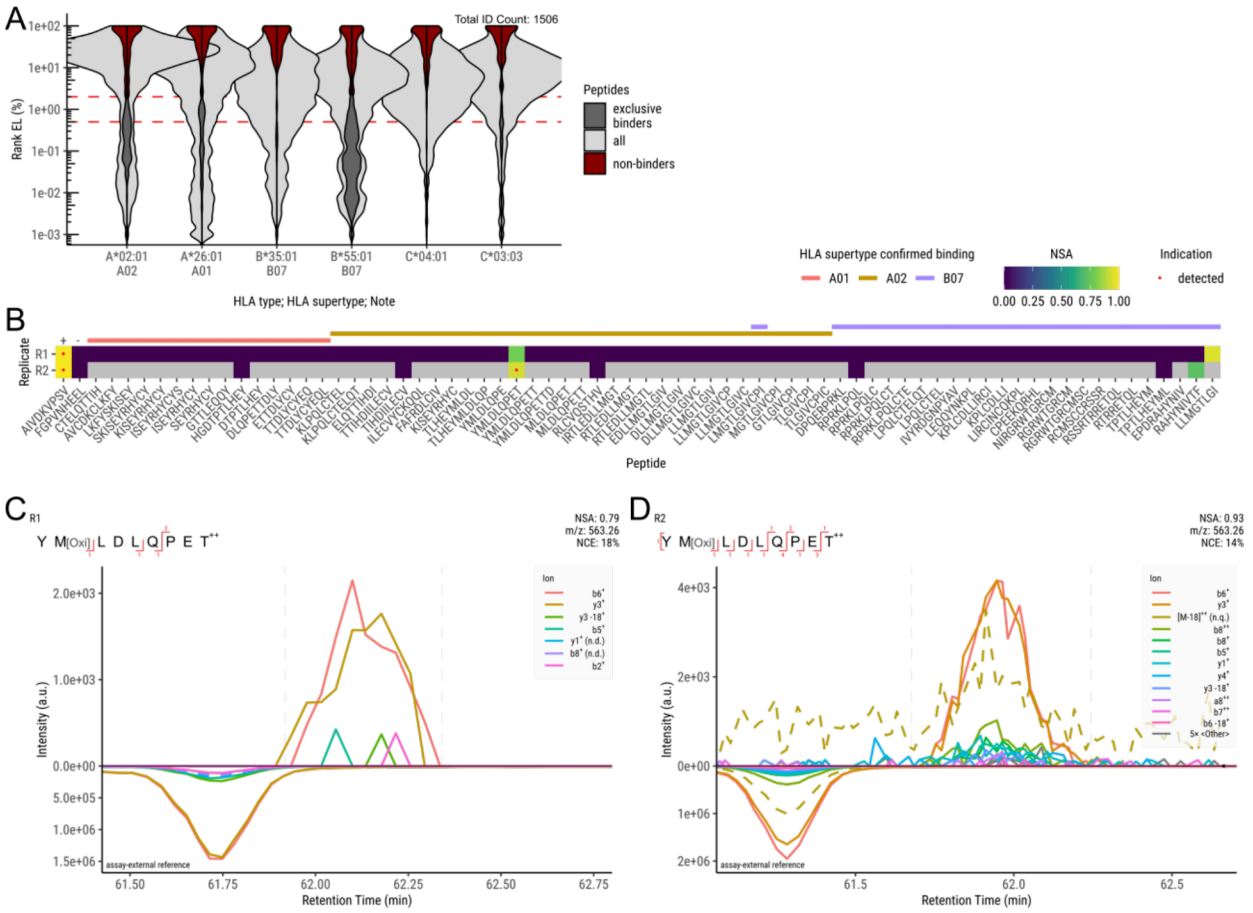


Figure 71. Immunopeptidomics results for patient tumor sample. (A) Untargeted evaluation of the sample immunopeptidome. (B) NSA results of targeted acquisition technical replicates R1 and R2. Where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the peptide to HLA supertype associations from experimental binding tests. Two control peptides are indicated with '+' and '-'. Peptides are sorted in the following order by: HLA association, source protein (E6/E7) and sequence position. (C) XICs for the targeted peptide acquisitions of precursor YM[Oxi]LDLQPET^{++} for peptide E7/11–19.

4 Discussion

4.1 Peptide detections

This thesis aimed at identifying HPV16 oncoprotein E6- and E7-derived peptides that are presented on naturally HPV-transformed cell lines in an effort to inform therapeutic vaccine design. This was enabled first, by the successful establishment of a sensitive targeted immunopeptidomics assay using the parallel reaction monitoring (PRM) acquisition scheme on the hybrid quadrupole-Orbitrap mass spectrometer. In comparison to the method of the preceding study from our lab (Blatnik et al., 2018), which used a time-consuming and labor-intensive MS3 technique, this allowed for the necessary scalability while maintaining specificity due to the high resolution and high mass-accuracy of the Orbitrap.

The application of the newly developed technique for the complete immunopeptidomics part of the HPV16 epitome map project allowed for the effective targeting of 239 distinct peptides in 20 cell lines. The MS detection of the targeted peptides was an overall successful endeavor. Of the 25 detected peptides, only two had clearly been MS-detected in prior experiments, namely E7/11–19 (YMLDLQPET) and E7/77–86 (RTLEDLLMGT) (Riemer et al., 2010; Keskin et al., 2011; Blatnik et al., 2018). Two further peptides, namely E6/29–38 (TIHDIILECV) (Riemer et al., 2010) and E7/11–20 (YMLDLQPETT) (Blatnik et al., 2018), for which the preceding studies could not establish clear evidence of detection, were clearly detected with the new technique. Therefore, the detection by immunopeptidomics of 23 of the peptides constitutes novel findings.

As is common with LC-MS and by extension immunopeptidomics, the obtained results are influenced by the capabilities and properties of LC separation and the MS technology in use, such as the choice of fragmentation method, performance of narrow precursor isolation and detector sensitivity. Still, the much lower success rate, when compared to the preceding study that targeted A02 epitopes on the cell line CaSki, was unexpected. It had clearly detected 11 of the 17 targeted peptides (Blatnik et al., 2018). When targeting the same 17 peptides on CaSki, only the two aforementioned detections were replicated with the new technique. This low rate of detection manifested also for the expansion in this thesis to more cell lines and more supertypes. In contrast, the peptide E7/11–20, which did not show with clear evidence in Blatnik et al. (2018) was clearly detectable with the new technique. The reasons for this could not be determined conclusively in work for this thesis and in related efforts by the group. The conclusion of the preceding study was followed by the decommissioning of the previous quadrupole linear ion trap (QTRAP) mass spectrometer and the connected LC apparatus. The main author of the study departed the group. The immunopeptidomics approach was re-established by Dr. Mogjib Salek who implemented changes and optimizations of the wet-lab technique that brought the method more in line with the state-of-the-art immunopeptidomics workflows at the time and were intended to reduce contamination of the samples with detergents that affect chromatographic separation performance and peptide detection in the MS (Salek & Förster et al., 2022). Such changes are known to introduce biases for immunopeptidomics peptide detections and it cannot be excluded that the performance of the method with regard to the target peptides was significantly different (Nicastri et al., 2020). The simultaneous switch to the new MS technology made for an overall stark difference in methodology. Due to this, later efforts to replicate the prior work were exceedingly difficult and attempts made by other members of the group did not result in detections beyond those achieved with the new technique. The MS detection in the preceding study was performed with a QTRAP instrument, which performs single reaction monitoring (SRM). In this method, the instrument performs, for each precursor, multiple fragmentation reactions and cycles through every single product of interest sequentially, acquiring the signal with low resolution. In contrast, parallel reaction monitoring (PRM), as used for the new

technique, acquires all ions produced from a fragmentation reaction in a single spectrum at high resolution. In general, these methods, when using MS2, are reported to achieve similar limits of detection with the PRM method providing benefits with regards to specificity in complex samples (Gallien et al., 2013; Schiffmann et al., 2014; Ronsein et al., 2015). However, the QTRAP was operated MS3 mode. MS3 methods rely on a further fragmentation step of selected MS2 fragments to enhance specificity in spite of the low resolution that the instrument provides (Hunter, 2010). Few publications on the performance characteristics of this method are available but it has been shown to provide high specificity and thereby improve performance in complex samples (Shi et al., 2012). Consequently, for the results in this thesis, I performed detailed analyses for all peptide detections and implemented an LC-MS sequence including negative control acquisitions for wet-lab (cell line C33A) and LC (LC controls), which were all devised in an effort to ensure reproducibility of the results and prevent false-positive detection. A crucial aspect of this is also the exclusive use of stable isotope-labeled (SIL) peptides for reference acquisition and LC-MS characterization of the targets on the LC-MS system. In comparison, the preceding study lacked a wet-lab control and used unlabeled synthetic target peptides on the LC-MS system in preparation of acquisition runs. Over the course of this thesis we developed the understanding that adjacent use of highly concentrated synthetic peptides can introduce trace-level contamination to immunopeptidomics samples either in the wet-lab or on the LC system, which likely made the previous method susceptible to spurious detection of peptides.

Beyond the preceding study, only three further publications on MS detection of HPV16-derived peptides are available (Riemer et al., 2010; Keskin et al., 2011; Peng et al., 2023). Riemer et al. (2010) detected the epitope E7/11–19 in the cell line CaSki. It could not establish clear evidence of detection for E6/29–38 (see below). Keskin et al. (2011) detected the epitope E7/11–19 in 7 of 9 cervical cancer (CxCa) patient tumor tissue samples. More recently Peng et al. (2023) reported the MS detection of HPV16-derived peptides using untargeted acquisition, both with cell lines and CxCa patient tumor tissue. There also, E7/11–19 was the only detection for HPV16 E6 and E7. In this thesis, E7/11–19 is the most frequently detected peptide and it was also detected in both preceding studies. It is also the peptide that I was able to detect in the patient tumor tissue. While this demonstrated that the method can be used to detect virus-derived HLA-presented peptides in tumor tissue, the finding itself is therefore not new. Interestingly, the tumor was a neuroendocrine carcinoma of the cervix, which is a rare and aggressive form of CxCa that was shown to also be heavily associated with high-risk HPV (Alejo et al., 2018). The finding suggests that targeted immunotherapies against E7/11–19 are compatible with this cancer subtype. The variant of E6 L90V (in literature commonly referred to as L83V, which reflects counting the E6 amino acids only from translation from the second start codon), that was identified from the patient whole-genome sequencing, is the most extensively studied single amino acid substitution of HPV16. It has been associated with increased oncogenicity as it was shown to enable more effective transformation of keratinocytes *in vitro* and is present in the highly oncogenic D lineage of HPV16 (Bletsa et al., 2021). As the variant of HPV16 carried by the patient had none of the other defining features of lineage D, it was likely of variant lineage A, which the reference sequence belongs to and which is known to frequently carry this mutation of E6 (Bletsa et al., 2021). In a global epidemiological study on HPV16 variants, variant lineage A was detected in 99 % of European CxCa samples (Clifford et al., 2019).

The detected peptide E7/11–20 (YMLDLQPETT) was not detected in Riemer et al. (2010), even though it was found to be immunogenic before (Ressing et al., 1995; Wentworth et al., 1996). It was therefore speculated that vaccination approaches based on this epitope would only induce T cell responses that fail to elicit an effective attack against HPV16-transformed cells. It was then not conclusively detectable in (Blatnik et al., 2018). With the results generated in this thesis, the epitope

is now clearly detected in combination with supertype A02, which highlights its potential as a target for immunotherapies.

The detected peptide E6/29–38 (TIHDIILECV) is a well-known HPV16-derived epitope. It was previously shown that peptide-specific $CD8^+$ T cells were able to kill CaSki cells (Thomas et al., 2008). In contrast to this clear indication of peptide presentation, it had also been reported that cervical cancer cell lines are deficient in presenting the peptide (Evans et al., 2001). The peptide was previously targeted for MS detection but not clearly detected (Riemer et al., 2010). Therefore, this thesis represents the first direct detection of the peptide. It was rather widely detectable on two CxCa-derived and three head and neck squamous cell carcinoma (HNSCC)-derived cell lines.

The HLA restrictions of the peptide detections are an important aspect for the design of immunotherapies. T cell receptor-based therapies are inherently HLA-restricted and are therefore developed for specific peptide:HLA complexes (Shafer et al., 2022). For vaccine design, the wide applicability of HLA supertypes allows for more flexibility, but understanding the restrictions of the epitopes enables complementary selection of epitopes that allows for broad applicability of the vaccine (Riemer et al., 2010; Nguyen et al., 2021). Of the 27 MS-detected peptide:HLA combinations, only 12 had previously been described in literature as a result of immunological tests. Many of these were only detected by indirect means, such as the detection of a peptide-specific IFNγ response from healthy donor or patient peripheral blood mononuclear cells (PBMCs). Due to this, the results of this thesis complement these results by providing direct proof of HLA-presentation of these peptides on naturally transformed cancer cells.

Several peptide detections were still indicated as non-binders by the employed binding predictor netMHCpan 4.1. This highlights the continued relevance of Bonsack et al. (2019), even though the study did not include this most recent version of the predictor. It found that many HLA binders are missed by the latest prediction tools at the time and that in many cases, allele-specific score thresholds, could be used to improve binder classifications. This is also in line with my observation that some non-binder populations, such as the one for the A*11 allele of cell line W12-20861 showed a bimodal distribution, with one mode clustering closely beyond the weak binder threshold. In addition, many C alleles showed score distributions that did not indicate any relevant quantity of binders. While this is plausible given the low expression of C alleles, the score distribution overall typically extended from the bulk of non-binders into the weak and strong binder classifications without showing exclusive binders. Similar observations were reported in Sarkizova et al. (2020), where this was attributed to a general promiscuity of C allotypes. However, underperformance of prediction tools for HLA-C is a well-recognized problem that arises due to limited datasets for the allotypes and this was also a concern already remarked in the publication of the prediction tool that I employed here (Reynisson et al., 2021). Due to this, I deemphasized binding prediction for C alleles.

Targeted MS detection is well accepted to be the most reliable method for valid peptide detection in immunopeptidomics (Admon, 2023; Chong et al., 2022; Fritsche et al., 2021). Yet, where full sequence coverage is not achieved, the method offers no information on the likelihood of another peptide producing the same signal. This is because, in the simplest case, any two amino acids for which no fragmentation of the connecting peptide bond is detected may actually be inverted in the sequence of the detected peptide. Another case is where combinations of amino acids can be substituted with identical mass, such as DT (Asp + Thr) and ES (Glu + Ser), for both of which the residues sum to $C_8H_{12}N_2O$ (216.0746 Da). Importantly, such sequence variants are likely to cause a shift of the fragment intensity pattern and the chromatographic retention times (Gessulat et al., 2019; Salz et al., 2021). Therefore, I put a high emphasis on these indications in the analysis of potential detections. Still, both appear with considerably more variance than the accurate mass measurements that the fragment detections are based on and there are no established criteria for setting suitable thresholds.

With this methodology, using assay-external references, I detected 24 peptides by targeted MS at varying levels of evidence. At this point, the main indication that false detection of peptides was rare was that no detections were encountered with the negative control cell line C33A. However, as a necessary compromise, this cell line came with its own HLA genotype that has only limited overlap with some of the other cell lines. In parallel, Dr. Mogjib Salek and I also successfully established the method for the detection of cancer neoepitopes, which are tumor-specific or tumor-associated epitopes that may be targeted by immunotherapies (Salek & Förster et al., 2023). In this application, sample is often much more limited and only one or two technical replicates are typically acquired. This is why I decided that another validation of the method with the cell line-based experiments of this thesis can provide valuable insight on the reliability of the technique that would translate also to further applications beyond the scope of this thesis.

The most potent indicator that was unavailable to my technique is the exact co-elution of the target and the synthetic reference on the retention time (RT) axis. This was because I decided not to spike in synthetic SIL variants of the targeted peptides due to the issues with the approach that became apparent during this thesis. Targeting such assay-internal references requires the allocation of instrument time that is best used on actual target acquisition. More crucial though, is our finding of common contamination of synthetic heavy SIL peptides with light peptides that are identical to the true targets of the assay (Salek & Förster et al., 2022). This makes spike-in methods targeting hundreds of peptides challenging to implement safely. Instead, I performed SIL spike-in experiments as an extra validation step for only the detected peptides. This provided an exact RT reference which is well suited to indicate peptides that have been misidentified. The confirmation of all 24 targeted detections in this manner shows that, with dozens of detections, there was no appreciable number of false positives. This also served to bring the method fully in line with what is considered in the field to be the gold-standard for valid peptide detection and will thereby preempt a likely request in peer review.

An aspect that limits how well the confidence in the robustness of the method can be transferred to cancer neoepitope detection is that peptides that are difficult to distinguish are usually highly homologous and with the earlier test case of virus-derived peptides within the immunopeptidome of human cells, such homology is uncommon. Indeed, in using the method for cancer neoepitope discovery, we have already observed cases where a neoepitope appeared highly similar to the endogenous peptide, and where the relatively high imposed normalized spectral contrast angle (NSA) threshold of 0.85 was a strict requirement to prevent false detection.

The decision to include HNSCC cell lines in the experiments did not introduce a significant bias as all but one peptide detection was achieved in at least one CxCa cell line. Rather, the HNSCC cell line results provided opportunities to back up several detections with more data and a different biological background. The large overlap of the detected peptides may indicate a potential for good compatibility of immunotherapies between different HPV16-associated malignancies. Interestingly, the peptide E7/66–74 (RLCVQSTHV) that was exclusively detected in HNSCC cell lines, is an A02 binder, for which many cell lines were available. In contrast, the A02 binder E7/11–19 was detected in nine cell lines. It is unclear why not more of these other A02 cell lines showed E7/66–74. Overall, I observed that many peptides were detected in only one cell line, even when sometimes several further cell lines with compatible HLA alleles were available. The peptide E7/77–86 was only detected in SNU-17, even though as an A02 binder it was targeted on several other cell lines. In most cases, these A02 cell lines also presented E7/11–19 or even E7/11–20 which SNU-17 did not. At the root may be a shortcoming of the HLA supertype concept as SNU-17 is the only cell line with the A*02:06 allotype as opposed to the A*02:01 allotype of other A02 cell lines, which is also the one that the experimental binding data stems from. The binding predictions do not support this hypothesis as

they are very similar for the two alleles. Several B15 peptides were only detected in the cell line W12-20861, even though two further cell lines 866 and SNU-1005 were available. Here, the overall exceptionally high level of HLA-peptide presentation for W12-20861 was likely a factor. I did not find that the sequence variants for E6 and E7 played an appreciable role in any of these observations. Furthermore, even though the untargeted evaluation showed that the overall yield of eluted HLA ligands varies significantly between the different cell lines, 16 of the 20 cell lines showed at least one HPV16-derived peptide in the targeted experiments. The lack of consistent detection of peptides in all compatible cell lines is therefore likely the result of one or more of several factors that influence peptide presentation and detectability. Reasons may include discrepancies between the exact HLA allele in question and the matched supertype, differences in HLA expression levels and many further biological factors of antigen presentation that either abrogate the presentation of a specific peptide or prevent it from reaching the threshold for MS detection. For vaccine development, peptides that were detected in several cell lines are of particular interest as this may indicate robust presentation. The A02 binders E6/29–38 and E7/11–19 showed the best results in that regard, with detection in five and nine cell lines, respectively. The A02 supertype is, however, together with B07, the most prevalent in the panel, with ten matching cell lines. As such, the 20 cell lines together only offer a small sample size for each of the different HLA supertypes that were tested, with the least prevalent being B15 in three cell lines. These factors have to be considered when selecting the best candidate peptide for further development.

The untargeted data-independent acquisition (DIA) acquisitions achieved a total of five detections of HPV-derived peptides. This illustrates the potential for untargeted acquisition to play a role in the discovery of viral epitopes. Nevertheless, the peptides detected in untargeted fashion were some of the most intense peptides overall. All of them reached peak areas of larger than 1×10^7 in one of the targeted experiments, whereas targeted detections were sometimes achieved on the order of 1×10^4. The susceptibility of untargeted immunopeptidomics experiments to false discovery is well known and accepted to some extent (Fritsche et al., 2021; Admon, 2023). Herein, I implemented strict quality control of untargeted detections, which helped reject the spurious detection of E7/45–52 AEPDRAHY due to in-source fragmentation. It should be noted that the employed method was not optimized for sensitivity. No method for ion-mobility separation in gas phase was used, which is a technique that has recently elevated the performance of DIA MS to new levels. Such a method was used in Peng et al. (2023) and allowed for the detection of several HPV-derived peptides stemming from multiple HPV types. This was demonstrated with cell lines and also with patient cancer tissue. Nevertheless, this study did not find any of the peptides detected in this thesis other than E7/11–19. Another recent study employed untargeted immunopeptidomics to comprehensively assess the immunopeptidome of HNSCCs from a cohort of 40 patients (Mühlenbruch et al., 2023). 22 of the cancers were determined to be HPV-positive. The authors identified various tumor-exclusive epitopes of endogenous origin but did not find any HPV-derived epitopes. These are all indications that there is still a significant sensitivity advantage to targeted MS.

For the future, the available HPV-transformed cell lines represent a treasure trove that may still enable the detection of further HPV epitopes. Further HLA binders have been validated experimentally and will be targeted next. A large potential for discovery of further peptides likely lies in the analysis using next-generation untargeted techniques. This may allow for the detection of further peptides that have not been considered for targeted detection. Among the reasons that some peptides may have been overlooked so far are presentation by an HLA type that was not targeted and possible underestimation of binding by all predictors that were used. With E6/8–17 one such peptide has already been detected here. I also did not include non-canonical sources in the search, such as the alternate reading frames targeted in Peng et al. (2023). Innovative new MS methods are currently

in development that promise improved sensitivity. For example, the midiaPASEF method allows for more informative spectra to be extracted, which promises to be a major boon for immunopeptidomics applications (Distler et al., 2023). In data analysis, peptide identification performance is continually being improved through application of deep learning techniques and this is especially efficacious for immunopeptidomics (Li et al., 2020; Wilhelm et al., 2021). Therefore, the DIA data generated and analyzed in this thesis may be reanalyzed in the future with the possibility that further peptides can be detected.

4.2 MS parameter optimization

The optimizations of the MS detection parameters confirmed that the Orbitrap can robustly detect peptides even in a complex background of coeluting peptides. This has previously been reported as one of the strengths of PRM on the Orbitrap (Peterson et al., 2012; Gallien et al., 2013). During the early time of establishment of the method, further studies also evaluated the effects of MS resolution, ion injection time and the isolation window and found that fine-tuning these to match the application is beneficial (Domon & Gallien, 2015; Gallien et al., 2014; Peterson et al., 2012). These studies all stopped short of tuning the parameters to the levels described here, likely due to the aim of establishing PRM as a highly quantitative method. Due to this, much of the available literature on PRM methods did not fully align with my aims, which were exclusively on the sensitive detection of peptides. There are also generational differences between the instruments used in these studies and the one used here. Therefore, the analysis performed here provided crucial information on how to achieve the best results in my experiments.

The development of collision energy optimization was similarly driven by the need to achieve optimal detection performance for all peptides of interest. Some recent studies on collision energy optimization for the Orbitrap focused on finding a generally optimal setting for untargeted acquisition using tryptic peptides (Révész et al., 2021; Szabó et al., 2021). These did not inform on whether there were outliers, which may benefit drastically from fine-tuned collision energy. Other literature was usually on domain specific optimization such as for the fragmentation of glycoproteins or other peptide modifications (Révész et al., 2023). Per-peptide collision energy optimization for a PRM method on another type of mass spectrometer was part of a recently published workflow with a focus on quantitation performance (Brzhozovskiy et al., 2022). No evaluation for immunopeptidomics was available and, due to the common focus on tryptic peptides, I was unsure whether the basic collision energy normalization technique performs well with HLA binders as they lack the consistent basic C-terminal side chains (K, R) that are a hallmark of tryptic peptides and drive the production of highly informative y-ion series. Over the course of this thesis, several deep learning-enabled tools for prediction of peptide fragmentation have been published (Gessulat et al., 2019; Gabriels et al., 2019; Zeng et al., 2019). These have been shown to perform very well and have had a major impact on proteomics in general (Xu et al., 2020). In trying to apply such tools for my work, I did however find that none of them offer a feature set that would allow them to replace the method established here. First, even though fragmentation can be predicted for different collision energies, the result was always a spectrum of normalized intensities. That means for each spectrum either the total intensity was 1 or the most dominant fragment intensity was 1. Due to this, it was not possible to compare intensities between different predicted spectra. Second, the tools did not support the full panel of non-standard transitions, beyond basic y and b series, that I was able to include in my analysis. Through the continued use of my optimization method, I found that some peptides can be exceptionally difficult to detect without collision energy optimization and that especially the non-tryptic peptides common in immunopeptidomics benefit from the optimization. For this thesis, I was developing the method throughout and adjusted the strategy for selection of the optimal collision energy over time. Also, I observed that detections are typically valid even when

based on few transitions and that the detection performance can improve drastically with full instrument focus on a precursor of interest. The second point was clearest with the detection of E7/11–19 in the cancer tissue sample. Due to this, I expect an optimal strategy would be to target each precursor with as few as three transitions in a first survey acquisition, followed by a second acquisition that aims to validate only the detected precursors with a higher sequence coverage. Purpose-tuned collision energies could then be used for both of these steps.

4.3 Comparing HLA binders, MS detections and peptide immunogenicity

The alignment of the detected peptides along the protein sequences revealed interesting trends. The MS-detected peptides showed two strong clusters, one for each of E6 and E7. These clusters showed up to four or five peptides at a single sequence position, respectively. Outside of these clusters, no more than two MS-detected peptides overlapped. Especially towards the E6 C-terminus only one binder was detected. It is not clear what is at the root of these distributions. I did not find that the two clusters showed sequences that would clearly boost the detectability with the mass spectrometer which suggests that the reasons underlying the clustering are biological. These clusters are reminiscent of the hypothesis of HLA presentation hotspots, which describes that within individual proteins, certain sequence regions are found to generate the majority of HLA-epitopes, while other regions are seemingly not efficiently processed for presentation (Bassani-Sternberg et al., 2016; Müller et al., 2017). Presentation of peptides from these hotspots are apparently only partially driven by promiscuous HLA binding of the local peptides. This is controversial however, as others found promiscuous HLA binding to be the dominant factor for hotspots (Jappe et al., 2018). The experimental binding data of the candidate epitopes that I illustrated in Figure 65 and Figure 67 showed multiple binding hotspots that did not translate into MS detection, whereas both hotspots of MS detection seem to derive from a location with a good amount of HLA binders. Therefore, no clear conclusion can be drawn.

Clear discrepancies are also apparent between the distributions of MS-detected peptides and immunogenic peptides. I devised the sequence alignment with a hierarchy in mind, where first a peptide would have to be a binder, second it would be HLA-presented as detectable in the immunopeptidomics experiments, and in a third step, it may elicit an immune reaction as detectable in the ELISPOT experiments. Instead, the immunogenicity results display an exceptional correlation with the HLA binder distribution and the MS-detected peptide deviate markedly. This does not conform well with the understanding of HLA-presentation as prerequisite for immunogenicity. The reason for this is likely due to the different source material for the two experiments. The immunogenicity results were generated from donors from the general population. Positive results would likely stem from past HPV infections of which the majority had been cleared. In contrast, the cell lines are derived from HPV-driven cancer and therefore represent late-stage disease. This implies that the cell lines have undergone immunoediting which can drastically alter the immunopeptidome and likely favors presentation of less immunogenic peptides (Zapata et al., 2023). Specifically, for CxCa and its precursors, increasing upregulation of the aminopeptidase ERAP1 with progression through disease stages, has been observed, which likely causes destruction of certain epitopes (Steinbach et al., 2017). According to the present data, 49 % of HLA binders are immunogenic, whereas 58 % of MS-detected peptides are immunogenic. This suggests a similar rate but one should be careful to draw this conclusion as the dataset is small and the immunogenicity results represent preliminary data with percentages likely to increase over time as more donor PBMCs are tested. In fact, it has been suggested that most viral peptides can be expected to be immunogenic. Croft et al. (2019) used immunopeptidomics to identify 170 vaccinia virus epitopes with mouse cells and found immunogenicity for more than 80 % of these peptides in virus-infected mice.

It remains unclear though, whether the overall low detection rate and the observed clustering are truly a result of the underlying biology. They may simply arise from systematic biases that accumulate from the HLA IP workflow and from the limitations of the LC-MS detection. The fact that further peptides were detected in the preceding study as well as the low correlation with the HLA binder and immunogenicity distribution could both be considered to indicate this.

The findings of this thesis could have implications for vaccine development. An effective therapeutic HPV-vaccine must be highly immunogenic and elicit a strong antigen-specific cytotoxic response. However, vaccine formulations that indiscriminately elicit immune responses to the proteins E6 and E7 are likely to elicit selective reactions to the most immunogenic epitopes that derive from them. Due to the obligatory forgone HPV infection, strong memory responses could also be favored due to their immunodominance. This may preclude the effective induction of responses to less dominant peptides and thereby prevent effective vaccination against progressive disease which has likely undergone immunoediting. Epitope-specific vaccines are uniquely suited to specifically direct immune responses to the most promising epitopes. I also found that some of the detected epitopes are altered by inactivating mutations for E6 and E7 that are commonly used in vaccine formulation containing/encoding the whole HPV16 oncoproteins. This is certainly an undesirable side-effect of necessary inactivation and represents another advantage of epitope-specific vaccines that do not require such mutations.

4.4 Cumulative population coverage

For each of the six HLA supertypes targeted, I found at least one peptide which allows for an expected worldwide cumulative population coverage 99.6 %. A limitation of this prediction is that the evaluation is based on allele frequency net database (AFND) data from 2013, even though the immune epitope database (IEDB) analysis tool is the latest version released in May 2023 (IEDB, 2023a). This became clear in correspondence with IEDB and no update could be provided. While the relative distribution of supertypes among the world population is not expected to have changed significantly in the span of these ten years, the AFND database has grown considerably with more submissions covering all geographical regions that are relevant for the precision of worldwide and ethnicity-based calculations (Gonzalez-Galarza et al., 2020). However, the evaluations that I performed showed that almost complete population coverage is expected with six supertypes and that the coverage is robust in case some of the supertypes eventually cannot be targeted effectively by a potential vaccine.

4.5 Conclusions and outlook

In this thesis, I developed a targeted immunopeptidomics workflow for the sensitive and highly specific detection of HLA class I-presented peptides. To this end, I developed effective quality control measures for the wet-lab technique and implemented chemical modification steps for peptide oxidation and alkylation that allow for optimal detection of the chemically diverse peptides. I optimized the LC-MS acquisition of the peptides so that each target could be acquired with high sensitivity. To achieve this, I optimized the LC gradient and the scheduling of the MS acquisition in a manner that minimizes sensitivity-limiting parallel acquisition by the MS instrument. Furthermore, I tested the role of various MS acquisition parameters and characterized synthetic variants of the targets in detail to allow the effective fine-tuning on a per-peptide basis for hundreds of peptides. The HLA-restrictions of the peptide detections were validated through the complementary use of untargeted LC-MS experiments. This enabled me to detect the presentation of 25 HPV16-derived peptides on naturally HPV-transformed cell lines, 23 of which had not been MS-detected before.

The peptide detections and indicated supertype restrictions will support the rational design of a therapeutic HPV16 vaccine by the research group. The data complete the immunopeptidomics part of the HPV16 E6 and E7 epitome map. Ongoing experiments of the complementary immunogenicity assessment functionally characterize the peptides regarding their immunogenicity, immunodominance and frequency of memory responses in healthy PBMC donors. The most promising peptides, as informed by these combined efforts, will be tested for their capability to induce specific killing of HPV-transformed cells. This will indicate the most promising epitopes to be included an epitope-specific vaccine formulation that is widely applicable through broad HLA supertype population coverage.

5 References

Abe, F., Van Prooyen, N., Ladasky, J. J., & Edidin, M. (2009). Interaction of Bap31 and MHC Class I Molecules and Their Traffic Out of the Endoplasmic Reticulum. *The Journal of Immunology*, *182*(8), 4776-4783. https://doi.org/10.4049/jimmunol.0800242

Admon, A. (2023). The biogenesis of the immunopeptidome. *Seminars in Immunology*, *67*, 101766-101766. https://doi.org/10.1016/j.smim.2023.101766

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, *422*(6928), 198-207. https://doi.org/10.1038/nature01511

Alejo, M., Alemany, L., Clavero, O., Quiros, B., Vighi, S., Seoud, M., Cheng-Yang, C., Garland, S. M., Juanpere, N., Lloreta, J., Tous, S., Klaustermeier, J. E., Quint, W., Bosch, F. X., de Sanjosé, S., Lloveras, B., & group, R. H. T. s. (2018). Contribution of Human papillomavirus in neuroendocrine tumors from a series of 10,575 invasive cervical cancer cases. *Papillomavirus research (Amsterdam, Netherlands)*, *5*, 134-142. https://doi.org/10.1016/j.pvr.2018.03.005

Allaire, J. J. (2022). *quarto: R Interface to `Quarto' Markdown Publishing System*. https://CRAN.R-project.org/package=quarto

Allaire, J. J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2023). *rmarkdown: Dynamic Documents for R*. https://github.com/rstudio/rmarkdown

Aphalo, P. J. (2023). *ggpp: Grammar Extensions to `ggplot2'*. https://CRAN.R-project.org/package=ggpp

Apps, R., Meng, Z., Del Prete, G. Q., Lifson, J. D., Zhou, M., & Carrington, M. (2015). Relative Expression Levels of the HLA Class-I Proteins in Normal and HIV-Infected Cells. *The Journal of Immunology*, *194*(8), 3594-3600. https://doi.org/10.4049/jimmunol.1403234

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., . . . Zaks, T. (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine*, *384*(5), 403-416. https://doi.org/10.1056/nejmoa2035389

Bandrowski, A. E., & Martone, M. E. (2016). RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods. *Neuron*, *90*(3), 434-436. https://doi.org/10.1016/j.neuron.2016.04.030

Barker, D. J., Maccari, G., Georgiou, X., Cooper, M. A., Flicek, P., Robinson, J., & Marsh, S. G. E. (2023). The IPD-IMGT/HLA Database. *Nucleic acids research*, *51*(D1), D1053-D1060. https://doi.org/10.1093/nar/gkac1011

Bassani-Sternberg, M. (2018). Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens. *Methods in molecular biology (Clifton, N.J.)*, *1719*, 209-221. https://doi.org/10.1007/978-1-4939-7537-2_14

Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., H Busch, D., Peschel, C., Rad, R., Cox, J., Mann, M., & Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications*, *7*, 13404-13404. https://doi.org/10.1038/ncomms13404

Bassani-Sternberg, M., Chong, C., Guillaume, P., Solleder, M., Pak, H., Gannon, P. O., Kandalaft, L. E., Coukos, G., & Gfeller, D. (2017). Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *PLoS computational biology*, *13*(8), e1005725-e1005725. https://doi.org/10.1371/journal.pcbi.1005725

Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J., & Mann, M. (2015). Mass spectrometry of human leukocyte antigen class i peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Molecular and Cellular Proteomics*, *14*(3), 658-673. https://doi.org/10.1074/mcp.M114.042812

Bates, D., Lai, R., & Byrne, S. (2023). *RCall.jl*. https://github.com/JuliaInterop/RCall.jl

Becker, J. P., & Riemer, A. B. (2022). The Importance of Being Presented: Target Validation by Immunopeptidomics for Epitope-Specific Immunotherapies. *Frontiers in Immunology*, *13*(April), 883989-883989. https://doi.org/10.3389/fimmu.2022.883989

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65-98. https://doi.org/10.1137/141000671

Blatnik, R., Mohan, N., Bonsack, M., Falkenby, L. G., Hoppe, S., Josef, K., Steinbach, A., Becker, S., Nadler, W. M., Rucevic, M., Larsen, M. R., Salek, M., & Riemer, A. B. (2018). A targeted LC-MS strategy for low-abundant HLA class I-presented peptide detection identifies novel human papillomavirus T-cell epitopes. *PROTEOMICS*, e1700390-e1700390. https://doi.org/10.1002/pmic.201700390

Blees, A., Januliene, D., Hofmann, T., Koller, N., Schmidt, C., Trowitzsch, S., Moeller, A., & Tampé, R. (2017). Structure of the human MHC-I peptide-loading complex. *Nature*, *551*(7681), 525-528. https://doi.org/10.1038/nature24627

Bletsa, G., Zagouri, F., Amoutzias, G. D., Nikolaidis, M., Zografos, E., Markoulatos, P., & Tsakogiannis, D. (2021). Genetic variability of the HPV16 early genes and LCR. Present and future perspectives. *Expert Reviews in Molecular Medicine*, *23*, e19-e19. https://doi.org/10.1017/erm.2021.18

Bonsack, M., Hoppe, S., Winter, J., Tichy, D., Zeller, C., Küpper, M. D., Schitter, E. C., Blatnik, R., & Riemer, A. B. (2019). Performance Evaluation of MHC Class-I Binding Prediction Tools Based on an Experimentally Validated MHC-Peptide Binding Data Set. *Cancer Immunology Research*, *7*(5), 719-736. https://doi.org/10.1158/2326-6066.CIR-18-0584

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., & zur Hausen, H. (1984). A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *The EMBO journal*, *3*(5), 1151-1157. https://doi.org/10.1002/j.1460-2075.1984.tb01944.x

Bourmaud, A., Gallien, S., & Domon, B. (2016). Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *PROTEOMICS*, *16*(15-16), 2146-2159. https://doi.org/10.1002/pmic.201500543

Bouvier, M., & Wiley, D. C. (1994). Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules. *Science*, *265*(5170), 398-402. https://doi.org/10.1126/science.8023162

Bouwmeester, R., Gabriels, R., Hulstaert, N., Martens, L., & Degroeve, S. (2021). DeepLC can predict retention times for peptides that carry as-yet unseen modifications. *Nature Methods*, *18*(11), 1363-1369. https://doi.org/10.1038/s41592-021-01301-5

Brady, C. S., Bartholomew, J. S., Burt, D. J., Duggan-Keen, M. F., Glenville, S., Telford, N., Little, A. M., Davidson, J. A., Jimenez, P., Ruiz-Cabelto, F., Garrido, F., & Stern, P. L. (2000). Multiple mechanisms underlie HLA dysregulation in cervical cancer. *Tissue Antigens*, *55*(5), 401-411. https://doi.org/10.1034/j.1399-0039.2000.550502.x

Brinkman, C. C., Peske, J. D., & Engelhard, V. H. (2013). Peripheral Tissue Homing Receptor Control of Naïve, Effector, and Memory CD8 T Cell Localization in Lymphoid and Non-Lymphoid Tissues. *Frontiers in Immunology*, *4*, 241-241. https://doi.org/10.3389/fimmu.2013.00241

Brisson, M., Kim, J. J., Canfell, K., Drolet, M., Gingras, G., Burger, E. A., Martin, D., Simms, K. T., Bénard, É., Boily, M. C., Sy, S., Regan, C., Keane, A., Caruana, M., Nguyen, D. T. N., Smith, M. A., Laprise, J. F., Jit, M., Alary, M., . . . Hutubessy, R. (2020). Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *The Lancet*, *395*(10224), 575-590. https://doi.org/10.1016/S0140-6736(20)30068-4

Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., & Garin, J. (2007). Isotope-labeled protein standards: Toward absolute quantitative proteomics. *Molecular and Cellular Proteomics*, *6*(12), 2139-2149. https://doi.org/10.1074/mcp.M700163-MCP200

Bruni, L., Albero, G., Rowley, J., Alemany, L., Arbyn, M., Giuliano, A. R., Markowitz, L. E., Broutet, N., & Taylor, M. (2023a). Articles Global and regional estimates of genital human papillomavirus prevalence among men : a systematic review and meta-analysis. *The Lancet Global Health*, *11*(9), e1345-e1362. https://doi.org/10.1016/S2214-109X(23)00305-4

Bruni, L., Albero, G., Serrano, B., Mena, M., Collado, J., Gómez, D., Muñoz, J., Bosch, F., & de Sanjosé, S. (2023b). *Human Papillomavirus and Related Diseases in the World. Summary Report 10 March 2023.* https://hpvcentre.net/statistics/reports/XWX.pdf

Brzhozovskiy, A., Kononikhin, A., Bugrova, A. E., Kovalev, G. I., Schmit, P. O., Kruppa, G., Nikolaev, E. N., & Borchers, C. H. (2022). The Parallel Reaction Monitoring-Parallel Accumulation–Serial Fragmentation (prm-PASEF) Approach for Multiplexed Absolute Quantitation of Proteins in Human Plasma. *Analytical Chemistry*, *94*(4), 2016-2022. https://doi.org/10.1021/acs.analchem.1c03782

Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics*, *7*, 1-5. https://doi.org/10.1186/1471-2105-7-153

Burk, R. D., Harari, A., & Chen, Z. (2013). Human papillomavirus genome variants. *Virology*, *445*(1-2), 232-243. https://doi.org/10.1016/j.virol.2013.07.018

Campitelli, E. (2023). *ggnewscale: Multiple Fill and Colour Scales in `ggplot2'*. https://CRAN.R-project.org/package=ggnewscale

Castelló, I. V. (2022). *Supressor.jl*. https://github.com/JuliaIO/Suppressor.jl

Chabeda, A., Yanez, R. J. R., Lamprecht, R., Meyers, A. E., Rybicki, E. P., & Hitzeroth, I. I. (2018). Therapeutic vaccines for high-risk HPV-associated diseases. *Papillomavirus research (Amsterdam, Netherlands)*, *5*, 46-58. https://doi.org/10.1016/j.pvr.2017.12.006

Chong, C., Coukos, G., & Bassani-Sternberg, M. (2022). Identification of tumor antigens with immunopeptidomics. *Nature Biotechnology*, *40*(2), 175-188. https://doi.org/10.1038/s41587-021-01038-8

Chong, C., Marino, F., Pak, H., Racle, J., Daniel, R. T., Müller, M., Gfeller, D., Coukos, G., & Bassani-Sternberg, M. (2018). High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferonγ-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. *Molecular & cellular proteomics : MCP*, *17*(3), 533-548. https://doi.org/10.1074/mcp.TIR117.000383

Choo, C. K., Rorke, E. A., & Eckert, R. L. (1994). Differentiation-independent constitutive expression of the human papillomavirus type 16 E6 and E7 oncogenes in the CaSki cervical tumour cell line. *Journal of General Virology*, *75*(5), 1139-1147. https://doi.org/10.1099/0022-1317-75-5-1139

Chujoh, Y., Sobao, Y., Miwa, K., Kaneko, Y., & Takiguchi, M. (1998). The role of anchor residues in the binding of peptides to HLA-A*1101 molecules. *Tissue Antigens*, *52*(6), 501-509. https://doi.org/10.1111/j.1399-0039.1998.tb03080.x

Clifford, G. M., Tenet, V., Georges, D., Alemany, L., Pavón, M. A., Chen, Z., Yeager, M., Cullen, M., Boland, J. F., Bass, S., Steinberg, M., Raine-Bennett, T., Lorey, T., Wentzensen, N., Walker, J., Zuna, R., Schiffman, M., & Mirabello, L. (2019). Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: Whole viral genome sequences from 7116 HPV16-positive women. *Papillomavirus research (Amsterdam, Netherlands)*, *7*, 67-74. https://doi.org/10.1016/j.pvr.2019.02.001

Coelho, A. C. M. F., Fonseca, A. L., Martins, D. L., Lins, P. B. R., Da Cunha, L. M., & De Souza, S. J. (2020). NeoANT-HILL: An integrated tool for identification of potential neoantigens. *BMC Medical Genomics*, *13*(1), 1-8. https://doi.org/10.1186/s12920-020-0694-1

Cohen, P. A., Jhingran, A., Oaknin, A., & Denny, L. (2019). Cervical cancer. *Lancet (London, England)*, *393*(10167), 169-182. https://doi.org/10.1016/S0140-6736(18)32470-X

Colbert, J. D., Cruz, F. M., & Rock, K. L. (2020). Cross-presentation of exogenous antigens on MHC I molecules. *Current Opinion in Immunology*, *64*, 1-8. https://doi.org/10.1016/j.coi.2019.12.005

Compagnone, M., Cifaldi, L., & Fruci, D. (2019). Regulation of ERAP1 and ERAP2 genes and their disfunction in human cancer. *Human Immunology*, *80*(5), 318-324. https://doi.org/10.1016/j.humimm.2019.02.014

Creasy, D. M., & Cottrell, J. S. (2004). Unimod: Protein modifications for mass spectrometry. *PROTEOMICS*, *4*(6), 1534-1536. https://doi.org/10.1002/pmic.200300744

Croft, N. P., Smith, S. A., Pickering, J., Sidney, J., Peters, B., Faridi, P., Witney, M. J., Sebastian, P., Flesch, I. E. A., Heading, S. L., Sette, A., La Gruta, N. L., Purcell, A. W., & Tscharke, D. C. (2019). Most

viral peptides displayed by class I MHC on infected cells are immunogenic. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(8), 3112-3117. https://doi.org/10.1073/pnas.1815239116

Csárdi, G., & Chang, W. (2023). *processx: Execute and Control System Processes*. https://CRAN.R-project.org/package=processx

de Martel, C., Georges, D., Bray, F., Ferlay, J., & Clifford, G. M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *The Lancet. Global health*, *8*(2), e180-e190. https://doi.org/10.1016/S2214-109X(19)30488-7

de Martel, C., Plummer, M., Vignat, J., & Franceschi, S. (2017). Worldwide burden of cancer attributable to HPV by site, country and HPV type. *International journal of cancer*, *141*(4), 664-670. https://doi.org/10.1002/ijc.30716

de Sanjosé, S., Brotons, M., & Pavón, M. A. (2018). The natural history of human papillomavirus infection. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *47*, 2-13. https://doi.org/10.1016/j.bpobgyn.2017.08.015

del Guercio, M. F., Sidney, J., Hermanson, G., Perez, C., Grey, H. M., Kubo, R. T., & Sette, A. (1995). Binding of a peptide antigen to multiple HLA alleles allows definition of an A2-like supertype. *Journal of immunology (Baltimore, Md. : 1950)*, *154*(2), 685-693. http://www.ncbi.nlm.nih.gov/pubmed/7529283

Dersh, D., Hollý, J., & Yewdell, J. W. (2021). A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nature Reviews Immunology*, *21*(2), 116-128. https://doi.org/10.1038/s41577-020-0390-6

DiGiuseppe, S., Bienkowska-Haba, M., Guion, L. G., & Sapp, M. (2017). Cruising the cellular highways: How human papillomavirus travels from the surface to the nucleus. *Virus research*, *231*, 1-9. https://doi.org/10.1016/j.virusres.2016.10.015

Distler, U., Łącki, M. K., Startek, M. P., Teschner, D., Brehmer, S., Decker, J., Schild, T., Krieger, J., Krohs, F., Raether, O., Hildebrandt, A., & Tenzer, S. (2023). midiaPASEF maximizes information content in data-independent acquisition proteomics. *bioRxiv*, 2023.2001.2030.526204-522023.526201.526230.526204. https://doi.org/10.1101/2023.01.30.526204

Domon, B., & Gallien, S. (2015). Recent advances in targeted proteomics for clinical applications. *Proteomics - Clinical Applications*, *9*(3-4), 423-431. https://doi.org/10.1002/prca.201400136

Draper, L. M., Kwong, M. L. M., Gros, A., Stevanović, S., Tran, E., Kerkar, S., Raffeld, M., Rosenberg, S. A., & Hinrichs, C. S. (2015). Targeting of HPV-16+ epithelial cancer cells by TCR gene engineered T cells directed against E6. *Clinical Cancer Research*, *21*(19), 4431-4439. https://doi.org/10.1158/1078-0432.CCR-14-3341

Dürst, M., Gissmann, L., Ikenberg, H., & zur Hausen, H. (1983). A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(12), 3812-3815. https://doi.org/10.1073/pnas.80.12.3812

Evans, M., Borysiewicz, L. K., Evans, A. S., Rowe, M., Jones, M., Gileadi, U., Cerundolo, V., & Man, S. (2001). Antigen Processing Defects in Cervical Carcinomas Limit the Presentation of a CTL Epitope from Human Papillomavirus 16 E6. *The Journal of Immunology*, *167*(9), 5420-5428. https://doi.org/10.4049/jimmunol.167.9.5420

Falk, K., Rötzschke, O., Stevanovié, S., Jung, G., & Rammensee, H. G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, *351*(6324), 290-296. https://doi.org/10.1038/351290a0

Faridi, P., Purcell, A. W., & Croft, N. P. (2018). In Immunopeptidomics We Need a Sniper Instead of a Shotgun. *PROTEOMICS*, *18*(12), 1-4. https://doi.org/10.1002/pmic.201700464

Feltkamp, M. C. W., Smits, H. L., Vierboom, M. P. M., Minnaar, R. P., De Jongh, B. M., Drijfhout, J. W., Schegget, J. T., Melief, C. J. M., & Kast, W. M. (1993). Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. *European Journal of Immunology*, *23*(9), 2242-2249. https://doi.org/10.1002/eji.1830230929

Foster, C. (2021). *Underscores.jl*. https://github.com/c42f/Underscores.jl

Fritsche, J., Kowalewski, D. J., Backert, L., Gwinner, F., Dorner, S., Priemer, M., Tsou, C. C., Hoffgaard, F., Römer, M., Schuster, H., Schoor, O., & Weinschenk, T. (2021). Pitfalls in HLA ligandomics—how to catch a li(e)gand. *Molecular and Cellular Proteomics*, *20*, 100110-100110. https://doi.org/10.1016/J.MCPRO.2021.100110

Gabriels, R., Martens, L., & Degroeve, S. (2019). Updated MS^2PIP web server delivers fast and accurate MS^2 peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. *Nucleic acids research*, *47*(W1), W295-W299. https://doi.org/10.1093/nar/gkz299

Gallien, S., Bourmaud, A., Kim, S. Y., & Domon, B. (2014). Technical considerations for large-scale parallel reaction monitoring analysis. *Journal of Proteomics*, *100*, 147-159. https://doi.org/10.1016/j.jprot.2013.10.029

Gallien, S., & Domon, B. (2015). Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. *Methods*, *81*, 15-23. https://doi.org/10.1016/j.ymeth.2015.03.015

Gallien, S., Duriez, E., Demeure, K., & Domon, B. (2013). Selectivity of LC-MS/MS analysis: Implication for proteomics experiments. *Journal of Proteomics*, *81*, 148-158. https://doi.org/10.1016/j.jprot.2012.11.005

Gallien, S., Kim, S. Y., & Domon, B. (2015). Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Molecular & cellular proteomics : MCP*, *14*(6), 1630-1644. https://doi.org/10.1074/mcp.O114.043968

Gan, L., Jia, R., Zhou, L., Guo, J., & Fan, M. (2014). Fusion of CTLA-4 with HPV16 E7 and E6 enhanced the potency of therapeutic HPV DNA vaccine. *PLoS ONE*, *9*(9), e108892-e108892. https://doi.org/10.1371/journal.pone.0108892

Garbett, S. P., Stephens, J., Simonov, K., Xie, Y., Dong, Z., Wickham, H., Horner, J., reikoch, Beasley, W., O'Connor, B., Warnes, G. R., Quinn, M., & Kamvar, Z. N. (2023). *yaml: Methods to Convert R Data to YAML and Back*. https://CRAN.R-project.org/package=yaml

Gehlenborg, N. (2019). *UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Vis ualizing Intersecting Sets*. https://CRAN.R-project.org/package=UpSetR

Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., Reimer, U., Ehrlich, H. C., Aiche, S., Kuster, B., & Wilhelm, M. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, *16*(6), 509-518. https://doi.org/10.1038/s41592-019-0426-7

Gevaert, K., Impens, F., Ghesquière, B., Van Damme, P., Lambrechts, A., & Vandekerckhove, J. (2008). Stable isotopic labeling in proteomics. *PROTEOMICS*, *8*(23-24), 4873-4885. https://doi.org/10.1002/pmic.200800421

Gfeller, D., & Bassani-Sternberg, M. (2018). Predicting Antigen Presentation-What Could We Learn From a Million Peptides? *Frontiers in Immunology*, *9*, 1716-1716. https://doi.org/10.3389/fimmu.2018.01716

Gonzalez-Galarza, F. F., McCabe, A., Santos, E. J. M. D., Jones, J., Takeshita, L., Ortega-Rivera, N. D., Cid-Pavon, G. M. D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., Middleton, D., & Jones, A. R. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic acids research*, *48*(D1), D783-D788. https://doi.org/10.1093/nar/gkz1029

Grabowska, A. K., & Riemer, A. B. (2012). The invisible enemy - how human papillomaviruses avoid recognition and clearance by the host immune system. *The Open Virology Journal*, *6*(1), 249-256. https://doi.org/10.2174/1874357901206010249

Gupta, S., Ahadi, S., Zhou, W., & Röst, H. (2019). DIAlignR Provides Precise Retention Time Alignment Across Distant Runs in DIA and Targeted Proteomics. *Molecular & Cellular Proteomics*, *18*(4), 806-817. https://doi.org/10.1074/mcp.TIR118.001132

Gusho, E., & Laimins, L. (2021). Human Papillomaviruses Target the DNA Damage Repair and Innate Immune Response Pathways to Allow for Persistent Infection. *Viruses*, *13*(7). https://doi.org/10.3390/v13071390

Haag, A. M. (2016). Mass Analyzers and Mass Spectrometers. *Advances in experimental medicine and biology*, *919*, 157-169. https://doi.org/10.1007/978-3-319-41448-5_7

Hall, E., Wodi, A. P., Hamborsky, J., Morelli, V., & Schillie, S. (2021). Human Papillomaviruses. In *Epidemiology and Prevention of Vaccine-Preventable Diseases*.

Hancock, G., Hellner, K., & Dorrell, L. (2018). Therapeutic HPV vaccines. *Best practice & research. Clinical obstetrics & gynaecology*, *47*, 59-72. https://doi.org/10.1016/j.bpobgyn.2017.09.008

Harris, H., DuBois, C., White, J. M., Bouchet-Valat, M., & Kamiński, B. (2023). *DataFrames.jl*. https://doi.org/10.5281/zenodo.7632427

Hassan, C., Kester, M. G. D., Oudgenoeg, G., de Ru, A. H., Janssen, G. M. C., Drijfhout, J. W., Spaapen, R. M., Jiménez, C. R., Heemskerk, M. H. M., Falkenburg, J. H. F., & van Veelen, P. A. (2014). Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes. *Journal of Proteomics*, *109*, 240-244. https://doi.org/10.1016/j.jprot.2014.07.009

Hebert, A. S., Thöing, C., Riley, N. M., Kwiecien, N. W., Shiskova, E., Huguet, R., Cardasis, H. L., Kuehn, A., Eliuk, S., Zabrouskov, V., Westphall, M. S., McAlister, G. C., & Coon, J. J. (2018). Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Analytical Chemistry*, *90*(3), 2333-2340. https://doi.org/10.1021/acs.analchem.7b04808

Hecht, E. S., Scigelova, M., Eliuk, S., & Makarov, A. (2019). *Fundamentals and Advances of Orbitrap Mass Spectrometry*. https://doi.org/10.1002/9780470027318.a9309.pub2

Hester, J., Wickham, H., & Csárdi, G. (2023). *fs: Cross-Platform File System Operations Based on `libuv'*. https://CRAN.R-project.org/package=fs

Hibma, M. H. (2013). The Immune Response to Papillomavirus During Infection Persistence and Regression. *The Open Virology Journal*, *6*(1), 241-248. https://doi.org/10.2174/1874357901206010241

Hillen, N., & Stevanovic, S. (2006). Contribution of mass spectrometry-based proteomics to immunology. *Expert Review of Proteomics*, *3*(6), 653-664. https://doi.org/10.1586/14789450.3.6.653

Hlevnjak, M., Schulze, M., Elgaafary, S., Fremd, C., Michel, L., Beck, K., Pfütze, K., Richter, D., Wolf, S., Horak, P., Kreutzfeldt, S., Pixberg, C., Hutter, B., Ishaque, N., Hirsch, S., Gieldon, L., Stenzinger, A., Springfeld, C., Smetanay, K., . . . Schneeweiss, A. (2021). CATCH: A Prospective Precision Oncology Trial in Metastatic Breast Cancer. *JCO Precision Oncology*(5), 676-686. https://doi.org/10.1200/po.20.00248

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Ng, K. F., Suen, M. W. M., & Tai, H. L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical biochemist. Reviews*, *24*(1), 3-12. http://www.ncbi.nlm.nih.gov/pubmed/18568044

Holy, T. (2023). *Revise.jl*. https://github.com/timholy/Revise.jl

Hu, A., Noble, W. S., & Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research*, *5*, 419-419. https://doi.org/10.12688/f1000research.7042.1

Hundal, J., Kiwala, S., McMichael, J., Miller, C. A., Xia, H., Wollam, A. T., Liu, C. J., Zhao, S., Feng, Y. Y., Graubert, A. P., Wollam, A. Z., Neichin, J., Neveau, M., Walker, J., Gillanders, W. E., Mardis, E. R., Griffith, O. L., & Griffith, M. (2020). PVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunology Research*, *8*(3), 409-420. https://doi.org/10.1158/2326-6066.CIR-19-0401

Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A. L., Appella, E., & Engelhard, V. H. (1992). Characterization of Peptides Bound to the Class I MHC Molecule HLA-A2.1 by Mass Spectrometry. *Science*, *255*(5049), 1261-1263. https://doi.org/10.1126/science.1546328

Hunter, C. (2010). MRM3 Quantitation for Highest Selectivity of Proteins in Complex Matrices. *J. Biomol. Tech*, *21*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2918007/

IEDB. (2023a). *IEDB Analysis Resource v2.27 release notes (25 May 2023)*. Immune Epitope Database (IEDB). Retrieved 2023-10-17 from https://help.iedb.org/hc/en-us/articles/15819624348059-IEDB-Analysis-Resource-v2-27-release-notes-25-May-2023-

IEDB. (2023b). *MHC-I binding predictions - Tutorial - "IEDB recommended is the default prediction method selection ... Currently for peptide: MHC-I binding prediction, NetMHCpan EL 4.1 is used*

across all alleles.". Immune Epitope Database (IEDB). Retrieved 2023-10-17 from http://tools.iedb.org/mhci/help/

Imkeller, K. (2021). *immunotation: Tools for working with diverse immune genes*. https://github.com/imkeller/immunotation

Irvine, D. J., Purbhoo, M. A., Krogsgaard, M., & Davis, M. M. (2002). Direct observation of ligand recognition by T cells. *Nature*, *419*(6909), 845-849. https://doi.org/10.1038/nature01076

Jappe, E. C., Kringelum, J., Trolle, T., & Nielsen, M. (2018). Predicted MHC peptide binding promiscuity explains MHC class I 'hotspots' of antigen presentation defined by mass spectrometry eluted ligand data. *Immunology*, *154*(3), 407-417. https://doi.org/10.1111/imm.12905

Jeon, S., & Lambert, P. F. (1995). Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: Implications for cervical carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(5), 1654-1658. https://doi.org/10.1073/pnas.92.5.1654

Jiang, J., Xia, M., Zhang, L., Chen, X., Zhao, Y., Zeng, C., Yang, H., Liang, P., Li, G., Li, N., Qi, H., Wei, T., & Ren, L. (2022). Rapid generation of genetically engineered T cells for the treatment of virus-related cancers. *Cancer Science*, *113*(11), 3686-3697. https://doi.org/10.1111/cas.15528

Jin, B. Y., Campbell, T. E., Draper, L. M., Stevanović, S., Weissbrich, B., Yu, Z., Restifo, N. P., Rosenberg, S. A., Trimble, C. L., & Hinrichs, C. S. (2018). Engineered T cells targeting E7 mediate regression of human papillomavirus cancers in a murine model. *JCI insight*, *3*(8). https://doi.org/10.1172/jci.insight.99488

Johnson, D. E., Burtness, B., Leemans, C. R., Lui, V. W. Y., Bauman, J. E., & Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers*, *6*(1). https://doi.org/10.1038/s41572-020-00224-3

Johnson, R. S., Martin, S. A., & Biemann, K. (1988). Collision-induced fragmentation of (M + H)+ ions of peptides. Side chain specific sequence ions. *International Journal of Mass Spectrometry and Ion Processes*, *86*, 137-154. https://doi.org/10.1016/0168-1176(88)80060-0

Joyce, S., & Ternette, N. (2021). Know thy immune self and non-self: Proteomics informs on the expanse of self and non-self, and how and where they arise. *PROTEOMICS*, e2000143-e2000143. https://doi.org/10.1002/pmic.202000143

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology*, *199*(9), 3360-3368. https://doi.org/10.4049/jimmunol.1700893

Kaba, G., Stevenson, A., Sakyi, S. A., Konney, T. O., Bhatia, R., Titiloye, N. A., Oppong, S. A., Agyemang-Yeboah, F., Cuschieri, K., & Graham, S. V. (2023). Diversity of cervicovaginal human papillomavirus (HPV) genotypes and naturally occurring E6/E7 DNA polymorphisms of HPV-16 in Ghana. *Tumour Virus Research*, *15*(April), 200261-200261. https://doi.org/10.1016/j.tvr.2023.200261

Kaufmann, A., Widmer, M., & Maden, K. (2010). Post-interface signal suppression, a phenomenon observed in a single-stage Orbitrap mass spectrometer coupled to an electrospray interfaced liquid chromatograph. *Rapid communications in mass spectrometry : RCM*, *24*(14), 2162-2170. https://doi.org/10.1002/rcm.4615

Keskin, D. B., Reinhold, B., Lee, S. Y., Zhang, G., Lank, S., O'Connor, D. H., Berkowitz, R. S., Brusic, V., Kim, S. J., & Reinherz, E. L. (2011). Direct identification of an HPV-16 tumor antigen from cervical cancer biopsy specimens. *Frontiers in Immunology*, *2*, 75-75. https://doi.org/10.3389/fimmu.2011.00075

Kessler, J. H., Benckhuijsen, W. E., Mutis, T., Melief, C. J. M., van der Burg, S. H., & Drijfhout, J. W. (2004). Competition-based cellular peptide binding assay for HLA class I. *Current protocols in immunology*, *Chapter 18*, Unit 18.12-Unit 18.12. https://doi.org/10.1002/0471142735.im1812s61

Khallouf, H., Grabowska, A. K., & Riemer, A. B. (2014). Therapeutic vaccine strategies against human papillomavirus. *Vaccines*, *2*(2), 422-462. https://doi.org/10.3390/vaccines2020422

Khan, A. M., Miotto, O., Heiny, A. T., Salmon, J., Srinivasan, K. N., Nascimento, E. J. M., Marques, E. T. A., Brusic, V., Tan, T. W., & August, J. T. (2006). A systematic bioinformatics approach for

selection of epitope-based vaccine targets. *Cellular Immunology*, *244*(2), 141-147. https://doi.org/10.1016/j.cellimm.2007.02.005

Koopmann, J. O., Post, M., Neefjes, J. J., Hämmerling, G. J., & Momburg, F. (1996). Translocation of long peptides by transporters associated with antigen processing (TAP). *European Journal of Immunology*, *26*(8), 1720-1728. https://doi.org/10.1002/eji.1830260809

Kota, U., & Stolowitz, M. L. (2016). Improving Proteome Coverage by Reducing Sample Complexity via Chromatography. *Advances in experimental medicine and biology*, *919*, 83-143. https://doi.org/10.1007/978-3-319-41448-5_5

Kovalchik, K. A., Ma, Q., Wessling, L., Saab, F., Duquette, J. D., Kubiniok, P., Hamelin, D. J., Faridi, P., Li, C., Purcell, A. W., Jang, A., Paramithiotis, E., Tognetti, M., Reiter, L., Bruderer, R., Lanoix, J., Bonneil, É., Courcelles, M., Thibault, P., . . . Sirois, I. (2022). MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- To Large-Scale Immunopeptidome Datasets. *Molecular and Cellular Proteomics*, *21*(1), 0-14. https://doi.org/10.1016/j.mcpro.2021.100178

Krishna, S., Ulrich, P., Wilson, E., Parikh, F., Narang, P., Yang, S., Read, A. K., Kim-Schulze, S., Park, J. G., Posner, M., Wilson Sayres, M. A., Sikora, A., & Anderson, K. S. (2018). Human papilloma virus specific immunogenicity and dysfunction of CD8+ T cells in head and neck cancer. *Cancer Research*, *78*(21), 6159-6170. https://doi.org/10.1158/0008-5472.CAN-18-0163

Kruse, S., Büchler, M., Uhl, P., Sauter, M., Scherer, P., Lan, T. C. T., Zottnick, S., Klevenz, A., Yang, R., Rösl, F., Mier, W., & Riemer, A. B. (2018). Therapeutic vaccination using minimal HPV16 epitopes in a novel MHC-humanized murine HPV tumor model. *OncoImmunology*, *8*(1), 1-12. https://doi.org/10.1080/2162402X.2018.1524694

Kuznetsov, A., Voronina, A., Govorun, V., & Arapidi, G. (2020). Critical Review of Existing MHC I Immunopeptidome Isolation Methods. *Molecules (Basel, Switzerland)*, *25*(22), 1-26. https://doi.org/10.3390/molecules25225409

Łącki, M. K., Valkenborg, D., & Startek, M. P. (2020). IsoSpec2: Ultrafast Fine Structure Calculator. *Analytical Chemistry*, *92*(14), 9472-9475. https://doi.org/10.1021/acs.analchem.0c00959

Lagerwerf, F. M., van de Weert, M., Heerma, W., & Haverkamp, J. (1996). Identification of oxidized methionine in peptides. *Rapid communications in mass spectrometry : RCM*, *10*(15), 1905-1910. https://doi.org/10.1002/(SICI)1097-0231(199612)10:15<1905::AID-RCM755>3.0.CO;2-9

Lange, V., Picotti, P., Domon, B., & Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular systems biology*, *4*(222). https://doi.org/10.1038/msb.2008.61

Larsson, J. (2022). *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*. https://CRAN.R-project.org/package=eulerr

Laydon, D. J., Bangham, C. R. M., & Asquith, B. (2015). Estimating T-cell repertoire diversity: Limitations of classical estimators and a new approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1675). https://doi.org/10.1098/rstb.2014.0291

Lee, M. Y., & Allen, C. T. (2021). Immunotherapy for HPV Malignancies. *Seminars in Radiation Oncology*, *31*(4), 361-370. https://doi.org/10.1016/j.semradonc.2021.02.008

Lenčo, J., Jadeja, S., Naplekov, D. K., Krokhin, O. V., Khalikova, M. A., Chocholouš, P., Urban, J., Broeckhoven, K., Nováková, L., & Švec, F. (2022). Reversed-Phase Liquid Chromatography of Peptides for Bottom-Up Proteomics: A Tutorial. *Journal of Proteome Research*, *21*(12), 2846-2892. https://doi.org/10.1021/acs.jproteome.2c00407

Li, C. (2019). JuliaCall: an R package for seamless integration between R and Julia. *The Journal of Open Source Software*, *4*(35), 1284. https://doi.org/10.21105/joss.01284

Li, J., Chen, S., Ge, J., Lu, F., Ren, S., Zhao, Z., Pu, X., Chen, X., Sun, J., & Gu, Y. (2017). A novel therapeutic vaccine composed of a rearranged human papillomavirus type 16 E6/E7 fusion protein and Fms-like tyrosine kinase-3 ligand induces CD8+ T cell responses and antitumor effect. *Vaccine*, *35*(47), 6459-6467. https://doi.org/10.1016/j.vaccine.2017.09.003

Li, K., Jain, A., Malovannaya, A., Wen, B., & Zhang, B. (2020). DeepRescore: Leveraging Deep Learning to Improve Peptide Identification in Immunopeptidomics. *PROTEOMICS*, *20*(21-22), 1-10. https://doi.org/10.1002/pmic.201900334

Loo, J. A., Udseth, H. R., & Smith, R. D. (1989). Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Analytical Biochemistry*, *179*(2), 404-412. https://doi.org/10.1016/0003-2697(89)90153-X

MacLean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S., Whiteaker, J. R., Paulovich, A. G., Carr, S. A., & MacCoss, M. J. (2010a). Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry. *Analytical Chemistry*, *82*(24), 10116-10124. https://doi.org/10.1021/ac102179j

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010b). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics (Oxford, England)*, *26*(7), 966-968. https://doi.org/10.1093/bioinformatics/btq054

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry*, *72*(6), 1156-1162. https://doi.org/10.1021/ac991131p

Malagón, T., Drolet, M., Boily, M. C., Franco, E. L., Jit, M., Brisson, J., & Brisson, M. (2012). Cross-protective efficacy of two human papillomavirus vaccines: A systematic review and meta-analysis. *The Lancet Infectious Diseases*, *12*(10), 781-789. https://doi.org/10.1016/S1473-3099(12)70187-1

Marsh, S. G. E., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Fernández-Viña, M., Geraghty, D. E., Holdsworth, R., Hurley, C. K., Lau, M., Lee, K. W., MacH, B., Maiers, M., Mayr, W. R., Müller, C. R., Parham, P., Petersdorf, E. W., Sasazuki, T., . . . Trowsdale, J. (2010). Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, *75*(4), 291-455. https://doi.org/10.1111/j.1399-0039.2010.01466.x

Marx, B., Miller-Lazic, D., Doorbar, J., Majewski, S., Hofmann, K., Hufbauer, M., & Akgül, B. (2017). HPV8-E6 Interferes with Syntenin-2 Expression through Deregulation of Differentiation, Methylation and Phosphatidylinositide-Kinase Dependent Mechanisms. *Frontiers in Microbiology*, *8*, 1724-1724. https://doi.org/10.3389/fmicb.2017.01724

McBride, A. A. (2022). Human papillomaviruses: diversity, infection and host interactions. *Nature reviews. Microbiology*, *20*(2), 95-108. https://doi.org/10.1038/s41579-021-00617-5

Milius, R. P., Mack, S. J., Hollenbach, J. A., Pollack, J., Heuer, M. L., Gragert, L., Spellman, S., Guethlein, L. A., Trachtenberg, E. A., Cooley, S., Bochtler, W., Mueller, C. R., Robinson, J., Marsh, S. G. E., & Maiers, M. (2013). Genotype List String: A grammar for describing HLA and KIR genotyping results in a text string. *Tissue Antigens*, *82*(2), 106-112. https://doi.org/10.1111/tan.12150

Mohan, N., Wellach, K., Özerdem, C., Veits, N., Förster, J. D., Foehr, S., Bonsack, M., & Riemer, A. B. (2022). Effects of hypoxia on antigen presentation and T cell-based immune recognition of HPV16-transformed cells. *Frontiers in Immunology*, *13*, 918528-918528. https://doi.org/10.3389/fimmu.2022.918528

Moody, C. A., & Laimins, L. A. (2010). Human papillomavirus oncoproteins: pathways to transformation. *Nature reviews. Cancer*, *10*(8), 550-560. https://doi.org/10.1038/nrc2886

Mühlenbruch, L., Abou-Kors, T., Dubbelaar, M. L., Bichmann, L., Kohlbacher, O., Bens, M., Thomas, J., Ezić, J., Kraus, J. M., Kestler, H. A., von Witzleben, A., Mytilineos, J., Fürst, D., Engelhardt, D., Doescher, J., Greve, J., Schuler, P. J., Theodoraki, M. N., Brunner, C., . . . Laban, S. (2023). The HLA ligandome of oropharyngeal squamous cell carcinomas reveals shared tumour-exclusive peptides for semi-personalised vaccination. *British Journal of Cancer*, *128*(9), 1777-1787. https://doi.org/10.1038/s41416-023-02197-y

Müller, K. (2020). *here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here

Müller, M., Gfeller, D., Coukos, G., & Bassani-Sternberg, M. (2017). 'Hotspots' of Antigen Presentation Revealed by Human Leukocyte Antigen Ligandomics for Neoantigen Prioritization. *Frontiers in Immunology*, *8*, 1367-1367. https://doi.org/10.3389/fimmu.2017.01367

Murphy, K. M., Weaver, C., & Berg, L. (2022). *Janeway's Immunobiology* (10 ed.). W.W. Norton.

Nagarsheth, N. B., Norberg, S. M., Sinkoe, A. L., Adhikary, S., Meyer, T. J., Lack, J. B., Warner, A. C., Schweitzer, C., Doran, S. L., Korrapati, S., Stevanović, S., Trimble, C. L., Kanakry, J. A., Bagheri, M. H., Ferraro, E., Astrow, S. H., Bot, A., Faquin, W. C., Stroncek, D., . . . Hinrichs, C. S. (2021). TCR-engineered T cells targeting E7 for patients with metastatic HPV-associated epithelial cancers. *Nature Medicine*, *27*(3), 419-425. https://doi.org/10.1038/s41591-020-01225-1

Nelde, A., Rammensee, H.-g., & Walz, J. S. (2021). The Peptide Vaccine of the Future. *Molecular & Cellular Proteomics*, *20*, 100022-100022. https://doi.org/10.1074/mcp.R120.002309

Nelson, C. W., & Mirabello, L. (2023). Human papillomavirus genomics : Understanding carcinogenicity. *Tumour Virus Research*, *15*, 200258-200258. https://doi.org/10.1016/j.tvr.2023.200258

Nguyen, A. T., Szeto, C., & Gras, S. (2021). The pockets guide to HLA class I molecules. *Biochemical Society Transactions*, *49*(5), 2319-2331. https://doi.org/10.1042/BST20210410

Nicastri, A., Liao, H., Muller, J., Purcell, A. W., & Ternette, N. (2020). The Choice of HLA-Associated Peptide Enrichment and Purification Strategy Affects Peptide Yields and Creates a Bias in Detected Sequence Repertoire. *PROTEOMICS*, *20*(12), e1900401-e1900401. https://doi.org/10.1002/pmic.201900401

Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., & Buus, S. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE*, *2*(8). https://doi.org/10.1371/journal.pone.0000796

Niessen, W. M. A., & Correa C, R. A. (2016). Interpretation of MS-MS Mass Spectra of Drugs and Pesticides. *Interpretation of MS-MS Mass Spectra of Drugs and Pesticides*, 1-388. https://doi.org/10.1002/9781119294269

Norberg, S. M., & Hinrichs, C. S. (2023). Engineered T cell therapy for viral and non-viral epithelial cancers. *Cancer Cell*, *41*(1), 58-69. https://doi.org/10.1016/j.ccell.2022.10.016

Öhlschläger, P., Pes, M., Osen, W., Dürst, M., Schneider, A., Gissmann, L., & Kaufmann, A. M. (2006). An improved rearranged Human Papillomavirus Type 16 E7 DNA vaccine candidate (HPV-16 E7SH) induces an E7 wildtype-specific T cell response. *Vaccine*, *24*(15), 2880-2893. https://doi.org/10.1016/j.vaccine.2005.12.061

Okunade, K. S. (2020). Human papillomavirus and cervical cancer. *Journal of obstetrics and gynaecology : the journal of the Institute of Obstetrics and Gynaecology*, *40*(5), 602-608. https://doi.org/10.1080/01443615.2019.1634030

Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., & Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods*, *4*(9), 709-712. https://doi.org/10.1038/nmeth1060

Omilusik, K. D., & Goldrath, A. W. (2017). The origins of memory T cells. *Nature*, *552*(7685), 337-339. https://doi.org/10.1038/d41586-017-08280-8

Papanicolaou, G. N., & Traut, H. F. (1997). The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. *Archives of pathology & laboratory medicine*, *121*(3), 211-224. http://www.ncbi.nlm.nih.gov/pubmed/9111103

Pedersen, T. L. (2022). *patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork

Peng, X., Woodhouse, I., Hancock, G., Parker, R., Marx, K., Müller, J., Salatino, S., Partridge, T., Nicastri, A., Liao, H., Kruppa, G., Hellner, K., Dorrell, L., & Ternette, N. (2023). Novel canonical and non-canonical viral antigens extend current targets for immunotherapy of HPV-driven cervical cancer. *iScience*, *26*(3), 106101-106101. https://doi.org/10.1016/j.isci.2023.106101

Peri, A., Salomon, N., Wolf, Y., Kreiter, S., Diken, M., & Samuels, Y. (2023). The landscape of T cell antigens for cancer immunotherapy. *Nature Cancer*, *4*(7), 937-954. https://doi.org/10.1038/s43018-023-00588-x

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics : MCP*, *11*(11), 1475-1488. https://doi.org/10.1074/mcp.O112.020131

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Frenck, R. W., Hammitt, L. L., . . . Gruber, W. C. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*, *383*(27), 2603-2615. https://doi.org/10.1056/nejmoa2034577

Pollock, S. B., Rose, C. M., Darwish, M., Bouziat, R., Delamarre, L., Blanchette, C., & Lill, J. R. (2021). Sensitive and Quantitative Detection of MHC-I Displayed Neoepitopes Using a

Semiautomated Workflow and TOMAHAQ Mass Spectrometry. *Molecular & cellular proteomics : MCP*, *20*, 100108-100108. https://doi.org/10.1016/j.mcpro.2021.100108

Prudden, H. J., Achilles, S. L., Schocken, C., Broutet, N., Canfell, K., Akaba, H., Basu, P., Bhatla, N., Chirenje, Z. M., Delany-Moretlwe, S., Denny, L., Gamage, D. G., Herrero, R., Hutubessy, R., Villa, L. L., Murillo, R., Schiller, J. T., Stanley, M., Temmerman, M., . . . Gottlieb, S. L. (2022). Understanding the public health value and defining preferred product characteristics for therapeutic human papillomavirus (HPV) vaccines: World Health Organization consultations, October 2021—March 2022. *Vaccine*, *40*(41), 5843-5855. https://doi.org/10.1016/j.vaccine.2022.08.020

Purcell, A. W., McCluskey, J., & Rossjohn, J. (2007). More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery*, *6*(5), 404-414. https://doi.org/10.1038/nrd2224

Purcell, A. W., Ramarathinam, S. H., & Ternette, N. (2019). Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nature protocols*, *14*(6), 1687-1707. https://doi.org/10.1038/s41596-019-0133-y

Quinn, J. (2023). *CSV.jl*. https://doi.org/10.5281/zenodo.7915668

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rammensee, H. G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A., & Stevanović, S. (1999). SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics*, *50*(3-4), 213-219. https://doi.org/10.1007/s002510050595

Ramos da Silva, J., Bitencourt Rodrigues, K., Formoso Pelegrin, G., Silva Sales, N., Muramatsu, H., de Oliveira Silva, M., Porchia, B. F. M. M., Moreno, A. C. R., Aps, L. R. M. M., Venceslau-Carvalho, A. A., Tombácz, I., Fotoran, W. L., Karikó, K., Lin, P. J. C., Tam, Y. K., de Oliveira Diniz, M., Pardi, N., & de Souza Ferreira, L. C. (2023). Single immunizations of self-amplifying or non-replicating mRNA-LNP vaccines control HPV-associated tumors in mice. *Science translational medicine*, *15*(686), eabn3464-eabn3464. https://doi.org/10.1126/scitranslmed.abn3464

Reche, P. A., & Reinherz, E. L. (2007). Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. *Methods in molecular biology (Clifton, N.J.)*, *409*, 163-173. https://doi.org/10.1007/978-1-60327-118-9_11

Ressing, M. E., Sette, A., Brandt, R. M., Ruppert, J., Wentworth, P. A., Hartman, M., Oseroff, C., Grey, H. M., Melief, C. J., & Kast, W. M. (1995). Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through in vivo and in vitro immunogenicity studies of HLA-A*0201-binding peptides. *Journal of immunology (Baltimore, Md. : 1950)*, *154*(11), 5934-5943. http://www.ncbi.nlm.nih.gov/pubmed/7538538

Révész, Á., Hevér, H., Steckel, A., Schlosser, G., Szabó, D., Vékey, K., & Drahos, L. (2023). Collision energies: Optimization strategies for bottom-up proteomics. *Mass Spectrometry Reviews*, *42*(4), 1261-1299. https://doi.org/10.1002/mas.21763

Révész, Á., Milley, M. G., Nagy, K., Szabó, D., Kalló, G., Csősz, É., Vékey, K., & Drahos, L. (2021). Tailoring to Search Engines: Bottom-Up Proteomics with Collision Energies Optimized for Identification Confidence. *Journal of Proteome Research*, *20*(1), 474-484. https://doi.org/10.1021/acs.jproteome.0c00518

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2021). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*(W1), W449-W454. https://doi.org/10.1093/NAR/GKAA379

Riemer, A. B., Keskin, D. B., Zhang, G., Handley, M., Anderson, K. S., Brusic, V., Reinhold, B., & Reinherz, E. L. (2010). A conserved E7-derived cytotoxic T lymphocyte epitope expressed on human papillomavirus 16-transformed HLA-A2+ epithelial cancers. *The Journal of biological chemistry*, *285*(38), 29608-29622. https://doi.org/10.1074/jbc.M110.126722

Roden, R. B. S., & Stern, P. L. (2018). Opportunities and challenges for human papillomavirus vaccination in cancer. *Nature Publishing Group*, *18*(4), 240-254. https://doi.org/10.1038/nrc.2018.13

Roepstorff, P., & Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry*, *11*(11), 601-601. https://doi.org/10.1002/bms.1200111109

Ronsein, G. E., Pamir, N., von Haller, P. D., Kim, D. S., Oda, M. N., Jarvik, G. P., Vaisar, T., & Heinecke, J. W. (2015). Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *Journal of Proteomics*, *113*(1), 388-399. https://doi.org/10.1016/j.jprot.2014.10.017

Rudis, B. (2020). *hrbrthemes: Additional Themes, Theme Components and Utilities for `ggp lot2'*. https://CRAN.R-project.org/package=hrbrthemes

Ruesch, M. N., & Laimins, L. A. (1998). Human papillomavirus oncoproteins alter differentiation-dependent cell cycle exit on suspension in semisolid medium. *Virology*, *250*(1), 19-29. https://doi.org/10.1006/viro.1998.9359

Salek, M., Förster, J. D., Becker, J. P., Meyer, M., Charoentong, P., Lyu, Y., Lindner, K., Lotsch, C., Volkmar, M., Momburg, F., Poschke, I., Fröhling, S., Schmitz, M., Offringa, R., Platten, M., Jäger, D., Zörnig, I., & Riemer, A. B. (2023). optiPRM: A targeted immunopeptidomics LC-MS workflow with ultra-high sensitivity for the detection of mutation-derived tumor neoepitopes from limited input material. *bioRxiv*, 2023.2008.2022.554248-552023.554208.554222.554248. https://doi.org/10.1101/2023.08.22.554248

Salek, M., Förster, J. D., Lehmann, W.-D., & Riemer, A. B. (2022). Light contamination in stable isotope-labelled internal peptide standards is frequent and a potential source of false discovery and quantitation error in proteomics. *Analytical and bioanalytical chemistry*, *414*(8), 2545-2552. https://doi.org/10.1007/s00216-022-03931-w

Salz, R., Bouwmeester, R., Gabriels, R., Degroeve, S., Martens, L., Volders, P. J., & Hoen, P. A. C. (2021). Personalized Proteome: Comparing Proteogenomics and Open Variant Search Approaches for Single Amino Acid Variant Detection. *Journal of Proteome Research*, *20*(6), 3353-3364. https://doi.org/10.1021/acs.jproteome.1c00264

Santegoets, S. J., Welters, M. J. P., Schrikkema, D. S., Freriks, M. R., Kok, H., Weissbrich, B., van den Branden, A., Linnemann, C., Schumacher, T. N., Adhikary, S., Bendle, G., & van der Burg, S. H. (2022). The common HLA class I-restricted tumor-infiltrating T cell response in HPV16-induced cancer. *Cancer immunology, immunotherapy : CII*. https://doi.org/10.1007/s00262-022-03350-x

Saper, M. A., Bjorkman, P. J., & Wiley, D. C. (1991). Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *Journal of Molecular Biology*, *219*(2), 277-319. https://doi.org/10.1016/0022-2836(91)90567-P

Sarkizova, S., Klaeger, S., Le, P. M., Li, L. W., Oliveira, G., Keshishian, H., Hartigan, C. R., Zhang, W., Braun, D. A., Ligon, K. L., Bachireddy, P., Zervantonakis, I. K., Rosenbluth, J. M., Ouspenskaia, T., Law, T., Justesen, S., Stevens, J., Lane, W. J., Eisenhaure, T., . . . Keskin, D. B. (2020). A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nature Biotechnology*, *38*(2), 199-209. https://doi.org/10.1038/s41587-019-0322-9

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., . . . Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic acids research*, *50*(D1), D20-D26. https://doi.org/10.1093/nar/gkab1112

Schifflers, C., Zottnick, S., Förster, J. D., Kruse, S., Yang, R., Wiethoff, H., Bozza, M., Hoppe-Seyler, K., Heikenwälder, M., Harbottle, R. P., Michiels, C., & Riemer, A. B. (2023). Development of an Orthotopic HPV16-Dependent Base of Tongue Tumor Model in MHC-Humanized Mice. *Pathogens*, *12*(2), 1-13. https://doi.org/10.3390/pathogens12020188

Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjosé, S., Fakhry, C., Monk, B. J., Stanley, M. A., & Franceschi, S. (2016). Carcinogenic human papillomavirus infection. *Nature reviews. Disease primers*, *2*, 16086-16086. https://doi.org/10.1038/nrdp.2016.86

Schiffmann, C., Hansen, R., Baumann, S., Kublik, A., Nielsen, P. H., Adrian, L., Von Bergen, M., Jehmlich, N., & Seifert, J. (2014). Comparison of targeted peptide quantification assays for reductive dehalogenases by selective reaction monitoring (SRM) and precursor reaction monitoring

(PRM). *Analytical and bioanalytical chemistry*, *406*(1), 283-291. https://doi.org/10.1007/s00216-013-7451-7

Schiller, J., & Lowy, D. (2018). Explanations for the high potency of HPV prophylactic vaccines. *Vaccine*, *36*(32), 4768-4773. https://doi.org/10.1016/j.vaccine.2017.12.079

Searle, B. C., Pino, L. K., Egertson, J. D., Ting, Y. S., Lawrence, R. T., MacLean, B. X., Villén, J., & MacCoss, M. J. (2018). Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07454-w

Sette, A., & Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, *50*(3-4), 201-212. https://doi.org/10.1007/s002510050594

Shafer, P., Kelly, L. M., & Hoyos, V. (2022). Cancer Therapy With TCR-Engineered T Cells: Current Strategies, Challenges, and Prospects. *Frontiers in Immunology*, *13*, 835762-835762. https://doi.org/10.3389/fimmu.2022.835762

Shapiro, I. E., & Bassani-Sternberg, M. (2023). The impact of immunopeptidomics: From basic research to clinical implementation. *Seminars in Immunology*, *66*, 101727-101727. https://doi.org/10.1016/j.smim.2023.101727

Sherwood, C. A. H., Gafken, P. R., & Martin, D. B. (2011). Collision energy optimization of b- and y-ions for multiple reaction Monitoring mass spectrometry. *Journal of Proteome Research*, *10*(1), 231-240. https://doi.org/10.1021/pr1004289

Shi, T., Su, D., Liu, T., Tang, K., Camp, D. G., Qian, W. J., & Smith, R. D. (2012). Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *PROTEOMICS*, *12*(8), 1074-1092. https://doi.org/10.1002/pmic.201100436

Shuken, S. R. (2023). An Introduction to Mass Spectrometry-Based Proteomics. *Journal of Proteome Research*, *22*(7), 2151-2171. https://doi.org/10.1021/acs.jproteome.2c00838

Sidney, J., Grey, H. M., Kubo, R. T., & Sette, A. (1996). Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. *Immunology today*, *17*(6), 261-266. https://doi.org/10.1016/0167-5699(96)80542-1

Sidney, J., Peters, B., Frahm, N., Brander, C., & Sette, A. (2008). HLA class I supertypes: a revised and updated classification. *BMC immunology*, *9*, 1-1. https://doi.org/10.1186/1471-2172-9-1

Smith, L., Cassetti, M. C., Pullen, J. K., & Mcelhiney, S. P. (2012). *Fusion peptides comprising the human papillomavirus e7 and e6 polypeptides and immunogenic compositions thereof* (EP1553966B1). E. P. Office.

Steinbach, A., & Riemer, A. B. (2018). Immune evasion mechanisms of human papillomavirus: An update. *International journal of cancer*, *142*(2), 224-229. https://doi.org/10.1002/ijc.31027

Steinbach, A., Winter, J., Reuschenbach, M., Blatnik, R., Klevenz, A., Bertrand, M., Hoppe, S., von Knebel Doeberitz, M., Grabowska, A. K., & Riemer, A. B. (2017). ERAP1 overexpression in HPV-induced malignancies: A possible novel immune evasion mechanism. *OncoImmunology*, *6*(7), 1-9. https://doi.org/10.1080/2162402X.2017.1336594

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660

Szabó, D., Schlosser, G., Vékey, K., Drahos, L., & Révész, Á. (2021). Collision energies on QTof and Orbitrap instruments: How to make proteomics measurements comparable? *Journal of mass spectrometry : JMS*, *56*(1), e4693-e4693. https://doi.org/10.1002/jms.4693

Tabb, D. L., Smith, L. L., Breci, L. A., Wysocki, V. H., Lin, D., & Yates, J. R. (2003). Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. *Analytical Chemistry*, *75*(5), 1155-1163. https://doi.org/10.1021/ac026122m

Ternette, N., Adamopoulou, E., & Purcell, A. W. (2023). How mass spectrometric interrogation of MHC class I ligandomes has advanced our understanding of immune responses to viruses. *Seminars in Immunology*, *68*, 101780-101780. https://doi.org/10.1016/j.smim.2023.101780

Thermo. (2012). *Product Support Bulletin 104: Normalized Collision Energy Technology* (Vol. 62641). Thermo Fisher Scientific Inc.

Thermo. (2021). *Orbitrap Exploris Series: Operating Manual* (Revision E ed., Vol. BRE0014471). Thermo Fisher Scientific Inc.

Thomas, K. J., Smith, K. L., Youde, S. J., Evans, M., Fiander, A. N., Borysiewicz, L. K., & Man, S. (2008). HPV16 E629-38-specific T cells kill cervical carcinoma cells despite partial evasion of T-cell effector function. *International journal of cancer*, *122*(12), 2791-2799. https://doi.org/10.1002/ijc.23475

Toprak, U. H., Gillet, L. C., Maiolica, A., Navarro, P., Leitner, A., & Aebersold, R. (2014). Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Molecular & cellular proteomics : MCP*, *13*(8), 2056-2071. https://doi.org/10.1074/mcp.O113.036475

Townsend, A., Bastin, J., Bodmer, H., Brownlee, G., Davey, J., Gotch, F., Gould, K., Jones, I., McMichael, A., & Rothbard, J. (1989). Recognition of influenza virus proteins by cytotoxic T lymphocytes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *323*(1217), 527-533. https://doi.org/10.1098/rstb.1989.0032

Trolle, T., McMurtrey, C. P., Sidney, J., Bardet, W., Osborn, S. C., Kaever, T., Sette, A., Hildebrand, W. H., Nielsen, M., & Peters, B. (2016). The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. *Journal of immunology (Baltimore, Md. : 1950)*, *196*(4), 1480-1487. https://doi.org/10.4049/jimmunol.1501721

Tsang, K. Y., Fantini, M., Fernando, R. I., Palena, C., David, J. M., Hodge, J. W., Gabitzsch, E. S., Jones, F. R., & Schlom, J. (2017). Identification and characterization of enhancer agonist human cytotoxic T-cell epitopes of the human papillomavirus type 16 (HPV16) E6/E7. *Vaccine*, *35*(19), 2605-2611. https://doi.org/10.1016/j.vaccine.2017.03.025

Ushey, K., & Wickham, H. (2023). *renv: Project Environments*. https://CRAN.R-project.org/package=renv

van Hateren, A., & Elliott, T. (2023). Visualising tapasin- and TAPBPR-assisted editing of major histocompatibility complex class-I immunopeptidomes. *Current Opinion in Immunology*, *83*, 102340-102340. https://doi.org/10.1016/j.coi.2023.102340

Wagih, O. (2017). *ggseqlogo: A `ggplot2' Extension for Drawing Publication-Ready Sequenc e Logos*. https://CRAN.R-project.org/package=ggseqlogo

Waldman, A. D., Fritz, J. M., & Lenardo, M. J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*, *20*(11), 651-668. https://doi.org/10.1038/s41577-020-0306-5

Wang, X., Sandberg, M. L., Martin, A. D., Negri, K. R., Gabrelow, G. B., Nampe, D. P., Wu, M. L., McElvain, M. E., Toledo Warshaviak, D., Lee, W. H., Oh, J., Daris, M. E., Chai, F., Yao, C., Furney, J., Pigott, C., Kamb, A., & Xu, H. (2021). Potent, Selective CARs as Potential T-Cell Therapeutics for HPV-positive Cancers. *Journal of Immunotherapy*, *44*(8), 292-306. https://doi.org/10.1097/CJI.0000000000000386

Wentworth, P. A., Vitiello, A., Sidney, J., Keogh, E., Chesnut, R. W., Grey, H., & Sette, A. (1996). Differences and similarities in the A2.1-restricted cytotoxic T cell repertoire in humans and human leukocyte antigen-transgenic mice. *European Journal of Immunology*, *26*(1), 97-101. https://doi.org/10.1002/eji.1830260115

Wickham, H. (2023). *conflicted: An Alternative Conflict Resolution Strategy*. https://CRAN.R-project.org/package=conflicted

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2023). *readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017). Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Frontiers in Immunology*, *8*, 292-292. https://doi.org/10.3389/fimmu.2017.00292

Wilhelm, M., Zolg, D. P., Graber, M., Gessulat, S., Schmidt, T., Schnatbaum, K., Schwencke-Westphal, C., Seifert, P., de Andrade Krätzig, N., Zerweck, J., Knaute, T., Bräunlein, E., Samaras, P.,

Lautenbacher, L., Klaeger, S., Wenschuh, H., Rad, R., Delanghe, B., Huhmer, A., . . . Kuster, B. (2021). Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23713-9

Woo, Y. L., Sterling, J., Damay, I., Coleman, N., Crawford, R., Van Der Burg, S. H., & Stanley, M. (2008). Characterising the local immune responses in cervical intraepithelial neoplasia: A cross-sectional and longitudinal analysis. *BJOG: An International Journal of Obstetrics and Gynaecology*, *115*(13), 1616-1622. https://doi.org/10.1111/j.1471-0528.2008.01936.x

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome biology*, *20*(1), 1-13. https://doi.org/10.1186/s13059-019-1891-0

Xie, Y., Allaire, J. J., & Horner, J. (2023). *markdown: Render Markdown with `commonmark'*. https://CRAN.R-project.org/package=markdown

Xiong, C., Huang, L., Kou, H., Wang, C., Zeng, X., Sun, H., Liu, S., Wu, B., Li, J., Wang, X., Wang, Z., & Chen, L. (2022). Identification of novel HLA-A∗11:01-restricted HPV16 E6/E7 epitopes and T-cell receptors for HPV-related cancer immunotherapy. *Journal for ImmunoTherapy of Cancer*, *10*(9). https://doi.org/10.1136/jitc-2022-004790

Xu, R., Sheng, J., Bai, M., Shu, K., Zhu, Y., & Chang, C. (2020). A Comprehensive Evaluation of MS/MS Spectrum Prediction Tools for Shotgun Proteomics. *PROTEOMICS*, *20*(21-22), 1-10. https://doi.org/10.1002/pmic.201900345

Yao, G., Qiu, J., Zhu, F., & Wang, X. (2022). Survival of Patients With Cervical Cancer Treated With Definitive Radiotherapy or Concurrent Chemoradiotherapy According to Histological Subtype: A Systematic Review and Meta-Analysis. *Frontiers in Medicine*, *9*, 843262-843262. https://doi.org/10.3389/fmed.2022.843262

Yewdell, J. W. (2022). MHC Class I Immunopeptidome: Past, Present, and Future. *Molecular & cellular proteomics : MCP*, *21*(7), 100230-100230. https://doi.org/10.1016/j.mcpro.2022.100230

Yewdell, J. W., Antón, L. C., & Bennink, J. R. (1996). Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? *Journal of immunology (Baltimore, Md. : 1950)*, *157*(5), 1823-1826. http://www.ncbi.nlm.nih.gov/pubmed/8757297

Youde, S. J., Dunbar, P. R., Evans, E. M., Fiander, A. N., Borysiewicz, L. K., Cerundolo, V., & Man, S. (2000). Use of fluorogenic histocompatibility leukocyte antigen-A*0201/HPV 16 E7 peptide complexes to isolate rare human cytotoxic T-lymphocyte-recognizing endogenous human papillomavirus antigens. *Cancer Research*, *60*(2), 365-371. http://www.ncbi.nlm.nih.gov/pubmed/10667589

Zapata, L., Caravagna, G., Williams, M. J., Lakatos, E., AbdulJabbar, K., Werner, B., Chowell, D., James, C., Gourmet, L., Milite, S., Acar, A., Riaz, N., Chan, T. A., Graham, T. A., & Sottoriva, A. (2023). Immune selection determines tumor antigenicity and influences response to checkpoint inhibitors. *Nature Genetics*, *55*(3), 451-460. https://doi.org/10.1038/s41588-023-01313-1

Zemmour, J., & Parham, P. (1992). Distinctive polymorphism at the HLA-C locus: Implications for the expression of HLA-C. *Journal of Experimental Medicine*, *176*(4), 937-950. https://doi.org/10.1084/jem.176.4.937

Zeng, W. F., Zhou, X. X., Zhou, W. J., Chi, H., Zhan, J., & He, S. M. (2019). MS/MS Spectrum prediction for modified peptides using pDeep2 Trained by Transfer Learning. *Analytical Chemistry*, *91*(15), 9724-9731. https://doi.org/10.1021/acs.analchem.9b01262

Zolg, D. P., Wilhelm, M., Yu, P., Knaute, T., Zerweck, J., Wenschuh, H., Reimer, U., Schnatbaum, K., & Kuster, B. (2017). PROCAL: A Set of 40 Peptide Standards for Retention Time Indexing, Column Performance Monitoring, and Collision Energy Calibration. *PROTEOMICS*, *17*(21), 1-5. https://doi.org/10.1002/pmic.201700263

Appendix A IP table

Appendix Table 1. Properties of the targeted IPs for the epitome map. For cell counts, the 'M' suffix is used to denote millions. The column "AB" specifies the trivial name of the respective anti-HLA AB that was used (see section 2.1.2.3). The column "IP Type" specifies which IP protocol was used. Batch or high-throughput 96-well plate-based (Plate). The column "IP cell load" specifies the how many cells the lysate input corresponded to. The column "Beads" specifies the type of beads used for the IP. Gammabind (Gb) or ProteinA (PA). The columns "Oxi." and "Alkyl." specify whether the oxidation and alkylation treatment were performed, respectively. The column "SIL spike-in" specifies whether assay-internal synthetic SIL peptides were used. The column "LC-MS load" specifies the cell load equivalent that was injected for analysis. The column "Grd." specifies the LC gradient. OptiGrd3 (OG3) or Grd4 (G4).

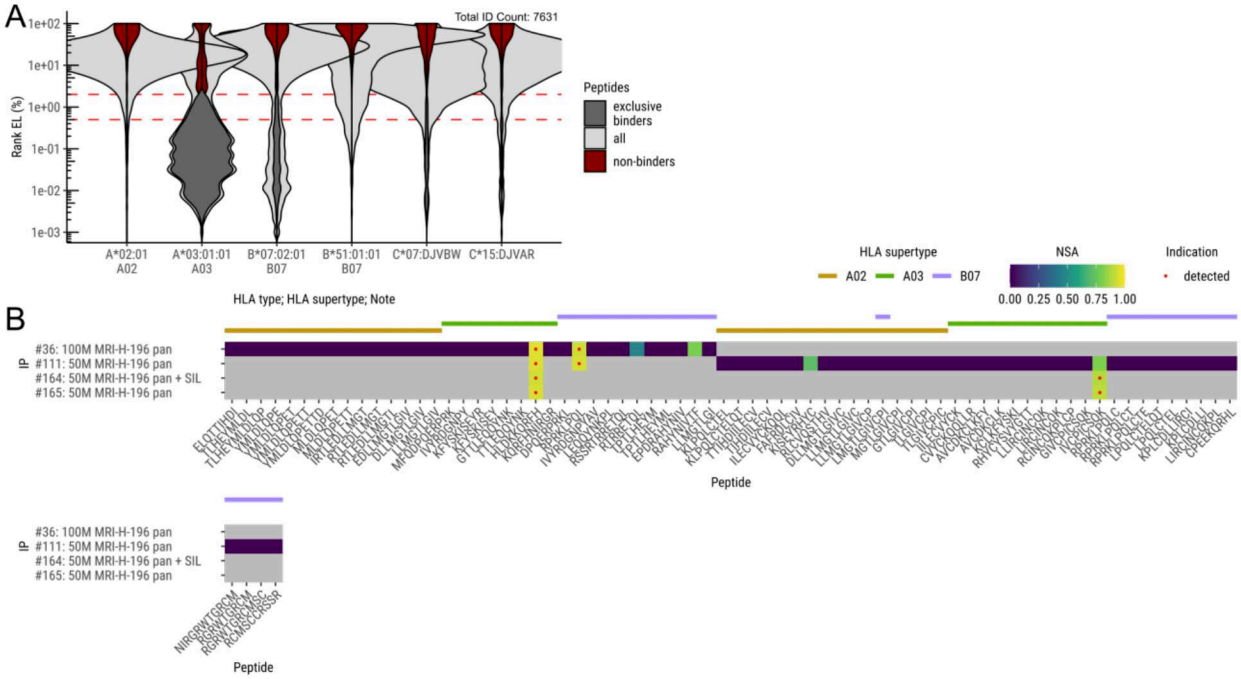
#	Cell Line	AB	IP Type	IP cell load	Beads	Oxi.	Alkyl.	SIL spike-in	LC-MS load	Grd.
10	CaSki2020	B7	Batch	100M	Gb	No	No	No	100M	OG3
11	SNU-902	B7	Batch	100M	Gb	No	No	No	100M	OG3
19	SNU-1005	A11A24	Batch	100M	Gb	No	No	No	100M	OG3
23	W12-20861	A11A24	Batch	100M	Gb	No	No	No	100M	OG3
27	SNU-703	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
28	SNU-1299	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
34	SNU-1005	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
35	SiHa	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
36	MRI-H-196	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
38	W12-20861	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
39	SNU-1000	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
40	UM-SCC-104	pan	Batch	100M	Gb	Yes	No	No	100M	OG3
41	CaSki2020	B7	Batch	100M	Gb	Yes	No	No	100M	OG3
43	SNU-902	B7	Batch	100M	Gb	Yes	No	No	100M	OG3
46	SNU-1299	pan	Batch	100M	PA	Yes	No	No	100M	OG3
47	SNU-1005	pan	Batch	100M	PA	Yes	No	No	100M	OG3
50	SNU-1000	pan	Batch	100M	PA	Yes	No	No	100M	OG3
58	CaSki2020	pan	Plate	2x50M	PA	No	No	No	100M	OG3
62	SNU-1005	pan	Batch	100M	PA	Yes	No	No	100M	OG3
73	CaSki2020	pan	Batch	200M	PA	Yes	No	No	100M	OG3
75	DoTc-2-4510	pan	Batch	100M	PA	Yes	No	No	100M	OG3
76	W12-20861	pan	Batch	300M	PA	No	No	No	100M	OG3
77	W12-20861	pan	Batch	300M	PA	No	No	No	200M	OG3
81	SNU-1005	A11A24	Batch	300M	Gb	No	No	No	200M	OG3

84	SNU-902	pan	Batch	300M	PA	No	No	No	300M	OG3
93	SNU-17	pan	Plate	50M	PA	Yes	No	No	50M	OG3
102	UM-SCC47	pan	Plate	50M	PA	Yes	No	No	50M	G4
103	UD-SCC2	pan	Plate	50M	PA	Yes	No	No	50M	G4
104	Goerke	pan	Plate	50M	PA	Yes	No	No	50M	G4
105	Cerv-215	pan	Plate	50M	PA	Yes	No	No	50M	G4
106	SCC152	pan	Plate	50M	PA	Yes	No	No	50M	G4
107	SCC154	pan	Plate	50M	PA	Yes	No	No	50M	G4
108	866	pan	Plate	50M	PA	Yes	No	No	50M	G4
109	879	pan	Plate	50M	PA	Yes	No	No	50M	G4
110	CaSki2020	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
111	MRI-H-196	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
112	SNU-902	pan	Plate	2x50M	PA	Yes	Yes	No	100M	G4
113	SiHa	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
114	866	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
115	SNU-17	pan	Plate	2x50M	PA	Yes	Yes	No	100M	G4
116	Goerke	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
117	W12-20861	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
118	DoTc-2-4510	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
120	UM-SCC47	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
121	SNU-703	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
122	SNU-1299	pan	Plate	2x50M	PA	Yes	Yes	No	100M	G4
123	UD-SCC2	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
124	SCC154	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
125	879	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
126	SCC152	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
127	Cerv-215	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
129	SCC090	pan	Plate	50M	PA	Yes	No	No	50M	G4
130	SCC090	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
131	SNU-1005	pan	Plate	2x50M	PA	Yes	Yes	No	100M	G4
132	SNU-1000	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
133	UM-SCC-104	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
135	CaSki2020	pan	Plate	50M	PA	Yes	Yes	Yes	50M	G4

136	CaSki2020	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
139	W12-20861	pan	Plate	50M	PA	Yes	Yes	Yes	50M	G4
140	W12-20861	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
141	SNU-1299	pan	Plate	2x50M	PA	Yes	Yes	Yes	100M	G4
142	SNU-1299	pan	Plate	2x50M	PA	Yes	Yes	No	100M	G4
143	SNU-1299	pan	Plate	2x50M	PA	Yes	Yes	No	100M	G4
144	SNU-902	pan	Plate	50M	PA	Yes	Yes	Yes	50M	G4
146	SNU-902	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
148	Goerke	pan	Plate	2x25M	PA	Yes	Yes	Yes	50M	G4
149	W12-20861	pan	Plate	35M	PA	Yes	Yes	Yes	3.5M	G4
150	SCC152	pan	Plate	50M	PA	Yes	Yes	Yes	50M	G4
151	SCC152	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
152	CaSki2020	B7	Batch	300M	PA	Yes	No	Yes	300M	G4
153	SNU-17	pan	Batch	200M	PA	Yes	No	Yes	200M	G4
154	SiHa	pan	Plate	50M	PA	Yes	Yes	Yes	50M	G4
155	SiHa	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
157	866	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
161	879	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
162	879	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
163	UM-SCC-104	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
164	MRI-H-196	pan	Plate	50M	PA	No	Yes	Yes	50M	G4
165	MRI-H-196	pan	Plate	50M	PA	No	Yes	No	50M	G4
167	SNU-1005	pan	Plate	100M	PA	No	Yes	No	100M	G4
169	SNU-1000	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
170	SiHa	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
172	W12-20861	pan	Plate	50M	PA	Yes	Yes	No	50M	G4
174	Goerke	pan	Plate	50M	PA	No	Yes	No	50M	G4
175	SCC152	pan	Plate	52M	PA	Yes	Yes	No	52M	G4
176	SCC090	pan	Plate	100M	PA	Yes	Yes	No	100M	G4
177	UM-SCC-104	pan	Plate	48M	PA	Yes	Yes	No	48M	G4
178	Goerke	pan	Plate	60M	PA	Yes	Yes	No	60M	G4
179	SNU-1299	pan	Plate	50M	PA	Yes	Yes	No	5M	G4
180	866	pan	Plate	50M	PA	Yes	Yes	No	5M	G4

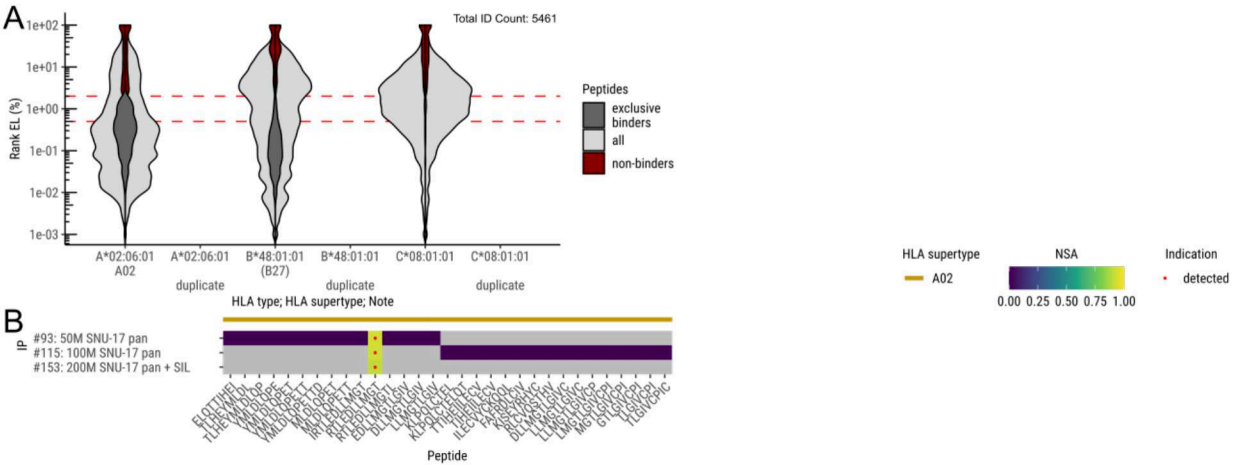
Appendix B Cell line results

MRI-H-196



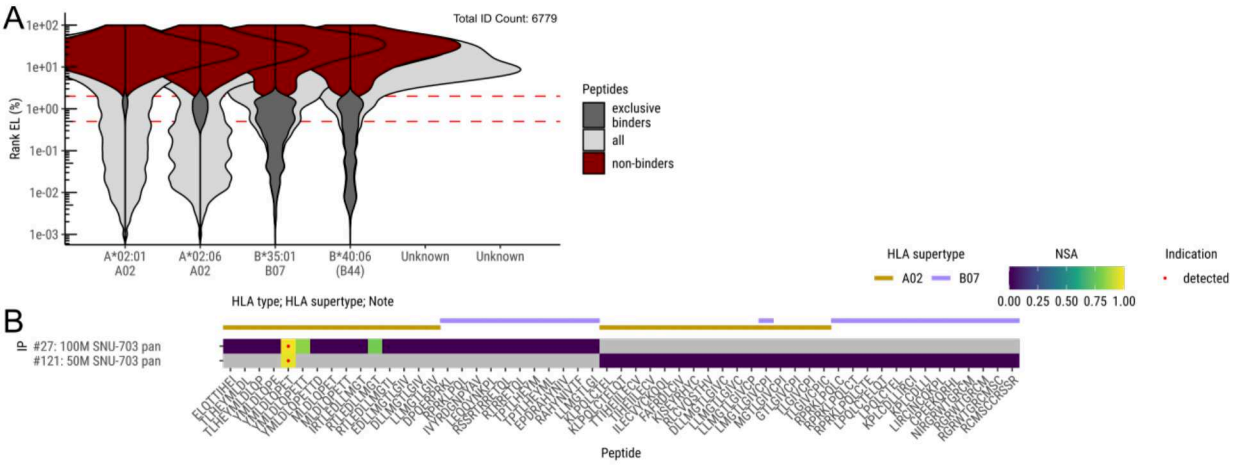
Appendix Figure 1. Cell line MRI-H-196 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SNU-17



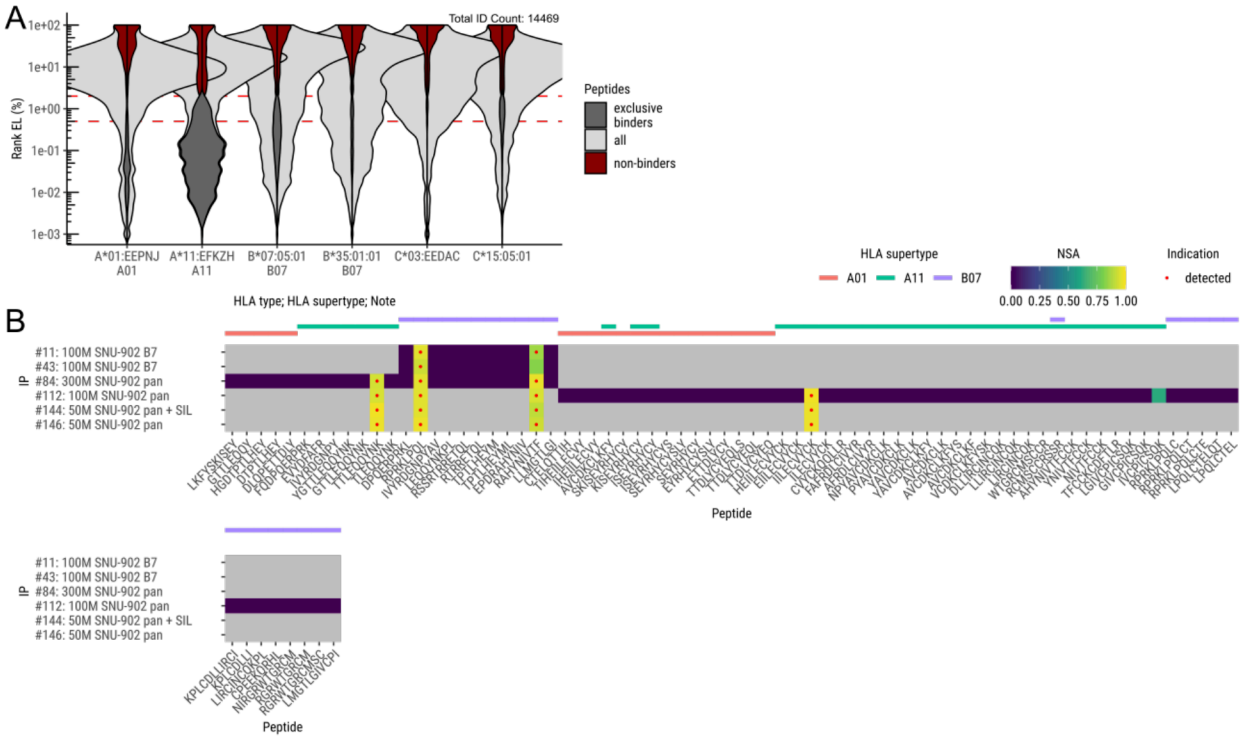
Appendix Figure 2. Cell line SNU-17 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SNU-703



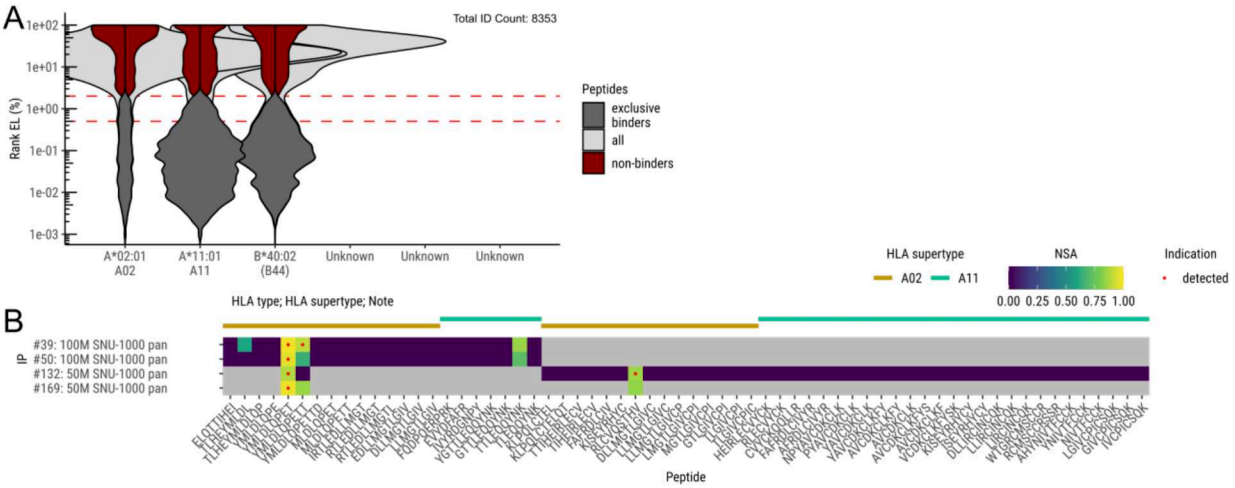
Appendix Figure 3. Cell line SNU-703 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SNU-902



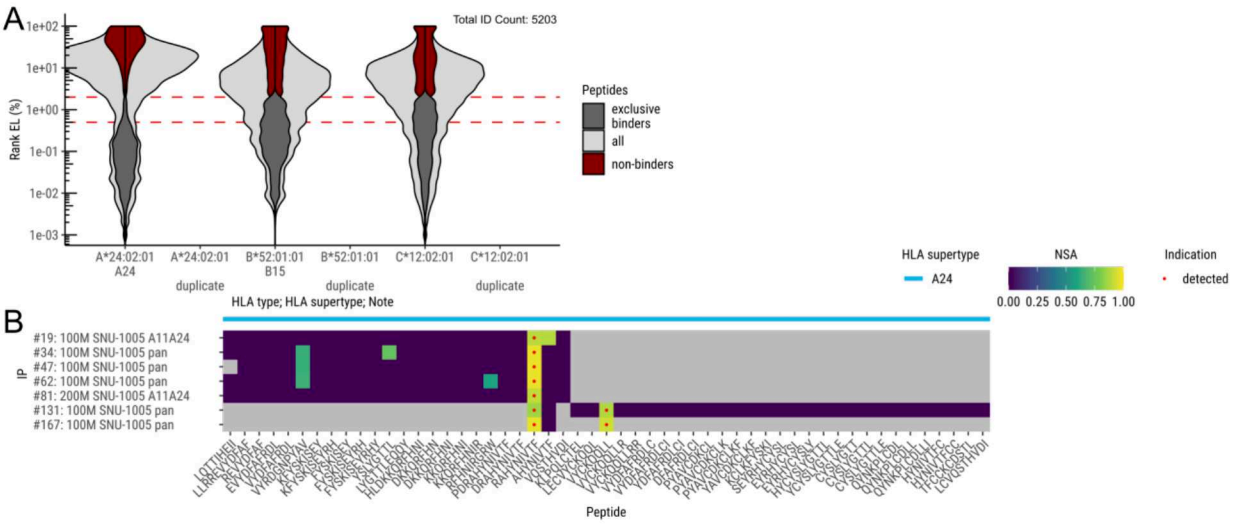
Appendix Figure 4. Cell line SNU-902 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SNU-1000



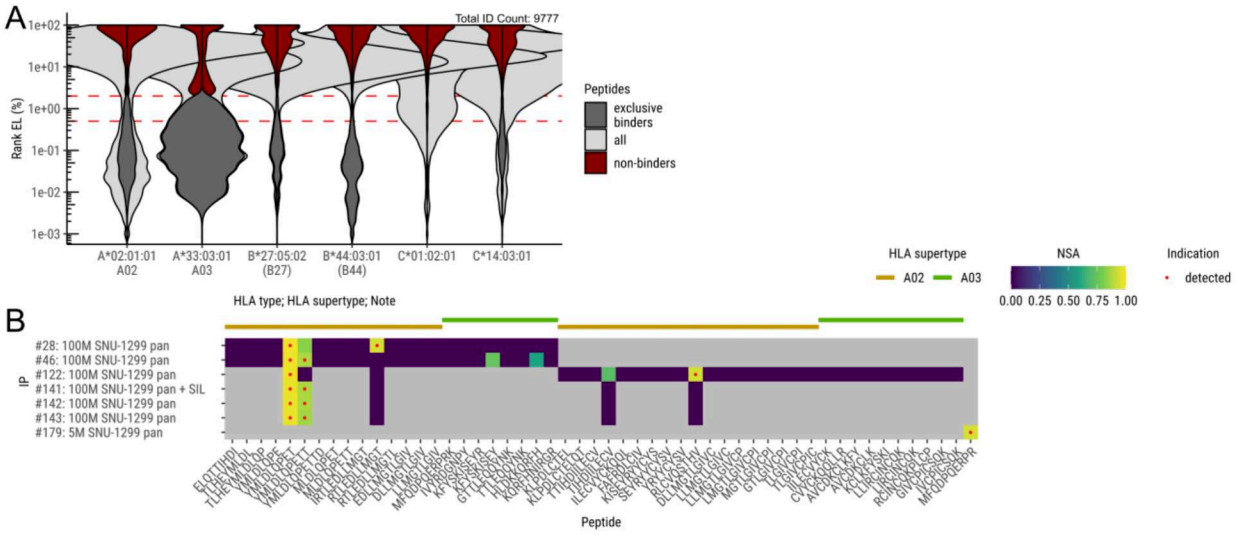
Appendix Figure 5. Cell line SNU-1000 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SNU-1005



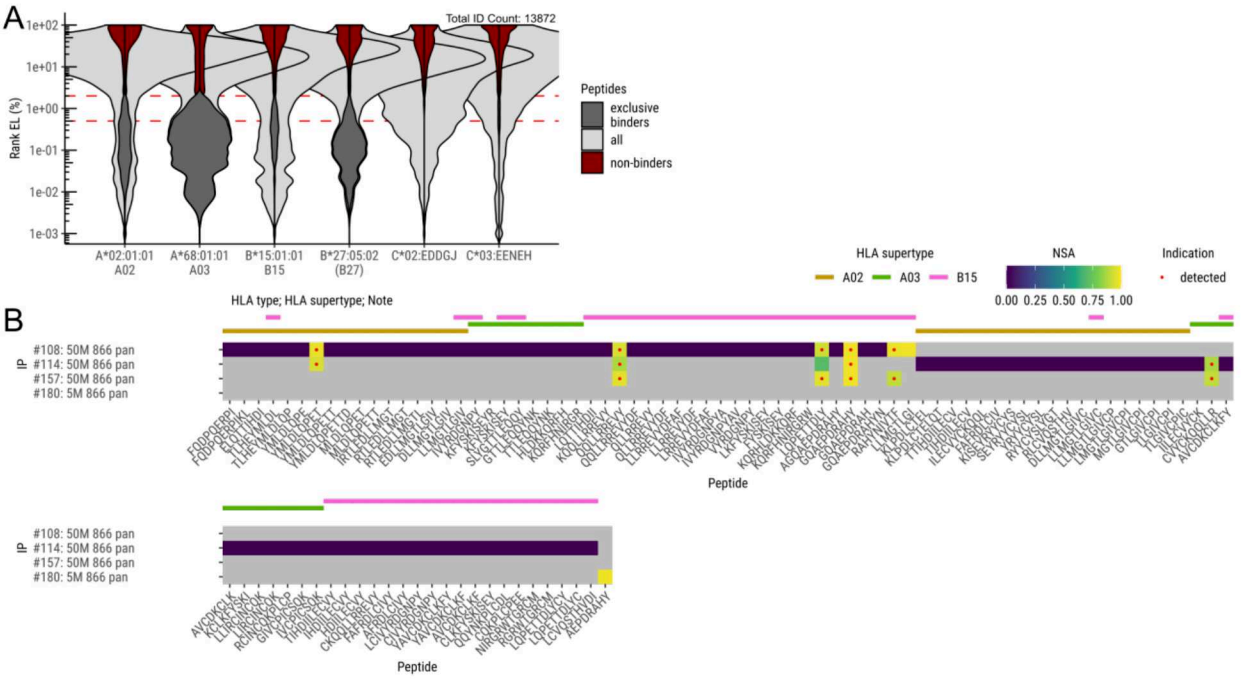
Appendix Figure 6. Cell line SNU-1005 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SNU-1299



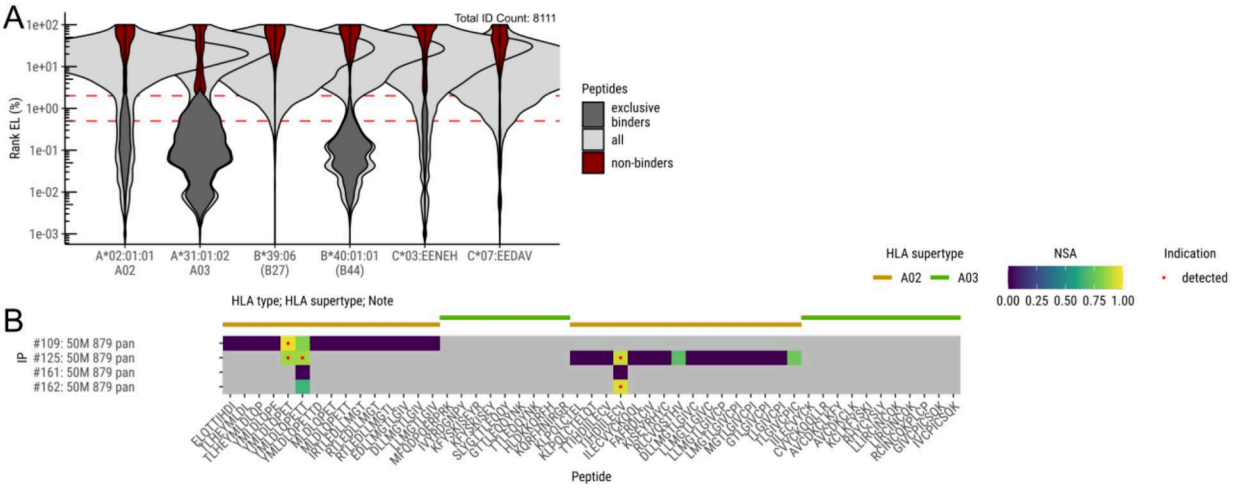
Appendix Figure 7. Cell line SNU-1299 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

866



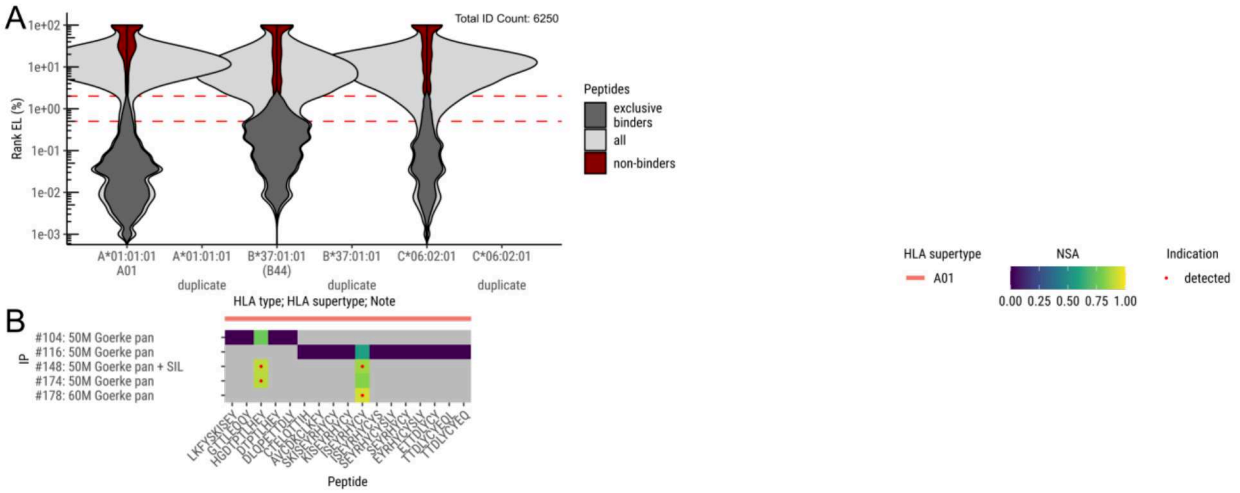
Appendix Figure 8. Cell line 866 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

879



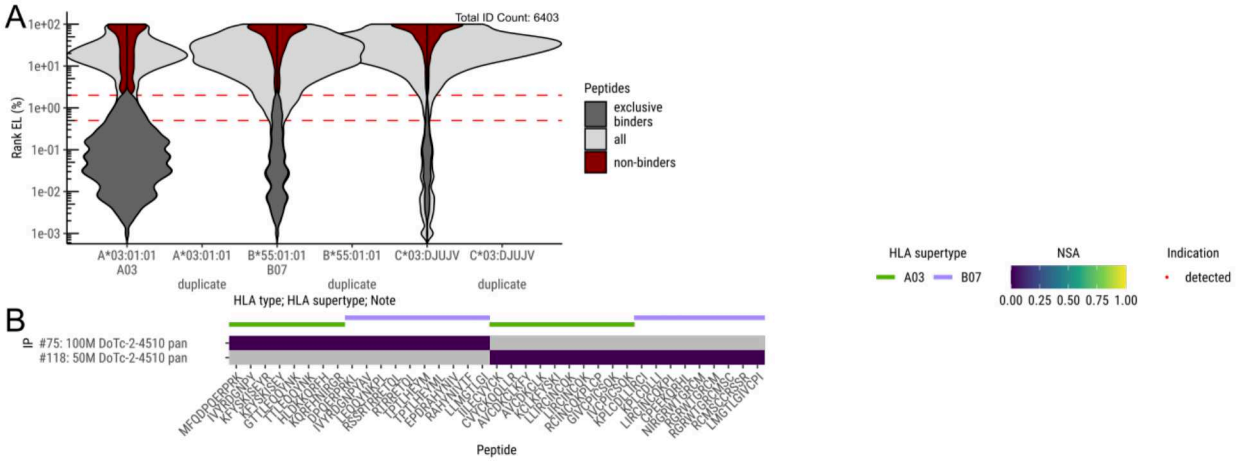
Appendix Figure 9. Cell line 879 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

Goerke



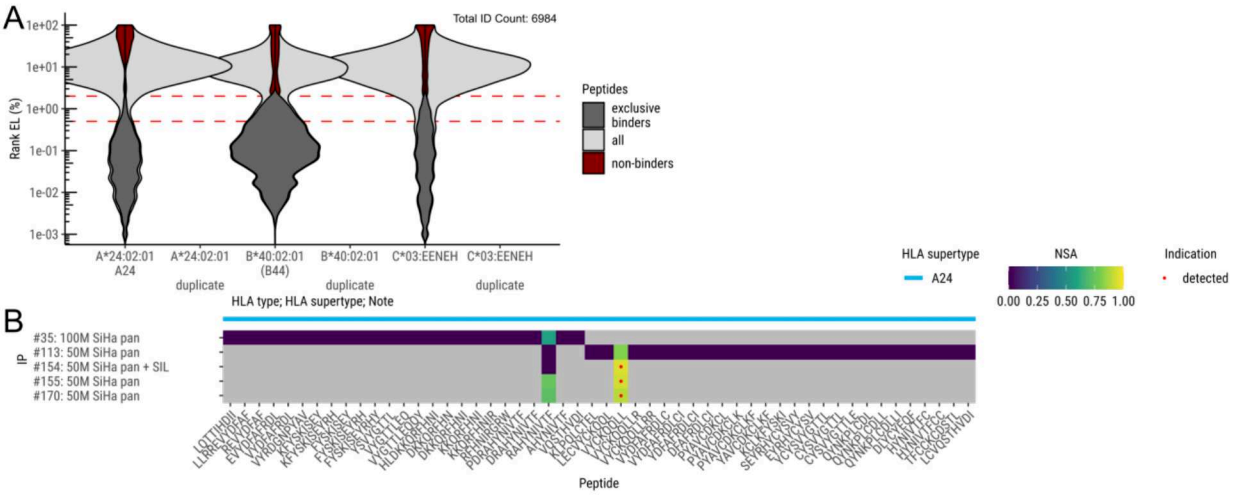
Appendix Figure 10. Cell line Goerke result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

DoTc2-4510



Appendix Figure 11. Cell line DoTc2-4510 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SiHa



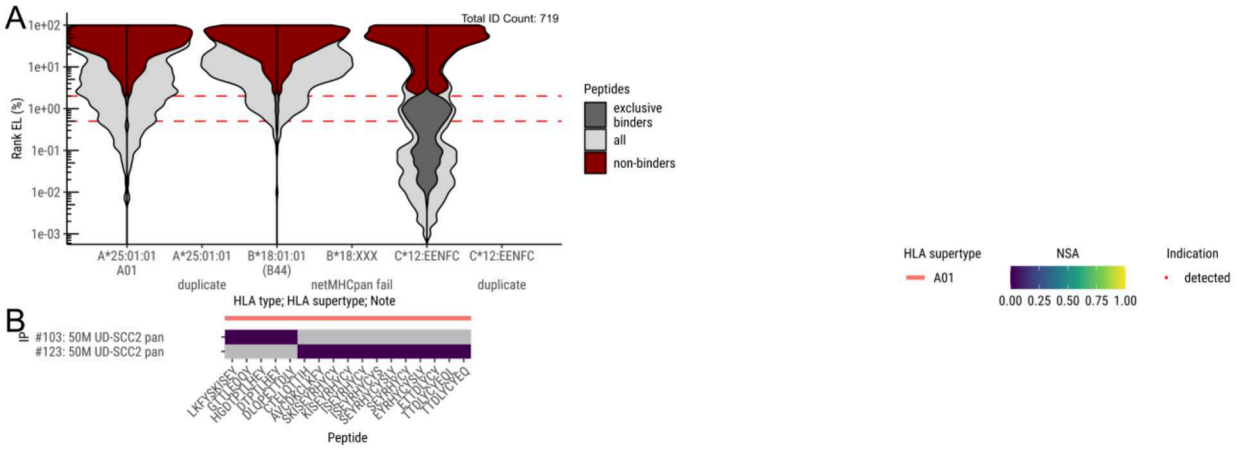
Appendix Figure 12. Cell line SiHa result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

SCC154



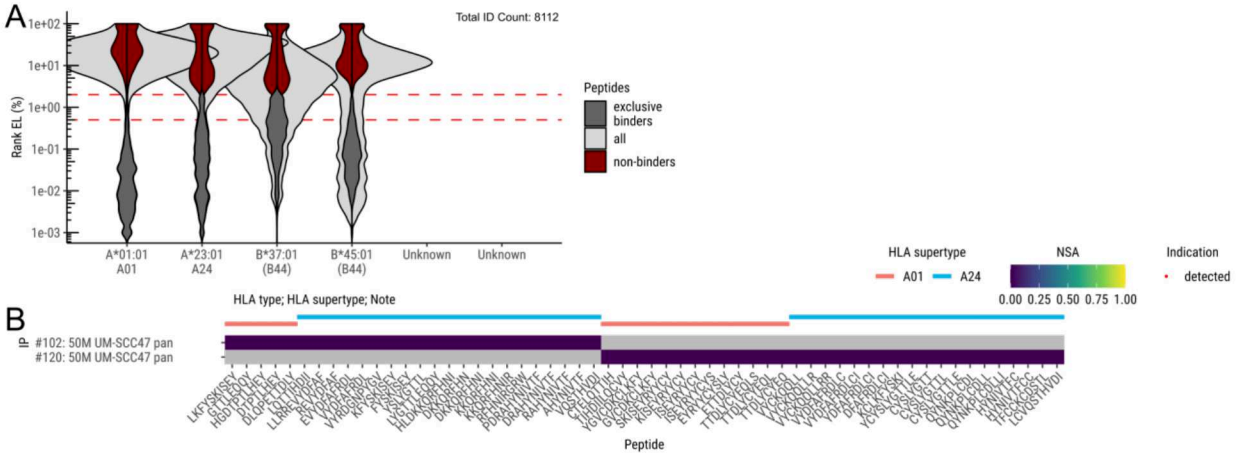
Appendix Figure 13. Cell line SCC154 result compilation. No untargeted evaluation was performed. NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

UD-SCC2



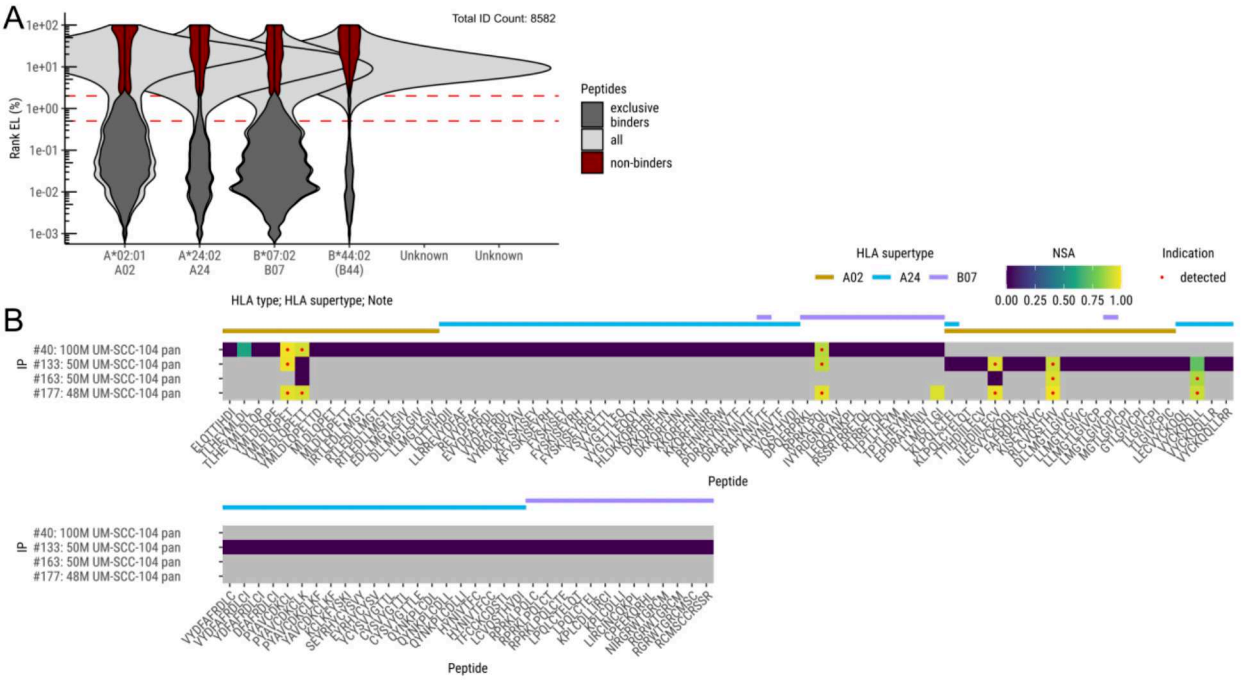
Appendix Figure 14. Cell line UD-SCC2 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

UM-SCC-47



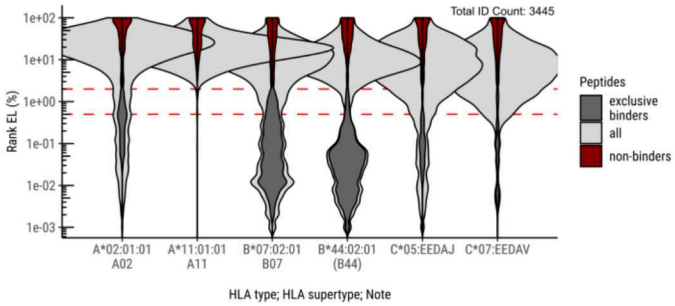
Appendix Figure 15. Cell line UM-SCC-47 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

UM-SCC-104



Appendix Figure 16. Cell line UM-SCC-104 result compilation. (A) Untargeted cell line immunopeptidome evaluation. (B) NSA results of targeted experiments. Specification for each IP contains: ID, cell load, cell line, antibody, SIL spike-in. For peptides where multiple precursors were targeted, the best result is shown. Colored (not gray) fields indicate that the peptide was targeted. Main criterion for detection indication (red dot) is NSA ≥ 0.85. Annotated above are the HLA supertype associations for each peptide from experimental binding tests. Peptides are sorted in the following order: First non-cysteine- and cysteine-containing peptides are split, then HLA association, source protein (E6/E7) and sequence position.

Negative Control: C33A



Appendix Figure 17. Cell line C33A result compilation. Untargeted cell line immunopeptidome evaluation.