

Dissertation

Submitted to the

Combined Faculty for Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

MSc Maria Bonsack

Born in Friedrichroda, Germany

Oral Examination: 23.09.2019

**Identification and biological validation of
HPV16 E6/E7-derived T cell target epitopes
and their use for performance assessment of
MHC class I binding predictors**

Referees: Prof. Dr. Ralf Bartenschlager
PD Dr. Dr. Angelika Riemer

The work described in this thesis was performed from January 2016 to June 2019 under scientific supervision of PD Dr. Dr. Angelika Riemer in the group Immunotherapy and Immunoprevention at the German Cancer Research Center (DKFZ) in Heidelberg, Germany.

Peer-reviewed publications based on this work:

Bonsack M*, Hoppe S*, Winter J, Tichy D, Zeller C, Küpper MD, Schitter EC, Blatnik R, and Riemer AB. (2019). **“Performance Evaluation of MHC Class-I Binding Prediction Tools Based on an Experimentally Validated MHC–Peptide Binding Data Set.”** *Cancer Immunology Research* 7, 719-736.

Blatnik R*, Mohan N*, **Bonsack M***, Falkenby LG, Hoppe S, Josef K, Steinbach A, Becker S, Nadler WM, Rucevic M, , Larsen MR, Salek M and Riemer AB. (2018). **“A Targeted LC-MS Strategy for Low-Abundant HLA Class-I-Presented Peptide Detection Identifies Novel Human Papillomavirus T-Cell Epitopes.”** *Proteomics* 18, 1700390.

O’Donnell TJ, Rubinsteyn A, **Bonsack M**, Riemer AB, Laserson U, and Hammerbacher J. (2018). **“MHCflurry: Open-Source Class I MHC Binding Affinity Prediction.”** *Cell Systems* 7, 129-132.e4.

Shao, XM., Bhattacharya, R., Huang, J., Sivakumar, A., Tokheim, C., Zheng, L., Kaminow, B., Omdahl, A., **Bonsack, M.**, Riemer, AB., Velculescu, V.E., Anagnostou, V., Pagel, K., Karchin, R. (*submitted*). **“High-throughput prediction of MHC Class I and Class II neoantigens with MHCnuggets.”**. *Cancer Immunology Research*, Preprint Server bioRxiv, ID:154757.

Ballhausen, A., Przybilla, MJ., Jendrusch, M., Haupt, S., Pfaffendorf, E., Draxlbauer, M., Seidler, F., Krausert, S., Ahadova, A., Kalteis, MS., Heid, D., Gebert, J., Bonsack, M., Schott, S., Bläker, H., Seppälä, T., Mecklin, J., Broeke, ST., Nielsen, M., Heuveline, V., Krzykalla, J., Benner, A., Riemer, AB., von Knebel Doeberitz, M., Kloor, M. (*in submission*). **“The shared neoantigen landscape of MSI cancers reflects immunoediting during tumor evolution.”** Preprint Server bioRxiv, ID: 691469.

*contributed equally

Conference and workshop presentations based on this work:

Hoppe S, Bonsack M, Winter J, Zeller C, Blatnik R, Riemer AB. Evaluation of T cell epitope prediction servers with experimental HLA binding assay data. (2016) Poster presentation at the Annual Meeting of the German Society for Immunology (DGfI) 2016, Hamburg, Germany

Bonsack M, Hoppe S, Winter J, Zeller C, Blatnik R, Riemer AB. Experimental performance evaluation of T cell epitope prediction tools. (2016) Oral presentation and poster presentation at the 8th German-Israeli Cancer Research School “Cancer Immunotherapy”, Neve Ilan, Israel

Bonsack M. Performance evaluation of *in silico* MHC class I prediction tools. (2017) Oral presentation at the Infection, Inflammation and Cancer (IIC) retreat 2017, Rastatt, Germany

- Bonsack M, Hoppe S, Winter J, Blatnik R, Riemer AB. Performance evaluation of *in silico* MHC class I binding prediction tools based on *in vitro* assessment of binding affinity. (2017) Poster presentation at the workshop “Data resources and bioinformatics tools for immunologists”, EMBL-EBI, Hinxton, United Kingdom
- Bonsack M, Hoppe S, Winter J, Blatnik R, Riemer AB. Performance evaluation and optimized decision thresholds for MHC class I binding prediction algorithms. (2017) Poster presentation at the 3rd CRI-CIMT-EATI-AACR international cancer immunotherapy conference, Mainz, Germany
- Bonsack M, Hoppe S, Winter J, Blatnik R, Riemer AB. Experimental performance evaluation and threshold recommendations for MHC class I binding prediction tools. (2017) Poster presentation at the Joint annual meeting of the German Society for Immunology (DGfI) and the German center for infection research (DZIF), Hamburg, Germany
- Bonsack M; Becker S; Hoppe S; Küpper MD; Dreßler L; Schessner J; Blatnik R; Salek M; Riemer AB. MS-based detection and functional characterization of HLA-A2-restricted HPV16 E6 and E7-derived epitopes as a basis for therapeutic vaccine design. (2018) Poster presentation at the National center for tumor diseases (NCT) retreat 2018, Radebeul, Germany
- Bonsack M; Mohan N, Becker S; Hoppe S; Küpper MD; Dreßler L; Schessner J; Salek M; Blatnik R, Riemer AB. Targeted LC-MS detection identifies novel immunogenic HLA-A2-restricted T cell epitopes derived from HPV16 E6 and E7. (2018) Poster presentation at the annual meeting of the association for cancer immunotherapy (CIMT) 2018, Mainz, Germany
- Bonsack M; Mohan N, Becker S; Hoppe S; Küpper MD; Blatnik R; Salek M, Riemer AB. Identification of HPV16 E6/E7-derived T cell epitopes by targeted LC-MS and immunogenicity testing. (2018) Oral presentation at the DKFZ PhD retreat 2018, Weil der Stadt, Germany
- Bonsack M; Mohan N, Becker S; Lotsch C, Hoppe S; Küpper MD; Dreßler L; Schessner J; Blatnik R, Salek M; Riemer AB. Identification of HPV16 E6 and E7-derived HLA class-I epitopes for the development of a therapeutic vaccine against HPV-induced malignancies. (2019) Oral presentation and poster presentation at the 1st Winter School of the association for cancer immunotherapy “CIMT Academy of Translational Cancer Immunology”, Obergurgl, Austria
- Bonsack M, Hoppe S, Winter J, Tichy D, Zeller C, Küpper MD, Schitter EC, Blatnik R, Riemer AB. Evaluation of MHC class-I binding prediction tools: Performance assessment and threshold recommendations based on experimentally validated ligands. (2019) Poster presentation at the annual meeting of the association for cancer immunotherapy (CIMT) 2019, Mainz, Germany
- Lotsch C; Bonsack M; Mohan N; Heinze J; Schmitt J; Hoppe S; Förster J; Salek M; Riemer, AB. Identification and immunological characterization of HLA-A3-restricted epitopes derived from HPV16 E6 and E7. (2019) Poster presentation at the annual meeting of the association for cancer immunotherapy (CIMT) 2019, Mainz, Germany

I. Abstract

Persistent infection with high-risk types of human papillomavirus (HPV) can cause several malignancies, in particular oropharyngeal and anogenital cancers. HPV16 has been identified as the most prevalent high-risk type, being related to 60% of cervical cancers, 75% of oropharyngeal cancers, 71% of anal cancers and the majority of precancerous lesions. As standard of care treatment is invasive, and harbors risks and side effects, there is a need for new approaches. For rationally designing a therapeutic vaccine against HPV-induced malignancies, it is essential to identify suitable target epitopes, which are presented on the surface of an HPV-transformed cell and induce immune responses that eventually mediate target cell death. The HPV16 oncoproteins E6 and E7 represent ideal targets for immunotherapy as they mediate the transforming potential of the virus and are constitutively expressed in all malignant cells.

In order to define HPV16 target epitopes, in this thesis several algorithms were used to predict potential HPV16 E6- and E7-derived binders of human leukocyte antigen (HLA) class I *in silico*. Predicted peptides were synthesized and HLA binding capacity was validated in competition-based cellular binding assays. To ensure broad population coverage, predictions and validations were performed for seven frequent HLA alleles: *A*01:01*, *A*02:01*, *A*03:01*, *A*11:01*, *A*24:02*, *B*07:02* and *B*15:01*. Including peptides derived from HPV16 E6/E7 variants containing amino acid changes, 271 peptides were experimentally assessed and 69 binders were identified. Combined with previous results, the total HPV16 E6/E7 dataset comprised 779 peptide-HLA measurements.

The HPV16 E6/E7 dataset was used to evaluate the performance of employed predictors. No single algorithm was outperforming other methods, but different predictors were found to be best for different settings, depending on investigated HLA type and peptide length. As applying commonly used decision threshold yielded only low sensitivity, criteria for optimal decision thresholds were defined and optimal thresholds were calculated for individual predictors, HLA-types and peptide lengths. Comparing threshold-dependent performance of predictors showed that using criteria-based thresholds allowed more sensitive prediction of HLA-binding peptides without a strong negative influence on prediction accuracy.

To identify T cell epitopes among the HPV16 E6- and E7-derived HLA ligands, their capacity to induce immune responses was investigated. To this end, peripheral blood mononuclear cells of healthy donors were HLA-typed and stimulated with respective peptides to generate epitope-specific T cell lines. By assessing interferon- γ -secretion of these T cells, 31 immunogenic peptides were identified. Further characterizing the functionality of epitopes in cytotoxicity assays, five of ten immunogenic HLA-A*02:01-peptides mediated specific killing of HPV16⁺ target cells by CD8⁺ T cells.

In conclusion, several immunogenic HPV16 E6-and E7-derived epitopes were identified, which are the basis for rational design of a therapeutic HPV vaccine. Additionally, this thesis provides an evaluation of peptide–HLA class-I binding prediction method and recommendations to increase prediction sensitivity to extend the number of potential epitopes as targets for immunotherapy.

II. Zusammenfassung

Anhaltende Infektionen mit Hoch-Risiko-Typen von humanen Papillomviren (HPVs) können diverse maligne Erkrankungen auslösen, insbesondere Krebs des Oropharyngeal-, Anal- und Genitaltrakts. HPV16 wurde als häufigster Hoch-Risiko-Typ identifiziert, verantwortlich für 60% aller Zervix-, 75% der Oropharynx- und 71% der Analkarzinome sowie die Mehrheit prä-kanceröser Läsionen. Da Standardbehandlungen invasiv und mit Nebenwirkungen wie auch Risiken verbunden sind, werden neue Therapiemöglichkeiten benötigt. Um einen Impfstoff gegen HPV-induzierte Krankheiten zu entwickeln, ist es essentiell geeignete Ziel-Epitope zu identifizieren, die auf der Oberfläche von HPV-transformierten Zellen präsentiert werden und Immunantworten auslösen, die zum Zielzelltod führen. Die HPV16 Onkoproteine E6 und E7 sind ideale Zielstrukturen für Immuntherapien, da sie dem Virus sein Transformations-Potenzial verleihen und konstitutiv in allen malignen Zellen exprimiert werden.

Um potenzielle HPV16 E6 und E7 Peptide mit Bindungskapazität zu humanen Leukozyten-Antigenen (HLA) Klasse I Molekülen *in silico* vorherzubestimmen, wurden verschiedene Algorithmen benutzt. Die Peptide wurden synthetisiert und ihre HLA-Bindung in kompetitiven Bindungstests validiert. Um einen Großteil der Bevölkerung abdecken zu können, wurden diese Experimente für sieben häufige HLA-Allele durchgeführt, A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, B*07:02 und B*15:01. Inklusiv von Peptiden, die von HPV16 E6/E7 Varianten mit Aminosäureaustausch stammen, wurden 271 experimentell validiert und 69 Binder identifiziert. Gemeinsam mit vorherigen Resultaten umfasst der HPV16 E6/E7 Datensatz 779 Peptid-HLA-Messungen.

Mithilfe des HPV16 E6/E7 Datensatzes wurde die Leistung der verwendeten Vorhersagemethoden evaluiert. Kein Algorithmus übertraf alle anderen, sondern verschiedene Methoden waren, abhängig von untersuchten HLA-Typen und Peptidlängen, die leistungsstärksten. Weil die Verwendung gebräuchlicher Grenzwerte in geringer Sensitivität resultierte, wurden Kriterien für optimale Grenzwerte definiert und optimale Grenzwerte individuell für Algorithmen, HLA-Typen und Peptidlängen berechnet. Im Vergleich mit gebräuchlichen Grenzwerten erreichte die Verwendung optimaler Grenzwerte eine höhere Sensitivität ohne starken negativen Einfluss auf die Genauigkeit.

Um T-Zell-Epitope unter den HLA-bindenden HPV16 E6/E7 Peptiden zu identifizieren, wurde deren Fähigkeit zur Induktion von Immunantworten ermittelt. Dazu wurden mononukleäre Zellen des peripheren Blutes gesunder Spender HLA-typisiert und mit entsprechenden Peptiden stimuliert, um Epitop-spezifische T-Zelllinien zu generieren. Durch Untersuchung der Interferon- γ -Sekretion von restimulierten T-Zelllinien wurden 31 immunogene Peptide identifiziert. In Zytotoxizitäts-Studien zur weiteren Charakterisierung der Funktionalität der Epitopevermittelten 5 von 10 immunogenen HLA-A*02:01-Peptiden den spezifischen Tod HPV16⁺ Zielzellen durch CD8⁺ T-Zellen.

Zusammenfassend wurden mehrere HPV16 E6/E7 Epitope identifiziert, die die Basis für die Entwicklung eines therapeutischen HPV-Impfstoffs bilden. Zusätzlich legt diese Arbeit die Evaluierung von Peptid-HLA Klasse I Bindungs-Vorhersagemethoden vor, sowie Empfehlungen um deren Sensitivität und die Anzahl potenzieller Epitope als Zielstrukturen für Immuntherapien zu erhöhen.

III. Acknowledgements

Ich bedanke mich bei allen Personen, die mich bei der Umsetzung und Erstellung dieser Doktorarbeit fachlich sowie persönlich unterstützt haben.

Zuerst möchte ich mich bei PD Dr. Dr. **Angelika Riemer** bedanken, die mir die Möglichkeit gab, diese Doktorarbeit in ihrer Arbeitsgruppe anzufertigen. Herzlichen Dank für dein Vertrauen und deine Unterstützung in wissenschaftlicher und auch in persönlicher Art. Den von dir gebotenen wissenschaftlichen Freiraum und die Chancen zur Weiterentwicklung schätze ich sehr.

Bei Prof. Dr. **Ralf Bartenschlager** möchte ich mich dafür bedanken, dass er die Rolle als Erstgutachter meiner Arbeit übernommen hat und mir als TAC Mitglied zur Seite stand. Des Weiteren danke ich Prof. Dr. **Martin Müller**, der mich ebenso als TAC Mitglied mit Ratschlägen unterstützte, und Dr. **Marco Binder** für ihre Teilnehmer als Prüfer an meiner Disputation.

Ein besonderer Dank gilt den Praktikantinnen **Eva Christine Schitter**, **Laura Castelletti**, **Julia Madeleine Heinze** und **Janika Schmitt**, sowie den Master-Studentinnen **Sara Becker** und **Catharina Lotsch** für ihren Beitrag zu binding assays, PBMC Isolationen und ELISpot assays.

Allen aktuellen und ehemaligen Mitgliedern der Arbeitsgruppe gilt meine größte Wertschätzung. Ich möchte **Stephanie Hoppe** und **Renata Blatnik** danken, für ihre Einleitung und Beiträge zu diesem Projekt. Bei **Alina Steinbach** und **Alexandra Klevenz** möchte ich mich für ihre Ratschläge und praktische Hilfe im Labor bedanken. **Sebastian Kruse**, **Nitya Mohan**, **Samantha Zotnick** und **Jonas Förster** danke ich vielmals, nicht nur für hilfreiche fachliche Diskussionen und Meinungen, sondern auch für den wertvollen gemeinsamen Spaß im Büro, auf dem Weg zum Uni-Shop, beim Mittagessen, bei Parties, auf der Wasn und dem Mathaise-Markt. Auch **Mogjib Salek**, **Rebecca Köhler**, **Sophia Föhr**, **Lena Postawa** und **Mine Özcan** gebührt an dieser Stelle mein Dank. Ein besonders großes Dankeschön gilt **Monika Bock**, nicht nur für ihre administrative Hilfe, sondern auch für ihr stets offenes Ohr und ihre verständnisvolle Art.

Ebenso bin ich vielen weiteren Kollegen am DKFZ dankbar. Ich danke den **Core Facilities für Durchflusszytometrie** und **Peptidsynthese** (insbesondere **Mario Koch**). Den **Arbeitsgruppen um Timo Bund**, **Felix Hoppe-Seyler** und **Mathias Heikenwälder** danke ich allgemein für tolle Nachbarschaft. Im Besonderen gilt AG Bund großer Dank, dass ich tagelang eure Zellkultur benutzen durfte, als es Probleme bei dem Austausch unserer Sterilbänke gab. Ich möchte **Diana Tichy** Danke sagen, für ihre fachliche Meinung im Bereich Statistik. Ich danke **Marleen Büchler** und **Vanessa Dieterle** für gemeinsame Kaffeepausen und **Kai Bates** für zahlreiche Aufmunterungen.

Ich danke den externen **Kollaborationspartnern aus den Laboren von Thomas Meyer** am MPI-IB in Berlin, **Jeff Hammerbacher** an der Icahn School of Medicine at Mount Sinai und **Rachel Karchin** an der Johns Hopkins Universität. Im Besonderen gilt mein Dank hier **Tim O'Donnell**, für hilfreiche Diskussionen und gemeinsames Abfahrts-Skifahren im Rahmen der CIMT Winterschool.

Außerhalb von Laboren danke ich unbedingt meinen Freunden und meiner Familie, die mein Leben bedeuten. An erster Stelle geht ein gigantisches Dankeschön an **Thomas**, dafür, dass er mich in allen guten und schlechten Augenblicken unterstützt, das Beste in mir sieht und hervorholt. Ich freue mich auf unsere Reisen durch die Welt. Wo du bist, ist zu Hause. Auch bei meinen Mitbewohnern **Allegra** und **Kai** möchte ich mich dafür bedanken, dass wir neben der Anschrift, und Miete für Küche, Bad und Balkon auch an vielen Abenden Mahlzeiten und, bei so manchen Serien und Filmen, die Couch teilen. Ihr seid meine Heidelberg-Familie. Vielen Dank auch an die (teilweise ehemaligen) Mitbewohner des Hauses: **Melanie, Carolin, Lea** und **Marcel**, die während gemeinsamer Koch- und Grillabende, Partys und vor allem Festivals zu guten Freunden geworden sind. Ich danke ebenso den Volleyballern der „**Bunsen-Brenner**“ (**Timo, Tillmann, Nina, Robert, Sandra, Steffi, Mario, Bernhard, Alex, Heike, Heiko, Tobi, Jan** und **Anna**) für Montagabende in der flachsten Halle, die ich kenne, und einige zusätzliche Abende im Sand des Heidelberger USZ oder des Neckarstrands. Für jede Menge spaßige Unterhaltungen und tolle Lehrstunden danke ich **Melina** und **Milan** von der Nuzinger Tanzschule.

Abschließend, dafür keineswegs weniger, danke ich meiner Familie. Zu dieser zähle ich auch **Sylvia** und **Michael Brauer**, die mich besonders während der Schreibphase motivierten. Ich danke **Stefan, Alma** und **Stefanie**, meiner herzallerliebsten Schwester. Ich freue mich über jeden Tag, den ich mit euch verbringen kann. Danke Steffi, dass du immer für mich da bist und für all unsere gemeinsamen Schwestern-Wochenenden und -Urlaube. Auch meinen Eltern, **Karin** und **Harald Bonsack** möchte ich danken. Vielen lieben Dank für eure Unterstützung während meines gesamten Studiums, für euer Zuhören und eure Ratschläge in allen Lebenslagen.

IV. Table of Contents

I.	Abstract	I
II.	Zusammenfassung	II
III.	Acknowledgements	III
IV.	Table of Contents	V
V.	Abbreviations	VIII
VI.	List of Figures	XI
VII.	List of Tables.....	XII
1	Introduction	1
1.1	Cancer and human papillomavirus infection.....	1
1.1.1	The organization of the HPV genome.....	4
1.1.2	HPV life cycle and protein expression.....	5
1.1.2.1	The early proteins E1, E2, E4, E5 and the oncoproteins E6 and E7	7
1.1.2.2	The late proteins L1 and L2	9
1.1.3	Prophylactic vaccination	9
1.1.4	Classification of HPV	10
1.2	The immune system and HPV.....	12
1.2.1	Innate and adaptive immunity	13
1.2.2	Antigen presentation by major histocompatibility complex molecules	14
1.2.3	Immune responses of T lymphocytes.....	16
1.2.4	Immune evasion mechanisms of HPV	19
1.2.4.1	Immunoediting of tumor cells.....	19
1.2.4.2	Immune evasion mechanisms of HPV and HPV-positive tumor cells	19
1.3	Cancer Immunotherapy.....	20
1.3.1	An overview of cancer immunotherapy approaches	20
1.3.2	<i>In silico</i> epitope prediction and its role in cancer immunotherapy	23
1.3.3	HPV-specific immunotherapies	26
2	Aims of this study.....	31
3	Materials and Methods	33
3.1	Materials.....	33
3.1.1	Laboratory equipment	33
3.1.2	Consumables	35
3.1.3	Chemicals and biological reagents.....	35
3.1.4	Buffers and solutions	37
3.1.5	Cell lines	37
3.1.6	Blood samples and buffy coats	38
3.1.7	Cell culture basal media and supplements	38
3.1.8	Cell culture media	39
3.1.9	Kits.....	40
3.1.10	Molecular Markers.....	40

Table of Contents

3.1.11 Oligonucleotides	40
3.1.12 Antibodies	40
3.1.12.1 Monoclonal antibodies for HLA typing.....	40
3.1.12.2 Monoclonal antibodies for ELISpot assays	41
3.1.12.3 Monoclonal antibodies for analysis of cytokine production in immune cells.....	41
3.1.13 Peptides	41
3.1.13.1 Control peptides.....	41
3.1.13.2 HPV16 E6 and E7 peptides.....	42
3.1.14 Software	42
3.2 Methods.....	43
3.2.1 <i>In silico</i> methods	43
3.2.1.1 MHC class I binding and T cell epitope prediction	43
3.2.1.2 Web Application MHCcombine	45
3.2.1.3 Sequence motif analysis.....	46
3.2.2 Molecular biological methods.....	46
3.2.2.1 Isolation of genomic DNA.....	46
3.2.2.2 Determination of DNA concentration.....	46
3.2.2.3 Polymerase chain reaction	47
3.2.2.4 PCR for sequencing of HPV16 E6 and E7	47
3.2.2.5 Agarose gel electrophoresis	48
3.2.2.6 Purification of DNA from agarose gels	49
3.2.2.7 Purification of DNA from PCR products.....	49
3.2.2.8 Sequencing of HPV16 E6 and E7 genes.....	50
3.2.2.9 Analysis of HPV16 E6 and E7 sequences	50
3.2.3 Cell culture methods	50
3.2.3.1 Thawing and freezing of cells.....	50
3.2.3.2 Culturing and passaging of cells	51
3.2.3.3 Counting of cells	52
3.2.4 Cellular assays.....	52
3.2.4.1 Flow cytometry	52
3.2.4.2 Competition-based peptide-HLA binding assays	53
3.2.4.3 Peripheral blood mononuclear cell (PBMC) isolation.....	54
3.2.4.4 HLA-type screening of PBMCs using flow cytometry.....	55
3.2.4.5 Generation of epitope-specific T cell lines	56
3.2.4.6 Generation of short-term epitope-specific T cells lines for cultured IFN γ -ELISpot assays	56
3.2.4.7 Generation of long-term epitope-specific T cell lines for Vital-FR cytotoxicity assays	57
3.2.4.8 IFN γ -ELISpot assay.....	58
3.2.4.9 Magnetic-activated cell sorting (MACS).....	59
3.2.4.10 Vital-FR cytotoxicity assay.....	60
3.2.4.11 Assessment of intracellular cytokine production of PBMCs	61

3.2.5	Statistical analysis	62
3.2.5.1	Binding affinity calculation based on IC ₅₀	62
3.2.5.2	Performance evaluation of prediction algorithms.....	63
3.2.5.3	Calculation and validation of criteria-based decision thresholds.....	63
3.2.5.4	Comparison of criteria-based and bootstrapping-validated thresholds.....	64
3.2.5.5	ELISpot analysis	64
3.2.5.6	Vital-FR analysis	65
4	Results	67
4.1	<i>In silico</i> predicted and <i>in vitro</i> validated HLA binding affinity of HPV16 E6 and E7 peptides. 67	67
4.1.1	MHC class I binding prediction methods predict numerous potential HLA-binding HPV16 E6- and E7-derived peptides.	67
4.1.2	Competition-based cellular binding assays identify HPV16 E6- and E7-derived ligands to investigated HLA types.....	71
4.1.3	MHC class I binding predictions do not match experimental binding results.	75
4.2	Performance evaluation of MHC class I binding prediction methods based on experimentally validated HPV16 E6- and E7-derived ligands.....	76
4.2.1	Prediction algorithm results discriminating binders from nonbinders vary depending on HLA type and peptide length.....	76
4.2.2	Commonly used decision thresholds result in low prediction sensitivity.	80
4.2.3	Novel individual decision thresholds increase prediction sensitivity.	82
4.2.4	Criteria-based thresholds and bootstrapping-validated thresholds result in similar prediction performance.....	85
4.2.5	Applying the recommended thresholds increases the number of predicted true binders.	86
4.3	HPV16 variants and their HLA-binding E6 and E7 peptides	88
4.3.1	HPV16 positive cell lines are infected with different genomic HPV variants.....	88
4.3.2	Genomic variants of HPV16 result in different HLA ligands.....	89
4.4	Evaluation of peptide immunogenicity using T cell lines.....	92
4.4.1	Flow cytometry characterizes HLA-type of PBMCs isolated from buffy coats.	92
4.4.2	IFN γ -ELISpot assays identify immunogenic HPV16 E6 and E7 peptides.	95
4.4.3	Epitope-specific T cell lines can induce specific lysis of HPV16 ⁺ target cells.....	99
5	Discussion	101
6	References	109
7	Annex	123

V. Abbreviations

Abbreviation	Meaning
°C	degree Celsius (unit)
A	absorbance (unit)
aa	amino acid
AA	Asian American (cluster of human papillomavirus sublineages)
ACK	ammonium-chloride-potassium
ACT	adoptive cell transfer
Af	African (cluster of human papillomavirus sublineages)
ANN	artificial neural network
APCs	antigen presenting cells
APM	antigen processing machinery
B-LCL	B-lymphoblastoid cell lines
bp	base pairs (unit genome size)
CCR	chemokine receptor
CEF	cytomegalovirus, Epstein-Barr virus, influenza virus
CFSE	carboxyfluorescein succinimidyl ester
ConA	Concanavalin A
csv	comma separated values (file format)
CTLs	cytotoxic T lymphocytes
Da	dalton (unit atomic mass)
DAMPs	damage-associated molecular patterns
DAPI	4',6-diamidino-2-phenylindole
DCs	dendritic cells
DKFZ	Deutsches Krebsforschungszentrum, German Cancer Research Center
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleoside triphosphate
E	European (cluster of human papillomavirus sublineages)
E-A	European-Asian (cluster of human papillomavirus sublineages)
EBV	Epstein-Barr virus
e.g.	<i>exempli gratia</i> , “for example”
EGF	epidermal growth factor
ELISpot	enzyme-linked immunospot
ER	endoplasmatic reticulum
ERAP	ER-associated aminopeptidase
FACS	fluorescence activated cell scanning/sorting
Fc	fragment crystallizable
FITC	fluorescein isothiocyanate
FR	far-red
FSC	forward scatter

Abbreviations

Abbreviation	Meaning
G	gram (SI unit mass)
GM-CSF	granulocyte-macrophage colony-stimulating factor
h	hour (unit time)
HLA	human leukocyte antigen
HPLC	high performance liquid chromatography
HPV	human papillomavirus
IC50	half maximal inhibitory concentration
i.e.	id est, “that is”, “namely”
IEDB	immune epitope data base
IFN	interferon
IL	interleukin
kb	kilo base (unit genome size, 1000bp)
l	liter (unit volume)
LC	liquid chromatography
LPS	lipopolysaccharide
m	meter (SI unit length)
M	molarity, also mol/l (unit amount of substance)
MACS	magnetic-activated cell sorting
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
min	minute (unit time)
mol	mole (SI unit amount of substance)
MS	mass spectrometry
NA	North-American (cluster of human papillomavirus sublineages)
NKs	natural killer cells
ORF	open reading frame
PAMPs	pathogen-associated molecular patterns
PBMCs	peripheral blood mononuclear cells
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PGE2	prostaglandin E2
PMA	phorbol myristate acetate
PRR	pattern recognition receptor
rhIL	recombinant human interleukin
RNA	ribonucleic acid
RNase	ribonuclease
RPMI	Roswell Park Memorial Institute
RT	room temperature
s	second (SI unit time)
SD	standard deviation
SFU	spot forming units

Abbreviations

Abbreviation	Meaning
SLP	synthetic long peptide
SM	scoring matrix
SMM	stabilized matrix method
SMMPMBEC	stabilized matrix method with a peptide:MHC binding energy covariance matrix
SSC	sideward scatter
SSP	synthetic short peptide
TAE	Tris base, acetic acid and EDTA (buffer)
TAP	transporter associated with antigen processing
Taq	Thermus aquaticus (bacterium)
TCR	T cell receptor
Th	T helper cell
TIL	tumor infiltrating lymphocyte
TLR	Toll-like receptor
TNF	tumor necrosis factor
TPBS	Tween PBS (0.05% [v/v] Tween20)
Treg	regulatory T cell
URR	upstream regulatory region
UV	ultraviolet
vs.	versus
x g	multiple of standard gravity of 9.8 m/s ² (g-force)

VI. List of Figures

Figure 1. Cancer cases attributable to pathogens in 2012.	3
Figure 2. HPV infection stages and population prevalence at different ages.	4
Figure 3. HPV and the organization of its genome.	5
Figure 4. Productive and non-productive HPV infection.	7
Figure 5. Simplified synergistic mechanisms of the E6 and E7 oncoproteins in driving transformation.	8
Figure 6. Phylogenetic tree representing 118 papillomavirus sequences.	11
Figure 7. Schematic taxonomy of <i>Papillomaviridae</i> focusing on HPV16.	12
Figure 8. Antigen-presentation pathways of MHC class I and MHC class II	15
Figure 9. T cell priming by DCs and induction of CD8 ⁺ T cell responses.	17
Figure 10. Key steps of MHC binding prediction methods	25
Figure 11. Schedule for generating short-term peptide-specific T cell lines.	57
Figure 12. Schedule for generating long-term epitope-specific T cell lines.	58
Figure 13. Competition-based binding assays identify HLA-ligands.	72
Figure 14. Evaluation of HLA class I binding predictions based on experimental binding results.	76
Figure 15. Receiver operating characteristics (ROC) curves of predictors for different HLA types and pooled and individual peptide lengths.	78
Figure 16. Area under the ROC curves (A_{ROC}) of predictors for different HLA types and pooled and individual peptide lengths.	79
Figure 17. Threshold-dependent sensitivity and specificity of predictors for different HLA types.	81
Figure 18. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended “validated” or	84
Figure 19. Comparison of predictor performance measures between applying criteria-based thresholds and validated thresholds	86
Figure 20. Classification of HLA binding prediction of HPV16 E6/E7 peptides to HLA A2 and A24 according to application of different thresholds.	87
Figure 21. Influences of amino acid changes in HPV16 E6/E7 variants on HLA binding affinity.	91
Figure 22. Analysis of HLA-specific antibodies for flow cytometry.	93
Figure 23. HLA-typing of healthy donors by flow cytometry.	94
Figure 24. HLA-specific immunogenic HPV16 E6-/E7-derived peptides identified by IFN γ -ELISpot assays.	97
Figure 25. HLA-ligands and immunogenicity detected in donors.	98
Figure 26. Cytotoxicity mediated by CD8 ⁺ T cells specific for HPV16 E6-/E7-derived A2-restricted epitopes.	100

VII. List of Tables

Table 1. Prediction methods used in this study.	44
Table 2. Reference amino acid sequences of the HPV16 E6 and E7 proteins.	44
Table 3. Amino acid changes in HPV16 E6 and E7 protein variants.	45
Table 4. Primer pairs used for HPV16 E6- and E7-amplifying PCR.	48
Table 5. Components of the PCR reaction mix amplifying of HPV16 E6 and E7.	48
Table 6. PCR program used for amplifying of HPV16 E6 and E7.	48
Table 7. HLA-alleles expressed by B-LCLs and required pH of elution buffer.	54
Table 8. Centrifugation conditions for fresh whole blood and buffy coat.	55
Table 9. Antibody dilutions used in HLA type screenings of PBMCs isolated from buffy coats.	55
Table 10. Antibody panel for immunophenotyping of PBMCs.	62
Table 12. List of experimentally validated HPV16 E6-/E7-derived nonbinders.	73
Table 13. List of amino acid changes prevalent HPV16 E6- and E7-variants and their potential relevance for HLA binding.	89
Table 14. Results for flow cytometry based HLA typing of healthy blood donors.	94

List of Tables

List of Tables

1 Introduction

1.1 Cancer and human papillomavirus infection

The term “cancer” describes a group of diseases characterized by uncontrolled growth of abnormal cells beyond their original boundaries and by invasion or spreading (metastasis) into different tissues and organs. Cancer encompasses malignant tumors, also called neoplasms, which can develop from premalignant dysplasia. It was the cause of 9.6 million deaths in 2018 and thus is globally the 2nd leading cause of death (WHO, 2019). As cancer cells represent altered cells of the own body, for therapeutic approaches it is crucial to distinguish cancerous from healthy cells.

Cancer cells acquired specific capabilities during their malignant transformation, which were described 2000 by Hanahan and Weinberg as the hallmarks of cancer (Hanahan and Weinberg, 2000). These are - next to tissue invasion and metastasis, as mentioned above - evasion of apoptosis, self-sufficiency in growth signals, insensitivity to anti-growth signals, limitless replicative potential and sustained angiogenesis. In 2011, two emerging hallmarks, deregulation of cellular energetics and avoiding immune destruction, and two enabling characteristics, genome instability and mutation and tumor-promoting inflammation, were added (Hanahan and Weinberg, 2011). Cancer causing factors are multifaceted and comprise inherited genetic predispositions and environmental factors such as diet, alcohol drinking, tobacco smoking, exposition to ultraviolet (UV)-light and infection with pathogens (Wu et al., 2016).

Pathogens are relevant cancer causing agents (Figure 1 A) and comprise for example the bacterium *Helicobacter pylori* in ~75-89% of non-cardia gastric cancer cases, hepatitis B and C viruses (HBV and HCV) accounting for ~76% of liver cancer cases, Epstein-Barr virus (EBV) related to lymphoid malignancies and nasopharyngeal tumors, and high-risk types of human papillomavirus (HPV) (Plummer et al., 2015; Maucort-Boulch et al., 2018; Maeda et al., 2009). Infections with HPVs are frequent in all world populations (Figure 1 B) and attributable to all cervical, 88% of anal, 50% of penile and ~31% of oropharyngeal cancer cases (de Martel et al., 2017). In 2012, 630,000 cancer cases were attributed to HPV infection (Figure 1 C), which represents 8.6% of all cancer cases in women and 0.8% in men (de Martel et al., 2017). Of all cancer cases attributable to pathogens, HPV is the cause of 5.6% in men and more than half (53.6%) in women (Figure 1 E and D). HPVs are associated with virtually all cervical carcinomas, which explains the pronounced difference in numbers between males and females (Plummer et al., 2016). In terms of incidence (~570,000) and mortality (~311,000) cervical cancer was the fourth most common cancer entity in women worldwide in 2018 (Figure 1 F) (Ferlay et al., 2019). In 2012, it was the second most common cancer and the third most frequent cause of cancer death in less developed regions (Ferlay et al., 2015). The prevalence of all HPV-attributable cancers is highest in these regions, but HPV represents a worldwide health problem (Figure 1 G).

Papillomaviruses are highly species-specific and infect epithelium of not only humans but also other mammals, birds and reptiles (Bravo et al., 2010). The connection between human PV infection and

cervical cancer was postulated and proven by groundbreaking research of Harald zur Hausen and his team (Gissmann et al., 1977). Since they isolated and sequenced the first HPV DNA of type 16 (HPV16) from cervical carcinoma samples in 1983, more than 220 HPV types were identified (Eklund and Dillner, 2019; Dürst et al., 1983). These were divided into low- and high-risk types according to their carcinogenic potential. Low-risk types mainly cause benign warts of skin and anogenital regions, whereas high-risk types can induce malignant transformation into anogenital (cervical, vaginal, vulvar, anal and penile) and head and neck (oropharyngeal, oral cavity, laryngeal) cancers and their precursor lesions (Chow et al., 2010). Other sites of infections are the conjunctiva of the eyes, ear canals and nasal sinuses.

So far, 12 high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) and 12 types which are probably carcinogenic (26, 30, 34, 53, 66, 67, 68, 69, 70, 73, 82 and 97) were defined (ICO/IARC HPV Information Centre, 2019). The majority of HPV-associated cancer cases are caused by HPV16 and HPV18 (de Martel et al., 2017). Together they contribute to 70.8% of cervical, 84.9% of head and neck and even 87% of anal cancers. HPV16 is the most prevalent genotype in all world regions accounting for 60.5% cervical, 75.2% oropharyngeal, and 71.4% anal cancer cases and the majority of precancerous lesions (Serrano et al., 2018; Castellsagué et al., 2016).

In the affected tissues, the virus infects the epithelia of cutaneous and mucosal surfaces. Infections at the transformation zones between squamous and columnar epithelium are at especially high risk for malignant transformation (Egawa et al., 2015). Epithelial cells get infected during sexual intercourse as primary route of transmission (ICO/IARC HPV Information Centre, 2019). This explains a peak in HPV incidence in people at the age of first sexual activity (Figure 2 A). Infections are typically asymptomatic and unnoticed HPV spreads rapidly in sexually active males and females. Fewer new sexual contacts with increasing age result in HPV incidence drops. Sexual behavior is also expected to influence the highly variable HPV prevalence by geographical region (Schiffman et al., 2016).

Persisting HPV infection with a high-risk type is referred to as intraepithelial neoplasia grade 1, named after the infection site: CIN1 (cervical), VIN1 (vulval), VAIN1 (vaginal) or AIN1 (anal). In >90% of the infected individuals, HPV infection is cleared within less than 2 years. If not cleared by the immune system these lesions might progress into precancerous moderate (e.g. CIN2) or severe (CIN3) dysplasia and eventually into cancer (Figure 2 B) (Schiffman et al., 2016). The slow transformation process leaves time for intervention.

Introduction

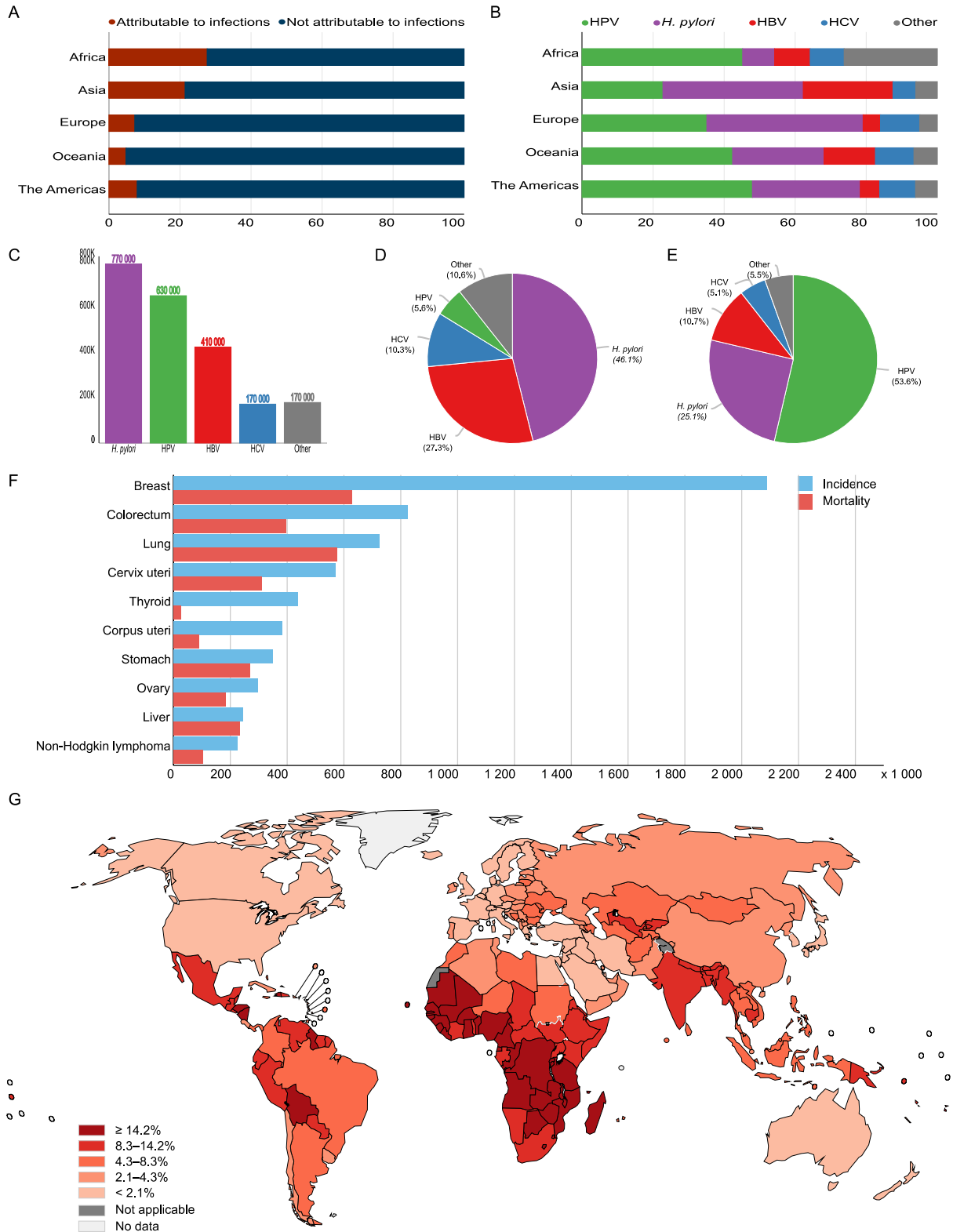


Figure 1. Cancer cases attributable to pathogens in 2012. (A) Bar graphs show the number of cancer cases per population among both sexes attributable to pathogens. (B) Contribution of major agents to pathogen-caused cancers. (C) Numbers of worldwide cancer cases per infectious agent for both sexes, (D) and (E) Frequencies of pathogen-related cancers in 1.1 million total cases for males and females, respectively. (F) Estimated worldwide incidence and death numbers (in thousands) in 2018 (GLOBCAN2018) of ten major cancer entities in women of all ages. (G) The proportion of cancer cases attributable to human papillomavirus infection among both sexes shown by country. Graphs were produced using tools provided by the Global Cancer Observatory and the International Agency for Research on Cancer (Plummer et al., 2016; Ferlay et al., 2019).

Preventive measures against HPV infection and associated diseases comprise prophylactic vaccination and cancer screening programs. Three prophylactic vaccines have been developed and represent the most effective long-term intervention for controlling high-risk infections (see section “Prophylactic vaccination”). National vaccination programs were introduced by 65 mostly high- and middle-income countries (by 2016) and typically target adolescent girls and, increasingly, boys. As most countries did not yet start vaccination programs, secondary prevention still plays a major role. Screening aims at detecting pre- and early stage cancers to minimize overtreatment. Methods comprise the screen, e.g. by Papanicolaou (Pap) testing or HPV DNA detection, triage, colposcopy and biopsy for histological examination (Schiffman et al., 2016).

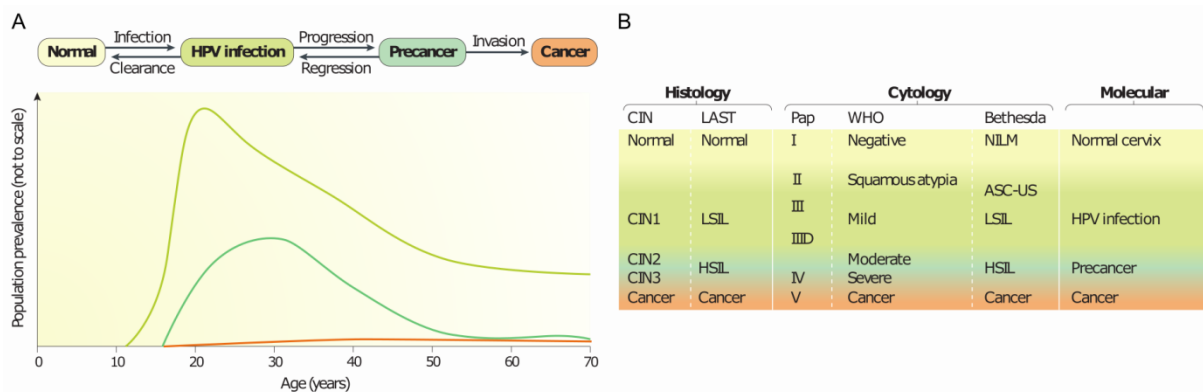


Figure 2. HPV infection stages and population prevalence at different ages. (A) Without clearance, persisting HPV infection can progress to precancerous lesions and, upon invasion, to a final cancer state. HPV prevalence peaks with beginning sexual activity in youth. Several years later the prevalence of precancer peaks. A plateau for invasive cancer is seen many years later starting from the age of 40. (B) The progressing transformation by HPV infection can be divided into different grades of severity according to cytological abnormalities found by histology and cytology. Figure adapted from (Schiffman et al., 2016).

Current treatment options for severe dysplasia and cancers involve surgery, radiotherapy and chemotherapy (mainly cisplatin-based), depending on the stage of the disease. As gold standard, cervical precancerous lesions are treated by invasive excisional procedures with a risk for subsequent obstetrical complications. Less invasive ablative technologies exist but do not provide tissue for subsequent histopathology. Especially surgery can damage surrounding tissue and carry the risk for losing fertility as a consequence of radical hysterectomy. Treatments of invasive cancers show the expected post-treatment morbidity (Schiffman et al., 2016). This highlights the need of new non-invasive interventions, which could be achieved by immunotherapy approaches such as therapeutic HPV vaccination (see section “HPV-specific immunotherapies”).

1.1.1 The organization of the HPV genome

HPVs are non-enveloped double-stranded DNA (dsDNA) viruses with a circular genome of a size between 7kb and 8kb with eight or nine open reading frames (ORFs) (McBride and Warburton, 2017; Doorbar et al., 2015). In order to fit the genome into the relatively small size of about 8kb, genes partially overlap, resulting in polycistronic transcripts that are separated by splicing of mRNA.

In case of HPV16, 7908bp contain six early genes (E1, E2, E4, E5, E6 and E7) and two late genes (L1 and L2) differentiated by their time of expression during the viral life cycle (see section "HPV life cycle and protein expression"). Figure 3 A shows the genomic organization of the early and late genes. A noncoding upstream regulatory region (URR) contains promoter and enhancer elements as well as the viral origin of replication (Stanley, 2012). A large region contains early genes which have functions at the level of viral replication and transcription and contain the transforming proteins E6 and E7 (see section "The early proteins E1, E2, E4, E5 and the oncoproteins E6 and E7"). The late gene region encodes for the major (L1) and minor (L2) capsid proteins (see section "The late proteins L1 and L2"). Together, these two proteins form 72 capsomers that build the icosahedral capsid structure (Figure 3 B), which is able to self-assemble and measures 55nm in diameter (Doorbar et al., 2015; Hagensee et al., 1993).

HPV gene products can further be divided into highly conserved core proteins (E1, E2, L1 and L2) that are directly involved in viral genome replication and into accessory proteins (E4, E5, E6 and E7) which show greater functional variability across HPV types (Schiffman et al., 2016).

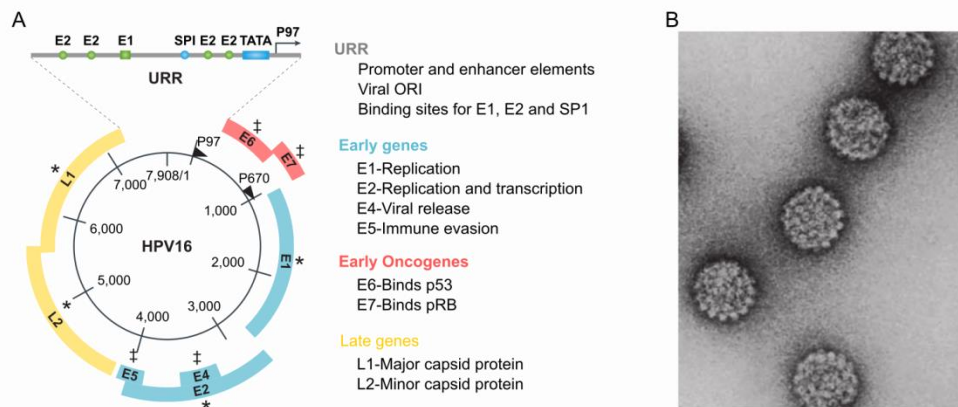


Figure 3. HPV and the organization of its genome. (A) The genomic organization of HPVs is representatively shown for HPV16. Arrowheads mark the early (P97) and late (P670) promoters. The genome is organized in three main regions. The non-coding upstream regulatory region (URR) contains promoter and enhancer elements, the origin (ORI) of viral replication and binding sites for the E1 and E2 proteins and SP1 transcription factor. The early gene region contains open reading frames (ORFs) for E1, E2, E4 and E7 (blue) and the transforming oncogenes E6 and E7 (red). The late gene region is comprised of the ORFs for L1 and L2, the major and minor capsid protein, respectively. Core viral proteins (*) are required for genome replication, viral assembly and release whereas accessory proteins (‡) provide functions of cell cycle entry and immune evasion. (B) Negatively stained transmission electron micrograph shows HPV particles. Figure and legend were adapted from (Schiffman et al., 2016).

1.1.2 HPV life cycle and protein expression

Infection with HPV occurs if virus particles access the epithelial basal layer through wounds or epithelial trauma. In order to become established, they infect basal keratinocytes with a stem cell-like phenotype (Egawa et al., 2015). As the virus is dependent on the host's replication machinery, it requires infecting an actively dividing cell (as in wound healing). In epithelia only basal cells are actively dividing and thus infected by HPV (Hoffmann et al., 2006; Pyeon et al., 2009). Internalization, in contrast to other viruses, takes several hours and is initiated by L1 binding to the basement membrane (Horvath et al., 2010). This triggers conformational changes of both capsid

proteins as a prerequisite for cellular uptake. Endocytosis of HPV16 happens via a clathrin-dependent pathway (Day et al., 2003). L2, on the other hand, is complexed to the viral DNA and is responsible for intracellular transport and nuclear accumulation (Campos, 2017). A recent study showed that virions remain largely intact during nuclear entry which is in contrast to a previous concept of disassembling capsids (Day et al., 2019).

An initial phase of viral genome replication is separate from the cell cycle but dependent on the cellular DNA synthesis machinery supported by the viral replication proteins E1 and E2 (Chiang et al., 1992). The viral copy number is amplified to 50-100 copies per cell (Egawa et al., 2012). Upon host cell division, the resulting episomes are distributed to daughter cells. The low episomal copy numbers are maintained as dividing cells transit the parabasal layers of the epithelium until they eventually enter suprabasal layers and undergo differentiation. Viral gene expression and DNA replication then is upregulated and increases the viral copy numbers to thousands per cell. This is achieved by expression of E5 and, primarily, the E6 and E7 proteins, which play an essential role in viral genome amplification, mainly by driving S-phase re-entry (Egawa et al., 2015). Pathological characteristics of E6 and E7, in addition to increased cell division are a lower sensitivity of the infected cells to cellular contact inhibition and inhibition of the normal cellular differentiation program (Schiffman et al., 2016). In high-risk types, persistent overexpression of these two oncogenes severely affects the integrity of the host cell by various mechanisms (see section “The early proteins E1, E2, E4, E5 and the oncoproteins E6 and E7”).

In the upper suprabasal layers of the epithelium, E4 and the late genes L1 and L2 are expressed (Doorbar et al., 2015). The E2 protein recruits the minor capsid protein L2 to regions of replication where amplified viral genomes get packed when L2 and the major capsid protein L1 self-assemble and form virions (Day et al., 1998). The virus finally matures in the superficial terminally differentiated keratinocytes, which undergo natural shedding. It is thought that E4 contributes to virion release, but it definitely plays a role in disrupting keratin structure and compromising the normal assembly of the cornified envelope by forming amyloid fibrils (Doorbar et al., 1991; Brown et al., 2006; McIntosh et al., 2008). Released virions can now infect further basal keratinocytes in compromised epithelium or can be transmitted to another host organism.

The whole viral lifecycle from infection to virus release takes a comparatively long time, ranging from 3 weeks to months (Stanley, 2012). In this time, the virus remains exclusively intraepithelial as it is dependent on and tailored to the complete keratinocyte differentiation. In high-risk types, productive infection can turn into abortive infection (Figure 4) if ordered expression of gene products is prevented by cell cycle deregulation mediated by E6 and E7. This can lead to the integration of the viral genome into the host cell's DNA. The thus induced instability of the host genome characterizes progressing dysplasia and, if persisting for several years, accumulation of genetic changes eventually leads to invasive cancer (Doorbar et al., 2015).

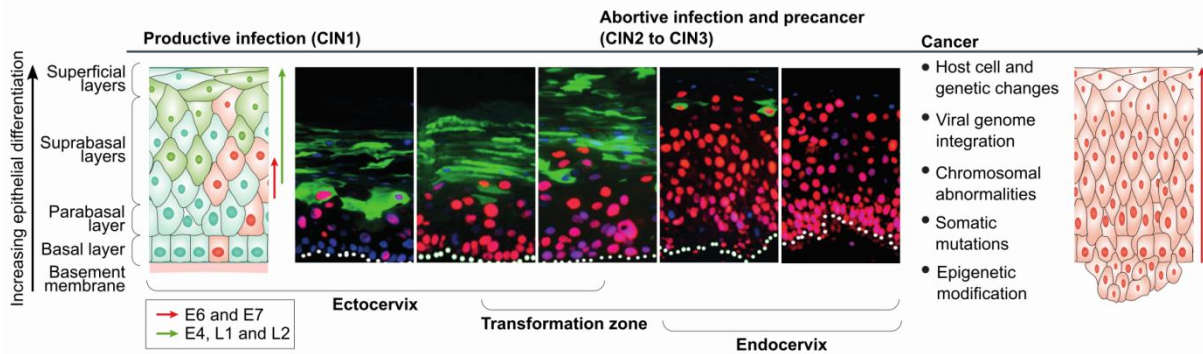


Figure 4. Productive and non-productive HPV infection. In virion-producing productive infection, E6 and E7 are expressed at low levels and do not stimulate excessive cell cycle entry in parabasal layers (absence of red cell cycle marker in first immunofluorescence image). E4 and late L1 and L2 gene products can be extensive (green on fluorescence images). Such lesions are labeled cervical intraepithelial neoplasia grade 1 (CIN1). Higher grade lesions (CIN2 to CIN3) are regarded to be abortive (non-productive) infections, where E6 and E7 activity is increased and late gene expression restricted. Progression to cancer can take several years, as the HPV genome integrates into the host genome and genetic changes accumulate in the host cell promoted by deregulated E6 and E7 expression. More severe disease grades are located in the transformation zone and endocervix rather than in the ectocervix. Figure and legend adapted from (Schiffman et al., 2016).

1.1.2.1 The early proteins E1, E2, E4, E5 and the oncoproteins E6 and E7

As outlined above, the expression of the early proteins is spatiotemporally regulated, which is reflected by their specific roles in the viral life cycle. After infection of a basal keratinocyte, the DNA helicase E1 and the transcription factor E2 are regulating early transcription of the viral episome. E2 initiates viral replication by binding to the non-coding URR, thereby recruiting E1 to the viral origin of replication (Dell et al., 2003; Abbate et al., 2004). E2 regulates E6 and E7 expression, as low E2 levels enhance and high E2 levels repress and transcription (Doorbar et al., 2015). Overexpression of E2 induces expression of the late genes in the course of keratinocyte differentiation (Johansson et al., 2012). The E2 protein furthermore anchors the viral episomes to cellular chromosomes and also plays a role in virus assembly as it recruits the L2 protein to replication sites (You, 2010; Day et al., 1998). Viral DNA integration into the host genome often disrupts the E1 and E2 genes which alleviates transcriptional repression of the E6 and E7 oncogenes (McBride and Warburton, 2017).

The early protein E4 is expressed as a splicing product in the upper epithelial layers where keratinocytes terminally differentiate (Doorbar, 2013). Its functions reflect adaptation of the HPVs to the keratinocytes, as it is known to disrupt the keratin cytoskeleton and to modify the cornified cell envelope, which is believed to support release of virions (Doorbar et al., 1991; Brown et al., 2006).

Viral replication is supported by the E5 transmembrane protein. It is mostly found in the endoplasmic reticulum (ER) where it interferes with endosomal trafficking, which leads to recycling of epidermal growth factor (EGF) receptors to the cell surface (Suprynowicz et al., 2010). This way, E5 enhances the activity of EGF receptors, promoting cellular replication.

Due to their potential in driving malignant transformation of host cells, the proteins E6 and E7 were designated as oncoproteins. In high-risk types these proteins are involved in immortalization and transformation of the host cell, whereas in low-risk types they simply stimulate replication (zur

Hausen, 1996; Egawa et al., 2015). To keep the host cell dividing, E6 and E7 reactivate cellular DNA synthesis, inhibit apoptosis and delay differentiation. Overexpression can lead to uncontrolled cell growth and the development of dysplasia and invasive cancer (Figure 5 A).

In HPV16, the E7 protein has a size of 98 amino acids (aa) and a molecular weight of 11kDa. Its multiple functions on the host cell cycle were thoroughly reviewed (Roman and Munger, 2013; Vande Pol and Klingelutz, 2013; Moody and Laimins, 2010). It binds to the retinoblastoma protein (pRB) and thereby in turn inhibits pRB binding to the transcription factor E2F (Figure 5 B). The free E2F induces activation of S-phase promoting genes such as cyclin A and E. Further, E2F mediates transcription of p16^{INK4A}, a surrogate marker for HPV infection, and p14^{Arf}, which increase the levels of cyclin kinases and p53, respectively (Klaes et al., 2001). The pRB-dependent actions are further promoted by E7 supporting the proteasome-mediated degradation of pRB (Schiffman et al., 2016). Additionally, E7 interferes with epigenetic pathways, cell cycle checkpoints and centrosome synthesis, promoting instability of the host genome (Moody and Laimins, 2010).

The E6 protein of HPV16 consists of 158aa with a molecular weight of 19kDa. E6 complements the function of E7 as it interacts with p53, levels of which are increased in response to E7 expression (Figure 5 C) (Vande Pol and Klingelutz, 2013) . Specifically, it binds to the ubiquitin-protein ligase E3A, also called E6 associated protein, thereby directing p53 poly-ubiquitination, which marks p53 for proteasomal degradation (Werness et al., 1990). The reduced presence of the p53 tumor suppressor prevents apoptosis of the host cell in response to E7-mediated cell cycle activation (Schiffman et al., 2016). Further, E6 upregulates expression of the reverse transcriptase component of telomerase (TERT) and thus activates telomerase (Klingelutz et al., 1996; Veldman et al., 2001). This decreases telomere shortening and prevents the cells from becoming senescent.

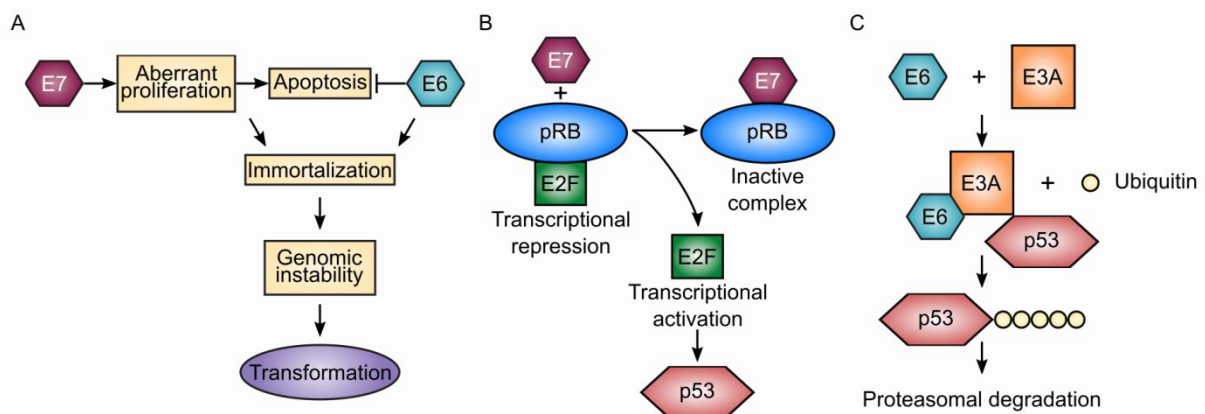


Figure 5. Simplified synergistic mechanisms of the E6 and E7 oncoproteins in driving transformation. (A) E7 and E6 together drive viral genome replication and host cell instability. (B) E7 achieves this by binding pRB, which leads to a release of the E2F transcription factor and transcriptional activation of many cell cycle activators, but also the tumor suppressor protein p53. (C) E6 binds to the ubiquitin ligase E3A, which ubiquitinizes p53 and marks it for proteasomal degradation. Figure is modified from (Moody and Laimins, 2010; Yim and Park, 2005).

1.1.2.2 The late proteins L1 and L2

The late gene products L1 and L2 represent the major and minor capsid proteins, respectively. Both are expressed in the late stage of the viral life cycle, when differentiated keratinocytes reach outer epithelial layers. The major capsid protein L1 has a weight of ~55kDa and forms pentamers, of which 72 can self-assemble in a non-covalent complex with up to 72 L2 proteins to build-d the capsid structure (Buck et al., 2013). The minor capsid protein has a weight of 55-78kDa and its role has been reviewed (Wang and Roden, 2013). It remains controversial whether either capsid protein interacts specifically with encapsulated viral DNA, but the current hypothesis is that L2 co-localizes with L1, E2 and the viral genome at the replication site before virions assemble. Additionally, L1 is involved in the infectious entry of the virion into the host cell, whereas L2 is supposed to have a role in intracellular and nuclear trafficking.

As the outer surface of the capsid is mainly comprised of L1, it represents the major antigen presented to the immune system. Thus, classification of different HPV types is mainly based on the L1 gene sequence (see section “Classification of HPV”). In contrast, it is still controversial if L2 is exposed on the virion surface. As the capsid proteins do not require viral DNA in order to self-assemble, empty non-infectious virus-like particles (VLPs) can be produced from purified L1 for immunization purposes (see section “Prophylactic vaccination”).

1.1.3 Prophylactic vaccination

Since HPV infection was first associated with cervical cancer, the development of preventive vaccines was actively investigated. The finding that the L1 protein can self-assemble into VLPs was exploited for vaccine design. Injected intramuscularly, VLPs traffic to the lymph nodes where they are taken up by antigen presenting cells (APCs) (Lenz et al., 2003). This leads to activation of APCs and subsequent initiation of an immune cascade that eventually leads to T cell-dependent B cell responses and high levels of L1-specific neutralizing antibodies (Deschuyteneer et al., 2010).

To date, three HPV type-specific prophylactic vaccines have been licensed by US Food and Drug Administration (FDA) and European Medicines Agency (EMA) (de Oliveira et al., 2019). The first was the quadrivalent vaccine Gardasil, developed by Merck in 2006, which contains VLPs of the most prevalent high-risk types HPV16 and HPV18, thus protecting against ~70% of cervical cancer. It additionally contains VLPs of the low-risk types HPV6 and HPV11 which are associated with close to 100% of all genital warts (Lacey et al., 2006). The second vaccine was the bivalent Cervarix developed by GlaxoSmithKline in 2007. Cervarix contains VLPs of HPV16 and HPV18. Both vaccines induce antibody-titers that are considerably higher than after natural infection, which is required in order to scavenge virions before they infect a host cell (Arbyn et al., 2018). The third and most recent vaccine, the nonavalent Gardasil9, was brought onto the market in 2014. In contrast to Gardasil, it additionally contains VLPs of HPV31, 33, 45, 52 and 58, and thus covering a broader range of high-risk HPV types, and protecting against 90% of cervical cancer (Huh et al., 2017). Other vaccine formulations, comprising also monovalent vaccines, are being tested in clinical trials.

A recent Cochrane review of 26 trials evaluated the possible harms and the protective effect of prophylactic HPV vaccines against cervical precancer and HPV16/18 infection in adolescent girls and women (Arbyn et al., 2018). The studies involved 1 monovalent, 18 bivalent and 7 quadrivalent vaccines and enrolled mostly women younger than 26 over a period from 6 months to 7 years. This period was not enough to evaluate cervical cancer outcomes, but addressed CIN2 and CIN3 lesions as well as adenocarcinoma-in-situ (AIS). The meta-analysis concludes that vaccination protected against cervical precancer with high certainty and without increased risk of serious adverse events, miscarriage or pregnancy termination. A higher efficacy was observed in association with HPV16/18 and in women negative for high-risk HPVs or HPV16/18 at the time of enrolment. However, effects against any CIN and AIS were shown, which is in line with other reports about cross-protection (Folschweiller et al., 2019). The nonavalent vaccine was not assessed in the reviewed trials. However, safety and efficacy of the nonavalent vaccine were demonstrated in direct comparison to the quadrivalent vaccine in a randomized double-blind trial (Huh et al., 2017).

Although prophylactic HPV vaccination is recommended in several countries for adolescent girls, and increasingly also for boys, vaccination coverage rates still greatly vary (de Oliveira et al., 2019). Especially in low and lower-middle income countries, access to HPV vaccines is almost non-existent (Bruni et al., 2016). In many cases, challenges for national HPV vaccine implementation are political will, vaccine costs, scarce knowledge about HPV-induced diseases and anti-vaccine movements (de Oliveira et al., 2019). It has been estimated that in 75 years from now, in the target population of 118 million women worldwide, 39.7% (444,627 of 1,120,178) expected cancer cases will be prevented by vaccination (Bruni et al., 2016). The remaining 60.3% of cancer cases will arise in unvaccinated women and result from non-16/18 HPV types. This estimation highlights a continuous need for therapeutic treatment options.

1.1.4 Classification of HPV

According to the latest taxonomy release, the family of Papillomaviridae has two subfamilies, the First- and Secondpapillomaviridae (ICTV International Committee on Taxonomy of Viruses, 2018). Whereas the latter subfamily only contains Alefpapillomavirus species, *Firstpapillomaviridae* are further categorized into different genera named after greek letters (Figure 6).

Alphapapillomaviridae have been researched extensively, as the cancer-causing high-risk types belong to this genus. They infect mucosal epithelia, whereas other genera primarily infect cutaneous sites. The α -genus is further distinguished into 15 species groups (**Fehler! Verweisquelle konnte nicht gefunden werden.**). For example, the high-risk type HPV16 belongs to the α -9 species, as do other types with carcinogenic potential. More than 220 different types of HPV were classified based on $\geq 10\%$ sequence difference of the L1 ORFs (Burk et al., 2013; Eklund and Dillner, 2019). Isolates of the same HPV type were formerly referred to as variants or subtypes when the sequence of the L1 ORF differed by 2% to $<10\%$. However, Burk and colleagues reviewed this terminology and established the terms lineages and sublineages based on complete genomic difference of 1% to $<10\%$)

and 0.5% to <1%, respectively (Burk et al., 2013). For HPV16, four lineages were distinguished based on regional prevalence. “A” (European) is divided into four sublineages, “B” (African-1) into two, lineage “C” (African-2) is not yet divided further and the lineage “D” (Asian-American) has three sublineages.

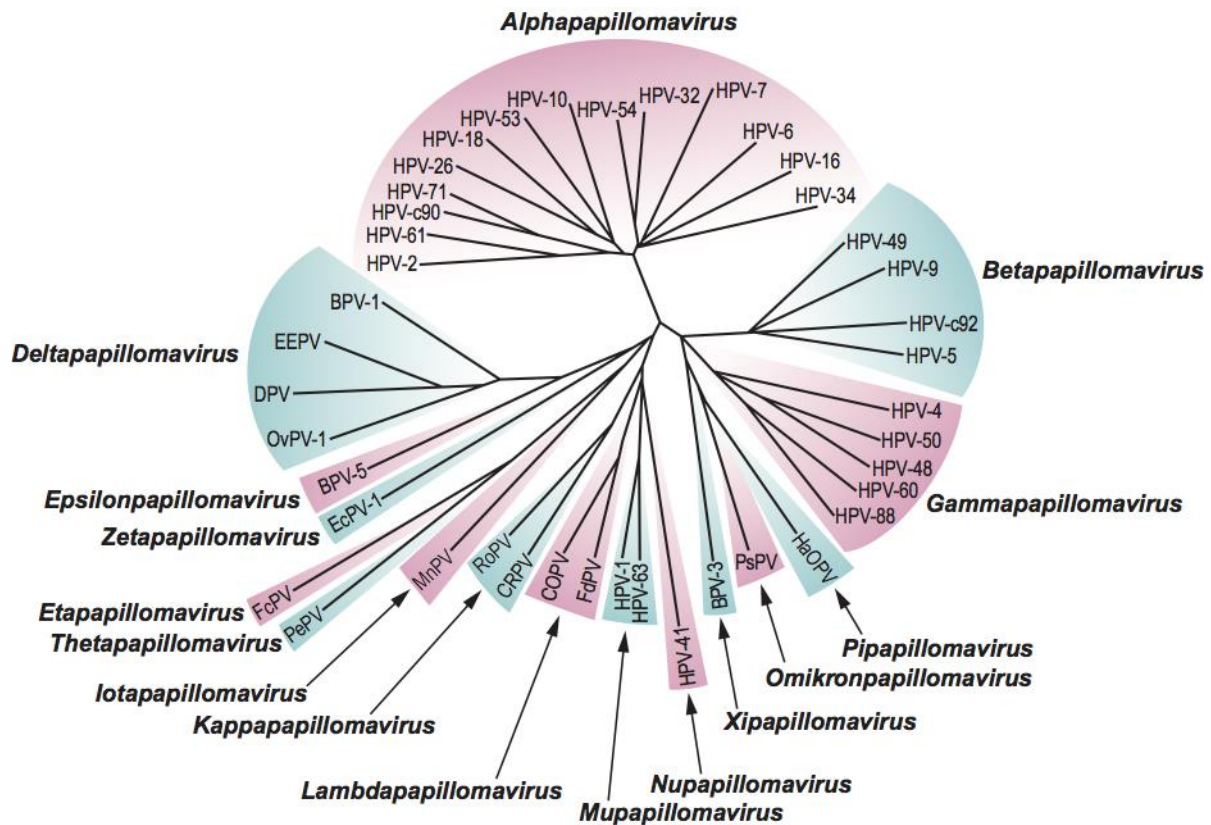


Figure 6. Phylogenetic tree representing 118 papillomavirus sequences. Papillomavirus sequences were grouped into different genera based on the L1 ORFs. Figure from (ICTV International Committee on Taxonomy of Viruses, 2018).

The first HPV16 genome was sequenced 1983 from a German patient and became the reference genome belonging to the European A1 sublineage (Dürst et al., 1983; Seedorf et al., 1985). Nowadays, this genomic sequence is known to be rather uncommon (Zehbe et al., 1998). Therefore, infections with genetic variants of HPV16 are likely to be the rule rather than an exception. Moreover, the genetic variations are possibly influencing HPV persistence, progression into cancer and survival of patients (Xi et al., 2007; Zuna et al., 2011; Clifford et al., 2019).

In a sublineage, HPV genomes can further differ in single nucleotide polymorphisms (SNPs). Such point mutations are described by the reference residue, the position in the gene and the new residue. For example, a nucleotide change in the E6 ORF from thymine to guanine at position 350 is designated as G350T. Similarly, aa changes in the protein are described. The SNP of the example results in an aa change from leucine to valine at position 90 of the E6 protein and is designated as L90V. In the scope of this thesis, this description of amino acid changes in comparison to the reference sequence was used to refer to HPV16 E6 and E7 proteins variants.

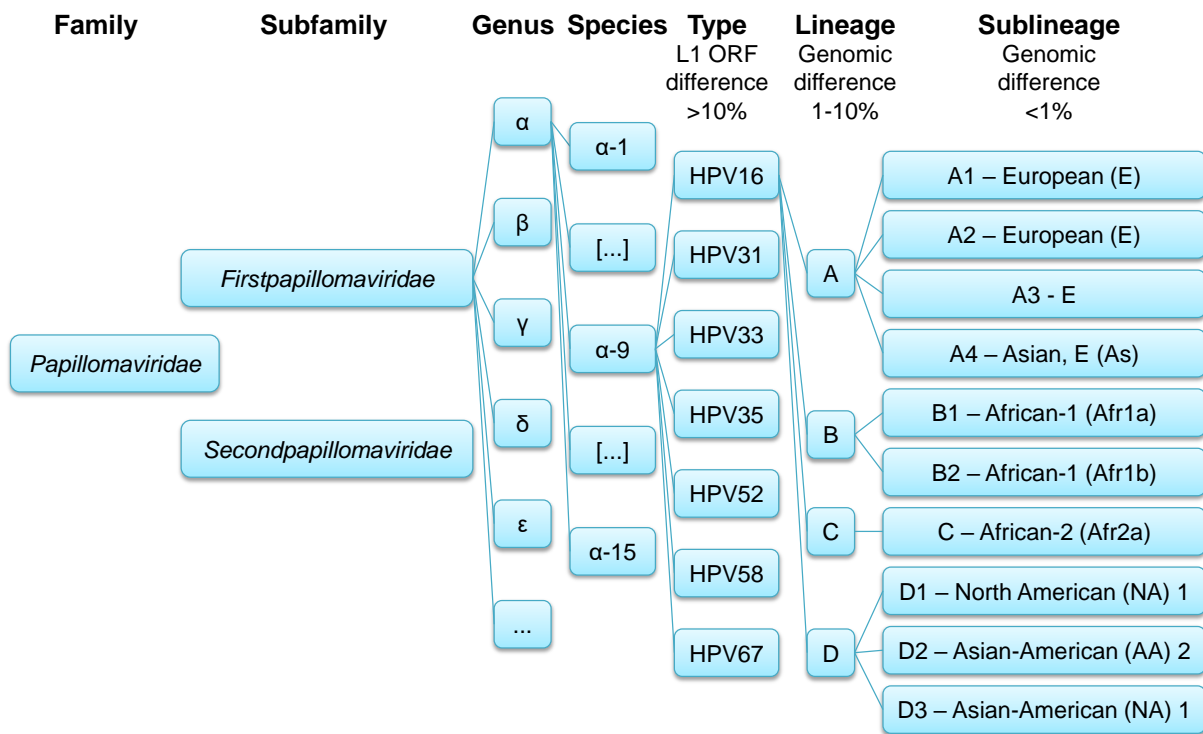


Figure 7. Schematic taxonomy of Papillomaviridae focusing on HPV16. Papillomaviridae are classified into families, subfamilies, genera, species and types. Types are distinguished based on a difference in the L1 ORF of more than 10% compared to all other types. Types are further divided into lineages and sublineages, based on genomic differences greater and smaller than 1%, respectively. Sublineages of HPV16 are related to regional prevalence which resulted in different clade names in prior phylogenetic analyses. This representation is based on the 2018b Taxonomy Release of the International Committee on Taxonomy of Viruses and the review by Burk and colleagues (Burk et al., 2013; ICTV International Committee on Taxonomy of Viruses, 2018).

1.2 The immune system and HPV

Constantly, the human body is contact with numerous pathogens like fungi, parasites, bacteria and viruses. In order to prevent and act against infections and disease organisms have developed multifaceted mechanisms to respond, which are summarized as the immune system. The field of science that aims to elucidate such responses is termed immunology. Immunology was born in 1796 as Edward Jenner proved his hypothesis that inoculation of healthy people with cowpox can protect from the deadly smallpox. This event represents the first described vaccination (from latin *vaccinus* – “from the cow”). Almost a hundred years later, Robert Koch and Louis Pasteur proved the connection between diseases and specific pathogens.

The immune system comprises lymphoid organs, which are classified into primary (bone marrow and thymus) and secondary (spleen, lymph nodes, mucosa-associated lymphoid tissues) lymphoid organs and tissues, and cellular and soluble (humoral) components. Cells of the immune system mainly originate from hematopoietic stem cells of the bone marrow (Murphy and Weaver, 2017).

1.2.1 Innate and adaptive immunity

The mechanisms of innate and adaptive immunity differ in mechanisms of pathogen recognition, kinetics of the immune response, involvement of cellular and soluble components and ability for generating immunological memory.

Innate immunity represents the immediate but less-specific response to pathogen encounter. On the humoral side, pathogens can be bound by a system of plasma proteins, called the complement system. They promote the uptake and destruction of pathogens by phagocytes. On the cellular side, the identification of pathogens depends on germline-encoded pattern recognition receptors (PRR) that recognize regular patterns of molecular structures shared by many microorganisms. These repetitive structures are also called pathogen-associated molecular patterns (PAMPs). In addition to PAMPs, there are also receptors that recognize damage-associated molecule patterns (DAMPs), which are released during cellular damage or death. One group of PRRs are Toll-like receptors (TLRs) and each of them engages with different ligands stimulating particular pathways. For example, surface-expressed TLR4 detects bacterial lipopolysaccharide (LPS), whereas endosomal TLR7 binds to single-stranded RNA of internalized viruses. Most TLRs interact with adaptor molecule myeloid differentiation factor 88 (MyD88), which leads to activation of transcription factor NF κ B. Upon binding of pathogen components, an immune response is instantly initiated. Such a response could be phagocytosis of the recognized particle, chemotaxis of immune cells to the site of infection, or production and release of effector molecules. These functions are performed by the innate immune cells such as neutrophils, natural killer (NK) cells, macrophages, monocytes and dendritic cells (DCs). DCs and macrophages act as a link between the innate and adaptive immune system, based on their ability to endocytose pathogenic components and to process and present resulting fragments to adaptive immune cells.

The adaptive immune cells are T and B lymphocytes. Both express specific antigen receptors, B cell receptors (BCRs) and T cell receptors (TCRs), respectively. In order to recognize virtually all pathogenic structures, receptor diversity is generated by random recombination of receptor genes, which is called V(D)J-recombination. Further, random nucleotides are added at the junction site of the recombined genes generating up to 10^{14} - 10^{18} different receptor specificities. This random mechanism leads to the unique specificity of a clonal cell. In a process called clonal deletion, self-reactive receptors get depleted to induce central tolerance.

If B and T cells encounter their specific antigen and get activated, they proliferate, expand clonally and differentiate into effector lymphocytes with special functionalities. The effector forms of B cells are plasma cells. They produce antibodies that share the antigen specificity of the BCR. Thus, these antibodies represent the humoral arm of the adaptive immune system. They target the pathogen that led to activation of the B cell. On the other hand, T cells can differentiate into one of several different effector T lymphocytes. They can be grouped by mediating three major functions, which are activation, regulation and killing.

Cytotoxic CD8⁺ T lymphocytes (CTLs) mediate cytolysis of target cells displaying specific antigens, which are cells of the body that are either infected with intracellular pathogens, such as viruses, or displaying mutated features of a tumor cells. CD4⁺ T helper (T_h) cells recognize extracellular antigens displayed on APCs and orchestrate other immune cells by secretion of different signature cytokines by which they can be distinguished into different T_h types. Their signals help activating antigen-stimulated B cells to differentiate and they promote the functions of macrophages and CTLs. Regulatory T cells (T_{reg}) control immune responses and induce tolerance by suppressing the activity of other lymphocytes. They play a role in preventing autoimmunity, but also downregulate anti-tumor immunity in the tumor microenvironment.

Most effector lymphocytes are eliminated when their specific antigen is successfully cleared. However, some of the cells that encountered antigen differentiate into so-called memory cells. If they encounter the same antigen in a secondary infection, they rapidly divide and differentiate into effector cells. Thus, the adaptive immune cells are responsible for long-lasting antigen specific immunity (Murphy and Weaver, 2017).

In the following, only the components of the immune system which are involved in antigen-specific and cytotoxic immune responses are described in greater detail as this work is focusing on induction of CTL-mediated immunity.

1.2.2 Antigen presentation by major histocompatibility complex molecules

Pathogen-derived peptide fragments will be recognized by T cells only if they are displayed on the cell surface in complex with special membrane glycoproteins. These complexes were first identified as determinants of histocompatibility in transplantations and are thus called major histocompatibility complexes (MHCs). In humans, MHC molecules are also called human leucocyte antigens (HLAs).

There are two main types of classical MHC molecules that differ in their properties for peptide binding and immune cell stimulation. MHC class I molecules are found on virtually all nucleated cells and present mainly 8-11aa long peptides derived from cytosolic antigens. The peptide-MHC class I complexes are preferentially bound by CD8⁺ T cells and stimulate cytotoxic responses in order to kill infected or tumor cells. In contrast, MHC class II molecules are mainly expressed by professional APCs such as B cells, macrophages and DCs. The displayed peptides are derived from proteins in intracellular vesicles, such as internalized pathogenic components, and typically vary in their length from 9aa to 25aa. MHC class II molecule-presented peptides are recognized by CD4⁺ T cells.

MHC class I molecules are heterodimers composed of two polypeptide chains. The membrane-spanning α -chain folds into three domains, α_1 - α_3 and non-covalently binds a smaller chain, β_2 -microglobulin. Together, α_1 and α_2 fold into a closed cleft or groove, which is the site for peptide binding. Hydrogen bonds and ionic interactions between the MHC chains and the atoms of the free peptide termini stabilize peptide binding. Additionally, other residues in the peptide serve as so-called “anchor residues”. The binding preferences at these anchor residues are specific for MHC alleles and

can be used to predict the MHC binding affinity of peptides (see section “*In silico* epitope prediction”).

As illustrated in Figure 8, the generation of peptide-MHC class I complexes starts in the cytosol where intracellular proteins are degraded into peptides by the proteasome. Stimulation with pro-inflammatory cytokines can induce expression of the immunoproteasome, which leads to a different enzymatic specificity and thus to changes in the presented peptide repertoire. After degradation, peptides are actively translocated to the ER by the transporter associated with antigen processing (TAP), a heterodimer that consists of TAP1 and TAP2. Translocated peptides can be subject to further trimming by ER-associated aminopeptidase 1 and 2 (ERAP1/ERAP2). In the ER, immature MHC class I molecules are part of the peptide loading complex, which additionally involves TAP, tapasin, calreticulin, Erp57 and calnexin. Calnexin stabilizes the MHC alpha chain until it finally assembles with β_2 -microglobulin. Calreticulin and Erp57 act as chaperones that stabilize the empty MHC molecule. Tapasin links the MHC molecule to TAP and facilitates peptide loading. The loaded, matured peptide-MHC class I complex leaves the ER and transits to the cell surface via the secretory pathways of the Golgi apparatus. All components involved in these processes are referred to as the antigen processing machinery (APM).

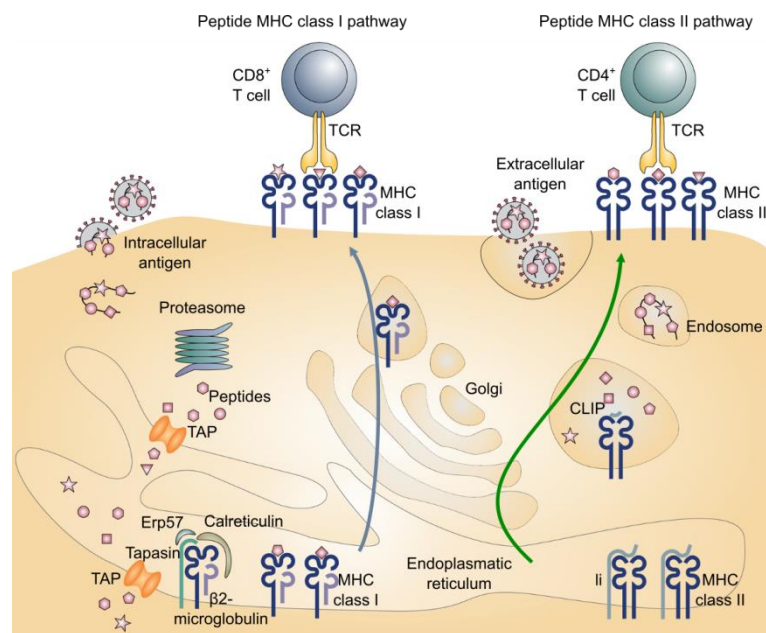


Figure 8. Antigen-presentation pathways of MHC class I and MHC class II. Intracellular antigens, as e.g. derived from viral proteins, are processed and presented via the MHC class I pathway (left, blue arrow). The proteasome degrades cytosolic proteins into peptides, which are translocated to the endoplasmic reticulum (ER) via the transporter associated with antigen processing (TAP). Immature MHC class I molecules form the peptide loading complex together with tapasin, Erp57 and calreticulin. Mature loaded MHC class I complexes transit through the Golgi apparatus to the surface where they present peptide to $CD8^+$ T cells. In the MHC class II pathway (right, green arrow), extracellular proteins are internalized into endosomes and digested by proteases. Immature MHC class II molecules are stabilized by the invariant chain (Ii). Ii is processed into a short peptide chain (CLIP), which is replaced by the peptide to be presented. Mature MHC class II complexes reach the cell surface and present peptides to $CD4^+$ T cells. Figure and legend were adapted from (Purcell et al., 2019).

Apart from this classical pathway of MHC class I presentation, it is possible that extracellular antigens get “cross-presented” on MHC class I if antigens escape from the endosome into the cytosol. However, this can only happen in APCs, which are capable of both MHC class I and class II antigen presentation.

The MHC class II heterodimer consists of two transmembrane chains, α and β , with two domains each. The domains α_1 and β_1 form the binding groove that, in contrast to MHC class I, forms hydrogen bonds all along the peptide length and allows the peptide to emerge from both ends. This enables binding of longer peptides. Similarly to MHC class I, certain aa residues of the peptide serve as anchor residues, but as bound peptides differ in their length, anchor residues are harder to define and prediction of binding peptides is more difficult.

Peptides presented by MHC class II are derived from extracellular proteins, which enter the APC through endocytosis (Figure 8). The resulting endosomes fuse with lysosomes containing acid proteases, such as cathepsins, that digest the internalized protein. Immature MHC class II molecules arising from the ER are stabilized by the invariant chain (Ii) that prevents binding of cellular peptides. During the transit of MHC class II in a vesicle, Ii gets cleaved, leaving a short class II-associated invariant chain peptide (CLIP) bound to the cleft. If the vesicle fuses with an endolysosome containing peptides from degraded proteins, CLIP is displaced by a peptide. Mature MHC class II complexes are delivered to the cell surface where they can be recognized by CD4⁺ T cells (Murphy and Weaver, 2017).

The MHC family is a large gene cluster. In humans, HLAs are encoded by >200 genes, of which the classical MHC genes HLA-A, -B, and -C (HLA class I) and HLA-DPA1, -DPB1, -DQA1, -DQB1, -DRA and -DRB1 (HLA class II) are studied best. These genes are highly polymorphic; each individual co-dominantly expresses two alleles for each of the MHC class I and II genes and to date 16,200 HLA class I and 6,162 HLA class II alleles are known (Robinson et al., 2015). However, some alleles are more frequent than others. For example, HLA-A*02:01 is one of the most frequent alleles in Europe, with 38.5-53.8% of individuals expressing that allele. Despite the polymorphism, HLA class I molecules can be clustered into sets of molecules that share peptide binding motifs. Such clusters of HLA class I molecules represent so-called supertypes (Sidney et al., 2008a). The combined phenotypic frequencies of the supertypes A2, A3, A24, B7 and B15 provide more than 95% population coverage, regardless of ethnicity. This indicates that for therapeutic vaccine development as few as five HLA class I-restricted epitopes may be enough to elicit CTL responses in the whole population (Reche and Reinherz, 2007).

1.2.3 Immune responses of T lymphocytes

Upon recognition of PAMPS or DAMPS, APCs internalize the triggering particle, via phagocytosis or macropinocytosis. APCs include macrophages, B cells and DCs. Macrophages are specialized in taking up particulate material, whereas B cells especially perform receptor-mediated endocytosis. DCs ingest extracellular fluid and its components by macropinocytosis. They are especially important for

initiating T cell responses. Immature DCs arise from hematopoietic progenitor cells in the bone marrow and migrate through the blood stream into tissues where they constantly screen their environment. Once DCs get activated, e.g. by sensing viral double-stranded RNA via TLR3 in an endosome, they switch to a mature phenotype. They lose the ability of endocytosis, secrete cytokines, increase the expression of MHC molecules and induce expression of chemokine receptor CCR7. This sensitizes DCs to the chemokines CCL19 and CCL21, which direct the cells to the draining lymph nodes. Additionally, maturing DCs induce expression of costimulatory molecules like CD70 and CD80/CD86 (B7 molecules) on their surface. In the lymph node, the mature DCs are able to activate naïve T cells by providing three signals as illustrated in Figure 9.

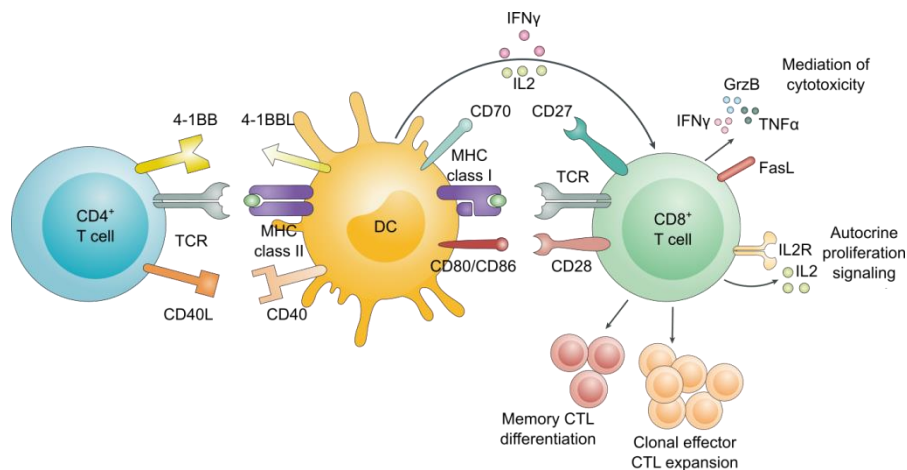


Figure 9. T cell priming by DCs and induction of CD8⁺ T cell responses. In the lymph node, mature DCs (yellow) provide three major signals to CD4⁺ (blue) and CD8⁺ T cells (green): 1) the cognate peptide-MHC complex recognized by the T cell receptor (TCR), 2) the costimulatory molecules CD70 and CD80/CD86 signaling to the receptors CD27 and CD28, respectively and 3) cytokines such as interferon γ (IFN γ) and interleukin-2 (IL-2) that direct effector differentiation when bound to IFN γ receptor 1 (IFNGR1) and IL2 receptor (IL2R), respectively. Additional costimulatory factors are provided by interaction between CD40 and 4-1BBL (DCs) with CD40L and 4-1BB (T cells), respectively. Activation of T cells leads to autocrine IL-2 signaling, driving proliferation and clonal expansion, and differentiation into effector and memory T cells. CD8⁺ T cells turn into cytotoxic T lymphocytes (CTLs) that mediate cytotoxicity by cytokines (IFN γ and TNF α), cytotoxins (GrzB) and FasL interaction. Figure adapted from (Borst et al., 2018).

First, the TCR of the T cell has to bind its cognate peptide presented by the respective MHC class I/II molecule on the DC in order for the T cell to get activated (signal 1). Second, the T cell is interacting with costimulatory molecules on the DC by their respective counterpart on the T cell surface, e.g. B7 by CD28 and CD70 by CD27. Without the costimulatory signal 2, the T cell would undergo anergy or apoptosis. Third, the milieu of secreted cytokines directs the differentiation of the naïve T cell into an effector cell. Together, these signals induce clonal expansion of the T cell. During this interaction, the T cell binds transiently to the DC via adhesion molecules, such as LFA1 on the T cell side binding to ICAM1 and ICAM2 on DCs, forming the ‘immunological synapse’. The T cell-APC-dialog is supported by other costimulatory factors, such as binding between CD40L and CD40 or 4-1BB and 4-1BBL.

A T cell that does not encounter its cognate peptide-MHC complex remains naïve, leaves the lymph node and re-enters the blood flow. If a T cell gets activated by the DC interaction, the recognized

cognate peptide is referred to as a T cell epitope. Once the T cell received all three signals, it starts to secrete interleukin 2 (IL-2), an autocrine proliferation factor for T cells, clonally expands and differentiates into different effector types. CD4⁺ T cells differentiate into different classes characterized by expressing signature cytokines. T_h1 cells activate macrophages and CTLs by producing interferon- γ (IFN γ), T_h2 cells stimulate humoral immune responses by B cells via IL-4, IL-5 and IL-13, and T_h17 cells recruit neutrophils through IL-17, IL-21 and IL-22. The effector type is directed by the third signal; IL-12 and IFN γ induce T_h1, IL-4 mediates T_h2 and tumor growth factor β (TGF β) and IL-6 lead to T_h17 differentiation. Importantly, a lack of IL-6 in abundance of TGF β favors development of adaptive T_{regs} expressing the forkhead box P3 (FoxP3) transcription factor. Primed CD8⁺ T cells differentiate into CTLs that produce IFN γ , tumor necrosis factor α (TNF α) and cytotoxins and express Fas ligand. The cytokines IFN γ and TNF α have antiviral, immunoregulatory and anti-tumor properties (Schroder et al., 2004). Cytotoxins produced by CTLs comprise perforins, granzymes and granulysin. They are stored in specific lysosomes called cytotoxic granules, which are released if a CTL engages its target. Upon degranulation, perforin induces the formation of pores in the target cell plasma membrane and allows granzymes to enter, which in turn activate caspases and trigger apoptosis. Granulysin has antimicrobial functions. The Fas ligand is a transmembrane protein that induces apoptosis in the target cells upon interaction with its receptor. Via these different mechanisms, CTLs can mediate cytotoxicity once they leave the lymph node and migrate into tissues. Once an infection is cleared, T cells start to express regulatory receptors. Examples for such coinhibitory interactions are cytotoxic T lymphocyte antigen 4 (CTLA4) and programmed cell death protein 1 (PD1). They can bind to B7 and PD1-ligand expressed on DCs, which induces anergy and apoptosis. Such immune checkpoint mechanisms are important to regulate T cell proliferation, maintain self-tolerance and prevent excessive immune responses after clearing an infection. T cell responses are ended either by activation-induced cell death or T lymphocyte exhaustion. Only a minority of specific effector T cells remains and constitutes memory T cells. They show a distinct phenotype. Whereas naïve T cells express chemokine receptor 7 (CCR7), the lymphoid homing marker, and the CD45RA isoform, central memory T cells (T_{CM}) switch to the CD45RO isoform. Effector memory T cells (T_{EM}) additionally lose CCR7 expression and are found in the periphery. They can rapidly mature into effector T cells and secrete large amounts of IFN γ , IL-4 and IL-5 after stimulation. A subset of T_{EM} re-expresses CD45RA and thus is called T_{EMRA}. Tissue-resident memory cells (T_{RM}) are another subset of memory T cells. They populate and reside in certain tissues and do not re-circulate. Such tissues represent primary barriers against pathogens, as for example intestinal, genital and respiratory mucosa or the skin (Murphy and Weaver, 2017).

1.2.4 Immune evasion mechanisms of HPV

1.2.4.1 Immunoediting of tumor cells

Tumor cells present DAMPS and antigens. Still, they can escape from detection by the immune system. Avoidance of immune destruction is one of the emerging hallmarks of cancer described in section “Cancer and human papillomavirus infection”. Tumor cells can achieve this in a process called immunoediting. Three phases of tumor editing are described. In the first “elimination” phase, the immune system recognizes target cells and eradicates them, leading to a prevention of tumor growth. However, if elimination is not complete, surviving tumor cells enter an “equilibrium” phase. The actions of the immune system are counterbalanced with immunosuppressive mechanisms. During this phase, the pressure of the immune system “selects” cells that have the ability to survive and remain undetected by immune cells – immunoediting in the strict sense. This eventually leads to the third phase of immune “escape” when tumors are established and progress (Murphy and Weaver, 2017).

The three immunoediting phases can be observed for HPV infections and HPV-associated tumors as well. In the elimination phase, macrophages, CD4⁺ and CD8⁺ T cells infiltrate the infection site (Woo et al., 2008). T cell responses eliminate infected cells and neutralizing antibodies are induced after natural infection (Stanley et al., 2008). However, HPVs can escape immune surveillance and induce persistent infection, as they employ several immune evasion strategies as reviewed (Grabowska and Riemer, 2012; Cicchini et al., 2016; Steinbach and Riemer, 2018).

1.2.4.2 Immune evasion mechanisms of HPV and HPV-positive tumor cells

The life cycle of the virus allows hiding from immune detection in several ways. The virus infects only cell layers above the basement membrane, which are less populated with immune cells. The hijacked cell produces viral proteins only at a low level and does not secrete them. Only in the upper epithelial layers, when keratinocytes terminally differentiate, expression of viral genes is upregulated and virions are formed. However, at this stage, keratinocytes are naturally shed and release virions away from the epithelial surface. Overall, HPVs remain completely intra-epithelial, without viremia and cause no lysis or death of the host cell and thus do not induce inflammation.

In order to prevent recognition HPVs actively interfere with immune recognition via the early proteins E5, E6 and E7. The proteins dysregulate gene expression, protein function and antigen processing. Major altering of gene expression is achieved by epigenetic changes. E7 associates with DNA methyltransferase 1 (DNMT1) (Burgers et al., 2007). The activated methyltransferase acts on different promoters of important immune components. For example, gene expression of CXCL14 and E-cadherin is downregulated, which leads to impaired attraction of Langerhans cells, the DCs of the skin (Laurson et al., 2010; Cicchini et al., 2016). Histone modification mediated by E7 leads to downregulation of TLR9, which senses viral dsDNA (Hasan et al., 2013). Further, E7 is able to interfere with interferon signaling by binding to transcription factors for interferon-induced genes (Um et al., 2002; Antonsson et al., 2006).

Components of the APM represent another class of targets. HPV⁺ cells show a repressed expression of immunoproteasome subunits and of TAP, which reduces the overall number of antigen-derived peptides available for HLA loading (Evans et al., 2001). In contrast, ERAP1 is overexpressed, which likely induces excessive trimming and progressive destruction of HPV epitopes (Steinbach et al., 2017). Further, the E5 protein interferes with MHC class I surface expression. It interacts with the transmembrane domain of the MHC α -chain, traps MHC class I in alkalized vesicles in the Golgi apparatus and scavenges calnexin that usually stabilizes the α -chain before binding β 2-microglobulin (Roman and Munger, 2013).

Moreover, HPV⁺ tumors generate an immunosuppressive microenvironment comprised of immunosuppressive cytokines and cell types such as tumor associated macrophages (TAMs), myeloid-derived suppressor cells (MDSCs), or T_{reg}s (Lepique et al., 2009).

Taken together, all of HPV's immune evasion mechanisms lead to a "low profile", a state at which only few immunogenic epitopes are available. The restricted epitope repertoire limits the points of attack for the immune system during progressive disease. Thus, it is important to identify available HPV target epitopes in order to develop novel immunotherapies against HPV-related malignancies.

1.3 Cancer Immunotherapy

1.3.1 An overview of cancer immunotherapy approaches

Cancer immunotherapy evolved over the past decades and today represents an emerging field of translational research and a promising strategy for cancer treatment or even cure. It aims at inducing or enhancing anti-tumor responses of the immune system and at overcoming immunosuppression. In the past decades, several unspecific and specific immunotherapeutic approaches showed efficacy and safety in clinical studies which led to approval by the FDA or the EMA (Riley et al., 2019).

The earliest approved **unspecific** immunological medications were cytokines for promoting lymphocyte proliferation such as recombinant IFN α 2 (1986) and IL-2 (1992) (Ahmed and Rai, 2003; Rosenberg, 2014). Another cytokine in clinical practice is granulocyte-macrophage colony-stimulating factor (GM-CSF). It promotes differentiation of myeloid cells and DCs, acts as adjuvant and contributes to the regulation of immunosuppression in the tumor microenvironment (Yan et al., 2017). Other cytokines, e.g. and TGF β receptor type 1 inhibitors, are investigated for clinical use (Uhl et al., 2004).

Latest advances in unspecific immunotherapy were the development of immune checkpoint inhibitors. The most common checkpoint inhibitors in use are blocking PD-1/PD-L1 and CTLA4. As explained earlier, immune checkpoint interaction represent coinhibitory signaling that leads to T cell anergy and apoptosis in order to maintain self-tolerance and downregulate excessive immune responses. In turn, when blocking this signaling, T cells remain active and can mediate tumor cell killing (Riley et al., 2019).

In contrast to generally activating the whole T cell pool, **specific** cancer immunotherapies target features that are predominantly expressed on tumor cells. Potential targets can be tumor-associated antigens (TAAs), cancer-testis antigens (CTs), onco-fetal antigens (OFs), tumor-specific antigens (TSA) or viral oncoproteins (Murphy and Weaver, 2017; Finn and Rammensee, 2018). TAAs are proteins that are also present on healthy tissue but are overexpressed by tumor cells, or proteins that are abnormal for the local tissue type but regularly expressed in other tissues. CTs represent tumor antigens, which are usually only expressed in male germ cells. OFs are proteins that are usually only expressed in fetal tissues, but re-expressed in tumors. In contrast, TSAs are solely expressed in tumor cells, such as mutation-derived neoepitopes. Viral oncoproteins represent a specific type of TSAs, as they are shared by all tumor cells but are derived from an infectious agent. These potential tumor rejection antigens have individual advantages and limitations considering different tumor entities. Therapeutic targets must be cautiously selected in order to induce an immune response directed against all tumor cells without harming healthy tissue.

The first approved specific immunotherapy approaches were antibody based therapies, which utilize the potential of the effector mechanisms of the immune system by targeting tumor antigens (Stamova et al., 2012). Monoclonal antibodies (Mabs) can trigger innate immune responses such as complement-mediated cytotoxicity and antibody-dependent cellular cytotoxicity or phagocytosis. Various tumor antigens are targeted by Mabs, e.g. CD20 (Rituximab), CD33 (Gemtuzumab) or human epidermal growth factor receptor 2 (Her2/neu) (Trastuzumab). Over the past decades, different bispecific antibody constructs have been developed. In order to specifically bind two targets, these constructs are engineered to provide two variable domains with different complementarity determining regions. For example, the bispecific antibody Blinatumomab is cross-linking CD3 on T cells and CD19 on B cells and was approved for treating B cell acute lymphocytic leukemia in 2014 (Krishnamurthy and Jimeno, 2018).

T cell-based therapies also aim at specific targets. Such therapies comprise adoptive cell transfer (ACT), e.g. of expanded tumor infiltrating lymphocytes (TILs) or of genetically engineered T cells with epitope-specific TCRs or chimeric antigen receptors (CARs) (Besser et al., 2013; Murphy and Weaver, 2017). CARs consist of antigen-recognition domains of antibodies coupled to intracellular signaling domains of the TCR. Thus, CAR T cells are not restricted to MHC-presented peptides. Currently, two CAR T cell therapies targeting CD19 are approved for treating B cell leukemia and lymphoma, but other antigens are investigated. Not yet approved but tested in clinical trials are TCR T cell therapies (NCT01352286, NCT01892293, NCT01343043, NCT01567891, NCT01350401 and NCT02588612) (Hughes et al., 2005). Here, T cells are genetically engineered to express TCRs with high target affinity. However, toxicities in clinical investigations demonstrated that TCR specificity is of utmost importance in order to prevent cross-reactivity against “self” (Morgan et al., 2013; Cameron et al., 2013). In contrast to CAR T cells, TCR T cells depend on matching peptide-MHC complexes.

A rather broad field of antigen-specific treatment strategies is referred to as cancer vaccines. These strategies aim at inducing a T cell-mediated immune response against tumor antigens. Cancer vaccines are dependent on delivery technologies in order to provide functionality at the tumor site. Delivery strategies comprise nanoparticle or liposome-based drug delivery, molecular conjugates, such as antibodies and albumin- or matrix-binding domains, depot-forming platforms, like Montanide ISA 51, implantable or injectable biomaterial scaffolds or transdermal microneedles, as reviewed (Riley et al., 2019). Types of cancer vaccines include tumor cell lysates, DC-based vaccines, nucleic acid vaccines and protein- or peptide-based vaccines. Tumor cell lysates supply a broad range of tumor antigens, which can be presented by any HLA type (Chiang et al., 2015). Tumor lysates can also be used for generating DC vaccines (Garg et al., 2017). The first approved DC vaccine is Sipuleucel-T for treatment of prostate cancer. The principle of DC vaccines is based on autologous immature DCs, which are *in vitro* supplied with tumor antigen and maturation factors. The mature DCs, loaded with tumor antigen, are transfused back into the patient where they can induce anti-tumor T cell responses. Similarly, immune responses can be induced *in vivo* when tumor proteins or peptides are injected. For example, vaccination with the whole tumor protein NY-ESO1 efficiently primed CD8⁺ T cell responses (Karbach et al., 2011).

Like the majority of whole antigens, peptides do not provide their own PAMPs or DAMPs and thus need adjuvants in order to stimulate strong and durable immune responses. However, in contrast to whole protein, peptide-based vaccinations do not pose the risk of introducing biological contaminations or DNA transforming functions. Moreover, peptide vaccines are attractive because fully characterized peptides can be synthetically produced in a fast, simple, cost-effective and reproducible manner. Freeze-drying of peptides facilitates storage and transport without the need of a cold chain (Skwarczynski and Toth, 2016). Another option to deliver peptide-based vaccines is encoding the peptide sequence in DNA or mRNA, which is internalized by APCs and translated into peptide, which is presented to epitope-specific T cells. Especially in the field of personalized medicine, mRNA-based vaccines encoding for individual maturation-derived neoepitopes represent promising treatment approaches (Sahin et al., 2017). Peptide vaccines consist of either long or short synthetic peptides (SLP or SSP, respectively). In contrast to protein-based approaches, a very narrow and targeted immune response against single epitopes can be induced, especially when SSPs are used (Chabeda et al., 2018). Synthetic long peptides are not HLA type-restricted and can contain sequences for both CD4⁺ and CD8⁺ T cell epitopes. They require uptake, antigen processing and presentation by DCs in order to activate CD4⁺ T cells and, via cross-presentation, CD8⁺ T cells. In contrast, short synthetic peptides can be externally loaded onto MHC class I as they contain only a single CD8⁺ T cell epitope. Thus, they lack the ability to induce CD4⁺ T cell responses if they are not accompanied by a T_h epitope. Due to the specific binding preferences of MHC molecules, short peptide vaccines are HLA-restricted and can thus only induce immune responses in HLA-matching patients (Van Hall and Van der Burg, 2012).

1.3.2 *In silico* epitope prediction and its role in cancer immunotherapy

Immunogenicity is dependent on several factors. Protein expression, antigen processing and transport, peptide-MHC binding affinity and competition, stability of the MHC complex and the available TCRs determine the repertoire of presented and recognized epitopes (Trolle and Nielsen, 2014). *In vitro* testing of all possible epitope candidates is not feasible, considering the variety of antigens that can be derived from a single pathogen or an established tumor cell, and the polymorphism of MHC alleles (Dendrou et al., 2018). Therefore, different computational methods have been developed to predict the T cell epitopes which are likely to result from the various steps involved in presentation.

Proteasomal cleavage of MHC class I peptides can be predicted by the algorithms PProC and NetChop 3.1 (Kuttler et al., 2000; Nielsen et al., 2005). Similarly, predictors for peptide-TAP binding likelihood, e.g. TAPPred, were developed and often integrated into combined prediction approaches for antigen processing such as NetCTL 1.2 and the Immune Epitope Database (IEDB) tools MHC-NP and MHC-I processing predictions (Bhasin and Raghava, 2004; Larsen et al., 2007; Giguère et al., 2013; Tenzer et al., 2005). The combined approaches also include predictions for MHC class I binding. The stability of peptide-MHC class I complexes can be estimated by methods like NetMHCstab 1.0 (Jørgensen et al., 2014). Additional algorithms exist for the prediction of T cell immunogenicity, e.g. the IEDB tool Class I immunogenicity, and were integrated into prediction chains that consider processing and MHC class I binding such as NetTepi 1.0 (Calis et al., 2013; Trolle and Nielsen, 2014).

MHC binding is the most crucial and selective step, as only peptides which harbor characteristic residues at specific anchor positions will bind to the MHC molecule (Yewdell and Bennink, 1999). The specific binding features characteristic for each MHC molecule can be deduced from experimentally determined MHC binders. Such experiments are for example *in vitro* competitive binding assays, which determine MHC affinity by the half-maximal inhibitory concentration (IC_{50}) of a test peptide vs. a reference ligand, or immunoprecipitation of endogenous MHC complexes and subsequent high performance liquid chromatography (HPLC) and/or mass spectrometry (MS) identification of eluted ligands. Using this strategy, peptide motifs of MHC-bound peptides have been intensively studied by the group of Hans-Georg Rammensee (Falk et al., 1991; Rammensee et al., 1993; Rammensee, 1995; Ghosh et al., 2019). Based on these motifs they developed SYFPEITHI, a database for MHC ligands and one of the first MHC binding predictors (Rammensee et al., 1999). The preferred binding chemistry was described for MHC class I and MHC class II molecules, and thus prediction methods are available for both. As explained before, MHC class II-associated peptides differ greatly in their length, which complicates precise prediction of ligands. In contrast, class I anchors are more conserved (Murphy and Weaver, 2017). Defined peptide motifs, uneconomical *in vitro* alternatives and the clinical value of CTL epitopes have been motivating the development of multiple different computational approaches mainly for MHC class I ligand prediction.

Existing MHC class I predictors use various strategies to predict binding. Approaches can be based on scoring matrices (SMs) and artificial neural network (ANN) machine learning methods. Additionally, there are consensus approaches that consider different methods to calculate new output. Machine learning of MHC predictors can be trained either allele-specific or pan-allelic.

Matrix-based algorithms calculate a binding likelihood score, considering sequence similarity and amino acid frequency in comparison to known motifs. Position specific effects are weighted as well (see Figure 10 A). Existing methods differ in their statistical scoring functions. Well-known and widely used predictors based on scoring functions are RANKPEP, PSSMHCpan and the tools evaluated in this study: SYFPEITHI, IEDB SMM, IEDB SMMPMBEC, PickPocket 1.1, and, MixMHCpred 2.0.2 (Reche et al., 2002; Liu et al., 2017; Rammensee et al., 1999; Peters and Sette, 2005; Kim et al., 2009; Zhang et al., 2009; Bassani-Sternberg et al., 2017). In detail, SYFPEITHI assigns each amino acid a position-dependent value of up to 10, based on frequency of occurrence in known MHC ligands, and adds the corresponding values to return a score for the evaluated peptide sequence. The stabilized matrix method (SMM) constructs a position-specific scoring matrix (PSSM) based on ligands in the training data (here IEDB), calculates the sum of residue contributions and transforms the result into a putative IC₅₀ value. The SMM method is also used in the SMMPMBEC tool, which uses a peptide:MHC-binding covariance matrix prior to the SMM method, and in the pan-allelic PickPocket 1.1 algorithm, that de-convolves the SMM data to pocket-specific binding events and generates a weighted average PSSM from all HLAs in the training data. MixMHCpred 2.0.2 was released in 2018 and represents the most recent scoring function-based predictor. In contrast to the other described methods, it is trained on a very large dataset of >115,000 MS-derived peptides associated to 123 HLA class I molecules. PSSMs are calculated based on peptide-associated MHC alleles and peptide lengths. Logarithms of the corresponding PSSMs at each position are summed to return a peptide score which is expressed as percentile rank, corresponding to the fraction of random 8-14-mer peptides that score higher than the test peptide.

In contrast to scoring function-based methods, predictors based on machine-learning are capable of identifying non-linear patterns in peptide binding data by using ANN algorithms. Generally, ANNs are composed of input, hidden and output layers, each containing different interconnected units (see Figure 10 B). The connections between units represent weights and biases. In a feed-forward structure, the signal of one unit can be used as input of a connected unit or a neighboring layer. Predictors based on ANNs are for example ConvMHC, HLA-CNN, MHCseqNet, DeepSeqPan and the methods described in this study: NetMHC 4.0, NetMHC 3.4, NetMHCpan 4.0, NetMHCpan 3.0, NetMHCpan 2.8, MHCflurry 1.2, and MHCnuggets 2.0 (Han and Kim, 2017; Vang and Xie, 2017; Phloyphisut et al., 2019; Liu et al., 2019; Nielsen et al., 2003; Andreatta and Nielsen, 2016; Lundegaard et al., 2008a, 2008b; Jurtz et al., 2017; Hoof et al., 2009; Nielsen and Andreatta, 2016; Nielsen et al., 2007; O'Donnell et al., 2018; Shao et al.). These machine-learning methods differ in their training data and allele-specificity. NetMHCpan 4.0 and MHCnuggets 2.0 are partially trained on MS-derived peptides.

MHCflurry 1.2 offers the optional inclusion of MS data which was dismissed in the analysis performed in this thesis. NetMHC 4.0 and NetMHC 3.4 return allele-specific prediction likelihoods, whereas NetMHCpan 4.0, NetMHCpan 3.0, NetMHCpan 2.8, MHCflurry 1.2, and MHCnuggets 2.0 perform pan-allelic predictions. All selected ANN methods express predicted binding affinity as putative IC_{50} values. Detailed differences between predictors are not explained here due to the complex nature of ANNs.

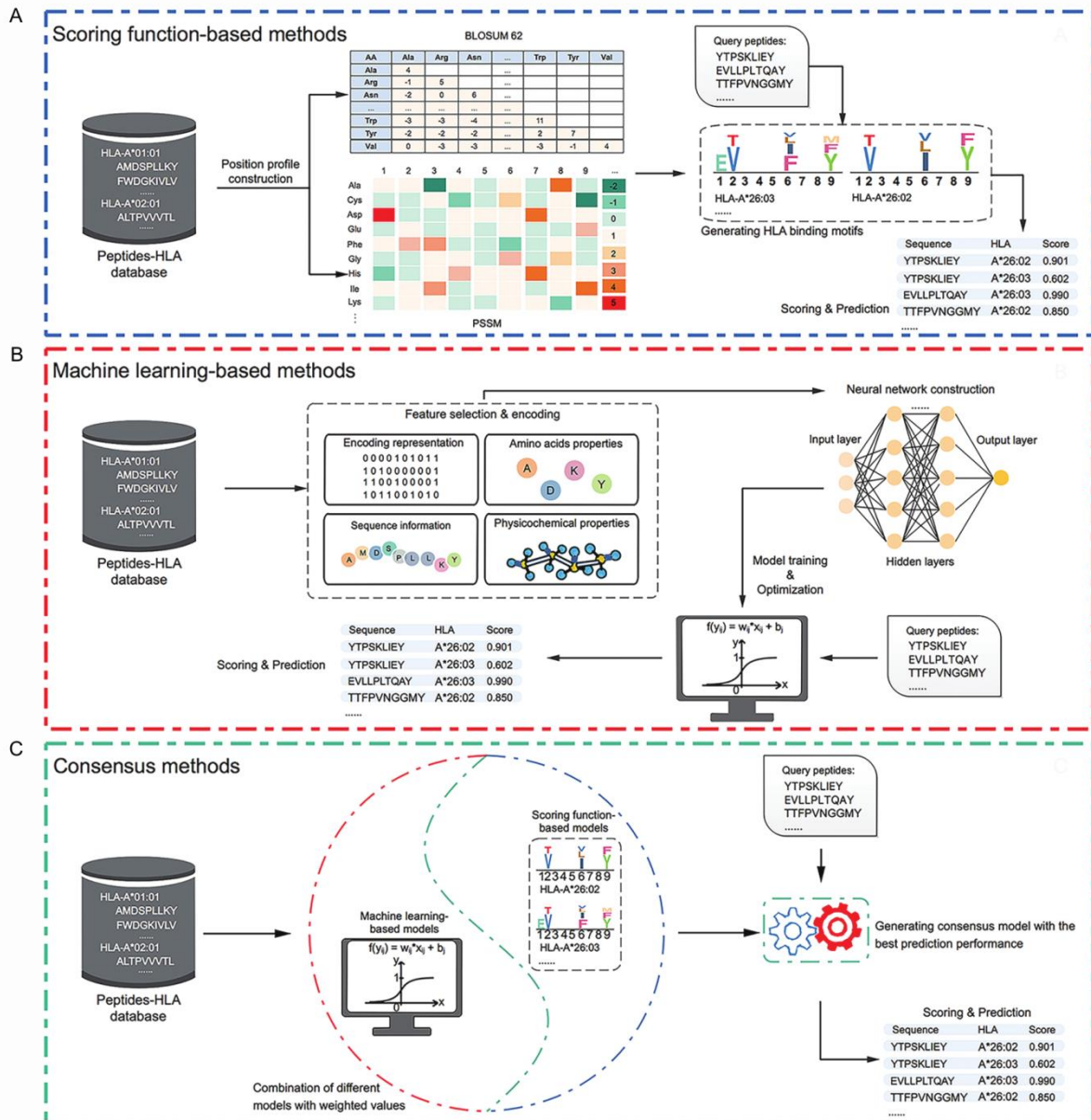


Figure 10. Key steps of MHC binding prediction methods. (A) Scoring functions are used to generate motifs of specific HLA-alleles and score query peptides accordingly. (B) Machine learning-based methods make use of models generated from training data interconnected in artificial neural networks. (C) Consensus methods integrate different prediction approaches to calculate a new binding likelihood score. Figure and legend adapted from (Mei et al., 2019).

Consensus approaches aim to improve prediction performance by considering the output of individual methods combined in a weighted score (see Figure 10 C). Examples for such methods are NetMHCcons 1.1 and the IEDB tools consensus and recommended (Karosiene et al., 2012; Moutaftsi

et al., 2006; Sidney et al., 2008b). NetMHCcons 1.1 integrates NetMHC 3.4, NetMHCpan 2.8 and PickPocket 1.1 to return a putative IC₅₀ value as binding affinity. IEDB consensus weighs the output of NetMHC 4.0, IEDB SMM and the CombLib predictor (Sidney et al., 2008b). IEDB recommended selectively decides for the best method available for a given MHC molecule, based on a previously observed performance ranking that favors IEDB consensus over NetMHC 4.0, SMM, NetMHCpan 3.0 and CombLib. Both IEDB methods result in a percentile rank as output.

Precise MHC binding prediction can be exploited to predict pathogen-derived peptides, which likely represent epitopes. Such epitopes demonstrate promising candidates for vaccination approaches. In a recent study, Croft et al. were using MS for the identification of vaccinia virus-derived epitopes in mice. They found that the majority of MHC-presented viral peptides were immunogenic (Croft et al., 2019). This finding highlights the role of MS profiling and its potential impact on improving MHC-epitope predictions as well as the relevance of MHC binding for the identification of T cell epitopes (Creech et al., 2018). Lately, MHC binding predictors have emerged as valuable tools in neoepitope identification strategies. Using next-generation sequencing methods, individual mutations are identified in tumor samples, RNA sequencing reveals expressed neo-antigens, and MHC binding predictions are performed to select promising neoepitopes. Infrequently predicted neoepitopes are validated for immunogenicity before they are used for the design of cancer vaccines (Ott et al., 2017; Sahin et al., 2017; Koşaloğlu et al., 2016). Thus, in the context of personalized medicine, MHC binding predictions greatly gained importance in order to identify patient-specific candidate targets for cancer immunotherapy.

1.3.3 HPV-specific immunotherapies

Compared to the conventional treatment options for HPV-induced malignancies, like surgery, radiation and chemotherapy, successful HPV-specific immunotherapy would provide a non-invasive alternative with low side effects and sustained immunity (Gulley, 2013). Although prophylactic vaccines against HPV exist, there is a continuous need for the development of immunotherapies against HPV-associated diseases as described above. HPV-specific immunotherapies are especially promising because of the expression of ideal target proteins, namely E6 and E7. They are constitutively expressed in all stages of progressing disease and carcinoma and are not subject of central tolerance because of their viral origin (zur Hausen, 2002). Therefore, E6 and E7 of the most prevalent high-risk HPV type HPV16 are the antigens of choice of most approaches (van der Burg et al., 2016). Generally, immunotherapy approaches can be divided into passive immunotherapies, which autonomously perform anti-tumor effects, and active immunotherapies, which stimulate the recipient's immune system to generate cellular and/or humoral anti-tumor responses.

In HPV-specific passive immunotherapies, HPV-targeting T cells are used in adoptive cell-based immunotherapy. For example, in a clinical phase II trial (NCT01585428), ACT of HPV-TILs was used to treat 18 patients with metastatic cervical cancer. In two cases, complete responses were observed and three patients showed partial responses (Stevanović et al., 2019). Another cellular

approach used against HPV-induced cancers are TCR-transgenic T cells. A first in human phase I/II clinical trial (NCT02280811) tested dose limiting toxicities of T cells expressing a TCR directed against E6/29-38 in twelve HLA-A*02:01⁺ patients with HPV16-related metastatic or recurrent cancers. They observed two partial responses in the group of 6 patients who received the highest dose of T cells (>10¹¹ up to 2x10¹¹ cells). A similar study using an E7/11-19 TCR is currently recruiting (NCT02858310).

The majority of anti-HPV immunotherapy approaches are active, i.e. cancer vaccinations. However, as described above, immune evasion strategies of HPV can complicate effective immunization. The mucosal location of high-risk HPV-induced lesions and carcinomas represents another challenge for vaccination strategies. Specific T cells that express a molecular code associated with mucosal homing have to be addressed (Nardelli-Haeffliger et al., 2013). This has been demonstrated to be linked to the route of vaccination, as T cells more efficiently migrate to mucosal tumor site if vaccination is applied orthotopically (Sandoval et al., 2013). Apart from orthotopic vaccination, T cells can be directed to the tumor site by specifically inducing mucosal T cells or by prime-pull approaches (Sun et al., 2015; Tan et al., 2018). As described above, various different approaches exist for cancer vaccination. Many of them are employed in the development of HPV vaccines. However, to date, no therapeutic HPV vaccines are approved by drug licensing authorities albeit numerous candidates are being tested in clinical trials (Khallouf et al., 2014; Liu et al., 2017; Chabeda et al., 2018). The most clinically advanced therapeutics are introduced in the following.

Investigation of a bacterial vector, ADXS11-001 expressing a fusion protein of listeriolysin O (LLO) and E7 in *Listeria monocytogenes*, showed increased IFN γ ⁺ T cells with E7-specificity and reduction in tumor size in patients with metastatic or advanced cervical cancer in a phase I/II study (Maciag et al., 2009). Based on these results, the study advanced into a phase III trial (NCT02853604).

A viral vector in clinical trials is TA-HPV, a recombinant modified vaccinia virus Ankara (MVA) vector expressing E6 and E7 of HPV16 and HPV18. In phase II studies, 8 of 29 cervical cancer patients showed serological and 4 of 29 showed CTL-based HPV-specific responses after a single dose, whereas 10 of 12 high-grade VIN or VAIN patients showed an average decrease in lesion size of 40% (Kaufmann et al., 2002; Baldwin et al., 2003). Another MVA vector is TG4001, expressing HPV16 E6/E7 and IL-2. In a phase II study, it was administered in 3 weekly subcutaneous injections and induced responses in ten HPV16-related CIN2/3 patients (48%), of which seven experienced regression (Brun et al., 2011).

A HPV-specific DNA-based vaccine is VGX-3100, which encodes HPV16/18 E6/E7 proteins and is injected intramuscularly with subsequent electroporation. It passed a phase I trial in CIN2/3 patients with promising CD8⁺ T cell responses (14 of 18 patients) and increased HPV16 and HPV18 antibody titers (17 and 18 of 18, respectively) (Bagarazzi et al., 2012). In a subsequent randomized, double-blind, placebo-controlled phase IIb trial, 49.5% of the vaccine recipients had histopathological regression of CIN lesions (Trimble et al., 2015). Several trials are currently recruiting to analyze the

VGX-3100 vaccine for use against anal neoplasia (NCT03603808 and NCT03499795), VIN (NCT03180684), and in phase III studies against CIN (NCT03185013 and NCT03721978). A combination of VGX-3100 and a recombinant IL-12 encoding molecular adjuvant (INO-9012) was recently demonstrated to induce HPV-specific CD8⁺ T cells in 18 of 21 head and neck cancer patients (Aggarwal et al., 2019).

Protein-based vaccines contain all epitopes of an antigen and are not MHC-restricted. However, they are internalized as extracellular material and thus promote T_h cell over CTL responses. Fusion proteins and adjuvants are used to increase T cell responses. A fusion protein of HPV16 L2, E6 and E7 used as subunit vaccine is called TA-CIN, which after 52 weeks showed complete regression of VIN in 12 of 19 patients treated in combination with imiquimod (Daayana et al., 2010).

One of the most intensively studied therapeutic HPV vaccination approaches is based on synthetic long peptides (SLPs), consisting of overlapping SLPs of HPV16 E6 and E7. This SLP mix was formulated with Montanide ISA-51 adjuvant and used to vaccinate 20 VIN patients subcutaneously in 3-week intervals. All patients showed vaccine-induced T cell responses, and 9 of 19 patients had a complete clinical response. Interestingly, these patients showed stronger CD4⁺ than CD8⁺ T cell responses (Kenter et al., 2009). The same vaccine was used to treat advanced or recurrent gynecological carcinoma in 20 patients. Although inducing HPV16-specific T cell responses, in 9 of 16 tested patients the tumors neither regressed nor were stopped from progressing (van Poelgeest et al., 2013). Currently, SLP mixes are tested in combination with chemotherapy and utomilumab (an antibody binding to 4-1BB, NCT03258008), or with the PD-1 checkpoint inhibitors cemiplimab (NCT03669718) and nivolumab (NCT02426892) in oropharyngeal cancer and various incurable HPV16⁺ tumors, respectively.

A peptide vaccine based on a synthetic short peptide (SSP) will be analyzed in a currently recruiting phase I/IIb trial for the treatment of incurable HPV 16-related oropharyngeal, cervical and anal cancer in HLA-A*02⁺ patients (NCT02865135). Herein, the safety and efficacy of DPX-E7, a HPV16-E7/11-19 nanomer with metronomic cyclophosphamide, will be evaluated. Previous studies on single HPV16 epitopes were conducted with the E7/11-20 and E7/86-93 peptides together with a pan-DR T helper epitope (PADRE) for treatment of cervical carcinoma patients. In these studies, no antigen-specific CTL responses were observed (Ressing et al., 2000). In contrast, formulations with lipidated E7/86-93 or E7/86-93 together with E7/12-20 did induce antigen-specific T cell responses in cervical carcinoma and CIN/VIN patients, respectively (Steller et al., 1998).

Results from the above mentioned conducted trials in HPV-mediated cancer patients indicate that whole protein and SLP vaccines can induce strong humoral and T_h immunity but do not trigger effective CTL responses which are needed for tumor regression and clinical success. Ongoing studies will show if SLPs in combination with checkpoint inhibitors overcome the immunosuppressive tumor microenvironment. In contrast, SSP vaccines would provide only epitope-specific stimuli for CTLs and thus mediate a tumor-directed cytotoxic immune response. However, clinical trials using SSP

vaccines to treat HPV16 showed that this approach is challenging. Knowledge about the actual presentation of the targeted peptide is important, as immune escape mechanisms can lead to presentation of only some HPV epitopes on the surface of tumor cells (see section Immune evasion mechanisms of HPV above). Other requirements for ideal epitopes are disease-specificity – given in the case of HPV; conserved sequences – possible for some HPV epitopes; and knowledge about HLA type-specific peptide recognition and immunogenicity.

The research group Immunotherapy and Immunoprevention of Angelika Riemer aims at overcoming the mentioned challenges and at developing a therapeutic vaccine against HPV16-induced malignancies based on the identification and validation of CTL epitopes. The identification of ideal HPV16 E6/E7-derived target epitopes is approached by first using computational prediction of potential HLA type-specific binders, followed by synthesis of predicted peptides and experimental validation. Subsequently, verified HPV E6 and E7 HLA ligands are investigated for HLA-restricted cell surface presentation on HPV16⁺ cancer cells using a targeted MS-strategy. In parallel, binding peptides are analyzed for immunogenicity in functional assays with HLA-matched peripheral blood mononuclear cells (PMBCs) of healthy donors. *Bona fide* presented immunogenic candidate epitopes are tested in different vaccine formulations in a preclinical tumor model in a HLA-humanized A2.DR1 mouse strain. In order to develop a HPV vaccine capable of immunizing >95% of the world population, the Riemer group is focusing on epitopes of the most prevalent HLA-types among supertypes.

2 Aims of this study

As outlined above, HPV-related dysplasia and cancer remain a major worldwide health burden, although prophylactic HPV vaccination and screening programs were introduced. As standard of care treatments are invasive and come with side-effects, alternative treatment strategies such as cancer vaccination are needed. For the development of an effective therapeutic vaccine and other epitope-targeting immunotherapeutic approaches against HPV16-associated malignancies, the selection of suitable target epitopes is crucial. T cell epitopes need to be presented in complex with HLA molecules. These have type-dependent binding preferences. Based on identified binding motifs, HLA ligands can be predicted *in silico* using online available computational methods.

To identify ligands to the HLA class I supertype representatives, A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, B*07:02 and B*15:01, prediction and experimental validation of potential HPV16 E6/E7 derived HLA class I binders was the **first aim** of this study. In order to avoid missing possible true ligands, the individual strengths of various prediction algorithms were exploited, by using 15 different prediction methods for peptides of 8aa to 11aa.in length.

Experimental validation of predicted peptides resulted in a comprehensive dataset of peptides and their associated predicted and actual binding affinities. Such a dataset represents a valuable resource for assessing binding prediction performance. Therefore, the **second aim** of this study was to evaluate the used 15 predictors based on the HPV16 E6/E7 peptide dataset.

As HPV16 evolved into genetically distinct variants, peptides derived from E6 and E7 variants with known amino acid substitutions were included in the HLA binding assessment. Regarding epitope-based vaccine design, such protein regions affected by variants have implications, as they are not conserved and not shared by all HPV16 patients. For the purpose of investigating the influence of amino acid exchanges in the HPV16 E6 and E7 peptides, **the third aim** of this thesis was to compare HLA binding affinities of reference sequence- and variant-derived peptides.

Only immunogenic HLA class I-presented HPV16 E6 and E7-derived peptides are suitable candidates for effective therapeutic vaccination. Accordingly, the **fourth aim** of this study was to discover functional T cell epitopes among the identified HPV16 E6/E7-derived HLA binders focusing on MS-detected *bona fide* presented peptides. Functionality was assessed by the ability of peptides to induce IFN γ -secretion and to mediate cytotoxicity against HPV16⁺ cancer cells.

Altogether, this thesis aimed at providing a widely usable performance evaluation of HLA class I ligand prediction methods, as well as at the definition of candidate epitopes for therapeutic HPV16 vaccine design.

3 Materials and Methods

3.1 Materials

3.1.1 Laboratory equipment

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™	Thermo Fisher Scientific, Waltham, USA
Automated cell counter	Nucleo Counter® NC-200™	ChemoMetec, Allerød, Denmark
Cell freezing container	Nalgene® Mr. Frosty® Cryo 1°C Freezing Container	Thermo Fisher Scientific, Waltham, USA
Centrifuge	Centrifuge 5417 R, 5418	Eppendorf, Hamburg
Centrifuge	Biofuge pico, fresco	Heraeus, Hanau
Centrifuge	Megafuge 16 R	Heraeus, Hanau
Centrifuge	Sunlab® Minizentrifuge SU1550	Labdiscount GmbH, Mannheim
Centrifuge rotor for microcentrifuge	F45-30-11	Eppendorf, Hamburg
Centrifuge rotor for plates	M-20, 75003624	Thermo Fisher Scientific, Waltham, USA
Centrifuge rotor for swinging buckets	TX-400, 75003629	Thermo Fisher Scientific, Waltham, USA
Electrophoresis chamber for agarose gels	Owl Easycast B2	Thermo Fisher Scientific, Waltham, USA
Electrophoresis chamber for agarose gels	PerfectBlue™ gel system, Mini L	Peqlab, VWR International GmbH, Darmstadt
Enzyme-linked immunospot (ELISpot) plate reader	CTL-Immunospot® S6 Ultra-UV	CTL Europe, Bonn
Flow cytometer	fluorescence activated cell scanner (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
Freezer (-20°C)	Mediline	Liebherr, Biberach an der Riss
Freezer (-20°C)		Bosch, Stuttgart
Freezer (-80°C)	U725 Innova	New Brunswick, Nürtingen
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, 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

Materials and Methods

Equipment	Product name	Company
Laminar flow hood	Maxisafe 2020	Thermo Fisher Scientific, Waltham, 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
Magnet for MACS	Quadro MACS	Miltenyi Biotech, Bergisch Gladbach
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, USA
Nano Drop	ND-8000	Thermo Fisher Scientific, Waltham, USA
polymerase chain reaction (PCR) cycler	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	2µl, 20µl, 200µl, 1000µl, 50µl-multichannel and 300µl-multichannel Finnpiptette F2	Thermo Fisher Scientific, Waltham, 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
Power supply	EPS3500	Pharmacia, Uppsala, Sweden
Power supply	MP 250V	MS Major Science, Saratoga, CA, USA
Refrigerator (4°C)	Mediline	Liebherr, Biberach an der Riss
Refrigerator (4°C)		Bosch, Stuttgart
Scale	Kern EG 4200-2NM	Kern & Sohn, Balingen
Surgical tweezer and scissors		Dimedda, Tuttlingen
Thermomixer	Thermomixer compact	Eppendorf, Hamburg
Vacuum pump	N86KT.18	KNF Neuberger, Freiburg
Vortexer	Vortex-Genie 2	Scientific Industries, Bohemia, USA
Water bath		GFL, Burgwedel

3.1.2 Consumables

Product	Company
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 (100mm x 20mm)	TPP, Trasadingen, Switzerland
Cell culture flask (25cm ² , 75cm ² , 125cm ²)	TPP, Trasadingen, Switzerland
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
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, USA
Cryogenic tubes Nalgene™	Thermo Fisher Scientific, Waltham, USA
FACS tubes (5ml Polystyrene round-bottom tube)	BD, Franklin Lakes, NJ, USA
Gloves	Microflex, Reno, NV, USA
Leucosep™ tubes Greiner Bio-One™	Thermo Fisher Scientific, Waltham, USA
MACS Separation Column LS	Miltenyi Biotec, Bergisch Gladbach
MultiScreen®-HA, 96-well plate, MAHAS4510	Merck Millipore, Cork, Ireland
Nucleo Counter Via1-Cassettes™	ChemoMetec, Allerod, Denmark
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
Reaction tubes (0.2ml, 0.5ml, 1.5ml and 2ml)	Starlab, Hamburg
Reaction tubes, black, flip cap (1.5ml)	NeoLab, Heidelberg
Scalpel	Feather, Osaka, Japan
Syringe (20ml, 50ml BD Plastipak Luer-Lok™)	BD, Drogheda, Ireland
Syringe filter (pore size 0.22µm)	TPP, Trasadingen, Switzerland
Test tube (15ml and 50ml)	nerbe plus GmbH, Winsen/Luhe
Vacuum Filter (bottle top, pore size 0.22µm)	TPP, Trasadingen, Switzerland

3.1.3 Chemicals and biological reagents

Product	Catalog number	Company
6x DNA Loading Dye	R0611	Thermo Fisher Scientific, Waltham, USA
Agarose	A8963	AppliChem GmbH, Darmstadt
Albumin, from bovine serum albumin (BSA)	A9418	Sigma-Aldrich, Taufkirchen
Ammonium chloride (NH ₄ Cl)	P726.1	Roth, Karlsruhe
Beta-2-(β ₂) microglobulin	153903	MP Biomedicals, Illkirch, France

Materials and Methods

Product	Catalog number	Company
Bovine serum albumin (BSA)	A9418	Sigma-Aldrich, Taufkirchen
Carboxyfluorescein succinimidyl ester (CFSE)	C1157	Invitrogen, Carlsbad, CA, USA
cytomegalovirus, Epstein-Barr virus, influenza virus (CEF) Peptide Pool HLA class I	PA-CEF-001	PANATecs - a brand of Protagen Protein Services GmbH, Heilbronn
Citric acid	X863.2	Roth, Karlsruhe
Concanavalin A (ConA) from <i>C. ensiformis</i>	C5275	Sigma-Aldrich, Taufkirchen
Dimethyl sulfoxide (DMSO)	D8418	Sigma-Aldrich, Taufkirchen
Distilled water deoxyribonuclease (DNase)/ribonuclease (RNase) free	821932	MP Biomedicals, Illkirch, France
Ethanol (absolute)		Sigma-Aldrich, Taufkirchen
Ethylenediaminetetraacetic acid (EDTA)	E6758	Sigma-Aldrich, Taufkirchen
Far Red (FR) (dimethyldodecylamine oxide-succinimidyl ester)	C34564	Invitrogen, Carlsbad, CA, USA
Ficoll-Paque™ PLUS	GE17-1440-03	Sigma-Aldrich, Taufkirchen
Formaldehyde, 37% in aqueous solution	0493-500ml	VWR International, Fontenay-sous-Bois, France
GelRed® Nucleic Acid Gel Stain in water	41003	Biotium, Fremont, USA
Hydrogen chloride (HCl)	30024.29	VWR International, Fontenay-sous-Bois, France
Ionomycin	10634	Sigma-Aldrich, Taufkirchen
Isopropanol		Sigma-Aldrich, Taufkirchen
NBT/BCIP-plus substrate for ELISpot	MAB 3650-10	Mabtech, Nacka Strand, Sweden
Paraformaldehyde (PFA)	335.3	Roth, Karlsruhe
Phorbol myristate acetate (PMA)	P8139	Sigma-Aldrich, Taufkirchen
Potassium bicarbonate (KHCO ₃)	X887.1	Roth, Karlsruhe
Potassium chloride (KCl)	6781.1	Roth, Karlsruhe
Potassium dihydrogen phosphate (KH ₂ PO ₄)	3904.1	Roth, Karlsruhe
Sodium acetate (C ₂ H ₃ NaO ₂)	6773.2	Roth, Karlsruhe
Sodium azide (NaN ₃)	A1430,0100	AppliChem GmbH, Darmstadt
Sodium chloride (NaCl)	10428420	Thermo Fisher Scientific, Waltham, USA
Sodium hydroxide (NaOH)	P031.2	Roth, Karlsruhe
Sodium phosphate dibasic dihydrate (Na ₂ HPO ₄ ·2H ₂ O)	12694947	Acros organics, Thermo Fisher Scientific Geel, Belgium
Streptavidin-Alkaline Phosphatase	3310-10	Mabtech, Nacka Strand, Sweden
Tris(hydroxymethyl)amino-methane (TRIS)	167620010	Acros organics, Thermo Fisher Scientific Geel, Belgium
Trypan blue stain (0.4%)	T10282	Thermo Fisher Scientific, Waltham, USA
Trypsin/EDTA (0.04%/0.03%)	C-41000	PromoCell GmbH, Heidelberg
Tween20 (Polysorbat 20)	Tween201	MP Biomedicals, Illkirch, France
Zombie Aqua Fixable Viability Dye	423101	BioLegend, San Diego, CA, USA

3.1.4 Buffers and solutions

Name	Ingredients
10x phosphate-buffered saline (PBS)	1.37M NaCl 27mM KCl 100mM Na ₂ HPO ₄ x 2H ₂ O 20mM KH ₂ PO ₄ pH 7.3
50x Tris base, acetic acid and EDTA (TAE) buffer	484g Tris 41g C ₂ H ₃ NaO ₂ 37g EDTA pH 7.8 ad. 2L ddH ₂ O
ACK lysis buffer	0.829g NH ₄ Cl (155mM) 0.1g KHCO ₃ (10mM) 0.38mg EDTA (0.1mM) pH 7.2-7.4 ad. 100mL ddH ₂ O
Elution buffer	50% 0.263M (v/v) citric acid 50% 0.123M (v/v) Na ₂ HPO ₄ x 2H ₂ O pH 2.9 or 3.1
FACS buffer	1x PBS 2% FCS 0.1% (v/v) NaN ₃
Fixation buffer	1xPBS 1% PFA
Flow cytometry fix buffer	1xPBS 1% FCS 2.5% formaldehyde
MACS buffer	1x PBS 0.5% (v/v) FCS 2mM EDTA
Staining buffer	1xPBS 0.1% (w/v) BSA 0.1% (v/v) NaN ₃
Tween PBS (TPBS)	1xPBS 0.05% (v/v)Tween20

3.1.5 Cell lines

Name	Description	HPV-status	Culture medium	Reference	Source
1341-8346	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA
BSM	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA
E481324	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA
EA	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA

Name	Description	HPV-status	Culture medium	Reference	Source
FH8	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA
LKT3	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA
WT100BIS	human, B-LCL, suspension	negative	B-LCL medium		IHWG Cell Bank, Seattle, WA, USA
C33A	human, cervical, adherent	negative	CXCA-D	(Auersperg, 1964)	Kindly provided by Felix Hoppe-Seyler, F065 DKFZ, Heidelberg
CaSki	human, cervical, adherent	HPV16 positive	CXCA-R	(Pattillo et al., 1977)	
MRI-H-186	human, cervical, adherent	HPV16 positive	CXCA-D	(Schmitt and Pawlita, 2011)	CLS, Eppelheim
UM-SCC 104	human, HNSCC, adherent	HPV16 positive	UM-SCC104 medium	(Tang et al., 2012)	EMD Millipore Corporation, Cat. # SCC072

3.1.6 Blood samples and buffy coats

Blood samples were taken from healthy donors after their written informed consent. Sampling and use of blood samples were in accordance with the Institutional Review Board at the DKFZ and the University of Heidelberg, Heidelberg, Germany. Blood buffy coats of anonymous, healthy donors were obtained from the German Red Cross (DRK) blood transfusion service Mannheim through the blood bank Institut für Klinische Transfusionsmedizin und Zelltherapie (IKTZ) Heidelberg.

3.1.7 Cell culture basal media and supplements

Product	Catalog number	Company
2-Mercaptoethanol	31350-010	Life Technologies Europe BV, Bleiswijk, Netherlands
Dulbecco's Modified Eagle Medium (DMEM), high glucose (-hi)	D5671	Sigma-Aldrich, Taufkirchen
Dulbecco's Modified Eagle Medium (DMEM), low glucose (-lo)	D5546	Sigma-Aldrich, Taufkirchen
Fetal Calf Serum (FCS)	10270	Thermo Fisher Scientific, Waltham, USA
GM-CSF, recombinant human	215-GM	Bio-technie, R&D Systems, Inc., Minneapolis, USA
HEPES	11560496	Gibco® by Life Technologies Europe BV, Bleiswijk, Netherlands
Human serum, type AB	H4522	Sigma-Aldrich, Taufkirchen
IL-2, recombinant human	200-02	PeptoTech, Rocky Hill, NJ, USA
IL-4, recombinant human	204-IL-050	Bio-technie, R&D Systems, Inc., Minneapolis, USA
IL-6, recombinant human	206-IL-050	Bio-technie, R&D Systems, Inc., Minneapolis, USA

Product	Catalog number	Company
IL-7, recombinant human	207-IL	PeproTech, Rocky Hill, NJ, USA
IL-1 β , recombinant human	201-LB-100	Bio-technie, R&D Systems, Inc., Minneapolis, USA
L-Glutamine	25030024	Thermo Fisher Scientific, Waltham, USA
Lipopolysaccharide (LPS)	tlrl-3pelps	Invivogen, Toulouse, France
MEM Non-essential amino acid solution (MEM-NEAA, 100X)	11140-035	Gibco® by Life Technologies Europe BV, Bleiswijk, Netherlands
Penicillin/Streptomycin-Solution (P/S) 10,000U penicillin and 10mg streptomycin per ml (100X)	P0781	Sigma-Aldrich, Taufkirchen
Prostaglandin E2 (PGE ₂)	14010	Cayman Chemical, Ann Arbor, USA
Rosswell Park Memorial Institute medium 1640 (RPMI)	R0883	Sigma-Aldrich, Taufkirchen
Sodium pyruvate	25-000-CIR	Corning GmbH, Kaiserslautern
TNF α , recombinant human	210-TA-100	Bio-technie, R&D Systems, Inc., Minneapolis, USA

3.1.8 Cell culture media

Name	Ingredients
Standard	10% (v/v) FCS 2mM L-Glutamine
Standard human	10% (v/v) human serum, AB 2mM L-Glutamine 1X P/S 10mM HEPES
B-LCL medium	RPMI 15% (v/v) FCS 2mM L-Glutamine 1mM sodium pyruvate
CXCA-D	DMEM-hi Standard 1X P/S
CXCA-R	RPMI Standard 1X P/S
DC medium	DMEM-hi Standard human
ELISpot medium	RPMI 5% (v/v) FCS 2mM L-Glutamine 1X P/S 10mM HEPES 0.1mM 2-mercaptoethanol
T cell medium	RPMI-1640 Standard human 0.1mM 2-mercaptoethanol

Name	Ingredients
UM-SCC104 medium	DMEM-10 Standard 1X MEM-NEAA

3.1.9 Kits

Name	Catalog number	Company
BD Accuri C6 Plus Flow Cytometer Fluidic Kit	661393	BD Biosciences, Franklin Lakes, NJ, USA
BD Cytofix/Cytoperm™ Kit	554715	BD Biosciences, Franklin Lakes, NJ, USA
CD8 ⁺ T Cell Isolation Kit, human	130-096-495	Miltenyi Biotec, Bergisch Gladbach
peqGOLD Gel Extraction Kit	732-2777	Peqlab, VWR International GmbH, Darmstadt
QIAamp DNA Mini Kit	51304	Qiagen, Hilden
QIAquick PCR Purification Kit	28106	Qiagen, Hilden

3.1.10 Molecular Markers

Name	Catalog number	Company
GeneRuler™ Ladder Mix, 100 bp -10000 bp	SM0333	Thermo Fisher Scientific, Waltham, USA

3.1.11 Oligonucleotides

Nucleotides in italics indicate T7 (forward) and T3 (reverse) sequencing primer sequences, respectively. Nucleotides in bold stand for actual PCR primer sequences.

Name	Sequence 5' - 3'	T _m [°C]
E6_T7_for	<i>TAATACGACTCACTATAGGGCGAAACCGGTTAGTATAA</i>	72.2
E6_T3_rev	<i>ATTAACCCTCACTAAAGGGAGTATCTCCATGCATGATT</i>	74.6
E7_T7_for	<i>TAATACGACTCACTATAGGGATAATATAAGGGGTCCGGTGG</i>	73.8
E7_T3_rev	<i>ATTAACCCTCACTAAAGGGACATTTTCGTTCTCGTCATCTG</i>	77.8

3.1.12 Antibodies

3.1.12.1 Monoclonal antibodies for HLA typing

Antigen	Clone	Host	Label	Isotype	Catalog number	Company
human HLA-A2	BB7.2	mouse	FITC	IgG2b	551285	BD Biosciences, San Diego, CA, USA
human HLA-A24	22E1	mouse	FITC	IgG2b	LS-C179736	LifeSpan BioSciences, Inc., Seattle, USA
human HLA-A3	GAP.A3	mouse	APC	IgG2a	17-5754-42	eBioscience, Thermo Fisher Scientific, Waltham, USA
human HLA-B7	BB7.1	mouse	PE	IgG1	372404	BioLegend, San Diego, CA, USA
Isotype Ctrl	MOPC-173	mouse	APC	IgG2a	400220	BioLegend, San Diego, CA, USA
Isotype Ctrl	MPC-11	mouse	FITC	IgG2b	400308	BioLegend, San Diego, CA, USA
Isotype Ctrl	MOPC-21	mouse	PE	IgG1	400114	BioLegend, San Diego, CA, USA

Antigen	Clone	Host	Label	Isotype	Catalog number	Company
USA						

3.1.12.2 Monoclonal antibodies for ELISpot assays

Antigen	Clone	Host	Label	Isotype	Catalog number	Company
Human IFN- γ	7-B6-1	mouse	Biotin	IgG1	3420-6-1000	Mabtech, Nacka Strand, Sweden
Human IFN- γ	1-D1K	mouse	-	IgG1	3420-3-1000	Mabtech, Nacka Strand, Sweden

3.1.12.3 Monoclonal antibodies for analysis of cytokine production in immune cells

Antigen	Clone	Host	Label	Isotype	Catalog number	Company
human CD3	REA613	recombinant human	PE-Vio770	IgG1	130-113-140	Miltenyi Biotech GmbH, Bergisch Gladbach
human CD4	RPA-T4	mouse	FITC	IgG1	555346	BD Biosciences, San Diego, CA, USA
human CD8	RPA-T8	mouse	PerCP-Cy 5.5	IgG1	560662	BD Biosciences, San Diego, CA, USA
human IFN γ	4S.B3	mouse	APC	IgG1	502512	BioLegend, San Diego, CA, USA
human TNF α	cA2	recombinant human	APC-Vio770	IgG1	130-120-491	Miltenyi Biotech GmbH, Bergisch Gladbach
granzyme B	GB11	mouse	PE	IgG1	561142	BD Biosciences, San Diego, CA, USA

3.1.13 Peptides

3.1.13.1 Control peptides.

X: cysteine residue with coupled fluorescein; **PMID**: PubMed ID for reference

Name	Source	Protein	Region	aa sequence	HLA	PMID
HLA A1 FL	consensus sequence			YLEPAXAKY	A1	12559627
HLA A2 FL	HBV	HBcg	18-27	FLPSDXFPSV	A2	12559627
HLA A3/11 FL	consensus sequence			KVFPXALINK	A3, A11	12559627
HLA A24 FL	HIV-1	gp41	583-591	RYLKXQQLL	A24	12559627
HLA B7 FL	Human	p53	84-93	APAPAPXWPL	B7	12559627
HLA B15 FL	Human	40S ribosomal protein S15	114-122	YLGEFSXTY	B15	12559627
HLA A1 binder	consensus sequence			YLEPAIAKY	A1	8047072
HLA A2 epitope	HIV-1	Nef	137-145	LTFGWCFKL	A2	11152503
HLA A2 epitope	HTLV	TAX	11-19	LLFGYPVYV	A2	1373197
HLA A3 binder	consensus sequence			KVFPYALINK	A3	8047072
HLA A11 binder	HIV-1	Nef	73-82	QVPLRPMTYK	A3,	8047072

Name	Source	Protein	Region	aa sequence	HLA	PMID
					A11	
HLA A24 binder	HIV-1	gp41	583-591	RYLKDQQLL	A24	1373204
HLA B7 binder	Human	p53	84-93	APAPAPSWPL	B7	12559627
HLA B15 binder	Human	40S ribosomal protein S15	114-122	YLGEFSITY	B15	7806292
CEF peptide pool	Influenza A	PB1	591-599	VSDGGPNLY	A1	11792386
CEF peptide pool	Influenza A	NP	44-52	CTELKLSDY	A1	11792386
CEF peptide pool	EBV	BMLF1	259-267	GLCTLVAML	A2	11792386
CEF peptide pool	Influenza A	Matrix 1	58-66	GILGFVFTL	A2	11792386
CEF peptide pool	HCMV	pp65	495-503	NLVPMVATV	A2	11792386
CEF peptide pool	Influenza A	NP	265-273	ILRGSVAHK	A3	11792386
CEF peptide pool	EBV	BRLF1	148-156	RVRAYTYSK	A3	11792386
CEF peptide pool	EBV	EBNA3A	603-611	RLRAEAQVK	A3	11792386
CEF peptide pool	EBV	EBNA3B	416-424	IVTDFSVIK	A11	11792386
CEF peptide pool	EBV	BRLF1	134-143	ATIGTAMYK	A11	11792386
CEF peptide pool	EBV	BRLF1	28-37	DYCNVLNKEF	A24	11792386
CEF peptide pool	Influenza A	NP	91-99	KTGGPIYKR	A68	11792386
CEF peptide pool	HCMV	pp65	417-426	TPRVTGGGAM	B7	11792386
CEF peptide pool	EBV	EBNA3A	379-387	RPPIFIRRL	B7	11792386
CEF peptide pool	EBV	EBNA3A	158-166	QAKWRLQTL	B8	11792386
CEF peptide pool	EBV	EBNA3A	325-333	FLRGRAYGL	B8	11792386
CEF peptide pool	EBV	BZLF1	190-197	RAKFKQLL	B8	11792386
CEF peptide pool	Influenza A	NP	380-388	ELRSRYWAI	B8	11792386
CEF peptide pool	EBV	EBNA3C	258-266	RRIYDLIEL	B27	11792386
CEF peptide pool	Influenza A	NP	383-391	SRYWAIRTR	B27	11792386
CEF peptide pool	EBV	EBNA3A	458-466	YPLHEQHGM	B35	11792386
CEF peptide pool	EBV	EBNA3C	281-290	EENLLDFVRF	B44	11792386
CEF peptide pool	HCMV	pp65	512-521	EFFWDANDIY	B44	11792386

3.1.13.2 HPV16 E6 and E7 peptides

All HPV16 E6- and E7-peptides used in the course of this thesis are listed in Supplementary Table S1 in the Annex.

3.1.14 Software

Name	Company/source
BD Accuri™ C6 Software	BD Biosciences, Franklin Lakes, NJ, USA
BD CSampler™ Software	BD Biosciences, Franklin Lakes, NJ, USA
BD FACS Diva Software	BD Biosciences, San José, CA, USA
BIMAS	http://bimas.dcrn.nih.gov/molbio/hla_bind/
CTL ImmunoSpot 5.1.36 Professional DC	CTL Europe, Bonn
EndNote X9	Thomas Reuter, Philadelphia, PA, USA
ExPASy compute pI/Mw tool	https://web.expasy.org/compute_pi/
FlowJo V10	TreeStar, Ashland, OR, USA
Fusion	Vilber Lourmat, Eberhardzell

Name	Company/source
GIMP 2	https://gimp.org
Immune epitope database (IEDB) MHC-I binding predictions	http://tools.iedb.org/mhci/
Inkscape 0.91	https://www.inkscape.org/
Mendeley 1.17.10	Mendeley Ltd., Elsevier Inc., New York, NY, USA
MHCcombine	http://mhccombine.dkfz.de/mhccombine/
MHCflurry 1.2	https://github.com/openvax/mhcflurry
MHCnuggets 2.0	https://github.com/KarchinLab/mhc-nuggets-2.0
MixMHCpred 2.0.2	https://github.com/GfellerLab/MixMHCpred
MS Office 2010 (German version)	Microsoft Corporation, Redmond, WA, USA
NanoDrop 1000 Software V3.8	Thermo Fisher Scientific, Waltham, USA
NetMHC 3.4	http://www.cbs.dtu.dk/services/NetMHC-3.4/
NetMHC 4.0	http://www.cbs.dtu.dk/services/NetMHC/
NetMHCcons 1.1	http://www.cbs.dtu.dk/services/NetMHCcons/
NetMHCpan 2.8	http://www.cbs.dtu.dk/services/NetMHCpan-2.8/
NetMHCpan 3.2	http://www.cbs.dtu.dk/services/NetMHCpan-3.0/
NetMHCpan 4.0	http://www.cbs.dtu.dk/services/NetMHCpan/
Notepad++ v7.4.1	https://notepad-plus-plus.org
NucleoView™	ChemoMetec, Allerød, Denmark
Pickpocket 1.1	http://www.cbs.dtu.dk/services/PickPocket/
PRISM® 7	GraphPad, La Jolla, CA, USA
R version 3.4.0	The R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/
R studio version 1.1.423	RStudio, Inc., Boston, USA, https://www.rstudio.com/
Seq2Logo	http://www.cbs.dtu.dk/biotools/Seq2Logo/
SigmaPlot 13	Systat Software, San José, CA, USA
SnapGene	GSL Biotech LLC, Chicago, USA
SYFPEITHI	http://www.syfpeithi.de/

3.2 Methods

3.2.1 *In silico* methods
















Different MHC class I binding prediction methods were used to predict the binding affinity of all possible HPV16 E6 and E7 peptides, including variant peptides, to the major HLA types A1 (HLA-A*01:01), A2 (HLA-A*02:01), A3 (HLA-A*03:01), A11 (HLA-A*11:01), A24 (HLA-A*24:02), B7 (HLA-B*07:02) and B15 (HLA-B*15:01). The results of the majority of the binding predictors were obtained using a new web application developed in the context of this project.

3.2.1.1 MHC class I binding and T cell epitope prediction

Prediction of MHC class I binding peptides derived from HPV16 E6 and E7 proteins was performed using 15 prediction methods. The selected predictors were NetMHC 4.0 and NetMHC 3.4 (allele-specific ANN), NetMHCpan 4.0, NetMHC 3.0, NetMHC 2.8, NetMHCflurry 1.2 and MHCnuggets 2.0 (pan-specific ANN), NetMHCcons 1.1 (consensus), PickPocket 1.1 (pan-specific SM), IEDB

recommended (allele-specific selective), IEDB consensus (consensus), IEDB SMM, IEDB SMMPBEC, SYFPEITHI and MixMHCpred 2.0.2 (allele-specific SM). These algorithms return predicted MHC binding likelihood as putative half maximal inhibitory concentration (IC₅₀) in nM, as percentile rank, corresponding to the fraction of random 8-14-mer peptides that score higher than the test peptide, or as affinity score (Table 1). Prediction results were obtained for HPV16 E6 and E7 reference protein (Table 2) and amino acid change variants (Table 3).

Table 1. Prediction methods used in this study.

Predictor	Approach	Online access	Score	Reference	Color
NetMHC 4.0	ANN (as)	DTU	IC ₅₀	(Nielsen and Andreatta, 2016)	
NetMHC 3.4	ANN (as)	DTU	IC ₅₀	(Lundegaard et al., 2008a)	
NetMHCpan 4.0	ANN (ps)	DTU	IC ₅₀	(Jurtz et al., 2017)	
NetMHCpan 3.0	ANN (ps)	DTU	IC ₅₀	(Nielsen and Andreatta, 2016)	
NetMHCpan 2.8	ANN (ps)	DTU	IC ₅₀	(Hoof et al., 2009)	
NetMHCcons 1.1	consensus	DTU	IC ₅₀	(Karosiene et al., 2012)	
PickPocket 1.1	SM (ps)	IEDB API	IC ₅₀	(Zhang et al., 2009)	
IEDB SMMPBEC	SM (as)	IEDB API	IC ₅₀	(Kim et al., 2009)	
IEDB SMM	SM (as)	IEDB API	IC ₅₀	(Peters and Sette, 2005)	
MHCflurry 1.2	ANN (ps)	GitHub	IC ₅₀	(O'Donnell et al., 2018)	
MHCnuggets 2.0	ANN (as)	GitHub	IC ₅₀	(Shao et al.)	
IEDB recommended	selective	IEDB API	p-rank	(Sidney et al., 2008b)	
IEDB consensus	consensus	IEDB API	p-rank	(Moutaftsi et al., 2006)	
MixMHCpred 2.0.2	SM (as)	GitHub	p-rank	(Bassani-Sternberg et al., 2017)	
SYFPEITHI	SM (as)	SYFPEITHI API	AU	(Rammensee et al., 1999)	

Approach: ANN: artificial neural network, (as): allele-specific, (ps): pan-specific, consensus: combination of methods, SM: scoring matrices, selective: returns IEDB consensus, NetMHC 4.0, SMM, NetMHCpan 3.0 or CombLib. The choice is based on the expected predictive performance: consensus > ANN > SMM > NetMHCpan > CombLib, CNN: convolutional neural network.

Online Access: DTU: Denmark Technical University, IEDB: Immune Epitope Database, API: application programming interface.

Score: IC₅₀: half maximal inhibitory concentration [nM], p-rank: percentile rank, AU: arbitrary units.

Table 2. Reference amino acid sequences of the HPV16 E6 and E7 proteins.

Protein	UniProtKB	Sequence
HPV16 E6	P03126	MHQKRTAMFQDPQERPR <u>R</u> KLP <u>Q</u> LCTELQTTIH <u>D</u> I <u>L</u> ECVYCKQQLLRR EVYDFAFRDLCIVYRDGN <u>N</u> PY <u>A</u> VCDKCLKFYISKISEYR <u>H</u> YCYS <u>L</u> YGT LEQQYNKPLCDLLIRCINCQKPLCP <u>E</u> EKQRHLDDKKQRFHNIRGRWTG RCMSSCRSSRTRRETQL
HPV16 E7	P03129	MHGDTPTLHEYMLDLQPETTDLYCYE <u>L</u> ND <u>S</u> SEEEDEIDGPAGQAE DRA <u>H</u> YNIVTFCC <u>K</u> CD <u>S</u> TLRLCVQSTHVDIRTLEDLLMGTLGIVCPICS QKP

UniProtKB: accession number in the UniProt knowledge base, bold and underlined: amino acids changed in protein variants.

Based on the E6 ORF the translated E6 protein would comprise 158aa. However, two in-frame ATG start codons exist and translation can be initiated from the second ATG, leading to a 151aa protein starting from the second methionine (Smotkin and Wettstein, 1986; Androphy et al., 1987).

Table 3. Amino acid changes in HPV16 E6 and E7 protein variants.

Variant	Sublineage	Amino acid changes in E6	Amino acid changes in E7
E-v1 (reference)	A1	-	-
E-v2	A2	L90V	-
E-v3		-	H51N
E-v4		L90V, E120D	L28F
E-2		R17T, L90V	-
E-Av1	A4	D32E	N29S
E-Av2		D32E, I34R	N29S
E-Av3		D32E	N29S, S63F
AA		Q21D, H85Y, L90V	-
Af-2v1	C	R17I, Q21D, H85Y	N29S
Af-2v2		R17I, Q21D, E36Q, A68G, H85Y	N29S
Af-2	B2	R17G, Q21D, H85Y	N29S
E	A3	N65S	-
Af-1	B1	R17T, Q21D, H85Y	-
NA, AA-2, AA-1	D1, D2, D3	Q21H, H85Y, L90V	-

Variant: nomenclature for amino acid changes as found in HPV16⁺ cell line samples, Sublineage: as defined by (Burk et al., 2013), bold: amino acid changes that were not present in HPV16⁺ cell line samples and therefore not included in the analysis.

For the predictors NetMHC 4.0, NetMHC 3.4, NetMHCpan 4.0, NetMHC 3.0, NetMHC 2.8, NetMHCcons 1.1, PickPocket 1.1, IEDB recommended, IEDB consensus IEDB SMM, IEDB SMMPMBEC and SYFPEITHI the web application MHCcombine (see below) was used to automatically retrieve output and combine it into a single comma-separated value (.csv) file. This file was converted into a sortable excel sheet. Output of other predictors was merged manually via the peptide's sequence.

To select peptides for synthesis and subsequent analysis, general binding affinity thresholds were applied as indicated by the respective algorithm ($IC_{50} \leq 500\text{nM}$, percentile rank ≤ 2 or affinity score >20 in SYFPEITHI). In the course of experiments, thresholds were lowered systematically as binders were found beyond the initial thresholds. Testing of peptides with lower predicted binding likelihood was stopped when only nonbinders were detected experimentally.

3.2.1.2 Web Application MHCcombine

To profit from the individual strengths of diverse MHC class I prediction methods, their output needs to be combined. To facilitate querying of many algorithms, a web application has been developed that systematically combines the prediction results. The user needs to enter the amino acid sequence of a protein or peptide, select HLA-allele, predictors and peptide lengths and optionally enter a filename identifier. After submitting, the tool returns a file containing comma-separated values that can be converted into a sortable excel file. The initial concept of this tool was developed by Stephanie Hoppe and Jan Winter and development was continued by Maria Bonsack. The script was programmed by Jan Winter, Christine Zeller and Cyril Mongis. The user interface, the webpage that enables the user to

communicate with the web application, was created by Maria Bonsack and integrated by Cyril Mongis. Tobias Reber helped to provide the Web Application on DKFZ webpages. This tool is available via “<http://mhccombine.dkfz.de/mhccombine/>”. The webpage provides additional information how to use, interpret and cite the application and provided predictors, and recommendations on decision thresholds.

3.2.1.3 Sequence motif analysis

Sequence motif analysis of the investigated HPV16 E6 and E7 peptides was performed to investigate the similarity of predicted, tested and MHC binding peptides to known binding motifs. Peptides with predicted binding likelihood within the general thresholds of $IC_{50} \leq 500nM$ or percentile rank ≤ 2 were considered “predicted”, peptides tested in competitive binding assays were considered “tested” and all validated ligands were considered “binders”. The motifs of 8-, 9-, 10-, and 11-mer peptides of these three groups were investigated separately. To generate sequence motifs, the web tool Seq2Logo 2.0 was used with default settings (Kullback-Leibler logo type, Hobohm1 clustering method with threshold 0.63). The motifs of the HPV16 peptides were compared to motifs of human linear epitopes from the immune epitope data base (IEDB).

3.2.2 Molecular biological methods

All centrifugations described for molecular biological methods were performed at room temperature (RT).

3.2.2.1 Isolation of genomic DNA

To extract genomic DNA from cells, the QIAamp DNA Mini Kit was used. Cells were harvested and counted as described in 3.2.3 Cell culture methods. Up to 5×10^6 cells were used for isolation of DNA. The procedure was performed according to the protocol of the manufacturer. Briefly, the cells were lysed with proteinase K in special buffer. The lysate was loaded onto a spin column containing a silica-gel membrane that provides specific DNA binding. During centrifugation the DNA is adsorbed while contaminants, such as protein and cations, pass through the column. Residuals were removed by two following washing steps using different washing buffers. The pure and concentrated DNA was eluted from the spin column by centrifugation after a short incubation in elution buffer or DNase- and RNase-free water. Concentration of isolated DNA was measured (see below) and DNA was stored at $-20^{\circ}C$.

3.2.2.2 Determination of DNA concentration

To measure the concentration of isolated DNA the spectrophotometers NanoDrop 1000 or 8000 were used. To generate a baseline of light absorbance, the eluent was measured as blank. The instrument calculated the DNA purity and concentration in ng of $1\mu l$ sample based on the absorbance at 260 and 280nm wavelength. The concentration of DNA was calculated based on the absorbances at 260nm and an extinction coefficient for DNA of $50(\text{ng}/\mu l)^{-1} \text{ cm}^{-1}$ using the Beer-Lambert law (Equation 1. Beer-

Lambert law). Purity of DNA can be identified by the ratio of absorbance as proteins would show maximum absorbance at 280nm. A ratio of absorbance $A(260\text{nm})/A(280\text{nm})$ of ~ 1.8 is generally identifying “pure” DNA.

Equation 1. Beer-Lambert law

$$\text{absorbance [A]} = \text{extinction coefficient} \left[\frac{1}{\text{M} * \text{cm}} \right] * \text{path length [cm]} * \text{concentration [M]}$$

3.2.2.3 Polymerase chain reaction

The principle of the polymerase chain reaction (PCR) was invented by Kary Mullis and Randall Saiki et al. to amplify DNA and specific DNA fragments for sequencing or proving their presence (Mullis et al., 1986). Short oligomers, called primers, are designed to complementary bind to the 3' ends of the sense and anti-sense strands of the DNA template at flanking sites of the fragment of interest. In the first step of the PCR both strands are exposed by denaturation of dsDNA, breaking of the hydrogen bonds between complementary bases, at 92-95°C for 20-30s. In the second step, lowering the temperature to 50-65°C for 20-40s allows annealing of the primers specifically to the complementary sites of the template strands. During this step, the primer-template hybrid is bound by a thermo-stable DNA polymerase. During the third step, the temperature is increased to the optimum activity temperature of the used polymerase (72-80°C) which synthesizes new DNA complementary to the template strands using free dNTPs from the reaction mix. The elongation time is dependent of the fragment size and the efficiency of the polymerase. A buffer solution provides optimal reaction conditions and Mg^{2+} ions stabilize the negatively charged phosphate backbone of DNA and dNTPs. The three steps of denaturation, annealing and extension are repeated for 20-45 cycles. In each cycle, the replicated DNA strands become new templates resulting in exponential amplification of the original sequence. Dependent on the polymerase, an initial hot start may be required to heat activate the enzyme at 92-95°C for 10min. Optionally, a final elongation can be performed at 72-80°C for 5-15min to ensure full extension of remaining single-stranded DNA. After PCR cycles the reaction chamber containing the samples is cooled to 4°C for an indefinite time until samples are removed and stored.

In this study, PCR reaction was used for sequencing of HPV16 E6 and E7 to identify variants with amino acid changes. Template DNA was used with specific primers in different reaction mixes. In negative control reactions template was substituted by DNase- and RNase-free water. PCR reaction mixes were prepared on ice and run on a pre-heated PCR cycler using different PCR programs. PCR products were assessed by gel electrophoresis (see below).

3.2.2.4 PCR for sequencing of HPV16 E6 and E7

Template DNA extracted from HPV16⁺ cell lines was used with HPV16 E6- and E7-specific primers (Table 4) to amplify the E6 and E7 genes for subsequent sequencing and determination of variants. The lists of components and the PCR program are given in Table 5 and Table 6, respectively.

Table 4. Primer pairs used for HPV16 E6- and E7-amplifying PCR.

Pair	Primers	Amplicon [bp]
E6	E6_T7_for E6_T3_rev	524
E7	E7_T7_for E7_T3_rev	505

Table 5. Components of the PCR reaction mix amplifying of HPV16 E6 and E7.

Component	Stock concentration	Final concentration	50µl
PCR buffer	10x	1x	5µl
dNTPs	10mM	200µM	4µl
MgCl ₂ solution	25mM	2.5mM	5µl
Forward primer	10µM	1µM	5µl
Reverse primer	10µM	1µM	5µl
DNA template		400ng	X µl
DNA Polymerase AmpliTaq Gold	5U/µl	1.25U	0.25µl
DNase- and RNase-free water			ad 50µl

X: The volume of DNA template varies depending on the DNA concentration.

Table 6. PCR program used for amplifying of HPV16 E6 and E7.

Step	Temperature [°C]	Duration [min]
Hot start	95	15
Denaturation	94	1
Annealing	55	1
Extension	72	2
Final elongation	72	10
Final hold	4	∞

3.2.2.5 Agarose gel electrophoresis

Agarose gel electrophoresis can be used to analyze DNA fragments. Due to the negatively charged phosphate backbone of the DNA strands, the DNA starts migrating through an agarose matrix towards the anode when an electric field is applied. Thereby, DNA fragments are separated by their length as shorter fragments migrate faster through the pores of the gel. A molecular marker that contains different DNA fragments of a defined size is co-loaded. The pore size, which can be determined by the concentration of agarose that polymerizes into the gel, influences the migration speed. To track the migration speed through the gel a loading dye can be added to the sample. To visualize the DNA fragments on the gel an intercalating dye, which is fluorescent under UV-light, is added before the gel polymerizes. The gel can be documented by taking a photograph during UV light exposure.

Agarose gel electrophoresis was performed by dissolving 1% [m/v] agarose in 1xTAE buffer by heating using a microwave. The agarose solution was gently cooled before the intercalating dye GelRed (10.000x) was added to a 1x final concentration. The gel was poured into an electrophoresis chamber (Owl Easycast B2 for 100ml gels and PerfectBlue™ gel system Mini L for 50ml gels) using differently sized separators and combs to generate the desired number of pockets. After the gel

polymerized, the chamber was filled with 1xTAE buffer until the gel was completely covered. The separators and the comb were removed to allow the pockets to fill with buffer and the gel was arranged in direction of the electric field. The pockets were loaded with 1µg DNA or PCR product dissolved in 12µl DNase- and RNase-free water containing 1x DNA Loading dye (6x TriTrack DNA loading dye). Ready-to-use GeneRuler DNA ladder mix was loaded as molecular marker. A power supply was attached to the chamber and 100V were applied for ~90min until the tracking bands of the loading dye reached the last centimeter of the gel. The gels were imaged and analyzed (Annex, Supplementary Figure S1).

3.2.2.6 Purification of DNA from agarose gels

To isolate specific DNA fragments, the respective bands in an agarose gel were excised and the DNA purified using the peqGOLD Gel Extraction. The kit uses the principle of DNA-binding silica-gel membrane columns, as described for isolation of genomic DNA. All steps were performed as described in the manufacturer's protocol.

In brief, specific bands were cut with a scalpel from the agarose gel under UV light. The slice was transferred into a reaction tube and weighed. The equal volume [V/M] of binding buffer was added and the mix was incubated for 7min at 60°C on a shaking Thermomixer until the gel completely melted. To accelerate melting, the tube was vortexed after 3min. The pH value of the mix was monitored by controlling the color of the pH indicator in the buffer. If the color changed to orange or red, pH was adjusted by adding 5µl 3M sodium acetate (pH 5) until the indicator turned yellow. If necessary, 750µl aliquots of the DNA gel mix were loaded onto PerfectBind DNA Columns in collection tubes. The columns were centrifuged for 1min at 10,000x g and the flow-through was discarded. The membrane-bound DNA was washed with 750µl CG Wash buffer and centrifuged as before. Again, the flow-through was discarded and the column dried by centrifugation. The column was placed in a reaction tube and the DNA was eluted from the column by adding 30-50µl elution buffer and centrifugation for 1min at 5,000x g. The yield was determined by measuring the DNA concentration as described before. Purified DNA fragments were stored at -20°C.

3.2.2.7 Purification of DNA from PCR products

The components of the PCR reaction mix can interfere with downstream applications such as sequencing or subsequent PCR. To purify the PCR product from chemicals, enzyme and primers contained in the PCR reaction mix, the QIAquick PCR Purification Kit was used. The principle of this kit is comparable with the DNA extraction kit. The method was performed according to manufacturer's protocol. Briefly, five volumes of PB buffer were added to the PCR sample and the pH checked by color indication. If the color changed to orange or violet the pH was adapted by adding 10µl of 3M sodium acetate (pH 5) until it turned yellow again. The sample solution was loaded on a spin column and centrifuged at 17.900x g for 1min. As the DNA was bound to the membrane, the flow through containing the contaminants was discarded. To wash the bound DNA 750µl PE buffer were

applied and the column was centrifuged again for 1min. The flow-through was discarded and the column was dried by 1min centrifugation to remove residual ethanol. The column was placed in a reaction tube and the DNA was eluted from the column by incubating the sample for 1min in 30µl elution buffer or water and finally centrifuging for 1min. The DNA concentration was measured on NanoDrop instruments as described. Purified PCR products were stored at -20°C.

3.2.2.8 Sequencing of HPV16 E6 and E7 genes

The products of PCR with template DNA from HPV16⁺ cell lines and HPV16 E6- and E7-specific primers were sequenced by GATC Biotech (Konstanz). The concentration of PCR products was set to 100ng/µl and 20µl of the samples were sent for sequencing with company-provided T7 primers.

3.2.2.9 Analysis of HPV16 E6 and E7 sequences

The nucleotide sequences for HPV16 E6 and E7 obtained from GATC Biotech (Konstanz) were *in silico* translated into protein sequences and compared to the HPV16 E6 and E7 reference sequences. If amino acid changes were detected, they were compared to the known changes in HPV16 sublineages to determine the HPV variant present in the respective HPV16⁺ cell line (Burk et al., 2013).

3.2.3 Cell culture methods

All cell culture methods were performed under sterile conditions in a laminar flow hood. Required equipment was sterilized by autoclaving or disinfected with ethanol. Cell culture solutions were sterilized by autoclaving or filtering through membrane filters with 0.22µm pore size. Consumables and basic cell culture medium and supplements were purchased sterilely packed. Cell lines were tested regularly to authenticate their identity and to rule out contamination with mycoplasma species, Squirrel monkey retrovirus (SMRV), EBV and cross-contamination with other cells (Multiplexion GmbH, Friedrichshafen). If not specifically mentioned otherwise, all centrifugations described for cell culture methods were performed for 5min at 350x g and RT.

3.2.3.1 Thawing and freezing of cells

Freezing and thawing procedures were carried out swiftly to prevent cell damage from toxic concentrations of cryo-protective additives such as DMSO.

To dilute cells, suspended in freezing medium, immediately after thawing, a 15ml tube was filled with 9ml pre-warmed cell culture medium. The cells, frozen in cryogenic tubes in a liquid nitrogen tank, were quickly thawed in a water bath heated to 37°C. Once the cell suspension had completely melted, the cells were transferred to the tube with pre-warmed medium. The cryogenic tube was washed with 1ml culture medium to collect remaining cells. The cells were re-suspended and centrifuged. The supernatant was removed and the cell pellet was re-suspended in 10ml culture medium. The suspension was centrifuged again and the supernatant was decanted. The pelleted cells were re-suspended in medium and seeded into a cell culture flask. Culture flasks were filled with 0.2ml medium per cm² surface area (5ml for 25cm², 15ml for 75cm² and 25ml for 125cm²) and a maximal

cell density of 1×10^6 cells/ml. After 12-24h or after adherent cells attached to the surface, the medium was exchanged (see below).

For freezing, a cell suspension was centrifuged and supernatant was removed to suspend the cell pellet in 1ml freezing medium. For cell lines, freezing medium consisted of the cooled respective culture medium supplemented with 10% DMSO. Peripheral blood mononuclear cells (PBMCs) isolated from fresh blood or buffy coat preparations from healthy donors were frozen in human AB serum supplemented with 10% DMSO. Freezing medium was always freshly prepared. Cells were re-suspended in a density ranging from 1×10^6 to 1×10^8 per ml and transferred into cryogenic tubes. The tubes were placed in a cell freezing container. Transferred into a -80°C freezer, the freezing container achieves consistent cooling of $-1^\circ\text{C}/\text{min}$ to prevent formation of intracellular ice crystals during cryopreservation. The frozen cryogenic tubes were transferred into a liquid nitrogen tank for long-term storage after a minimal period at -80°C of 4h.

3.2.3.2 Culturing and passaging of cells

Cell lines were cultured in a cell culture incubator providing a temperature of 37°C , a concentration of 5% CO_2 and 95% relative humidity. To contain cells, different sterile plastic consumables such as culture flasks (25cm^2 , 75cm^2 and 125cm^2) and well-plates (6-, 12-, 24-, 48- and 96-well plate formats) were used. Flasks were placed horizontally to culture adherent cells and vertically to culture suspension cells. Cells were monitored for cell layer growth and color of medium. The pH indicator in medium changes the color from pink to yellow as acidic metabolites change the pH. This reflects the consumption of nutrients in the medium. If adherent cells were not confluent when the color changed, the used medium was replaced with fresh. When adherent cells reached 80-100% confluency and when medium of suspension cells turned yellow, cultures were passaged. Passaging describes subculturing of cells to maintain or expand the cells in culture. Thus, cells are harvested and placed in fresh pre-warmed culture medium. Cell culture flasks were exchanged after three (adherent cells) to five (suspension cells) passages at the latest.

For harvesting adherent cells the used medium was removed and cells were washed with 1xPBS to remove residual medium and calcium and magnesium ions. The cell layer was treated with trypsin/EDTA (0.04%/0.03%) solution and incubated at 37°C for several minutes until all cells rounded up and detached upon tapping the flask. The enzyme reaction was stopped by adding fresh culture medium. Cells were re-suspended and transferred into a 50ml test tube. After the suspension was centrifuged, the supernatant was discarded and the pellet re-suspended in fresh medium. A subculture of 1:2 to 1:10 or a defined number of cells was seeded into fresh cell culture flasks.

Suspension cells were harvested by removing the used medium containing the cells. The suspension was transferred into a 50ml test tube and centrifuged as described for adherent cells. The cell pellet was re-suspended and a subculture of 1:2 to 1:20 or a defined number of was seeded into culture flasks containing fresh medium. To expand suspension cells, the same volume of fresh medium was added to the culture without passaging. As the suspension cells B-lymphoblastoid cell lines (B-LCLs) naturally

form aggregates, cultures were carefully shaken every other day to isolate cells and ensure supply with medium.

3.2.3.3 Counting of cells

Cells were counted using automated cell counters. The suspension of harvested cells was diluted in 1xPBS or in culture medium. To measure cell concentration with the Countess, diluted cells were stained with the same volume trypan blue stain (0.4%) and 10 μ l were loaded onto counting chamber slides. The slides were pushed into the slit of the instrument. The display showed the cells of an area of 1mm² in the middle of the slide. The view was focused and cells were automatically counted. The instrument returns the concentration of live and dead cells in the sample, calculated as cells/ml in the measured dilution. Viable cells were discriminated based on the exclusion of trypan blue stain by an intact cell membrane. The accuracy of the automated counting is dependent on the focal plane that was manually adjusted. To achieve more accurate counting results, the Nucleo Counter NC-200 was used. Diluted cells were loaded into a specialized cassette on which the fluorophores acridine orange and 4',6-diamidino-2-phenylindole (DAPI) were immobilized. In contact with cells, acridine orange as cell-permeable dye stains all cells giving a total cell count whereas DAPI, that passes membranes less efficiently, stains mainly dead cells. After loading the cassette with approximately 60 μ l cell dilution the cassette was immediately inserted and counted. The instrument returns the concentration of living and dead cells in cells/ml. After measurements with both instruments, the cell concentrations were multiplied with the respective dilution factor.

3.2.4 Cellular assays

If not specifically mentioned otherwise, all centrifugations described for cellular assays were performed for 5min at 350x g and RT.

3.2.4.1 Flow cytometry

Flow cytometry is used to measure the frequency of cells that have a specific property that is highlighted by a fluorescence signal. This feature is also called fluorescence activated cell scanning (FACS). The fluorescence can be acquired by genetic modification resulting in expression of a fluorescent protein or by labelling with a fluorescently conjugated ligand. A stained single cell solution is measured on a FACS instrument. Each cell passes one or more laser beams of different wavelengths. The light of the laser is scattered by the cells based on their properties and it excites the fluorescent dyes that emit light of a different wavelength. Forward scatter (FSC) is influenced by cell size and sideward scatter (SSC) is based on cell granularity. The emitted light is filtered and collected based on its wavelength and the amount of collected photons is used to calculate the fluorescence signal. After scanning the cells are usually discarded but the FACS principle can also be used to sort cells into tubes or plates.

To measure cells on a flow cytometer, they were fluorescently labelled as described for the specific assays and washed with FACS buffer at least once. Cells were centrifuged and the supernatant was

discarded. Washed cells were fixed in a solution of 1% PFA/PBS for 15min at 4°C. After fixation, cells were washed again and finally re-suspended in FACS buffer. If not immediately analyzed, the cells were kept at 4°C for a maximum of 2 days. For each assay, autofluorescence of cells was determined by measuring unstained cells. If necessary the voltages of photomultipliers of different fluorescence channels were adjusted based on single color staining. Based on FSC and SSC characteristics, cell subpopulations were gated and analyzed for fluorescence signals. Flow cytometry was performed on a FACS Canto II instrument with FACS Diva software provided by the DKFZ flow cytometry core facility, or on an Accuri C6 equipped with the CSampler accessory kit and respective operating software provided by the division of Chronic Inflammation and Cancer (F180). The obtained flow cytometry data was analyzed using Flow Jo V10 software.

3.2.4.2 Competition-based peptide-HLA binding assays

Peptides were synthesized with >95% purity by the GMP unit of the DKFZ (D210) using the F-moc strategy in a fully automated multiple synthesizer and the product was characterized by analytical HPLC and MS (Merrifield, 1963; Carpino and Han, 1972). After delivery of the lyophilized peptides they were stored at -20°C until they were dissolved in DMSO in a concentration of 10mg/ml, aliquoted and frozen at -80°C for long-term storage.

HPV16 E6- and E7-derived peptides were assessed for their binding affinity to selected HLA types in competitive cellular binding assays as previously described (Kessler et al., 2003, 2004). The principle of this assay is based on the competition for binding to specific HLA molecules between a test peptide and a fluorescently labeled reference peptide with known high binding affinity. The reference peptide is labelled with fluorescein isothiocyanate (FITC) coupled to a cysteine in the sequence. B-LCLs expressing the HLA type of interest were cultured and expanded and the test peptides were purchased and prepared as described. Peptides known to be ligands were chosen as positive controls and known nonbinders were negative controls.

On a U-bottom 96-well plate, control and test peptides were titrated in 1xPBS in a 1:2 dilution series in 8 steps ranging from 600-4.69µM. On a second plate 25µl of 900nM fluorescently labeled reference peptide in 1xPBS were distributed. Using a multichannel pipette, 25µl of the titrated test and control peptides were transferred from the titration plate to the second plate. This incubation plate was stored at 4°C. In the meantime, B-LCLs were harvested and counted. A minimum of 6×10^4 cells was needed per well. The required amount of cells was transferred into a 15ml test tube and washed twice with 1xPBS centrifuging at 400x g. One tube was used for a maximum of 1.5×10^7 cells. After washing, the cell pellet was loosened by tapping of the tube and the cells were rested on ice for 5-10min. To strip the cells from their bound endogenous peptides, they were treated with ice cold elution buffer with HLA specific pH (see Table 7) for exactly 90s. Instantly, the treated cells were washed twice with culture medium and centrifugation at 400x g. The cell pellet was re-suspended in culture medium supplemented with 2µg/ml β_2 -microglobulin to reconstitute the HLA class I complex. Finally, 100µl of the stripped cells were added to each wells with peptide mix and incubated over night at 4°C. The

next day, the fluorescence of cells was measured by flow cytometry using the Accuri C6 instrument. The B-LCLs were gated and mean fluorescence intensity (MFI) in the FITC channel was analyzed. Background and maximum fluorescence were measured from cells incubated without any added peptide and with only reference peptide, respectively. To determine the binding affinity of peptides, the MFI data was analyzed statistically (see Binding affinity calculation based on IC₅₀).

Table 7. HLA-alleles expressed by B-LCLs and required pH of elution buffer.

B-LCL	HLA class I alleles	Elution buffer [pH]
1341-8346	<i>A*02:01; B*07:02; C*15:01</i>	3.1
BSM	<i>A*02:01; B*15:01; C*04:01</i>	3.1 (HLA-A) and 2.9 (HLA-B)
E481324	<i>A*01:01; B*52:01; C*12:02</i>	3.1
EA	<i>A*03:01; B*07:02; C*07:02</i>	2.9 (HLA-A) and 3.1 (HLA-B)
FH8	<i>A*11:01, A*34:02; B*82:01, B*27:05; C*03:02, C*01:02</i>	3.1
LKT3	<i>A*24:02; B*54:01; C*01:02</i>	3.1

Bold: HLA-alleles of interest

3.2.4.3 Peripheral blood mononuclear cell (PBMC) isolation

Blood cells that have a round-shaped nucleus are referred to as PBMCs. They contain lymphocytic cells as well as monocytes and DCs and are therefore critical components of immunological assays. Human PBMCs are frequently isolated from different sources such as whole blood, buffy coats, buffy cones, leukapheresis products, cord blood or bone marrow.

In this study, PBMCs were isolated either from fresh human blood samples or from buffy coats purchased from the DRK blood transfusion service Mannheim via the IKTZ Heidelberg. In both cases, blood was drawn from voluntary healthy blood donors after their written informed consent. The isolation of PBMCs using density gradient centrifugation was performed under sterile conditions.

To create a density gradient, Ficoll-Paque™ PLUS was used. To prepare for PBMC isolation, Ficoll was equilibrated to RT and 15ml were distributed to leucosep tubes. The tubes were centrifuged for 1min to have the Ficoll layer below the porous membrane in the tubes. The blood sample was diluted with 1xPBS 1:1 (fresh whole blood) or 1:5 (buffy coat) and up to 35ml diluted blood was distributed to each prepared leucosep tube. Next, the tubes were centrifuged using different conditions for whole blood or buffy coats (see Table 8). After initial separation, the blood fractions formed layers from top to bottom consisting of yellow serum and opaque white PBMCs above the membrane and clear Ficoll and a pellet of granulocytes and erythrocytes below the membrane. The serum was removed as far as possible without disturbing the PBMC layer. The remaining liquid above the layer from two leucosep tubes was collected in a 50ml tube, which was filled with 1xPBS up to 50ml to perform an initial wash step. After centrifugation, the supernatant was discarded and the cell pellet was inspected. If the pellet contained erythrocytes, e.g. stemming from aggregated cells that did not pass the porous membrane, red blood cell lysis was performed. To do so, cells were treated with 5ml ACK lysis buffer for 5min. This leads to ammonium diffusing through the cell membranes and establishing an equilibrium with ammonia, generating free OH⁻ ions. These OH⁻ ions are consumed by reacting with CO₂ to HCO₃⁻ that

is exchanged against Cl⁻ through a specific membrane transporter, a feature of red blood cells. The net influx of NH₄Cl leads to swelling of these cells until they finally burst from osmotic pressure. After ACK treatment, the tube was filled up to 50ml with 1xPBS and washed twice. During these wash steps, two tubes were combined into one. Finally, one tube per donor remained. The pellet was re-suspended in 50ml 1xPBS and cells were counted. The tube was centrifuged again and the pelleted cells were immediately used or prepared for cryopreservation.

Table 8. Centrifugation conditions for fresh whole blood and buffy coat.

Centrifugation step	Whole blood	Buffy coat
1 – PBMC separation	260x g, 10min, RT, acc 9/9, dec 0/9	600x g, 25min, RT, acc 5/9, dec 0/9
2 – initial wash	260x g, 10min, RT, acc 9/9, dec 9/9	300x g, 15min, RT, acc 9/9, dec 5/9
3 – two repeated washes	260x g, 5min, RT, acc 9/9, dec 9/9	300x g, 10min, RT, acc 9/9, dec 9/9
4 – final centrifugation	260x g, 5min, RT, acc 9/9, dec 9/9	300x g, 5min, RT, acc 9/9, dec 9/9

acc: acceleration, dec: deceleration

3.2.4.4 HLA-type screening of PBMCs using flow cytometry

As buffy coats are typically not HLA-typed, flow cytometry was used to perform a quick screening of the HLA type of PBMC donors. Commercially available HLA type-specific antibodies with coupled fluorophores were purchased and stored at 4°C. Handling antibodies, all centrifugations and incubations were performed at 4°C and cold FACS buffer was used for washings. Specificity of antibodies and optimal dilutions were determined on B-LCLs of known HLA-type (Table 9).

For HLA type screening, aliquots of 1x10⁶ PBMCs per donor were frozen as described (see Thawing and freezing of cells) directly after isolation from buffy coats. After isolating PBMCs of up to 8 donors, the aliquots were thawed and the cells were prepared in a V-bottom 96-well plate, 5x10⁵ cells/well were seeded and washed. After discarding the supernatant, cells were re-suspended in staining dilutions consisting of antibody diluted in FACS buffer. Specificity of α-HLA antibody signals were controlled with unspecific isotype control antibodies coupled to the same fluorophore to exclude unspecific antibody binding via the Fc-receptor. Cells were incubated with staining dilutions for 25min in the dark. After incubation, cells were washed twice and fixed in 1% PFA in 1xPBS for 10min at 4°C in the dark. Fixed cells were washed again and finally re-suspended in 100µl FACS buffer for flow cytometry using the Accuri C6 instrument. Background fluorescence was determined on unstained cells.

Table 9. Antibody dilutions used in HLA type screenings of PBMCs isolated from buffy coats.

Specificity	Fluorophore	Isotype	Working dilution
α-HLA A2	FITC	IgG2b	1:400
α-HLA A24	FITC	IgG2b	1:400
- (control)	FITC	IgG2b	1:200
α-HLA A3	APC	IgG2a	1:200
- (control)	APC	IgG2a	1:200

α -HLAB7	PE	IgG1	1:200
- (control)	PE	IgG1	1:200

3.2.4.5 Generation of epitope-specific T cell lines

A substantial amount of T cells is needed to investigate immunogenicity of the many HPV16 E6 and E7 peptides. Therefore, T cells from PBMCs of healthy donors were expanded in peptide-specific T cell lines to allow memory T cells to develop effector function. T cells were expanded in short-term cell lines for the use in IFN γ -ELISpot assays and in long-term cell lines for the use in Vital-FR cytotoxicity assays. All cell culture steps were performed under sterile conditions as explained in the section Cell culture methods.

3.2.4.6 Generation of short-term epitope-specific T cells lines for cultured IFN γ -ELISpot assays

Short-term epitope-specific T cell lines were generated from isolated PBMCs of healthy blood donors (Figure 11). After thawing, the isolated cells were washed in T cell medium. Cells were centrifuged, the supernatant discarded and the cells re-suspended in T cell medium for counting. The suspension was set to a cell concentration of 1×10^6 cells/ml. Per test condition, one well of a 24-well plate was filled with 1ml T cell medium supplemented with 20ng recombinant human interleukin-7 (rhIL-7, final concentration in test 10ng/ml). Test epitopes were thawed and added to the wells. Next, the cell suspension was added to each well á 1ml per well per condition. The final concentrations of test epitopes was 10 μ g/ml peptides, 1:1000 diluted solvent (DMSO) served as unspecific negative control and 1:400 diluted CEF peptide pool as specific positive control. A T cell line to an immunologically foreign HLA-specific control peptide was added as epitope-specific negative control.

The plates were covered with cling film and incubated. At the third day of culture, cells were fed with 20U/ml rhIL-2. To this end, 4000U/ml rhIL-2 was supplemented to T cell medium and 10 μ l were added to each T cell line. Feeding was repeated at the seventh day of culture. If the medium turned yellow, a half-medium change was performed: T cell lines were re-suspended and centrifuged, 1ml (half) of the supernatant was discarded and replaced with 1ml fresh medium supplemented with 40U rhIL-2, and cells were re-suspended again. On the twelfth day of culture, cells were harvested and used for setting up IFN γ -ELISpot assays.

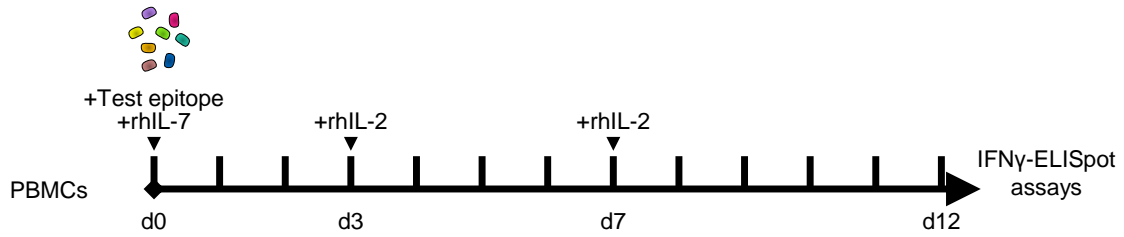


Figure 11. Schedule for generating short-term peptide-specific T cell lines. PBMCs were incubated with rhIL-7 and peptide to stimulate peptide-specific T cells. The T cells were fed twice with rhIL-2 to promote differentiation of memory T cells into effector T cells. After twelve days, T cell lines were harvested and used in IFN γ -ELISpot assays.

3.2.4.7 Generation of long-term epitope-specific T cell lines for Vital-FR cytotoxicity assays

Long-term T cell lines were generated from isolated PBMCs of healthy blood donors with the aim to expand present memory T cells with specificity to a HPV16 E6/E7-derived peptide (Figure 12). Cytotoxicity assessment was only performed with PBMCs from donors, which showed memory responses to specific epitopes in IFN γ -ELISpot assays. The generation of epitope-specific T cell lines requires repeated stimulation by autologous APCs. DCs are potent APCs that can activate naïve and memory T cells. The number of DCs in a blood sample can be increased by *in vitro* differentiation of the more plentiful monocytes resulting in so called mo-DCs or MDDCs (Feuerstein et al., 2000; Steinman, 1991). Therefore, prior to starting the T cell lines, an aliquot of donor PBMCs was thawed and used for isolation of monocytes. Monocytes were isolated by exploiting their ability to adhere to plastic.

In brief, thawed PBMCs were washed in 1xPBS, centrifuged and supernatant was discarded. The pellet was re-suspended in DC medium, cells were counted and cell density was set to 5×10^6 cells/ml. In a 6-well plate, 1×10^7 cells/well were seeded and incubated for 3h. After incubation, the supernatant of the cells was removed and attached cells were gently washed twice with DC medium. All supernatant was collected as it contained all non-adherent lymphocytes. The adherent monocytes were supplied with 2ml DC medium supplemented with 1000U/ml GM-CSF and 500U/ml rhIL-4 and cultured for 6 days with a 500 μ l medium feeding on day 3. Thereby, monocytes differentiated into immature DCs. To mature DCs a stimulating cocktail consisting of 1000U/ml TNF- α , 10ng/ml rhIL-1 β , 10ng/ml rhIL-6, 1 μ g/ml PGE2 and 1U/ml LPS was added (Jonuleit et al., 1997; Feuerstein et al., 2000). After 36-48h, the adherent matured DCs were harvested by gently scraping the bottom of the well with a cell scraper before transferring the suspension into a 50ml tube. Wells were washed twice with DC medium to collect any remaining cells. The cells were washed in serum-free DC medium, counted and set to a density of 1×10^6 cells/ml serum-free DC medium. To pulse DCs with peptide, 1ml of the cell suspension was transferred into a 15ml tube and 10 μ g/ml peptide was added. One tube per test peptide was prepared and incubated for 3h with loose tube lid. During incubation, tubes were gently shaken every hour. The pulsed DCs were washed in DC medium, centrifuged and the supernatant discarded. The cell pellet was re-suspended in T cell medium supplemented with 10ng/ml rhIL-7 to be used for stimulation of T cells in a ratio of 200-50:1 (T cells/DCs).

The generation of a T cell line starts at the harvest day of the first generation of autologous DCs which is also the starting day of a second autologous DC generation. The collected supernatant of the monocyte isolation containing the non-adherent lymphocytes was used to set up T cell lines. The cell suspension was centrifuged and supernatant was discarded. The cell pellet was re-suspended in T cell medium and cells were counted. Cells were centrifuged again and the pellet was re-suspended in T cell medium supplemented with 10ng/ml rhIL-7 to set a cell density of 1×10^7 cells/ml. In a well of a 24-well plate, 1ml T cell suspension was seeded and 1ml pulsed DCs of the first generation were added to set a T cell/DC ratio of 200:1 or better (usually up to 50:1). The T cells were incubated and fed with 50U/ml rhIL-2 every second day. This feeding was performed as half-medium change as described for short-term T cell lines. By the eighth day of T cell culture, the second generation of autologous DC were matured, pulsed and used for a second stimulation of the T cells as described. The T cell line was harvested after 14 days of culture and used for Vital-FR cytotoxicity assays.

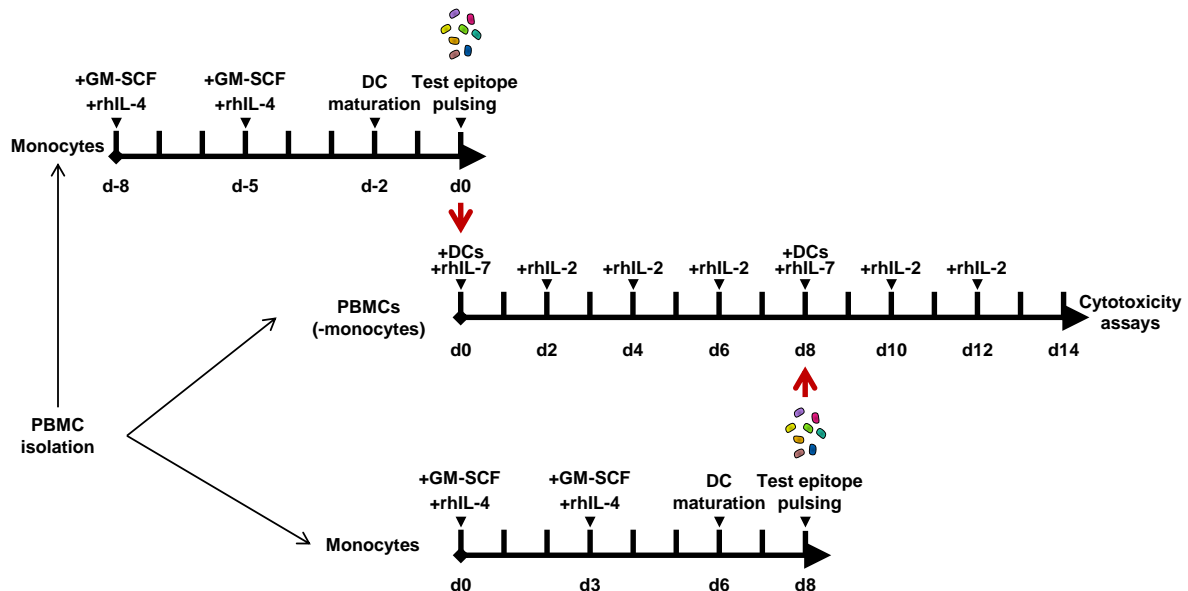


Figure 12. Schedule for generating long-term epitope-specific T cell lines. PBMCs isolated from the blood of healthy donors were used to generate long-term epitope-specific T cell lines. Prior to starting the T cell culture, autologous monocytic cells were derived by adherence. The monocytic cells were cultured in the presence of GM-CSF and rhIL-4 to promote differentiation into DCs. Two days before starting the T cell lines, DCs were matured by adding a cocktail containing TNF α , IL-1 β , IL-6, PGE2 and LPS. The matured DCs were harvested and pulsed with the respective specific epitope. Pulsed DCs were used to stimulate PBMCs in the presence of rhIL-7 for the generation of long-term epitope-specific T cells lines. The used PBMCs were depleted for naïve DCs as the same aliquot was used to derive monocytic cells to start a second DC culture at the same time. The T cells were fed with rhIL-2 every other day to promote proliferation of effector T cells. At day eight, a second stimulation of T cells was performed with matured pulsed DCs. After 14 days of cultivation, the T cell line was harvested and used in Vital-FR cytotoxicity assays.

3.2.4.8 IFN γ -ELISpot assay

The ELISpot assay is a sensitive technique that allows qualitative and quantitative assessment of immune cells secreting a certain protein. Antibodies to that target protein are coated on a membrane and capture the secreted specific target. A different antibody coupled to an enzyme binds to the very same target. Added substrate is processed by the coupled enzyme and stains the membrane with a colored precipitated product. Thereby, every cell secreting the target protein can be traced by a spot on

the membrane. An ELISpot assay assessing IFN γ secretion can detect CD4⁺ helper T cells, specifically T_h1, and CD8⁺ cytotoxic T cells that recently recognized specific epitopes via the TCR (Schoenborn and Wilson, 2007).

One day prior to setting up an ELISpot assay, the required amount of mixed cellulose ester plates were coated with 100 μ l of 2 μ g/ml IFN γ -capture antibody in sterile 1xPBS. The plate was sealed with Parafilm and incubated over night at 4°C. The next day, the antibody dilution was discarded and the membrane plates were washed three times with sterile 1xPBS. To reduce unspecific binding, the coated plates were blocked with ELISpot medium for 1h in a cell culture incubator. Epitopes were added to 1ml ELISpot medium and 100 μ l of the epitope dilution was distributed to the wells. Additionally to the HPV16 E6- and E7-derived peptides (10 μ g/ml final concentration) tested in sextuples, control conditions were added in triples. Solvent (DMSO, 1:1000 final dilution) was added as unspecific negative control and mitogen (Concanavalin A (ConA), 2 μ g/ml final concentration) was used as unspecific positive control for each short-term epitope-specific T cell line tested, to determine background activation and general IFN γ -secretion capacity. The T cell lines were re-suspended and 1x10⁵ cells were added per well. The plates were incubated for 20-26h without moving them, to prevent formation of blurred spots. The next day, plates were washed twice with 1xPBS and twice with TPBS, which removes all cells. After each washing, plates were dragged over paper towels to remove liquid. Biotinylated α -human IFN γ antibody was diluted 1:1000 in sterile PBS and 100 μ l antibody solution was distributed to each well. Plates were covered with aluminum foil and incubated for 2h at room temperature. Next, plates were washed with TPBS four times and dried. Streptavidin-ALP was diluted 1:2000 in sterile PBS and 100 μ l were dispensed per well. Again, plates were covered and incubated for 1.5h at room temperature. In the meantime, the NBT/BCIP substrate solution was equilibrated to room temperature. Directly before use, the substrate was filtered through a 0.22 μ m filter to remove precipitates. The plates were washed with TPBS four times and dragged over paper towels. The filtered substrate solution was distributed á 100 μ l/well and incubated at room temperature for 15-20min until spots developed on the membrane. The reaction was stopped by washing both sides of the membrane with tap water. The plates were air dried overnight and stored until analysis with the CTL-Immunospot S6 Ultra-UV analyzer. The operation software was used to scan and automatically count spots and to perform quality control. Automated counting allowed reducing subjectivity by setting spot morphology parameters applied to each well of the same donor. Quality control was used to exclude artefacts from counting, such as fiber, strong background staining or damaged membrane. Raw data resulting from scan, count and quality control were saved and analyzed as described below.

3.2.4.9 Magnetic-activated cell sorting (MACS)

To specifically analyze CTLs, T cell lines were sorted using MACS to isolate CD8⁺ T cells. The CD8⁺ T cell isolation Kit was used with LS columns and the QuadroMACS according to manufacturer's instructions. During the whole MACS procedure, cells were kept cold and pre-cooled (4°C) magnets and solutions were used.

The T cell lines were prepared and the cell number was determined. Cells were washed in 1xPBS and after centrifugation the supernatant was carefully removed without disturbing the pellet. Per 1×10^7 cells, the pellet was re-suspended in 40 μ l MACS buffer and 10 μ l Biotin-antibody cocktail was added. This cocktail contained biotin-conjugated monoclonal antibodies to CD4, CD15, CD16, CD19, CD34, CD36, CD56, CD123, TCR γ/δ and CD235a. The cell suspension was mixed and incubated at 4°C for 5min. After incubation, 30 μ l MACS buffer and 20 μ l CD8⁺ T cell microbead cocktail were added per 1×10^7 cells. The microbead cocktail contained magnetic microbeads conjugated to monoclonal antibodies to CD14, CD61 and biotin. Cells were mixed and incubated for 10min at 4°C. Non-target cells (CD4⁺ T cells, monocytes, neutrophils, eosinophils, B cells, stem cells, DCs, NK cells, granulocytes, γ/δ T cells and erythroid cells) were magnetically labelled, whereas CD8⁺ T cells remained untouched. The volume was filled up to 500 μ l with MACS buffer. A LS column was placed into the magnetic field of the QuadroMACS and equilibrated by rinsing with 3ml MACS buffer. The buffer was collected and discarded. The collection tube was replaced and the cell suspension was applied onto the column. The flow-through containing the enriched CD8⁺ T cells was collected. The columns were washed three times with 3ml MACS buffer and flow through was combined with the effluent. As the non-target cells were not required for any experiment the LS column was discarded. The collected suspension was centrifuged and the supernatant discarded. The CD8⁺ T cells were re-suspended in medium and counted for further use in Vital-FR cytotoxicity assays.

3.2.4.10 Vital-FR cytotoxicity assay

The Vital-FR assay is a flow cytometry based assay developed by Stanke et al. (Stanke et al., 2010). It was designed for limited sample size and increased sensitivity in analysis of antigen-specific CTL-mediated cytotoxicity. In contrast to typical cytotoxicity assays, the effector cells are co-cultured in the presence of specific and unspecific target cells that are labeled with different fluorophores. Thereby, the number of effector cells can be decreased and killing of target cells can be measured by flow cytometry.

For studying cytotoxicity to HPV16⁺ cervical carcinoma cells, the HLA-A2⁺ HPV16⁺ cell line CaSki was chosen as specific target, whereas HLA-A2⁺ HPV16⁻ C33A cells were selected as unspecific targets. One day prior to setting up the Vital-FR assay, the specific and unspecific target cells were harvested and counted. In a 50ml tube, 1×10^6 cells in 1ml plain RPMI were stained with fluorescent dyes. The specific target cells, presenting the cognate peptide, were labeled with CFSE in a concentration of 5 μ M, and unspecific target cells, presenting irrelevant peptide, were labeled with 0.25 μ M FarRed. The cells were incubated for 10min in a 37°C water bath and shaken every 3min. The tubes were filled up to 50ml with RPMI 10% FCS and centrifuged. The supernatant was discarded, cells were re-suspended in 15ml of their respective medium and cultured in T75 cell culture flasks until needed. The next day, target cells were harvested and re-suspended in T cell medium supplemented with 10U/ml rhIL-2 at a concentration of 6×10^4 cells/ml. If the number of CD8⁺ T cells was low, target cells were diluted to a minimal concentration of 4×10^4 cells/ml. The long-term epitope-

specific T cell lines were harvested and CD8⁺ T cells were isolated using MACS as described above. Isolated CD8⁺ T cells were counted and re-suspended in T cell medium supplemented with 10U/ml rhIL-2. A 1:2 dilution series of CD8⁺ T cells was prepared in order to set effector:target ratios from minimally 40:1 - 1.25:1 to maximally 100:1 – 3.13:1. The T cells were dispensed in 100µl/well of a 48-well plate. Per well, the same number of specific and unspecific target cells were added; 3000 down to 2000 cells/well. Additionally, a sample of target cell co-culture without effector cells was prepared to determine the ratio between specific and unspecific target cells after incubation for 48h. One day after setting up the Vital-FR assay, cells were carefully re-suspended by pipetting a volume of 100µl once. After 48h incubation, cells were harvested and transferred into 1.5ml tubes. The wells of the 48-well plate were washed with T cell medium and remaining cells were collected in the respective reaction tubes. In case of adherent cells, the supernatant, the PBS wash and the trypsinized cells were collected. The cells were pelleted by centrifugation and supernatant was discarded. The cells were fixed with 1% PFS in PBS for 10min at 4°C in the dark. The fixed cells were washed with FACS buffer and re-suspended in 100µl FACS buffer for flow cytometry using a FACS Canto II with FACS Diva software. Single cultures of CFSE- and Far Red-labelled cells were used for compensation. Flow cytometry data of Vital-FR assays was analyzed using Flowjo V10. First, a time gate was set that excluded the last seconds of cell acquisition in order to avoid unspecific cell counts resulting from emptying samples. Next, the target cells were gated based on FSC and SSC properties. On this population, the cell counts for CFSE⁺ and FarRed⁺ cells were determined based on respective negative single stained control samples.

3.2.4.11 Assessment of intracellular cytokine production of PBMCs

The production of cytokines is a surrogate marker for T cell activation. CTLs intracellularly store the cytotoxin granzyme B in granules and further produce IFN γ and/or TNF α upon antigen stimulation. Inhibition of the Golgi apparatus-mediated cytokine secretions keeps the cytokines in the cell. Permeabilization of the cell membrane allows fluorescently-labeled antibodies to enter the cell and to stain intracellular cytokines. Thus, activated T cells can be quantified. Additionally, antibodies to surface markers allow investigation of distinct T cell subsets, such as CD8⁺ CTLs.

To compare activation of freshly isolated and frozen PBMCs, 2x10⁶ cells per sample were stimulated. A mix of 10ng/ml PMA and 1mg/ml ionomycin is used to stimulate response in all T cells, as these compounds diffuse through the T cell membrane and directly induce intracellular signaling cascades involved in proliferation and cytokine production. Frozen PBMCs were additionally stimulated with CEF peptide pool and the E6/25-33 epitope (both at 10µg/ml final concentration) to test epitope-specific activation. PBMCs were incubated in 200µl GolgiStop solution from the BD Cytotfix/Cytoperm Kit under cell culture conditions. After 5h, cells were centrifuged at 4°C and supernatant was discarded. The cell pellet was stained in 50µl staining buffer containing 1:200 diluted Zombie Aqua dye for dead cell exclusion and the fluorescently-labeled antibodies for cell surface antigens listed in Table 10. Cells were incubated for 30min at 4°C in the dark. Subsequently, cells

were washed three times in a total volume of 200µl staining buffer with all centrifugation steps carried out at 4°C. Cells were re-suspended in 100µl fixation/permeabilization solution of the BD Cytofix/Cytoperm Kit and incubated for 10min at 4°C. Next, cells were washed three times at 4°C in a total volume of 200µl Cytofix/Cytoperm Kit wash buffer. Supernatant was discarded and the cell pellet was re-suspended in 50µl staining buffer with a mix of the cytokine antibodies (Table 10). After 30min incubation at 4°C in the dark, cells were again washed three times in 200µl wash buffer. After washing, the cell pellet was re-suspended in 200µl flow cytometry fix buffer and measured at the FACS Canto II flow cytometer.

Table 10. Antibody panel for immunophenotyping of PBMCs.

Antigen	Specificity	Fluorophore	Working dilution
Cell surface	human CD3	PE-Vio770	1:50
Cell surface	human CD4	FITC	1:50
Cell surface	human CD8	PerCP-Cy 5.5	1:50
Intracellular cytokine	human IFN γ	APC	1:50
Intracellular cytokine	human TNF α	APC-Vio770	1:50
Intracellular cytokine	human granzyme B	PE	1:50

3.2.5 Statistical analysis

3.2.5.1 Binding affinity calculation based on IC₅₀

The MFI values obtained from flow cytometry of the competitive binding assays were analyzed statistically to determine the binding affinity of HPV16 E6- and E7-derived peptides. The cells' background fluorescence was subtracted from all MFI values. The resulting value obtained for every test peptide condition was expressed relative to the maximal fluorescence to calculate the percentage of reference peptide inhibition (see Equation 2). The relative inhibition (y) was plotted against the test peptide concentration (x) (SigmaPlot V13.0, Systat Software). Non-linear regression analysis was used to determine the test peptide concentration at which 50% of the reference peptide is inhibited (Equation 3). This value is defined as half maximal inhibitory concentration (IC₅₀) and describes the binding affinity. Test peptides showing an IC₅₀ ≤ 100µM were considered HLA ligands, with high (≤ 5µM), intermediate (5 < X ≤ 15µM) and low binding affinity (15 < X ≤ 100µM). Peptides showing a binding affinity > 100µM were classified as nonbinders. The assay was performed twice for nonbinders and at least three times for binders.

Equation 2. Relative inhibition

$$\text{Inhibition [\%]} = \left(1 - \frac{\text{MFI}(\text{test}) - \text{MFI}(\text{background})}{\text{MFI}(\text{max}) - \text{MFI}(\text{background})} \right) \times 100$$

Equation 3. Non-linear regression

$$y = a + \frac{b}{1 + \left(\frac{x}{c}\right)^d}$$

a-d: constants in this equation

3.2.5.2 Performance evaluation of prediction algorithms

The experimentally determined binding affinities were used to evaluate the prediction performance of algorithms. First, tested peptides were sorted from high to low predicted binding likelihoods. Next, predictions were assessed to be true (T) or false (F) based on the experimentally validated binding affinity. Receiver operating characteristics (ROC) curves were plotted for every prediction method, HLA type and peptide length using Prism 7 software. In ROC curves, the rate of true positives (TPR, the frequency of true predicted binders among all binders) is shown on the y-axis dependent on the rate of false positives (FPR, the frequency of nonbinders falsely predicted to be binders among all nonbinders) on the x-axis. Each point on the ROC curve corresponds to a peptide with a discrete predicted binding likelihood value. The potential sensitivity and specificity of this value are represented by the associated TPR and 1-FPR, respectively. Thereby, the predictors' capacity to discriminate between binders and nonbinders was illustrated and the suitability of thresholds was estimated. A perfect prediction method would have a ROC curve with a coordinate of (0;1). Consequently, the area under the ROC curve (A_{ROC}) would be 1. The A_{ROC} is commonly used to evaluate and compare prediction methods (Bradley, 1997). For MHC binding prediction algorithms, A_{ROC} values of ≥ 0.9 were described to indicate excellent, values of 0.8-0.9 intermediate and < 0.8 poor performance (Lin et al., 2008).

By applying thresholds to the predicted binding likelihood, peptides were categorized into predicted positives (P, binders) and predicted negatives (N, nonbinders). Based on the experimental assessment the predictions were further categorized into true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). The frequencies of these fractions were used to calculate threshold-specific sensitivity, specificity, accuracy (Equation 4) and positive predictive value (PPV, Equation 5). Different thresholds were compared based on these statistics.

Equation 4. Prediction accuracy

$$\text{accuracy} = \frac{TP + TN}{P + N}$$

Equation 5. Positive predictive value (PPV)

$$\text{PPV} = \frac{TP}{TP + FP}$$

3.2.5.3 Calculation and validation of criteria-based decision thresholds

Threshold-specific sensitivity, specificity and accuracy were used to characterize and identify optimal decision thresholds for each set of predictor, HLA type and peptide length. Criteria for optimal

thresholds were defined by 1) specificity ≥ 0.66 (FPR ≤ 0.33 , respectively), 2) TPR $\geq 2 \times \text{FPR}$ and 3) highest possible sensitivity within the limits of 1) and 2). Such “criteria-based” thresholds were calculated for each set. To statistically validate sensitivity, specificity and accuracy obtained by the thresholds, bootstrapping was performed. To do so, the peptide dataset was randomly split into 2/3 training data and 1/3 test data. From training data, the optimal threshold was determined using the described criteria. This optimal threshold was applied to the test data and threshold-specific sensitivity, specificity and accuracy were derived. The procedure was repeated 100 times and the median optimal threshold as well as the confidence interval for sensitivity, specificity and accuracy was calculated based on this bootstrapping. To compare the median optimal threshold to the generally used high ($\text{IC}_{50} \leq 50\text{nM}$, percentile rank ≤ 0.5) intermediate ($\text{IC}_{50} \leq 500\text{nM}$, percentile rank ≤ 2) and low ($\text{IC}_{50} \leq 5000\text{nM}$) binding affinity thresholds, a second bootstrapping was performed. Again, the peptide dataset was randomly sampled into 1/3 of test data. The different thresholds were applied to the test data, and for each threshold sensitivity, specificity and accuracy were calculated. The differences of mean values of 100 resampling runs were compared by one-way ANOVA for repeated measures followed by Dunnett’s multiple comparisons test using Prism 7 software. The basic bootstrapping algorithm for R software was established by Diana Tichy from the Department of Biostatistics (C060). In order to analyze multiple predictors and decision thresholds at the same time, the original bootstrapping script was altered by Maria Bonsack.

3.2.5.4 Comparison of criteria-based and bootstrapping-validated thresholds

The calculation of bootstrapping-validated thresholds was only meaningful for the analysis of pooled peptide lengths as sample size of individual peptide lengths was limited. For pooled peptide length bootstrapping-validated thresholds were compared to criteria-based thresholds to investigate if prediction accuracy obtained by using criteria-based thresholds is representative for the whole statistical population, e.g. peptide sets different from the HPV16 E6-/E7-derived peptides. The significance of the difference of means was analyzed performing two-tailed Student’s *t* tests (significance, $P < 0.05$).

3.2.5.5 ELISpot analysis

IFN γ -ELISpot assays were analyzed using Microsoft Excel Software. The numbers of spot forming units (SFU) after quality control were obtained from the raw data of the CTL Immunospot Analyzer. In order to have dividable numbers, one count was added to all SFU values. All short-term T cell line samples were controlled for responding to the unspecific stimulation with the mitogen ConA. Only if a donor responded to the positive control stimulation with CEF peptide pool, the assay was regarded as successful. The mean SFU count was calculated for the sextuples stimulated with cognate peptide and the triplicates stimulated with solvent (background). Outliers were identified by Grubbs’ test (significance level $\alpha = 0.05$) and excluded from calculation of the mean SFU. First, the stimulation index (SI) was determined as the ratio between the mean SFU of cognate peptide and background.

Second, the mean SFU per 1×10^6 cells was calculated. Only if stimulation resulted in $SI \geq 2$ and ≥ 200 SFU per 1×10^6 cells, the used epitope was considered immunogenic. If peptides were found to be immunogenic in multiple donors, mean and standard deviation (SD) of SI was determined. ELISpot assay results were illustrated using Prism 7 software.

3.2.5.6 Vital-FR analysis

The FACS cell counts of specific and unspecific target cells were analyzed using Flowjo V10 software. Calculations were performed using Microsoft Excel Software. Per replicate, the frequency of specific and unspecific target cells was calculated relative to the sample without effector cells. The percentage of specific target cell killing was calculated based on the ratio between specific and unspecific target cells normalized to the co-culture without effector cells. Prism 7 software was used for graphical illustration.

4 Results

4.1 *In silico* predicted and *in vitro* validated HLA binding affinity of HPV16 E6 and E7 peptides

4.1.1 MHC class I binding prediction methods predict numerous potential HLA-binding HPV16 E6- and E7-derived peptides.

As outlined, for the development of a therapeutic vaccine it is crucial to know distinct T cell target epitopes on the target cell that can be used to trigger immune responses. In case of HPV16-associated disease, the oncoproteins E6 and E7 represent ideal target antigens. They are pivotal for inducing and maintaining the carcinogenic phenotype and are constitutively expressed in all stages of HPV16-mediated malignancy. MHC displaying of these viral antigens renders an infected cell distinct from healthy tissue. As both proteins play essential roles in driving cell transformation, their sequence is highly conserved, which reduces the possibility of immune escape once a specific T cell target is recognized (Mirabello et al., 2017). Thus, the first part of this study aimed to identify T cell target epitopes derived from HPV16 E6 and E7 proteins.

Translation of the full E6 and E7 ORFs of the HPV16 reference genome would result in 158 and 98 amino acid (aa) long proteins, respectively. However, translation of E6 can start at the second methionine and result in a 151aa protein (Smotkin and Wettstein, 1986; Androphy et al., 1987). Together, both proteins comprise 956 theoretically possible 8-, 9-, 10- and 11-mer peptides. When including amino acid change variants of E6 and E7, 1506 different peptide sequences can be derived, which would have to be tested for binding for all HLA types of interest. The experimental assessment of MHC binding affinity of many peptides is time consuming and expensive as it requires every peptide to be synthesized and repeatedly tested. In order to reduce the number of peptides assessed experimentally, *in silico* MHC class I binding prediction methods were employed. As only HLA class I-presented peptides have the potential to be CTL epitopes, the binding affinities of HPV16 E6-/E7-derived peptides to seven frequent HLA types were predicted: A1 (*A*01:01*), A2 (*A*02:01*), A3 (*A*03:01*), A11 (*A*11:01*), A24 (*A*24:02*), B7 (*B*07:02*) and B15 (*B*15:01*). A total of 15 different online available prediction algorithms were used in order to exploit the individual strengths of each method: NetMHC4.0, NetMHC 3.4, NetMHCpan 4.0, NetMHCpan 3.0, NetMHCpan 2.8, NetMHCcons 1.1, PickPocket 1.1, IEDB SMMPMBEC, IEDB SMM, MHCflurry 1.2, MHCnuggets 2.0, IEDB recommended, IEDB consensus, MixMHCpred 2.0.2 and SYFPEITHI. The algorithms differ in their prediction approach, training data and unit of binding likelihood (Table 1, in section MHC class I binding and T cell epitope prediction). Not all methods allow to predict 8-, 9-, 10- and 11-mer peptides to each HLA type. IEDB SMMPMBEC and IEDB SMM do not predict 8- and 11-mers for B15. SYFPEITHI does not allow making predictions for B15 nor any 8-mer, and prediction of peptides of 11aa length is only available for A1.

All prediction results were collected and analyzed. A summary of the predicted binding likelihood of analyzed peptides is shown in Supplementary Table S1 in the Annex. To facilitate handling the prediction output, the web application MHCcombine was developed in the lab. It allows automatic querying of selected prediction methods and returns the joined prediction results in .csv-format, which can be opened and sorted in Excel. During the course of this project, MHCcombine was modified in order to allow querying of up to 12 of the selected predictors, and it was provided online via DKFZ webpages. Data obtained by MHCcombine was used for prediction analysis.

First, the generally used decision threshold of a predicted high binding affinity (IC_{50} -value ≤ 50 nM or a percentile rank ≤ 0.5) was applied to discriminate positively and negatively predicted peptides. This resulted in very low numbers of positives (Table 11, columns “High binding affinity threshold (BAT)”). However, MixMHCpred 2.0.2, a predictor that was recently developed and trained on mass spectrometry data, was observed to predict more peptides. For example, A2 ligand prediction using MixMHCpred 2.0.2 resulted in 12 positives, whereas other predictors returned only 2-5 predicted binders (Table 11). For A1, A3 and A24, applying the high binding affinity decision threshold led to no predicted positives for the majority of predictors. Therefore, the general intermediate binding affinity threshold (IC_{50} -value ≤ 500 nM or a percentile rank ≤ 2) was used next. This uniformly resulted in more predicted peptides (Table 11, columns “Intermediate BAT”). Counting the amount of different positively predicted peptides, it was obvious that not all methods predict the same potential HLA ligands. For example, only 38 of 77 (49.4%) different predicted A1 ligands were predicted by a single predictor. By comparing the predicted binding likelihoods of individual peptides, it was clear that even highly ranked candidates of one predictor sometimes scored beyond the decision threshold for another method. For example, the peptide E7/78-86 (TLEDLLMGT) was predicted to bind A2 with a binding affinity ≤ 500 nM only by PickPocket 1.1, whereas other methods scored it ≤ 5000 nM (Annex, Supplementary Table S1). This resulted in a considerable fraction of peptides predicted to be binders by one or more predictors, but lying beyond the intermediate binding affinity threshold of other predictors. To investigate whether a less stringent threshold would show more homogenous prediction results, the general decision threshold for low binding affinity (IC_{50} -value ≤ 5000 nM) was applied. Naturally, the amount of positively predicted peptides increased further. However, the discrepancies between prediction methods remained (Table 11, columns “Low BAT”).

The number of positively predicted peptides ranged between 15 (B7) to 87 (B15) for the high, 51 (B7) to 155 (B15) for the intermediate and 207 (B7) to 444 (A2) for the low binding affinity threshold. Thus, peptide prediction did reduce the number of peptides to be tested. However, applying the low binding affinity threshold led to numbers of potential binders that can hardly be handled in validation experiments. Therefore, the intermediate binding affinity threshold was used, with a focus on peptides predicted by more than one method, to select peptides for initial HLA binding validation and epitope identification. In subsequent iterations of experimental assessment of binding affinity, the decision thresholds were lowered stepwise until only additional nonbinders were found by less stringent thresholds.

Table 11. Predicted peptides as per indicated decision thresholds.

Predictor	High BAT	Intermediate BAT	Low BAT	High BAT	Intermediate BAT	Low BAT
	A1			A2		
NetMHC 4.0	1	2	15	4	19	85
NetMHC 3.4	0	4	15	5	15	83
NetMHCpan 4.0	0	1	16	3	18	70
NetMHCpan 3.0	0	2	14	3	17	65
NetMHCpan 2.8	0	3	24	5	22	84
NetMHCcons 1.1	0	3	20	5	19	81
PickPocket 1.1	0	0	65	4	38	242
IEDB SMMPMBEC	0	3	53	3	27	226
IEDB SMM	0	5	57	3	38	258
MHCflurry 1.2	1	4	20	2	11	86
MHCnuggets 2.0	1	5	24	2	25	152
IEDB recommended	2	19	-	2	9	-
IEDB consensus	5	21	-	2	14	-
MixMHCpred 2.0.2	11	38	-	12	44	-
Different peptides	22	77	214	15	99	444
	A3			A11		
NetMHC 4.0	0	11	54	3	14	66
NetMHC 3.4	0	15	58	6	26	76
NetMHCpan 4.0	0	12	48	1	13	67
NetMHCpan 3.0	0	12	58	3	14	70
NetMHCpan 2.8	0	15	66	3	30	86
NetMHCcons 1.1	0	15	66	6	27	80
PickPocket 1.1	0	6	124	0	5	131
IEDB SMMPMBEC	0	20	173	1	37	211
IEDB SMM	0	16	180	3	39	218
MHCflurry 1.2	1	14	80	6	29	132
MHCnuggets 2.0	2	32	89	5	34	79
IEDB recommended	0	16	-	2	20	-
IEDB consensus	2	21	-	8	26	-
MixMHCpred 2.0.2	14	30	-	9	25	-
Different peptides	17	63	338	17	80	346

Results

Predictor	High BAT	Intermediate BAT	Low BAT	High BAT	Intermediate BAT	Low BAT
	A24			B7		
NetMHC 4.0	0	6	58	1	7	36
NetMHC 3.4	0	11	51	1	11	33
NetMHCpan 4.0	0	7	39	2	9	35
NetMHCpan 3.0	0	9	43	1	7	39
NetMHCpan 2.8	0	13	59	5	13	47
NetMHCcons 1.1	0	8	56	3	11	39
PickPocket 1.1	0	6	123	0	14	101
IEDB SMMPMBEC	0	3	134	0	7	98
IEDB SMM	0	11	145	1	25	135
MHCflurry 1.2	0	21	152	3	13	47
MHCnuggets 2.0	1	23	182	2	18	51
IEDB recommended	0	25	-	0	17	-
IEDB consensus	3	35	-	2	19	-
MixMHCpred 2.0.2	21	67	-	13	29	-
Different peptides	24	100	321	15	51	207
	B15					
NetMHC 4.0	3	17	106			
NetMHC 3.4	5	47	119			
NetMHCpan 4.0	3	18	110			
NetMHCpan 3.0	4	18	111			
NetMHCpan 2.8	4	33	121			
NetMHCcons 1.1	4	39	118			
PickPocket 1.1	0	2	115			
IEDB SMMPMBEC	13	47	202			
IEDB SMM	17	61	230			
MHCflurry 1.2	6	40	181			
MHCnuggets 2.0	21	64	176			
IEDB recommended	9	69	-			
IEDB consensus	59	85	-			
MixMHCpred 2.0.2	21	46	-			
Different peptides	87	155	410			

BAT: binding affinity threshold

4.1.2 Competition-based cellular binding assays identify HPV16 E6- and E7-derived ligands to investigated HLA types.

In order to validate HLA binding affinity, selected peptides were synthesized and binding affinity was determined experimentally in cellular competition-based binding assays as described (Kessler et al., 2003, 2004). B-LCL cells expressing the investigated HLA molecules were stripped of endogenous peptide by acid treatment. To reconstitute the HLA complex, β_2 -microglobulin and a mix of two peptides was supplied. The mix consisted of a fixed concentration of a fluorescently-labeled reference peptide with a known high HLA affinity and a test peptide at a range of concentrations. If the test peptide bound to the HLA in question, it competitively replaced the reference peptide in the cleft of the HLA molecule. Measuring the fluorescence of bound reference peptide resulted in decreased fluorescence intensity with increasing concentrations of test peptide. Analyzing a dilution series of test peptide allowed defining the test peptide concentration at which 50% binding of the reference peptide was replaced, referred to as experimental IC_{50} -value (Figure 13 A). Binders were grouped into peptides with strong ($IC_{50} \leq 5\mu M$), intermediate ($5\mu M < IC_{50} \leq 15\mu M$) and weak ($15\mu M < IC_{50} \leq 100\mu M$) experimental binding affinity. A nonbinder was characterized by an $IC_{50} > 100\mu M$ (Figure 13 B). Thus, a low IC_{50} value is representative of high binding affinity. However, determined IC_{50} -values greatly depend on the chosen reference peptide. High-affinity reference peptides, as used in this study, result in higher IC_{50} values than assays performed with reference peptides of lower affinity.

Using the outlined binding assay protocol, 508 HPV16 E6-/E7-derived peptides were assessed in the lab before the start of this thesis project, resulting in 224 binders and 284 nonbinders. However, progress in the development of prediction methods and inclusion of E6/E7 variants revealed new candidates for HLA binding validation and epitope identification. Hence, HLA binding of 271 peptides was additionally tested in the scope of this thesis. This resulted in the identification of 69 new binders. Figure 13 C shows the individual experimental IC_{50} results of these 69 validated HLA ligands (25.5% of tested peptides), the 202 nonbinders are listed in Table 12. Nonbinders were tested at least twice, whereas IC_{50} of binders was minimally determined three times (refer to column “n” in Figure 13 C and Table 12).

The majority of the tested peptides (58) were analyzed for binding to A1, and 20 HLA ligands were identified. Five peptides showed strong, three intermediate and twelve weak experimental binding affinities. Of 39 peptides tested for A2-binding, only 5 weak binders could be identified. Three out of 35 predicted ligands to A3 were actual binders with strong affinity. For A11, 52 peptides were analyzed for binding, and of 10 identified ligands five each showed intermediate and weak affinities. Because A24 had already been studied extensively in previous work, only 16 new candidate peptides were assessed. Surprisingly, the majority of these were true binders. Of these ten binders, five bound strongly, one intermediately and four weakly to the HLA molecule. Only six out of 22 peptides examined for B7 affinity were binders, three of them with strong, one with intermediate and two with weak binding. Of 49 tested peptides for B15, 15 were actual ligands. For three ligands a strong binding affinity was determined, two showed intermediate and the remaining ten weak binding.

Results

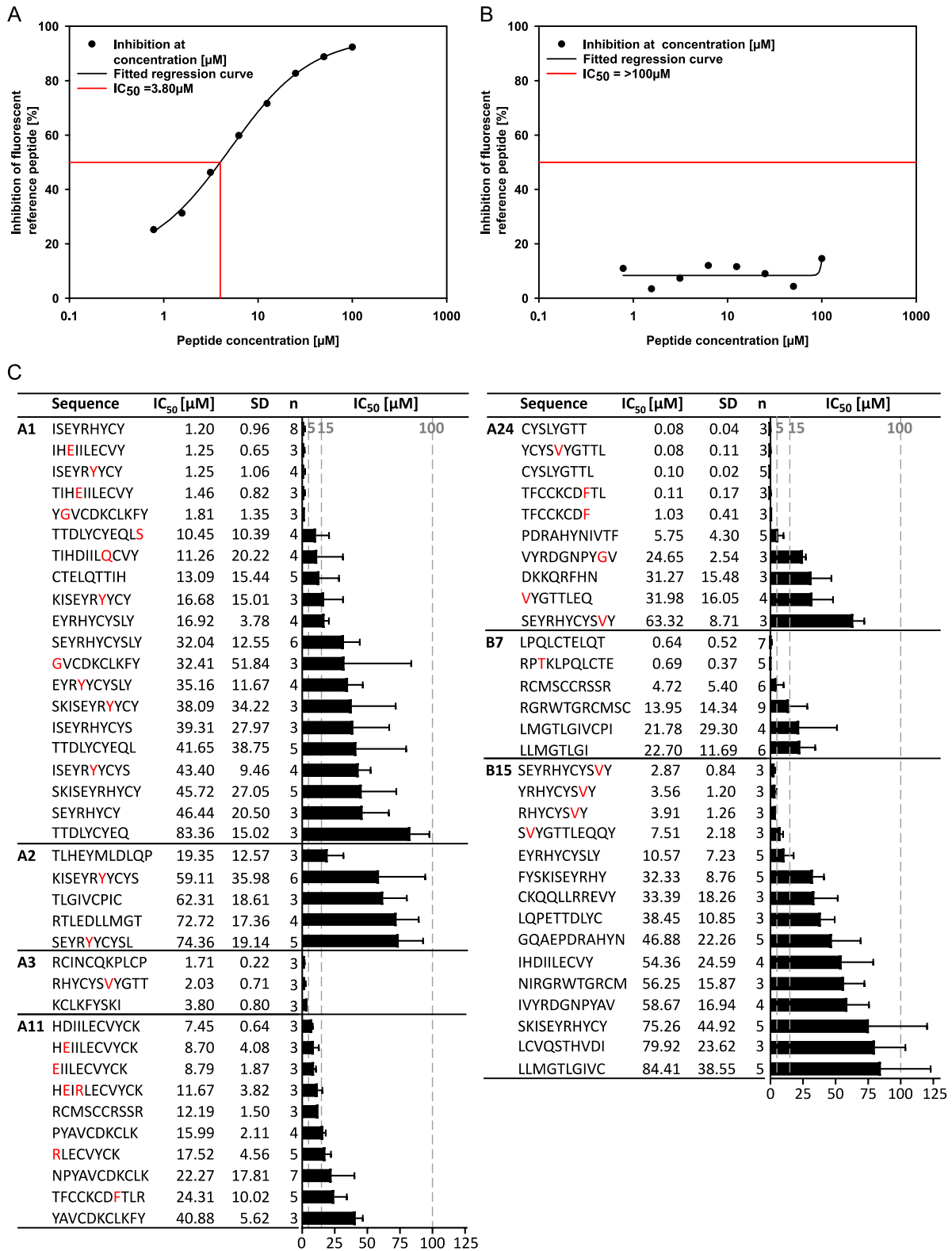


Figure 13. Competition-based binding assays identify HLA-ligands. The fluorescence of cells incubated with a labeled reference ligand and different concentrations of test peptide was measured by flow cytometry. The test peptide concentration at which 50% of the reference peptide binding was inhibited, referred to as IC₅₀ value, was calculated. Representative examples of a binder (A) and a nonbinder (B) are shown. Detailed IC₅₀ results of all binders are grouped by HLA type (C). Amino acid changes in the sequence of HPV16 E6/E7 variant-derived peptides are highlighted in red. Bar graphs represent the mean IC₅₀-values and SD. Dashed grey lines indicate different binding strength levels ($\leq 5\mu\text{M}$ =high, $5\mu\text{M} < \text{IC}_{50} \leq 15\mu\text{M}$ =intermediate, $15\mu\text{M} < \text{IC}_{50} \leq 100\mu\text{M}$ =low binding affinity).

Table 12. List of experimentally validated HPV16 E6-/E7-derived nonbinders.

Sequence	n	Sequence	n	Sequence	n
A1					
IHDII <u>LC</u> QCVY	4	SKISEYRHY	4	AFRDLCIVY	4
IVYRDGNPY	4	SLYGTTLEQQY	4	AGQAEPDRAHY	4
KFYISKISEY	4	TTDLYCYE	3	CIVYRDGNPYA	5
LCIVYRDGNPY	4	VYRDGNPY <u>G</u> V	4	CLKFYISKISEY	4
LDLQPETTDLY	4	YRHYCYSLY	5	CTELQTTIHDI	3
LQPETTDLY	4	YSKISEY <u>R</u> Y	4	DFAFRDLCIVY	4
LQPETTDLYCY	4	YSLYGTTL	3	FAFRDLCIVY	4
<u>L</u> S <u>D</u> SSEEEDEI	3	YSKISEYRHY	4	FYISKISEYRHY	2
NIRGRWTGRM	5	QAEPDRAHY	5	FYISKISEY <u>R</u> Y	4
QLLRREVDFA	4	TIHDIILECVY	4	FYISKISEY <u>R</u> Y <u>Y</u>	4
QPETTDLYCY	5	TTDLYCYEQLN	3	HDIILECVY	4
QQLLRREVY	4	YSKISEY <u>R</u> Y <u>Y</u>	2	IHDIILECVY	6
SEY <u>R</u> Y <u>Y</u> CY	4	YAVCDKCLKFY	2		
A2					
AVCDKCLKFY	3	KISEY <u>R</u> Y <u>Y</u> C	5	TELQTTIHDI	2
CKCDSTLRRCV	2	LKFYISKIS	2	TIHEIILEC	5
DIRTLEDLL	2	LMGTLGIVCP	3	TKLPQLCTEL	2
DKCLKFY	2	PQLCTELQTTI	2	TLEDLLMGT	4
DLQPETTDL	2	PTLHEYMLDL	2	TLEQQYNKPL	4
EYMLDLQPET	5	QER <u>P</u> I <u>K</u> L <u>P</u> D <u>L</u>	2	TLHEYMLDLQ	2
FQDPQER <u>P</u> T <u>K</u> L	4	QER <u>P</u> T <u>K</u> L <u>P</u> Q	2	TLRLCVQS	2
GIVCPICS	2	RLCVQSTHVDI	2	TLRLCVQST	2
HVDIRTLEDL	2	<u>R</u> Y <u>Y</u> CYS <u>V</u> YGT	2	VDIRTLEDL	2
IHDIILECV	2	SEEEDEIDGPA	2	YS <u>V</u> YGTTL	2
IILECVYCKQQ	2	STHVDIRTL	2		
<u>I</u> K <u>L</u> <u>P</u> <u>D</u> LCTEL	5	STLRRCVQST	2		
A3					
AFRDLCIVY	3	IVCPICSQKP	2	QLLRREVD	3
CTELQTTIHDI	2	IVYRDGNPYA	3	QQLLRREVY	3
CVYCKQQLL	2	KCLKFYISK	2	RGRWTGRMCMSC	2
DLLMGTLGI	2	KQRFHNIRGRW	3	RHYCYSLYGT	2
DLQPETTDLY	2	KQRHLKQRF	2	RLCVQSTHVD	2
EVYDFAFRD	2	LLIRCINCQ	3	<u>S</u> Y <u>Y</u> GTTLLEQQY	3
EVYDFAFRDL	2	LLRREVDFA	2	TLEDLLMGT	2
GTLGIVCPI	3	MLDLQPETT	2	TLRLCVQSTH	2
HNIRGRWTGR	2	QAEPDRAHY	2	TTLEQQYNKPL	2
HVDIRTLED	2	QLCTELQTT	2	YGTTLLEQQYNK	3
ISEYRHYCY	3	QLCTELQTTI	2		

Results

Sequence	n	Sequence	n	Sequence	n
A11					
AFRDLCIVYRD	3	HYCYSLYGTTL	3	SEYRHYCYSLY	5
AGQAEPDRA	2	IILECVYCKQ	2	SKISEYRHYCY	2
CDKCLKFY	3	IRCINCQK	2	SLYGTTLQY	3
CPEEKQRHLDK	3	IVCPICSQKP	2	<u>S</u> YGTTLQ	3
CTELQTTIHDI	2	KISEYRHYC	3	<u>S</u> YGTTLQY	2
CTELQTTI <u>H</u> EI	3	KISEYRHYCYS	2	TFCKCDSTLR	2
CVYCKQQL	2	KPLCPEEK	3	TLHEYMLDL	3
DEIDGPAG	2	LKFYSKISEYR	2	TTDLYCYEQLN	2
DIILECVYCKQ	2	MSCCRSSRTRR	3	TTI <u>H</u> EILE	3
FRDLCIVY	3	PAGQAEPDRA	2	TTLEQQYNKP	2
FYSKISEYR	2	RDGNPYAVCDK	2	TTLEQQYNKPL	2
GIVCPICS	2	REYDFAFR	4	VCDKCLKFY	2
GIVCPICSQKP	2	RLCVQSTHVD	2	VQSTHVDIR	3
GTTLEQQYNKP	2	RTLEDLLMGT	2	YSKISEYRH	2
A24					
CYEQL <u>S</u> DSSE	2	LMGTLGIV	3	RDGNPYAV	3
DFAFRDLCIVY	3	LYGTTLQ	3	TTLEQQYNK	3
B7					
CKQQLRREV	8	NPYAVCDKCLK	6	SSRTRRETQL	6
CPEEKQRHLD	8	QERPRKLPQL	7	TLEDLLMG	9
DPQER <u>P</u> TKL	11	QPETTDLYCY	6	YCYSLYGTTL	7
EPDRAHYNIVT	8	QYNKPLCDLL	11	YCYS <u>Y</u> GTTL	6
FAFRDLCIV	6	RGRWTGRCMS	6		
LLRREVYDFAF	9	RKLPQLCTEL	6		
B15					
AEPDRAHY	2	KQRFHNIRG	2	RLCVQSTHVD	2
AGQAEPDRAH	2	LECVYCKQQL	2	RSSRTRRETQL	4
CLKFYISKISE	2	LGIVCPICSQ	2	RTLEDLLM	2
CQKPLCPEEK	3	LQTTIHDIIL	2	STLRRCVQST	2
DFAFRDLCIV	3	LYGTTLQY	3	TLEDLLMGT	3
DIILECVY	2	MLDLQPETTD	2	TLGIVCPICS	2
DLLMGTGIV	2	PQLCTELQTT	2	TLRRCVQSTH	2
DRAHYNIVTF	3	QLCTELQTTI	2	YMLDLQPETT	2
ELQTTIHDI	2	QQLRREVYD	2	YSLYGTTL	3
HGDTPTLHEY	2	QQYNKPLCDLL	2	YS <u>Y</u> GTTL	2
IILECVYCKQ	4	QYNKPLCDLL	5		
IVCPICSQKP	2	RAHYNIVTFC	2		

bold and underlined: amino acid changes in E6-/E7-variant-derived peptides

4.1.3 MHC class I binding predictions do not match experimental binding results.

To assess how well predicted binding affinity matched reality, all binding data of HPV16 E6/E7-derived peptides was analyzed, including prior work in the light of present prediction methods. The commonly used intermediate binding affinity thresholds (IC_{50} -value ≤ 500 nM or a percentile rank ≤ 2) were applied and predicted binding was compared to experimental binding.

The complete dataset comprised 779 peptides analyzed for binding to one of the selected HLA types: 70 for A1, 156 for A2, 105 for A3, 137 for A11, 129 for A24, 55 for B7 and 127 for B15 (Figure 14 and Annex, Supplementary Table S1). For 78 peptides, binding affinities to the analyzed HLA types were reported before (Annex, Supplementary Table S2). Overall, 374 peptides were predicted to be binders by any algorithm, but only 293 peptides were actual binders. Moreover, there was only a partial overlap between predicted and experimentally validated HLA ligands. For example, the predictions indicated 55 peptides positive for A24 binding (TP+FP), but only 42 of these were truly binding (TP) and 13 were false positives (FP). Further, 24 peptides with experimentally validated binding affinity were falsely predicted to be nonbinders (FN). Fifty actual nonbinders were correctly identified to be negatives (TN). For all analyzed HPV16 E6/E7-derived peptides, 78 of 293 (26.6%) actual binders were predicted negatives and only 215 of 374 (57.5%) predicted positives were true HLA ligands.

Overall, the disparity between predicted and real HLA binding warranted a thorough prediction performance evaluation.

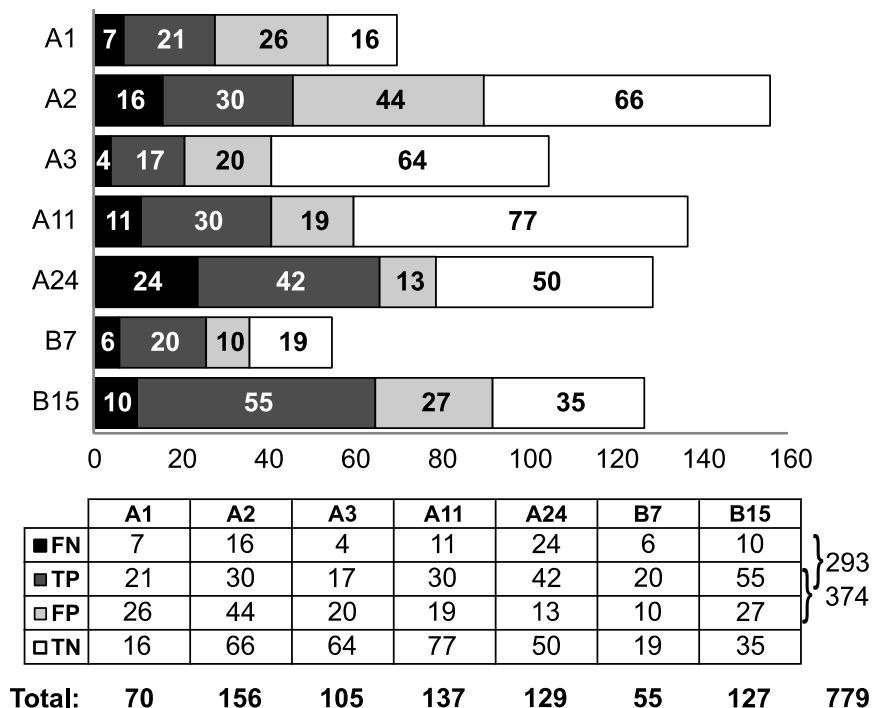


Figure 14. Evaluation of HLA class I binding predictions based on experimental binding results. Bar graphs represent the numbers of tested HPV16 E6 and E7 peptides (including aa variants) per HLA type. Tested peptides are grouped according to true and predicted binding affinity. Binders are represented by black and dark grey, nonbinders by light grey and white bars. Peptides were considered to be predicted positively (P, dark and light grey) if the predicted binding affinity of any predictor was below $IC_{50} \leq 500nM$ or percentile rank ≤ 2 , respectively. Beyond this threshold, peptides were considered to be negatively predicted (N, white and black). Prediction results were classified to be true (T) or false (F) based on experimental validation of binding. The table below the bar graph summarizes the numbers of peptides by group. To the right of the table the numbers of all binders (293), all predicted positives (374) and all tested peptides (779) are indicated.

4.2 Performance evaluation of MHC class I binding prediction methods based on experimentally validated HPV16 E6- and E7-derived ligands

4.2.1 Prediction algorithm results discriminating binders from nonbinders vary depending on HLA type and peptide length.

In order to find the most suitable prediction method, the prediction performance of the algorithms was analyzed in detail. Peptides were sorted by their predicted binding affinity and predictions were classified into true and false based on experimental binding of peptides. Each binder was considered a true positive event and each nonbinder a false positive event. The rates of true positives (TPR, also sensitivity) and false positives (FPR, also 1-specificity) were plotted as receiver operating characteristics (ROC) curves (Figure 15), allowing evaluation of the predictive strength of an algorithm independent of decision thresholds. ROC curves were analyzed for each HLA type and pooled as well as individual 8-, 9-, 10- and 11-mer peptide lengths. To compare ROC curves, the area under the curve (A_{ROC}) was calculated (Figure 16). A perfect ROC curve would show a maximal TPR (1) and a minimal FPR (0). Thus, A_{ROC} corresponding to perfect discrimination between binders and nonbinders equals 1. According to Lin et al. $A_{ROC} > 0.9$ indicate excellent performance and $A_{ROC} < 0.8$ demonstrate poor predictive capability (Lin et al., 2008).

The prediction performance analysis of ROC curves and A_{ROC} based on the dataset of HPV16 E6-/E7-derived peptides revealed that none of the analyzed predictors perfectly discriminated binders from nonbinders. Only slight differences in prediction performance between algorithms were observed when peptides of pooled lengths were considered. However, when analyzing peptide lengths individually for each HLA type, differences became more pronounced.

Overall, the prediction of 9-mers was observed to be most precise, which is reflected by reaching excellent A_{ROC} values of >0.9 (Figure 16). For 9-mers and 10-mers, the predictors performed quite uniformly. However, for the HLA types with smaller datasets, A1 and B7, the performance of 10-mer prediction was poor, with mean A_{ROC} of ~ 0.5 for A1 and ~ 0.7 for B7. Performances for 8-mer and 11-mer predictions were generally less precise but here precision depended on the analyzed HLA molecule. Best performance of 11-mer prediction was observed for A24 with quite consistent A_{ROC} values between algorithms, but an only intermediate mean A_{ROC} of ~ 0.8 . A3 showed the poorest A_{ROC} values for 11-mer predictions with a range of 0.32 (IEDB consensus) to 0.54 (IEDB SMMPMBEC and MixMHCpred 2.0.2). For 8-mer predictions, the most pronounced differences between prediction methods were seen. Multiple algorithms reached perfect discrimination (A_{ROC} 1) for A11 binding predictions. On the other hand, the A_{ROC} values for A24 ranged between 0.18 (IEDB SMMPMBEC and IEDB SMM) and 0.86 (NetMHCpan 4.0 and NetMHCpan 3.0). For other HLA types, similar but less extreme discrepancies between predictors were observed.

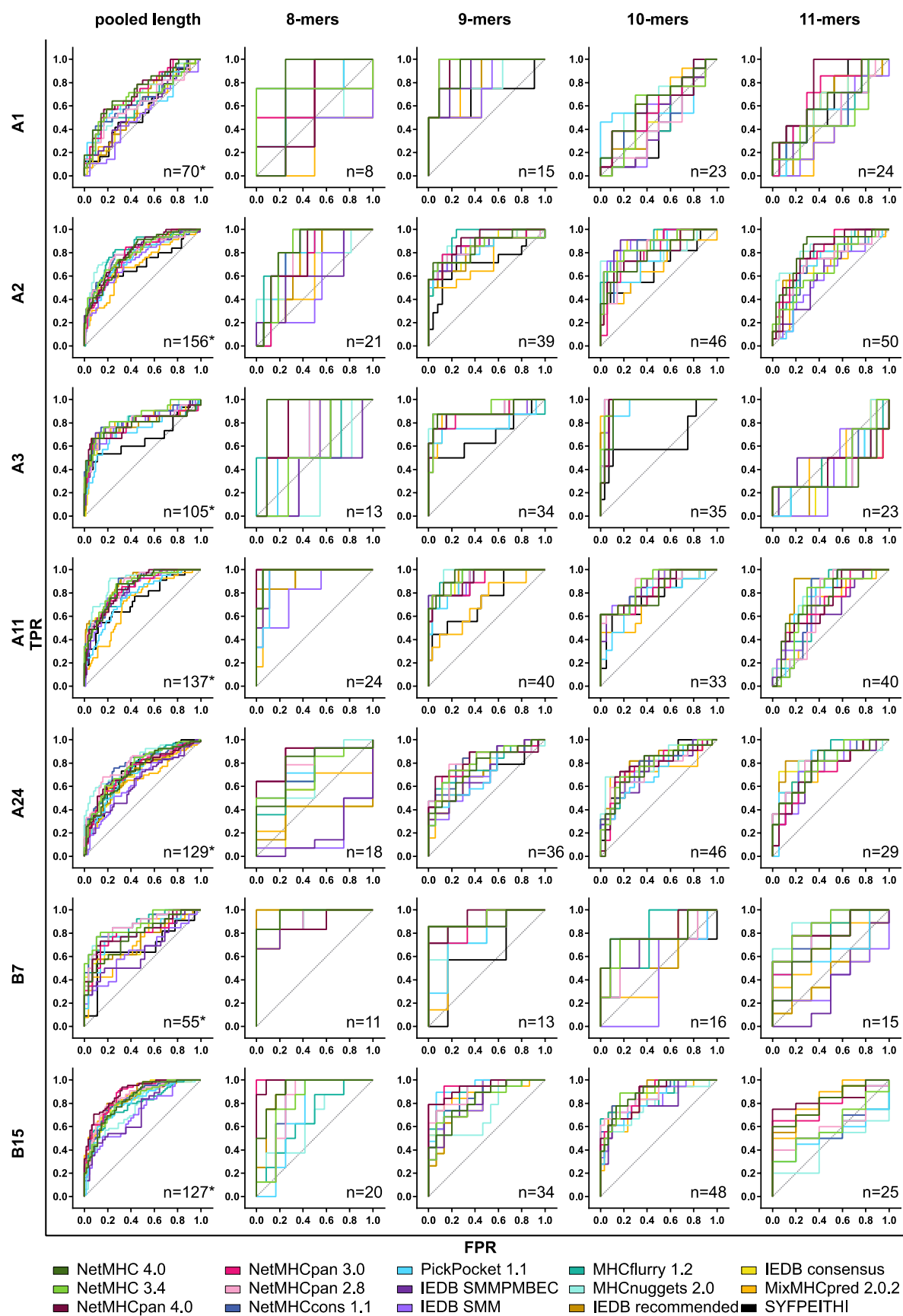


Figure 15. Receiver operating characteristics (ROC) curves of predictors for different HLA types and pooled and individual peptide lengths. The rate of true (TPR) and false (FPR) positive predictions of each predictor were analyzed for each HLA type and peptide length. The gray diagonal line represents a 50:50 chance of correct prediction. Exact numbers are given in Supplementary Table S3 in the Annex. (n) Number of peptides per group. (*) Numbers of pooled peptides differ for some predictors because prediction for some peptide lengths was not available.

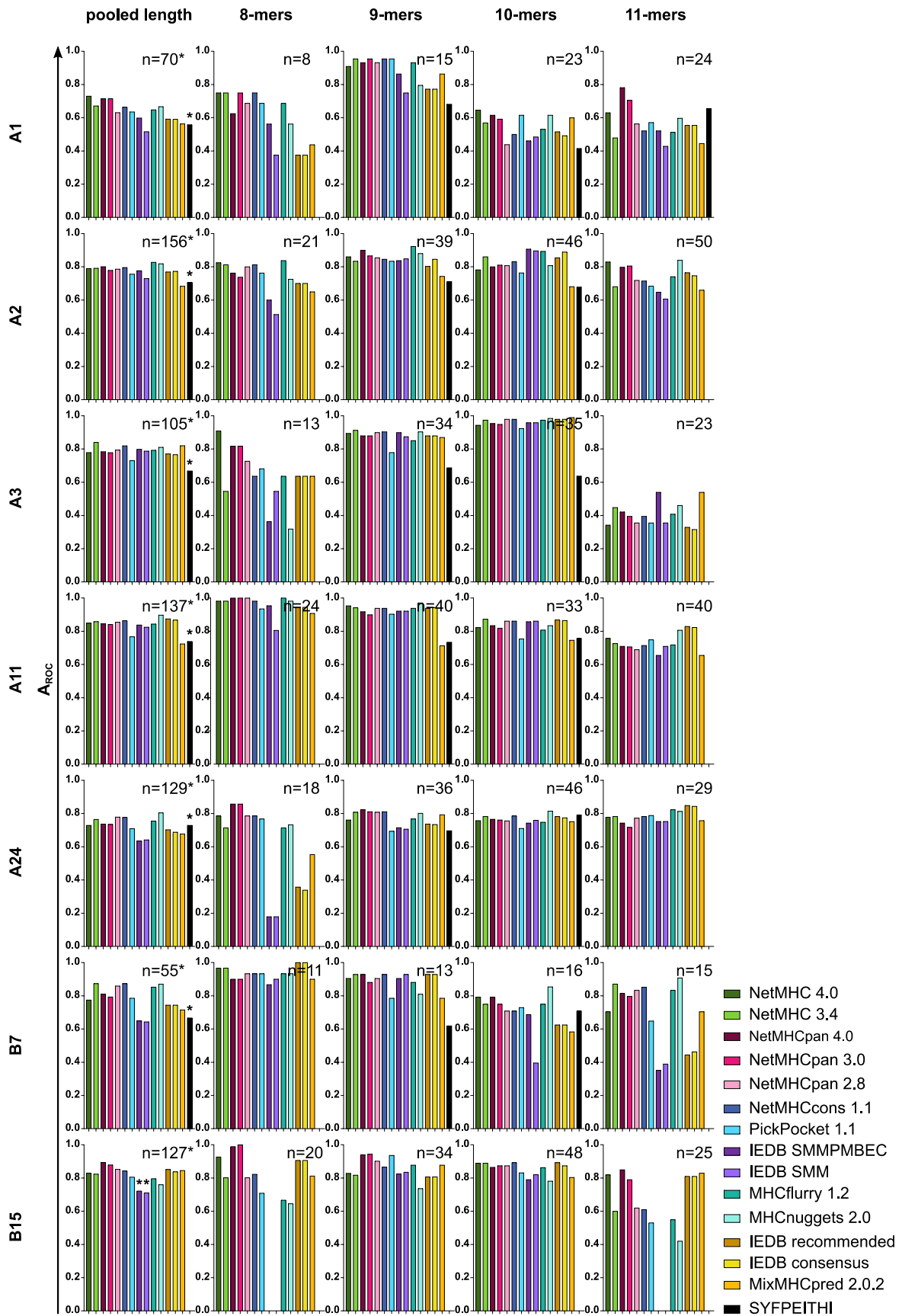


Figure 16. Area under the ROC curves (A_{ROC}) of predictors for different HLA types and pooled and individual peptide lengths. Prediction performance was measured by A_{ROC} . Perfect discrimination between binders and nonbinders results in $A_{ROC}=1$. Exact numbers are given in Supplementary Table S3 in the Annex. (n) Number of peptides per group. (*) Numbers of pooled peptides differ for some predictors because prediction for some peptide lengths was not available.

Overall, prediction methods based on artificial neural networks (ANN) using a pan-specific approach showed the best performances. Algorithms of the NetMHC family were always found among the well performing predictors. In particular for A1 and B15, predictors NetMHCpan 4.0 or NetMHCpan 3.0 were found to be the best choice. The MHCflurry 1.2 method showed the best results for binding predictions to A2, but for other HLA types it was average. MHCnuggets 2.0 can be recommended for all HLA types except A1 and B15. In contrast, it is generally not advisable to use PickPocket 1.1 for any other prediction than A1. MixMHCpred 2.0.2 only showed very convincing performance for A3. The algorithms IEDB SMMPMBEC, IEDB SMM and SYFPEITHI were mostly found among the poor performing prediction methods. However, due to partly confined datasets, comparability in the pooled peptide length analysis is limited for affected predictors (marked with asterisks in Figure 15 and Figure 16).

It became clear that none of the HLA binding predictors distinctively outperformed other methods, as no single method showed outstandingly well A_{ROC} values. However, for specific HLA types and peptide lengths, some prediction methods indeed showed superior performances. Thus, evaluation of ROC curves and A_{ROC} demonstrated that it is advisable to always select the most suitable algorithm, depending on the analyzed HLA type and peptide length.

4.2.2 Commonly used decision thresholds result in low prediction sensitivity.

The decision threshold is decisive for the statistical power of a prediction algorithm. To analyze the influence of different thresholds, accuracy (ratio of true predictions over all data points), sensitivity (TPR) and specificity (1-FPR) of the high (IC_{50} -value ≤ 50 nM or a percentile rank ≤ 0.5), intermediate (IC_{50} -value ≤ 500 nM or a percentile rank ≤ 2) and low (IC_{50} -value ≤ 5000 nM) binding affinity thresholds were compared. Predictors expressing binding likelihood as percentile rank only recommend two decision thresholds, which were here analyzed with the high and intermediate thresholds, respectively. The predictor SYFPEITHI does not recommend any decision thresholds and the scoring system differs from all other predictors. Thus, it was excluded from this analysis.

Analysis of accuracy revealed comparable results for predictors when high and intermediate binding affinity thresholds were used (Annex, Supplementary Figure S2). In contrast, the low binding affinity thresholds led to greater accuracy differences among prediction methods. The general dependence on HLA type and peptide length that had been detected by ROC curve analysis was also reflected by accuracy values. The high binding affinity threshold only yielded high accuracy for A2 8-mers and A3 11-mers. The least stringent low binding affinity thresholds were able to increase accuracy for the HLA types A1, A24, B7 and B15. Otherwise, the intermediate binding affinity threshold resulted in best accuracy values.

As accuracy is defined as the rate of all true predictions, a low capability to predict true positives may be masked by a high number of correct predictions of true negatives. Therefore, threshold-dependent sensitivity and specificity were calculated to evaluate prediction performance. Figure 17 shows the sensitivity and specificity of predictors using the high, intermediate and low binding affinity

thresholds for HLA specific binding prediction. The most favorable results would be 1 (represented by a fully colored pie) for both specificity and sensitivity, meaning that all and only binders would be predicted.

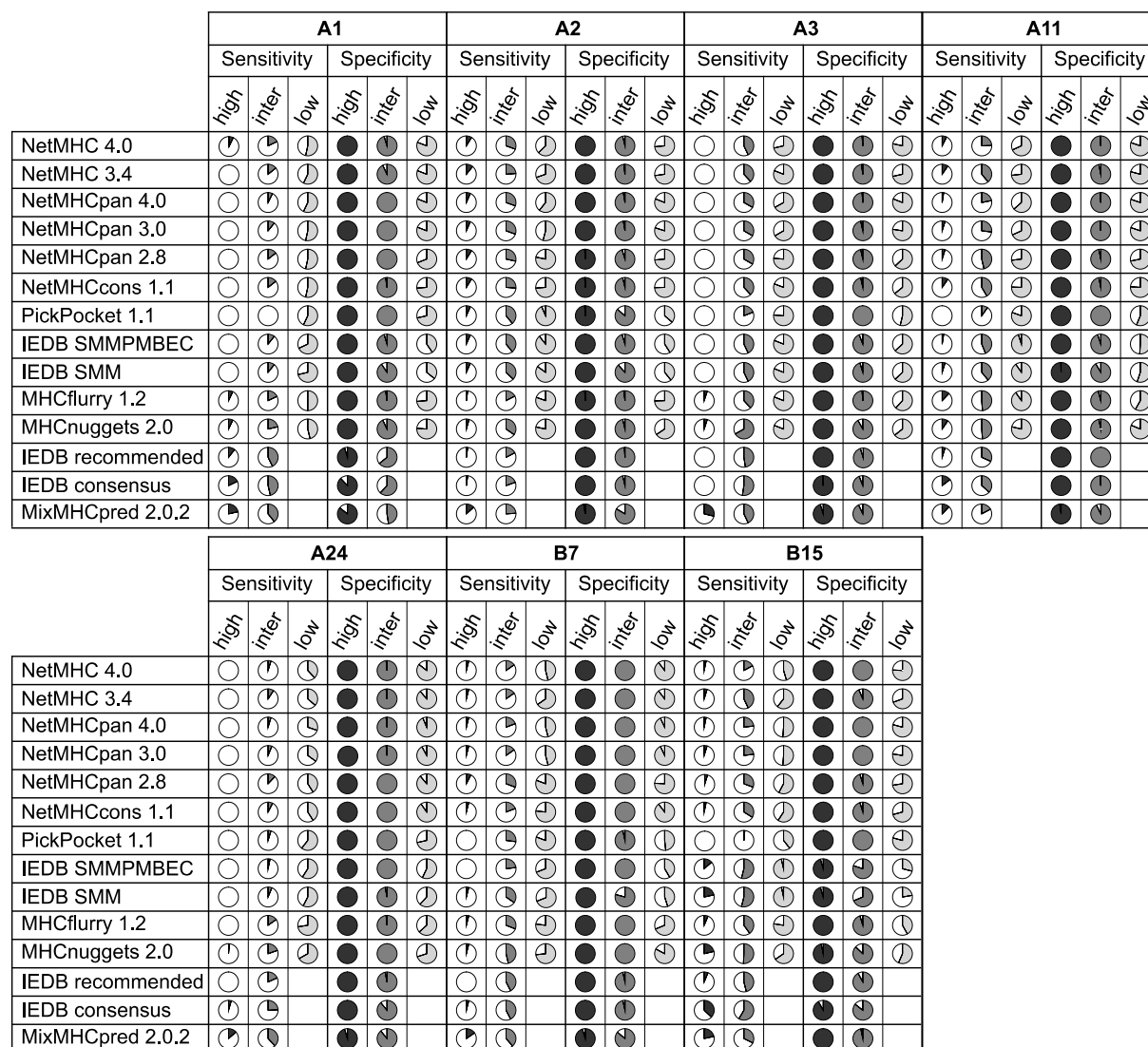


Figure 17. Threshold-dependent sensitivity and specificity of predictors for different HLA types. Grouped by HLA type and predictor, pie charts represent the sensitivity and specificity of a predictor by applying different decision thresholds (different tones of grey): high ($IC_{50} \leq 50nM$ or percentile rank ≤ 0.5), intermediate (inter, $IC_{50} \leq 500nM$ or percentile rank ≤ 2) or low ($IC_{50} \leq 5000nM$) binding likelihood. Pies completely filled with grey would indicate perfect values.

As expected, the lowest sensitivities and highest specificities were observed when applying the high binding affinity thresholds (columns “high” in Figure 17). Naturally, the use of less stringent thresholds led to an increase in sensitivity accompanied by a loss of specificity (columns “inter” and “low” in Figure 17). For the well-studied A2 allele applying the commonly used intermediate binding affinity thresholds yielded a maximum sensitivity of only 0.39 (PickPocket 1.1 and IEDB SMMPMBEC), whereas using the low binding affinity thresholds resulted in increased sensitivity of 0.54 (NetMHCpan 3.0) up to 0.93 (PickPocket 1.1). Likewise, the less stringent thresholds appeared to decrease specificity, ranging from 0.36 (PickPocket 1.1) to 0.81 (NetMHCpan 4.0) compared to 0.84

(MixMHCpred 2.0.2) to 0.99 (IEDB recommended) obtained with intermediate binding affinity thresholds. Similar results were observed for other HLA types.

Overall, sensitivity resulting from intermediate binding affinity thresholds was surprisingly low. As the low binding affinity thresholds were able to increase sensitivity without resulting in generally poor specificity, the use of less stringent decision thresholds should be considered to increase finding true HLA ligands.

4.2.3 Novel individual decision thresholds increase prediction sensitivity.

In order to find suitable thresholds, which balance the gain in sensitivity against a loss in specificity, new optimal thresholds were calculated individually for each analyzed predictor, HLA type and peptide length. To describe characteristics of an optimal decision threshold, three criteria were defined in the scope of this thesis: (1) the minimum specificity obtained should be ≥ 0.66 (equal to a maximum FPR ≤ 0.33), (2) there should always be minimally twice as many true positives than false positives ($TPR \geq 2 \times FPR$), and (3) within these first two criteria the TPR should be as high and FPR as low as possible (Annex, Supplementary Figure S3). Optimal “criteria-based” thresholds were calculated by applying these criteria.

For single length analysis of 8-, 9-, 10- and 11-mers, these thresholds and associated measures of performance (A_{ROC} , PPV, specificity and sensitivity) are listed in Supplementary Table S3 in the Annex. In general, criteria-based thresholds often resulted in values even less stringent than low binding affinity thresholds in case of predictors with a good performance as indicated by a high A_{ROC} . For poor performing prediction methods, calculation of criteria-based thresholds resulted rather in values between intermediate and low binding affinity thresholds.

As the criteria-based thresholds were calculated only based on the HPV16 E6-/E7-derived peptide dataset, it was necessary to statistically validate if the obtained performance measures are representative for any other dataset. In order to test applicability of thresholds, a bootstrapping method was performed. Bootstrapping allowed to resample the HPV dataset 100-times and to calculate the criteria-based threshold based on random 2/3 of data (“training data”). On the other third (“test data”), the calculated threshold was applied, and associated sensitivity, specificity and accuracy were derived to determine confidence intervals. The mean threshold determined of 100 bootstrappings was termed “validated threshold”. To compare the sensitivity, specificity and accuracy achieved by the individual validated thresholds to the generally used high (IC_{50} -value ≤ 50 nM or a percentile rank ≤ 0.5), intermediate (IC_{50} -value ≤ 500 nM or a percentile rank ≤ 2) and low (IC_{50} -value ≤ 5000 nM) binding affinity thresholds, a second bootstrapping was performed. Again, the HPV dataset was resampled 100-times, each threshold was applied to a random third of the HPV dataset and performance measures were calculated as confidence intervals (Figure 18).

A

A2

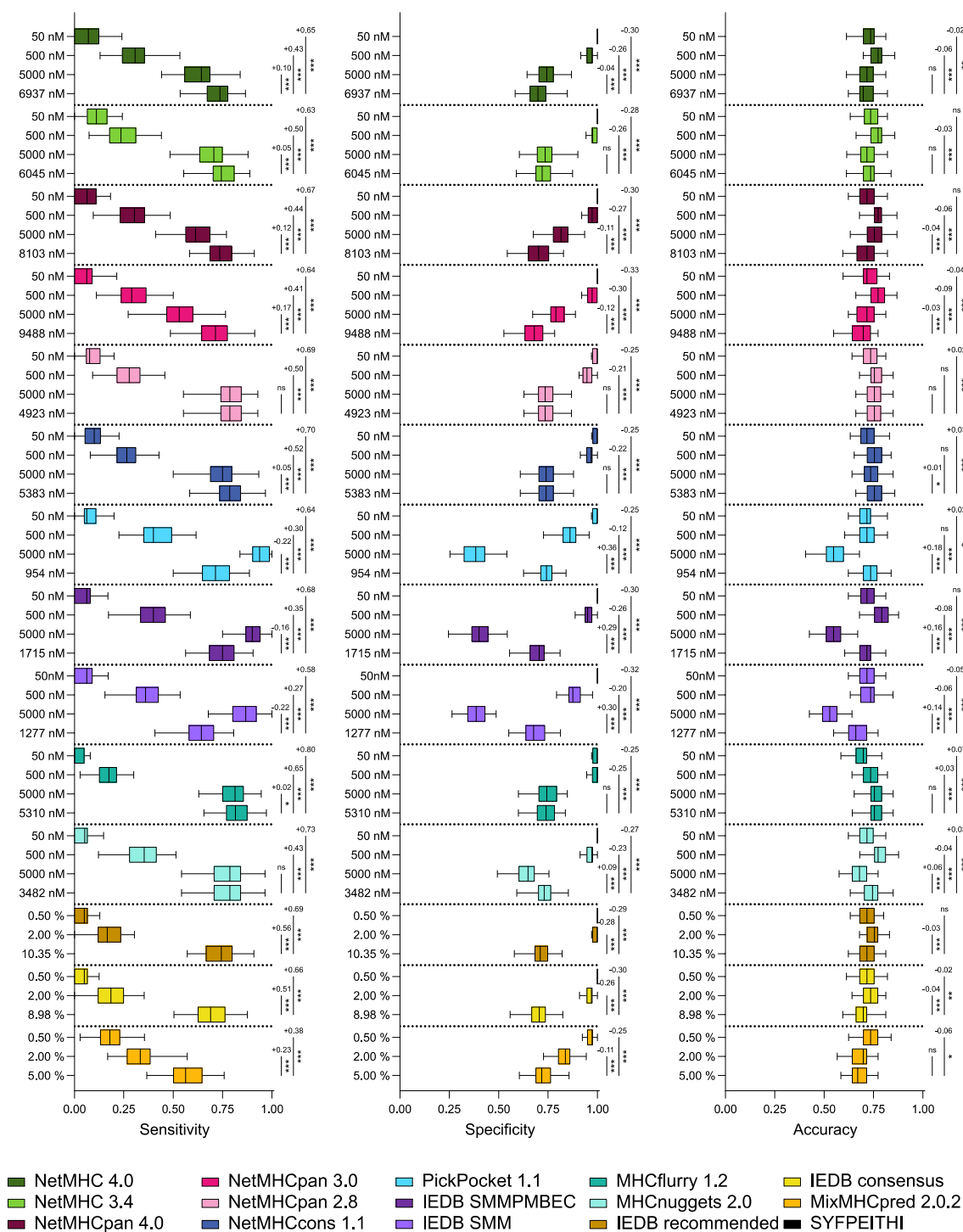


Figure 18. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended “validated” or generally used thresholds.

(Figure continues, see Figure legend on the following page)

B

A24

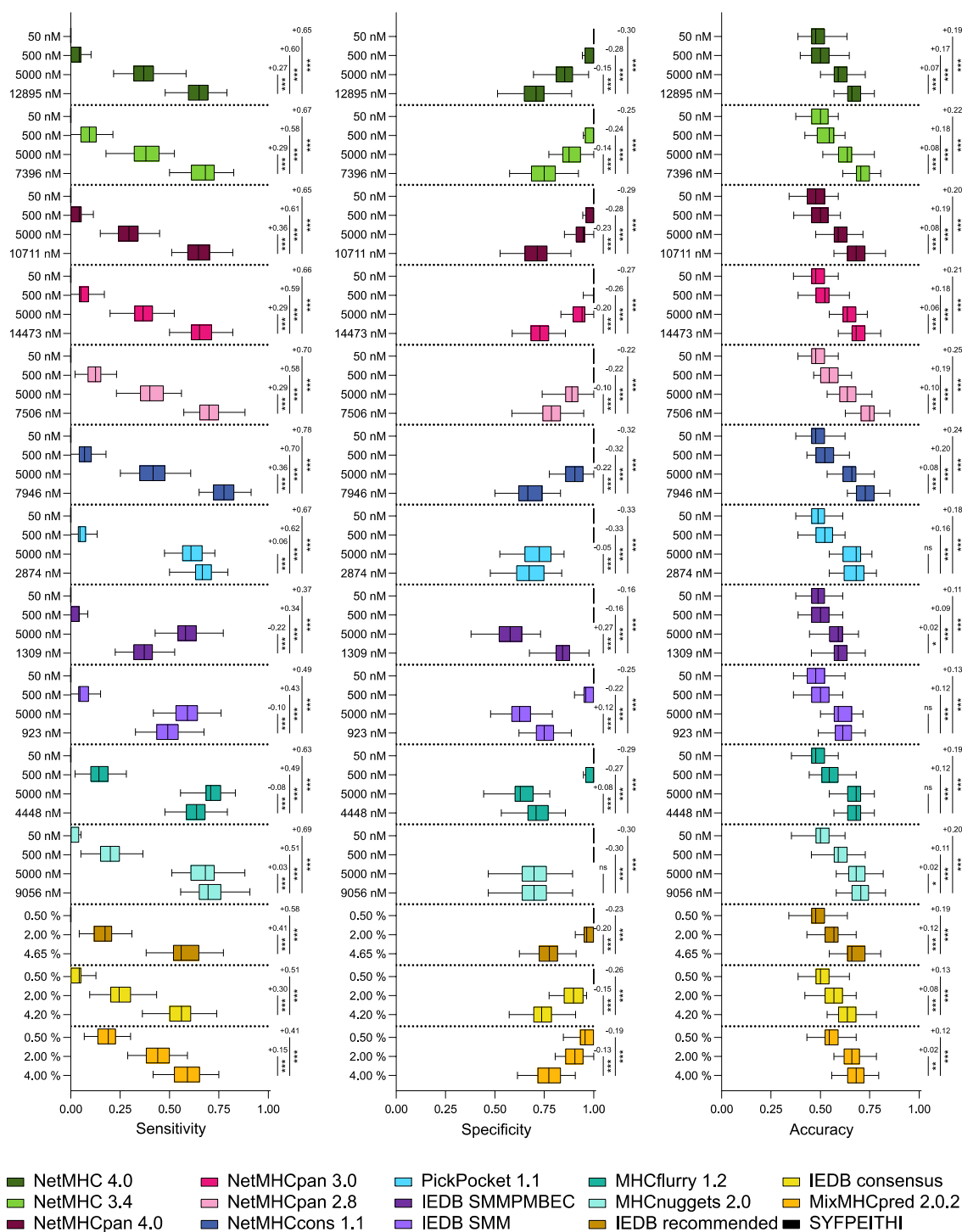


Figure 18. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended “validated” or generally used thresholds. Results are shown for (A) HLA A2 and (B) HLA A24. Recommended thresholds were calculated and validated by bootstrapping as described. In a second bootstrapping, sensitivity, specificity and accuracy of predictors applying the recommended thresholds were compared to general thresholds for predicting high ($IC_{50} \leq 50$ nM or percentile rank (%) ≤ 0.5), intermediate (inter, $IC_{50} \leq 500$ nM or percentile rank (%) ≤ 2) or low ($IC_{50} \leq 5000$ nM) binding likelihood. Box plots and whiskers show bootstrapping quartiles and the 95% confidence interval of data, respectively. Numbers indicate significant differences of means, which were determined by one-way ANOVA followed by Dunnett multiple comparisons test (significance, $p < 0.05$). Asterisks indicate p-values. (***) $p < 0.001$, (**) $p < 0.01$, (*) $p < 0.05$, (ns) not significant

Figure 18 shows the confidence intervals and means for threshold-dependent sensitivity, specificity and accuracy for A2 and A24. Analysis for A1, A3, A11, B7 and B15 is shown in Supplementary Figure S4 in the Annex. Overall, the use of the validated threshold led to a significant increase in sensitivity, and to a relevant increase (change by ≥ 0.1) compared to the high and intermediate binding affinity thresholds. In comparison to the low binding affinity thresholds, the validated threshold also significantly increased sensitivity for most predictors, especially in case of the NetMHC family. However, for the prediction algorithms IEDB SMMPMBEC and IEDB SMM sensitivity was decreased, because validated thresholds were more stringent than the low binding affinity thresholds. If sensitivity was lost, specificity was gained and *vice versa*. However, the improved sensitivity surpassed the specificity deficit. This is documented by the concomitant increases in accuracy. For example, when comparing validated vs. low binding affinity thresholds for A2, the accuracy was relevantly increased (for NetMHCcons 1.1, PickPocket 1.1, IEDB SMMPMBEC, IEDB SMM and MHCnuggets 2.0), insignificantly changed (NetMHC 4.0, NetMHC 2.4, NetMHCpan 2.8 and MHCflurry 1.2) or only slightly lowered (NetMHCpan 4.0 and NetMHCpan 3.0). Overall, accuracy was either not affected or even increased for other HLA types (A1, A11, A24, B7, B15). For A3, differences between accuracy resulting from validated and low binding affinity thresholds varied.

The calculation of individual thresholds based on defined criteria allowed more sensitive prediction of HLA-binding peptides without a strong negative influence on prediction accuracy. In contrast, accuracy was even increased to a minor extent for many prediction settings. Thus, systematically using less stringent validated thresholds increased the number of correctly predicted HLA ligands and limited false positives to a tolerable fraction.

4.2.4 Criteria-based thresholds and bootstrapping-validated thresholds result in similar prediction performance.

Due to the limited sample size of datasets for individual peptide lengths, the calculation of validated thresholds could only be performed for pooled peptide lengths. In order to provide an estimate of how representative criteria-based thresholds are for the whole statistical population, their associated performance measures were directly compared with the respective validated thresholds (Figure 19 and Annex, Supplementary Figure S5).

In the majority of cases validation resulted in the same binding likelihood as for criteria-based thresholds. Thus, sensitivity, specificity and accuracy were rarely affected. A greater divergence between validated and criterion-based threshold led to significant changes in the performance measures. In the few cases where significant changes were detected, the difference was < 0.1 . Relevant accuracy differences > 0.1 were only observed for A3 binding prediction by NetMHCcons 1.1 and IEDB recommended. However, variation between the validated and criteria-based threshold did not necessarily follow a change in accuracy, as can be seen for A2 prediction by NetMHCpan 3.0 and IEDB consensus. Therefore, individual criteria-based thresholds were fair to use for analysis of individual peptide lengths where validated thresholds were not available.

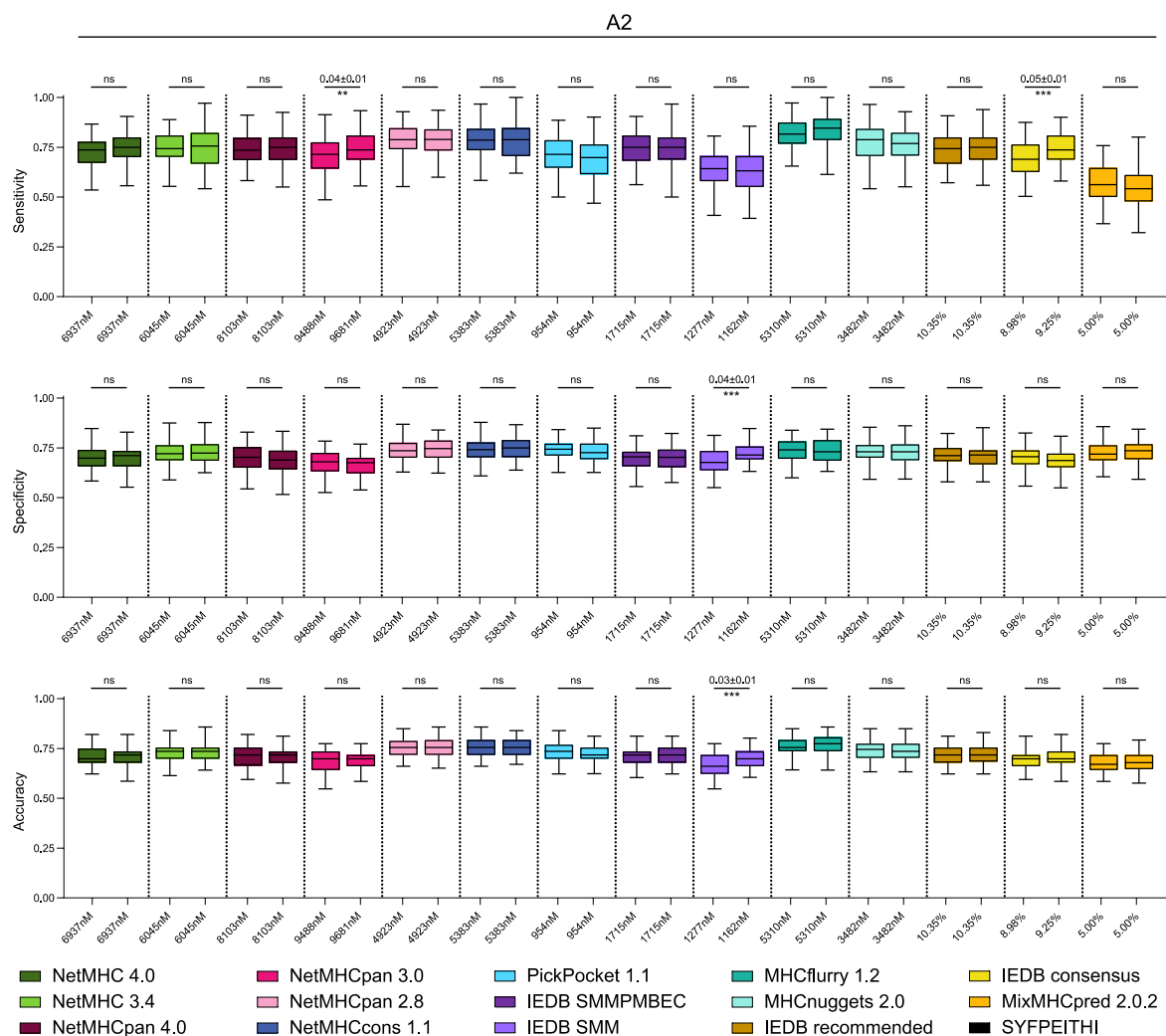


Figure 19. Comparison of predictor performance measures between applying criteria-based thresholds and validated thresholds. Validated thresholds (left) were calculated by bootstrapping as described. Criteria-based thresholds (right) were calculated by applying criteria ($FPR \leq 0.33$ (specificity ≥ 0.66), TPR (sensitivity) $\geq 2 \times FPR$ and the highest possible sensitivity within the first two criteria) to the respective complete HPV16 E6/E7 peptide set. In a second bootstrapping, the two thresholds were applied and confidence intervals of and sensitivity, specificity and accuracy were calculated for 100 samplings. Box plots and whiskers show bootstrapping quartiles and the 95% confidence interval of data, respectively. Significant differences of means were determined using Student's t test (significance, $p < 0.05$). (***) $p < 0.001$, (**) $p < 0.01$, (*) $p < 0.05$, (ns) not significant

4.2.5 Applying the recommended thresholds increases the number of predicted true binders.

Comparison of decision threshold-dependent prediction performance showed that using the individualized and more tolerant validated and criteria-based thresholds increased prediction sensitivity. The effect of applying these thresholds to the predictions for the HPV dataset is shown for A2 and A24 in Figure 20 and for A1, A3, A11, B7 and B15 in Supplementary Figure S6 in the Annex. Peptides were sorted according to their experimental binding affinity (first column) and separated into binders (blue) and nonbinders (red). Predicted binding likelihoods of predictors (following columns) are shown for individual (left) and pooled peptide lengths (right) using the intermediate binding affinity thresholds (dark colors; $IC_{50} \leq 500nM$ or percentile rank ≤ 2) or the criteria-based or validated thresholds (light colors), respectively.

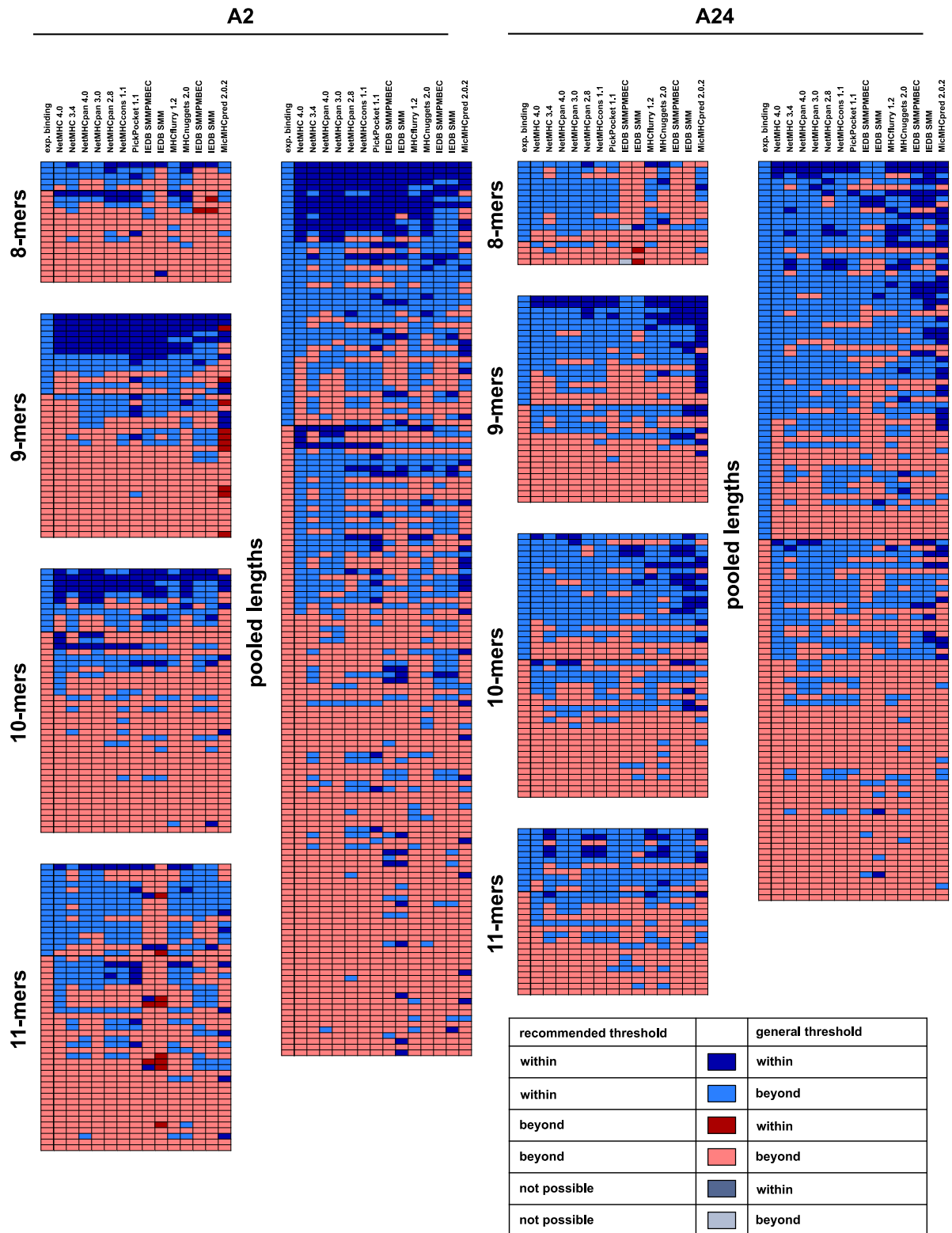


Figure 20. Classification of HLA binding prediction of HPV16 E6/E7 peptides to HLA A2 and A24 according to application of different thresholds. HLA-ligands derived from HPV16 E6/E7 were validated by experimental assessment (first column) and categorized into binders (blue) and nonbinders (red). Following columns indicate the peptides' predicted binding likelihood classified by different thresholds. Criteria-based (single lengths) or recommended thresholds (pooled lengths) separate peptides into predicted (blue), not predicted (red) or threshold calculation not possible (grey). Further, predictions were classified into predicted within (dark shade) or beyond (light shade) the general threshold of $IC_{50} \leq 500nM$ or percentile rank ≤ 2 .

For A2, applying the commonly used intermediate binding affinity thresholds the predictions yielded only about a third of all true binders and did not prevent prediction of some false positives (dark blue). However, for A24, the generally used thresholds positively predicted only a minority of actual binders. Only MixMHCpred 2.0.2 correctly predicted more than half of the actual binders. Detailed performance measures for applying individual criteria-based thresholds and validated thresholds are listed in Supplementary Table S3 in the Annex.

Overall, the generally used thresholds are more suitable for 9- and 10-mer prediction, whereas only few binders were detected among 8- and 11-mer peptides. Criteria-based (for single peptide lengths) and validated thresholds (for pooled lengths) increased the numbers of predicted true binders (light blue) or reduced prediction of false positives where intermediate binding affinity thresholds were too tolerant (dark red). However, whenever less stringent thresholds were used, this increased the amount of false positive predictions. Essentially, this applies for all analyzed HLA types. However, for HLA types where A_{ROC} analysis revealed less precise prediction performance, e.g. A24, using intermediate binding affinity thresholds led to very poor results. Here, recommended thresholds were superior in predicting actual HLA ligands.

4.3 HPV16 variants and their HLA-binding E6 and E7 peptides

4.3.1 HPV16 positive cell lines are infected with different genomic HPV variants.

Based on differences within the L1 ORF HPV16 is differentiated into several lineages and sublineages. Sequences of E6 and E7 are more conserved than L1, albeit not free of mutational changes. Mutations can lead to aa substitutions that affect the possible E6 and E7 epitope repertoire. Therefore, all HPV16-positive cell lines available in the lab were characterized for mutations in E6 and E7 sequences (listed in Annex, Supplementary Table S4). For the majority of cell lines in our collection, this was already performed by Stephanie Hoppe, a previous group member. However, the two cell lines UM-SCC104 and MRI-H-186 were newly acquired and their E6 and E7 variants needed to be determined.

E6 and E7 sequences were obtained as described in the section “Sequencing of HPV16 E6 and E7 genes” and compared to the HPV16 reference sequence (sublineage A1 European, variant E-v1). Sequence alignment revealed nucleotide changes and associated aa substitutions (Annex, Supplementary Table S5). UM-SCC104 showed a single mutation in the E6 gene. The T371G mutation resulted in the aa change L90V. The E7 sequence was completely identical to the reference sequence. For MRI-H-186, both genes were identical to the reference sequence. Thus, the two cell lines were identified to be of the A2 and A1 European sublineages (variant E-v2 and E-v1), respectively.

4.3.2 Genomic variants of HPV16 result in different HLA ligands.

Experimental binding affinities of analyzed variant peptides were compared to reference sequence counterparts to evaluate the effect of aa changes on the repertoire of HLA ligands. Considering known binding sequence motifs for the different HLA types, not all aa changes occurring in HPV16 E6 and E7 are equally likely to interfere with HLA binding. For HLA class I, the second and last positions have been shown to serve as anchor residues. Thus, it was expected that especially aa changes in anchor positions alter binding capacity. With regards to the binding motifs of the analyzed HLA types, prevalent HPV16 E6 and E7 variants were assessed for potentially relevant aa changes at anchor positions (Table 13). Substitutions were considered relevant if a beneficial amino acid was exchanged for a disadvantageous one and *vice versa*. Based on this assessment, four mutations, E6 E120D, E6 D32E E6 A68G and E7 N29S, were not expected to influence binding to any investigated HLA type. The substitution H51N in E7 should only affect A3-ligands. Half of the prevalent aa changes were estimated to influence binding to three or more HLA types. Some HLA types were expected to be more affected than others if aa substitutions would occur in anchor residues. For example, five or more variants could hypothetically alter binding to A1, A3, A24 and B15, whereas only a single change might have implications for B7 binding affinity.

Table 13. List of amino acid changes prevalent HPV16 E6- and E7-variants and their potential relevance for HLA binding.

Amino acid change		A1	A2	A3	A11	A24	B7	B15	Count
E6	R17I		X	X	X	X			4
	R17T	X		X	X				3
	Q21D	X						X	2
	D32E								-
	I34R		X	X		X			3
	E36Q	X	X					X	3
	A68G								-
	H85Y	X					X	X	3
	L90V						X	X	2
	E120D								-
E7	L28F		X	X	X	X			4
	N29S								-
	H51N			X					1
	S63F	X					X	X	4
Total		5	4	5	3	6	1	5	

X: potentially changing binding capacity if occurring at anchor position

Figure 21 shows binding affinities of variant peptides side-to-side with their reference sequence-derived counterparts, if available, and peptide sequence motifs for each HLA type. The exact reference counterpart was not analyzed for all variant peptides, e.g. if HLA binding predictions were negative. Among the 86 analyzed variant peptides, of which eight showed double aa changes, 18 changes were detected in anchor residues (marked with asterisks in Figure 21), 10 changes occurred at the N-terminus, 12 in the second last position (C-terminus -1) and 54 times any middle residue was affected. Among the 18 peptides with residue changes at anchor positions, five were expected to alter binding capacity to HLA types; three were supposed to have a positive effect and two a negative effect. Indeed, the A24 peptide E6 L90V/81-90 (SEYRH^YCYS^VL) was found to be a nonbinder in contrast to the reference peptide SEYRH^LCYS^LL (11.98 μ M) and the peptide E6 L90V/82-90 EYRH^YCYS^VV showed only intermediate (8.70 μ M) compared to high binding capacity (0.32 μ M) seen for EYRH^LCYS^LL. For the variant peptides with expected positive effects (FYSKISEYR^YY and YSKISEYR^YY (A1), FQDPQERPI^II (A2)) exact were predicted to be nonbinders. Although exact counterparts were not among tested peptides, it can be expected that aa substitutions were indeed beneficial for HLA-binding. Additionally, three more variant-derived peptides showed affected binding affinity when compared to the exact reference-derived counterpart, albeit their substitutions at anchor positions were considered irrelevant. The aa changes E6 A68G and E7 N29S increased the binding affinity of the peptides E6/67-77 and E7/19-29 to HLA A1. Binding affinity to A3 was unexpectedly reduced for the E6 L90V/89-99 peptide.

In summary, the comparison of HPV16 E6 and E7 reference sequence- and variant-derived peptides revealed that certain variant peptides do exhibit different binding affinities than their counterparts. Overall, 15 aa changes were found to have a positive effect (affinity gained) and 18 a negative effect (affinity reduced). However, most substitutions were located in the middle part of the peptides and not at anchor positions.

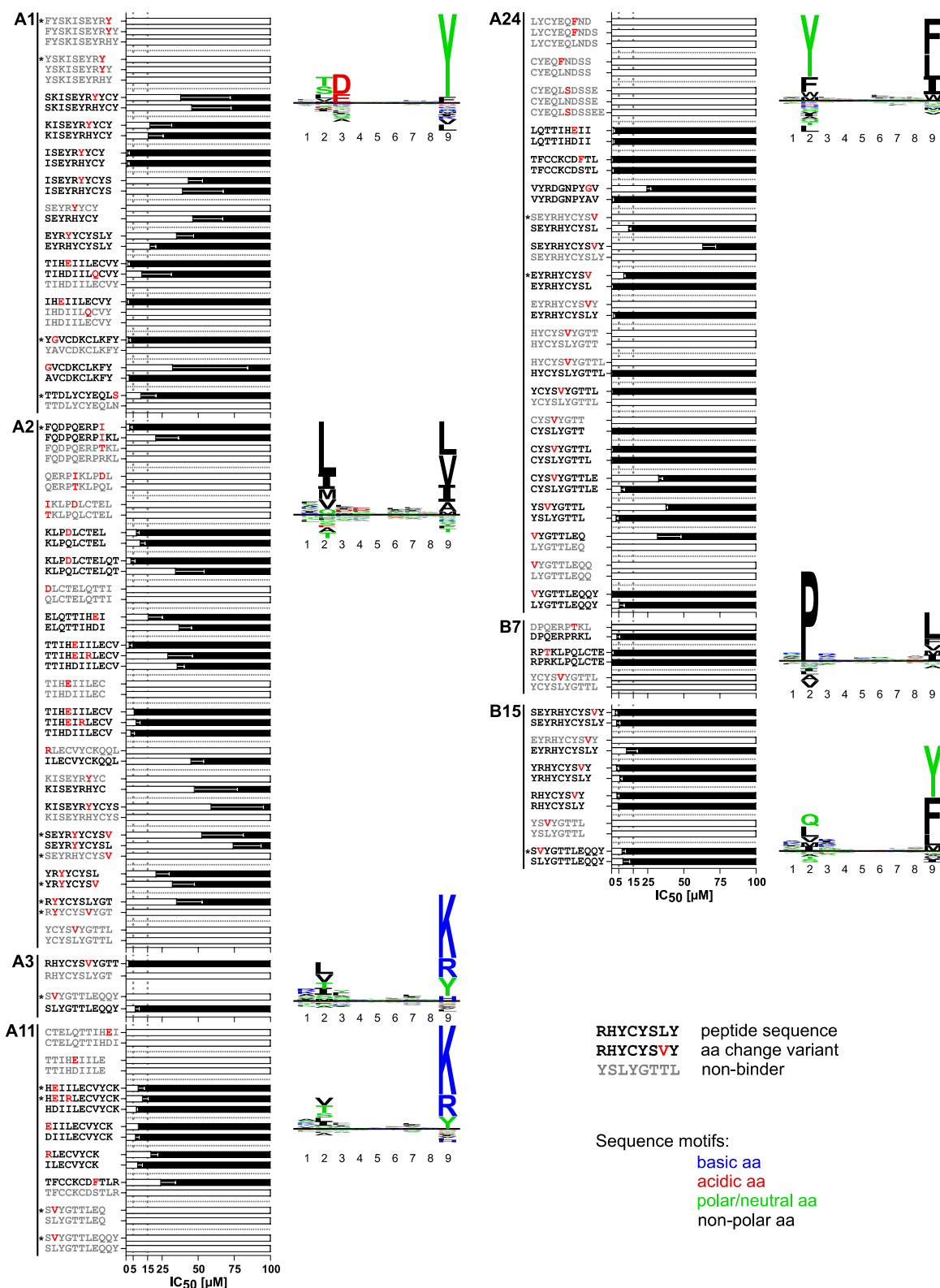


Figure 21. Influences of amino acid changes in HPV16 E6/E7 variants on HLA binding affinity. IC₅₀-values of HPV16 E6-/E7-derived peptides binding to specific HLA types have been determined in cellular competitive binding assays. HPV16 E6/E7 variant peptides, harboring amino acid changes (red), were compared to the reference sequence-derived counterparts. Affected anchor residues were marked with asterisks. Bar graphs illustrate mean ± SD of the IC₅₀ concentration on a black background to visually highlight binders. Dotted grey lines indicate levels of high (≤5µM) intermediate (5µM<IC₅₀≤15µM) and low (15µM< IC₅₀≤100µM) binding affinity. Nonbinder sequences are written in gray color. For each HLA type, sequence motifs of HLA ligands are given to the right of the bar graph (derived from the Immune Epitope Database). Amino acid (aa) characteristics are highlighted in different colors. Motifs display abundance of aa in letter size. Underrepresented aa are below the horizontal black line. Motifs were adapted from the Seq2Logo web tool.

4.4 Evaluation of peptide immunogenicity using T cell lines

4.4.1 Flow cytometry characterizes HLA-type of PBMCs isolated from buffy coats.

In the context of this study, buffy coat preparations were used to isolate PBMCs from the blood of 49 healthy donors. The yield of PBMCs ranged from 80 million to 1350 million cells with a mean PBMC yield of 425 ± 227 million cells. Typically, healthy blood donors are not characterized for HLA expression. Thus, HLA typing was required. As large numbers of PBMCs were needed for functional assessment of HLA ligands, a HLA typing methods with minimal cell input was desired.

HLA typing is typically performed by PCR with HLA type-specific primers. However, simultaneous characterization of many alleles requires substantial amounts of template DNA. Compared to PCR, fewer input cells (~200,000 per sample) are required for FACS analysis with fluorescently labeled HLA-specific antibodies, and several HLA-molecules can be investigated simultaneously by staining with differently marked antibodies. Four HLA-specific and directly labeled antibodies against A2, A3, A24 and B7 were purchased and tested in cells of known HLA type. Figure 22 shows the results of antibody characterization by flow cytometry. For each antibody, testing of different dilutions, isotype binding and binding to unspecific HLA types is shown. All concentrations show separate positive peaks. For anti-HLA-A3-APC, anti-HLA-B7-PE and anti-HLA-A24-FITC the 1:200 dilution was found to be the highest dilution able to stain a distinct positive population, whereas for anti-HLA-A2-FITC the 1:400 dilution was chosen. None of the antibody-matching isotype controls induced any background signal. The anti-HLA-A24-FITC antibody showed unspecific binding to the HLA types expressed by both mismatched control cell lines E481324 (*A*01:01; B*52:01*) and FH8 (*A*11:01; A*34:02; B*82:01; B*27:05*) and was therefore excluded from the antibody panel for HLA assessment.

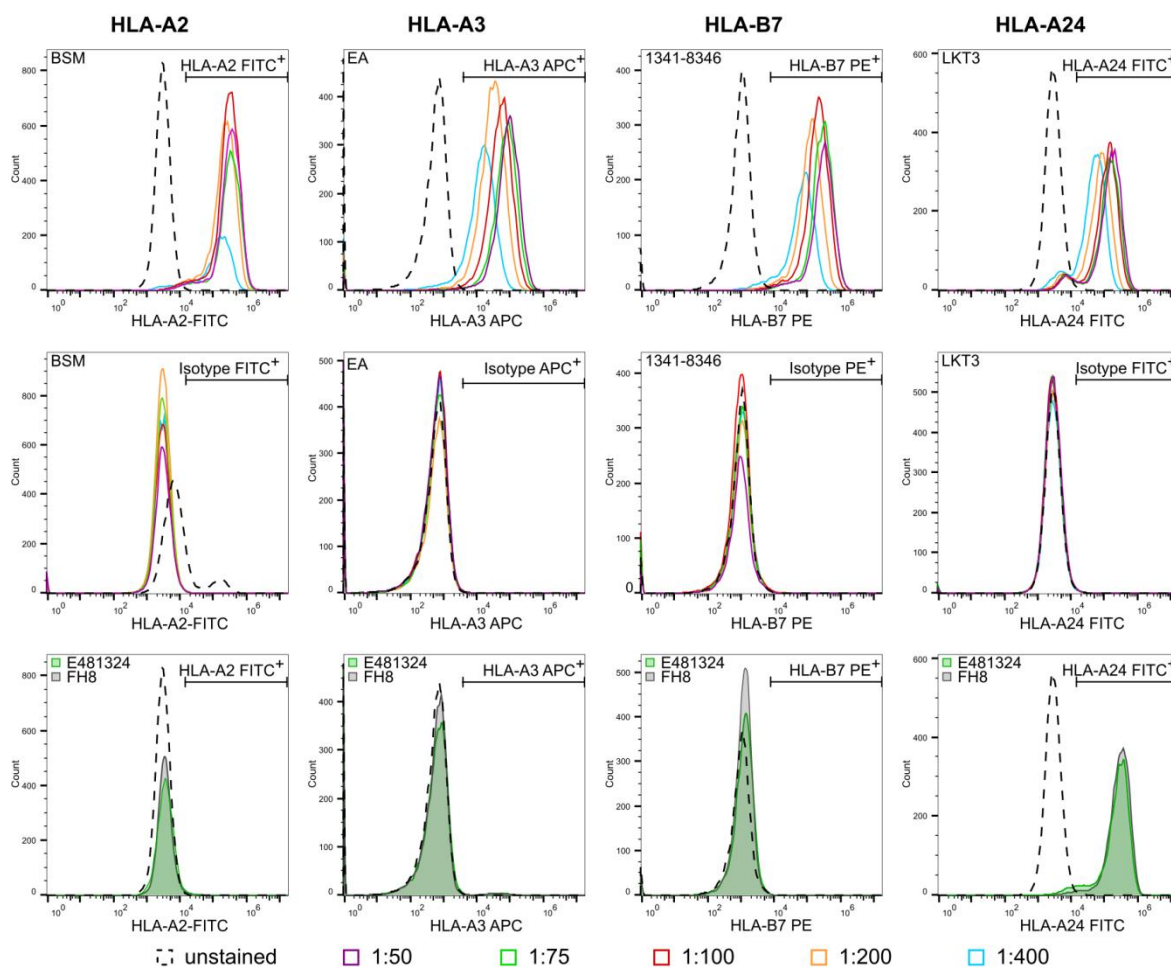


Figure 22. Analysis of HLA-specific antibodies for flow cytometry. Antibody-stained cells of known HLA type were analyzed using flow cytometry. Histograms show numbers and fluorescence intensities of cells. Upper panels: cell lines positive for the respective HLA type were stained with fluorescently-coupled HLA-specific antibody at different dilutions (indicated below the figure) and compared to unstained cells. Middle panels: HLA-independent unspecific binding of antibody was analyzed by staining with a respective antibody isotype control coupled to the same fluorophore. Lower panels: HLA-dependent unspecific binding was assessed by staining of the cell lines E481324 (green) and FH8 (grey), expressing HLA types different from the respective antibody specificity.

The three antibodies to A2, A3 and B7 were used for HLA typing of 49 healthy blood donors, 16 of whom were typed only for A2. Representative results of two donors are shown in Figure 23. Donor D25 showed a positive peak distinct from the isotype control antibody only for anti-HLA-A2-FITC and was thus determined to be A2⁺ A3⁻ B7⁻. In contrast, donor D30 showed no FITC signal for the A2-specific antibody, but specific peaks for anti-HLA-A3-APC and anti-HLA-B7, and was thus defined to be A2⁻ A3⁺ B7⁺. The results for all donors are listed in Table 14. Overall, 18 (36.7%) of tested donors expressed A2, 10 (30.3%) A3 and 6 (18.2%) B7.

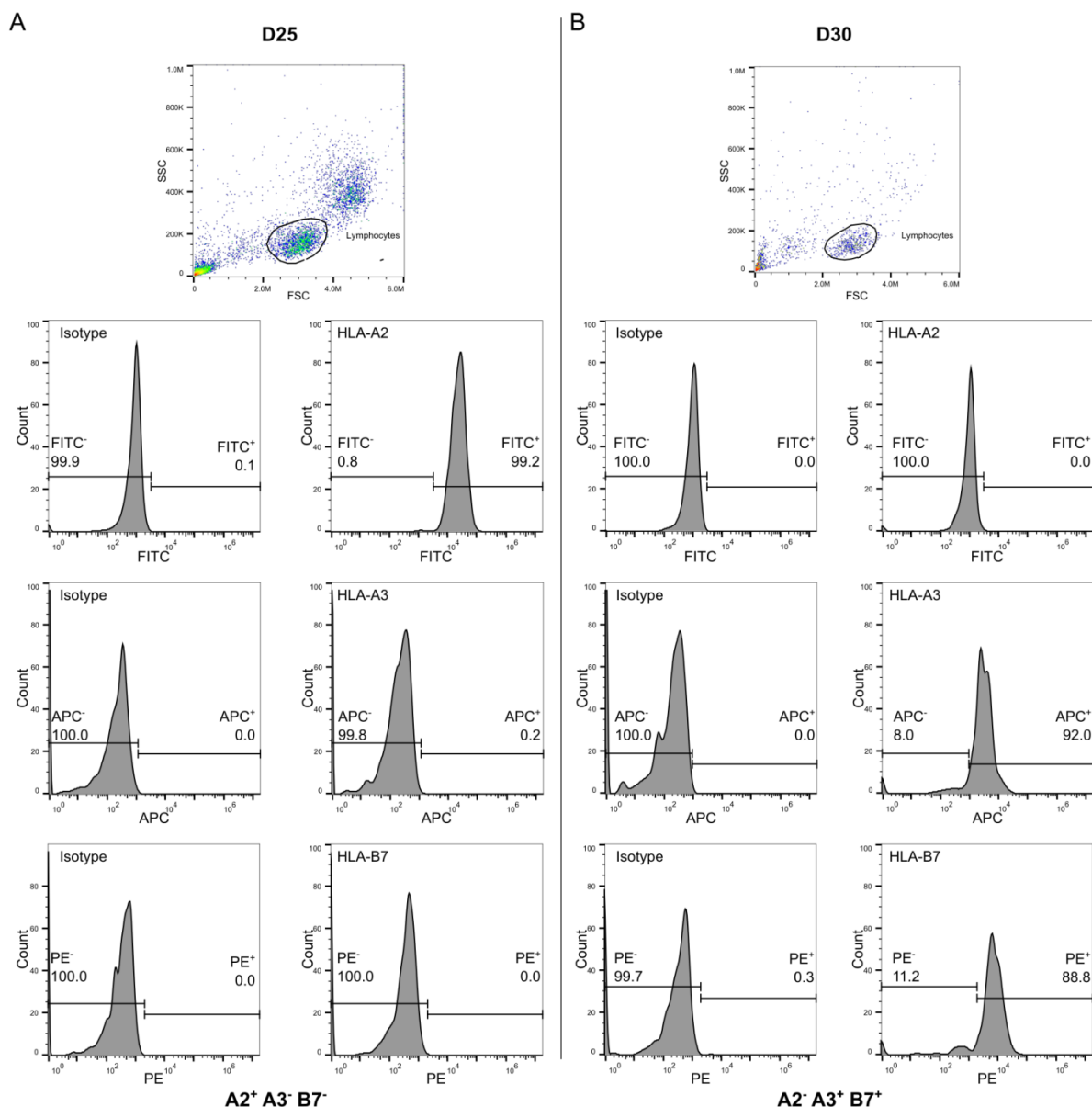


Figure 23. HLA-typing of healthy donors by flow cytometry. Isolated PBMCs of healthy donors were stained with specific fluorescently-coupled anti-HLA antibodies and analyzed by flow cytometry. Representative results of two donors are shown. Lymphocytes were gated based on forward (FSC) and sideward scatter (SSC) properties. Lymphocytes were further analyzed for fluorescence intensity of the respective fluorophore coupled to the used antibody. Markers for positivity were set based on the fluorescence intensities measured for staining with isotype controls (left). The HLA typing result is summarized below the figure.

Table 14. Results for flow cytometry based HLA typing of healthy blood donors.

Donor	HLA-A2	HLA-A3	HLA-B7
D1	+		
D2	-		
D3	-		
D4	-	+	+
D5	+		
D6	+		
D7	+		
D8	-	-	-
D9	+		
D10	-		
D11	-		

Results

Donor	HLA-A2	HLA-A3	HLA-B7
D12	+		
D13	-		
D14	+		
D15	-	-	-
D16	-		
D17	-		
D18	-	+	+
D19	-	-	-
D20	-		
D21	+		
D22	-	-	-
D23	+	-	-
D24	-	+	-
D25	+	-	-
D26	+	+	+
D27	+	-	-
D28	-	-	-
D29	-	-	-
D30	-	+	+
D31	-	-	-
D32	+	-	-
D33	+	-	-
D34	-	-	-
D35	-	+	-
D36	-	-	-
D37	-	+	-
D38	-	-	-
D39	-	-	-
D40	-	-	-
D41	-	-	+
D42	-	-	-
D43	-	+	-
D44	-	+	-
D45	+	+	-
D46	+	-	-
D47	-	-	-
D48	+	-	-
D49	+	-	+
Total	18/49 (36.7%)	10/33 (30.3%)	6/33 (18.2%)

+: positive, -: negative, blank: not analyzed

4.4.2 IFN γ -ELISpot assays identify immunogenic HPV16 E6 and E7 peptides.

In order to identify true T cell epitopes as candidates for vaccine development, it is necessary to confirm that HLA ligands induce T cell reactivity. To this end, healthy donors were screened for HPV16 E6- and E7-specific memory responses by IFN γ -ELISpot assays. As frequencies of HPV16-specific T cells are known to be low in the peripheral blood, isolated PBMCs were cultured for twelve days in the presence of a HLA-specific ligand to stimulate and expand T cells recognizing this peptide prior to setting up the assay. A response was considered positive if the stimulation index (fold-change over unspecific background with solvent DMSO) was ≥ 2 and if the mean number of SFUs was ≥ 200 per million cells. General ability of the cells to produce IFN γ was tested by unspecific stimulation with the mitogen Concanavalin A (ConA). The mean SFU of three wells for background and six wells for specific stimulation was calculated. Positive and negative epitope control stimulations were included.

Overall, 24 donors were screened for reactions to identified HLA-A2 and HLA-A3 ligands. Fifteen donors were A2⁺, eight were A3⁺ and one donor was A2⁺ A3⁺. Figure 24 shows representative examples for an A2⁺ Donor (D6) and an A3⁺ donor (D24). In donor D6, six A2-binding HPV16 E6-/E7-derived peptides elicited responses that fulfilled one of the two criteria, but only four fulfilled both and were thus considered positive (E7/11-19, E7/12-19, E7/80-90 and E7/81-90). For D24, nine peptides were found that reached SI or SFU values above either respective threshold, and five which excited responses exceeding both (E6/68-77, E6/75-83, E6/106-115, E6/107-115 and E6/129-138).

The immunogenicity results for all analyzed donors are summarized in Figure 24 C. In total, 18 A2- and 13 A3-associated epitopes were identified. For both HLA types one peptide was found to induce responses in the maximal number of four donors, which were E7/11-19 (A2) and E6/68-77 (A3). E7/11-19 was also the peptide inducing the strongest immune reactions. The two peptides E7/11-20 and E7/77-87 were reactive in three A2⁺ donors. Six more A2 peptides were reactive in two donors and nine induced positive responses only in single donors. Also for HLA-A3, two peptides were found to be reactive in three donors and two by two donors. Responses against eight A3 epitopes were only detected in single donors. Comparing the intensity of the responses against A2-restricted peptides with A3-ligands, A3-mediated responses clearly involved fewer IFN γ -positive cells.

In order to identify epitope-responses shared by multiple donors, previous experiments on immunogenicity of HPV16 E6- and E7-derived peptides were re-assessed using the same ELISpot analyzer and results were added to the current dataset (Figure 25). A list of all IFN γ -responses can be found in Supplementary Table S6 in the Annex. The previous analysis included investigation of other HLA types. Peptides which were found to be ligands to several HLA types were treated as potential epitopes to the binding HLA molecules in reactive donors with incomplete HLA typing. Thus, the promiscuous epitope E6/68-77, which is binding to A1, A3, A11 and B15, was found to be immunogenic in donors positive for each of these HLA molecules, although being detected primarily in context of HLA-A3 assays. Moreover, Figure 25 highlights immunogenic sequence hotspots. Especially in the extensively investigated HLA types A2 and A3, regions with multiple immunogenic peptides can be found. The regions from aa 7 to 21 and 78 to 90 in the E7 protein are hotspots for A2-restricted epitopes, whereas for A3 the regions from 68 to 83 and 106 to 115 in the E6 protein harbor many immunogenic peptides.

In IEDB, 34 HPV16 E6- and E7-derived peptides were reported to induce IFN γ -responses (Annex, Supplementary Table S2). Of thirteen reported A2 epitopes, eight were validated in the course of IFN γ -ELISpot assays and one peptide was not investigated, as it did not show HLA-A2 binding in competition-based binding assays. Eleven identified A2 epitopes were not reported before. Only one of two reported A3 epitopes experimentally showed binding and IFN γ -responses. Half of the twelve IFN γ -inducing A3 epitopes were known to be HLA A3 ligands.

Results

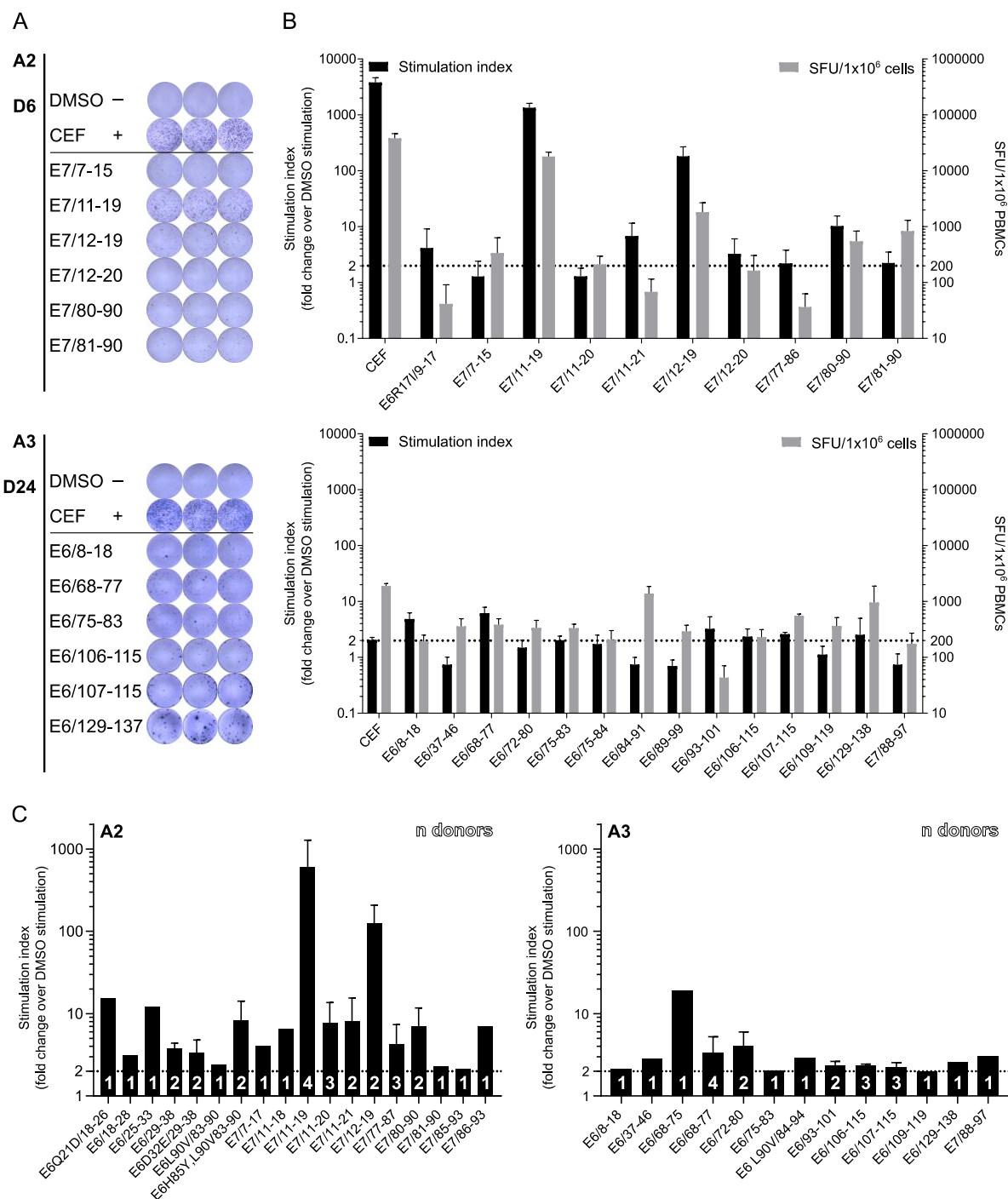


Figure 24. HLA-specific immunogenic HPV16 E6/E7-derived peptides identified by IFN γ -ELISpot assays. ELISpot assays allow for investigation of IFN γ -producing cells upon re-stimulation with specific peptide. Results of representative donors are shown for HLA-A2⁺ (D6) and A3⁺ (D24) (A). The mean spot forming units (SFU) of wells were scanned and counted (B). Responses were considered positive if the stimulation index (fold-change over background) was ≥ 2 and mean SFU was ≥ 200 per million cells. Immunogenic peptides were assessed for all donors resulting in mean SI (C) (n, numbers of responding donors).

Results

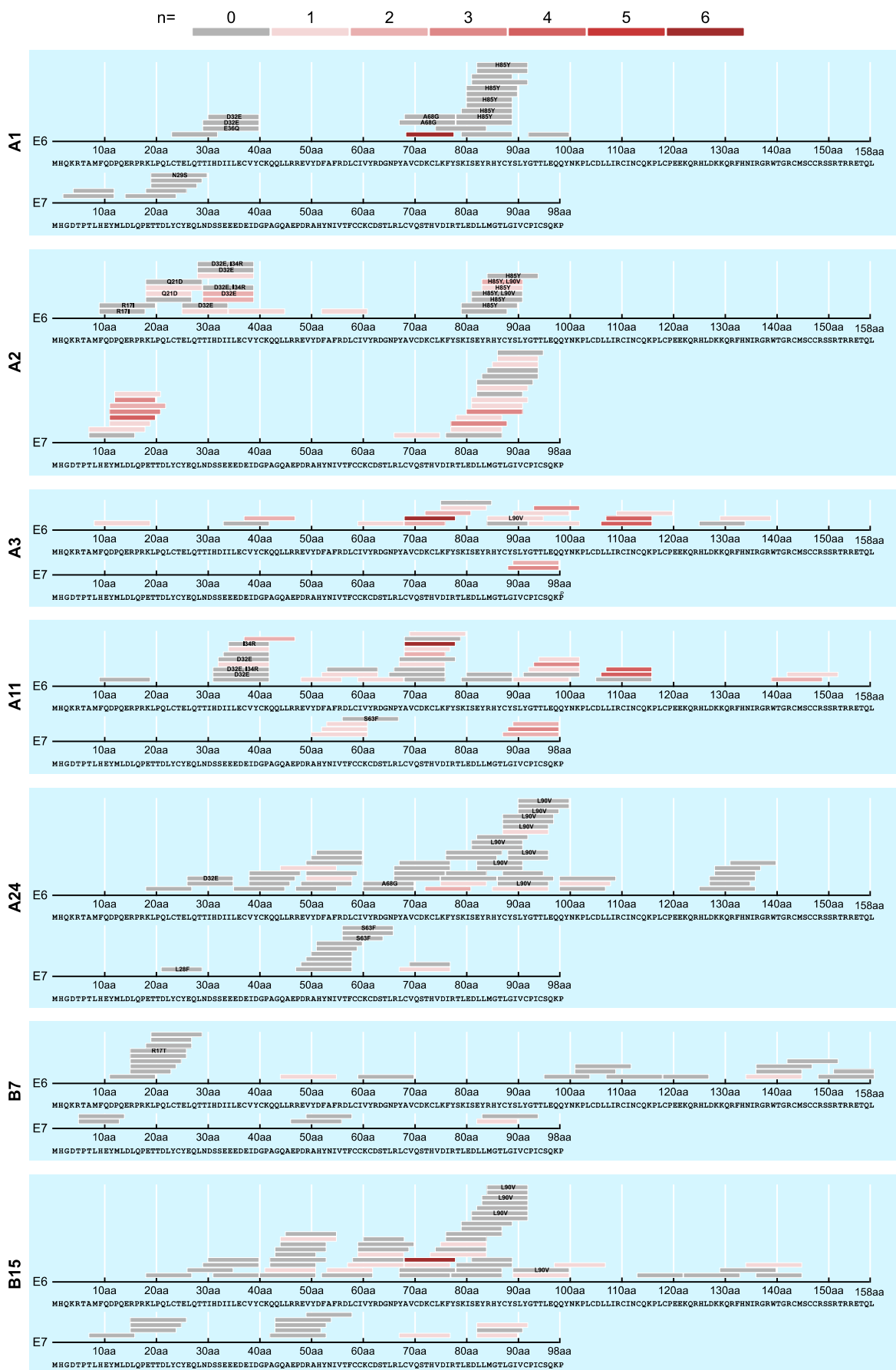


Figure 25. HLA-ligands and immunogenicity detected in donors. Determined HLA ligands are displayed at their respective position within the E6 or E7 protein. Shades of red illustrate immunogenicity in donors. Ligands that did not exhibit immunogenicity are shown in grey. As HLA typing of donors was incomplete, immunogenicity of ligands binding to multiple HLA types is shown for each associated molecule.

4.4.3 Epitope-specific T cell lines can induce specific lysis of HPV16⁺ target cells.

To investigate the potential of immunogenic epitopes to mediate target cell lysis, the killing capacity of MACS-purified CD8⁺ T cell lines was assessed in a cytotoxicity assay. As cell numbers were limited, the flow cytometry-based Vital-FR assay was chosen. It allows simultaneous assessment of differently labeled specific and unspecific target cells, which reduced the required number of effector cells in contrast to typically used radioactive chromium-release assays (Stanke et al., 2010).

Experiments were performed for HLA-A2⁺ healthy blood donors who showed a memory response in the IFN γ -ELISpot assays. PBMCs were expanded in long-term epitope-specific T cell lines for 15 days by two stimulations with peptide-pulsed autologous DCs in an interval of seven days. From these cell lines, CD8⁺ T cells were isolated by untouched magnetic cell sorting. The effector cells (E) were co-cultured with same numbers of CFSE-labeled A2⁺ HPV16⁺ CaSki (specific) and FarRed-labeled A2⁺ HPV16⁻ C33A (unspecific) target cells (T) at different ratios (E:T).

Effector cells from five donors were examined for specific killing induced by in total nine different immunogenic HPV16 E6-/E7-derived peptides as shown in Figure 26. Numbers of replicates and E:T ratios varied, as experiments had to be adapted to low CD8⁺ T cell numbers yielded after isolation from long-term epitope-specific T cell lines. Surprisingly, all epitopes assessed for one donor showed similar results. Either all or none were capable of mediating specific target cell killing. A lack of specific killing of epitope-specific T cells was observed in donors D6, D21 and D25, albeit all analyzed epitopes had been shown to induce IFN γ -responses in these donors in earlier ELISpot assays. In order to investigate if liquid nitrogen storage time between IFN γ -ELISpot and Vital-FR cytotoxicity assays influenced functionality of cells, and may thus explain the discrepancy between ELISpot and cytotoxicity assay results, cytokine production capacity of freshly isolated PBMCs and frozen PBMCs from donor D25 was compared. Production of the intracellular cytokines INF γ , TNF α and granzyme B was measured by flow cytometry (Annex, Supplementary Figure S7. Upon stimulation with PMI/Ionomycin, cytokine production by freshly isolated PBMCs was readily observed, whereas frozen PBMCs exhibited only a very weak production of IFN γ .

The well-known epitope E7/11-19 (Riemer et al., 2010) was analyzed in four out of five donors, of which half had specific T cells with the capability to kill CaSki cells. Among the remaining eight epitopes, the four that shared a core sequence with E7/11-19 (E7/11-20, E7/11-21, E7/12-19 and E7/12-20) were able to induce cytolysis, whereas the other half (E7/80-90, E7/81-90, E6/25-33 and E6 H85Y, L90V 83-90) was not. The three epitopes E7/11-19, E7/11-20 and E7/12-20 were previously demonstrated to induce cytolysis (Annex, Supplementary Table S2). These results indicate that HLA-A2 epitopes from the E7 hotspot region 7-21 may represent good candidates for vaccine development.

Results

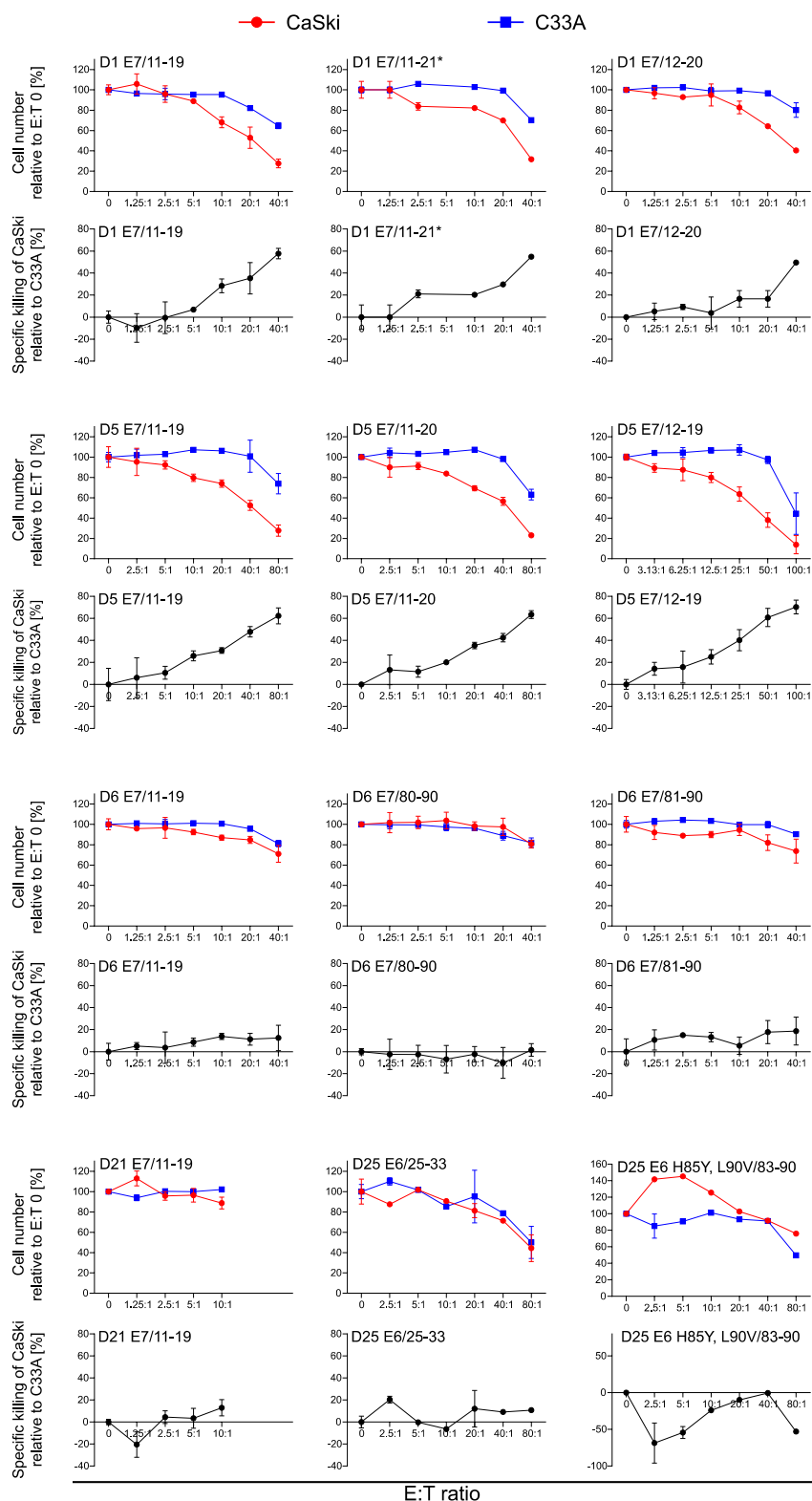


Figure 26. Cytotoxicity mediated by CD8⁺ T cells specific for HPV16 E6-/E7-derived A2-restricted epitopes. Frequencies of specific and unspecific target cells and specific killing after co-culture with epitope-specific CD8⁺ T cells, as assessed by flow cytometry-based Vital-FR assays. CD8⁺ T cells were isolated from long-term T cell lines stimulated weekly with epitope-pulsed autologous DCs. Isolated effector cells (E) were co-cultured at different ratios with 3000 specific A2⁺ HPV16⁺ CFSE⁺ CaSki cells (red) and 3000 unspecific A2⁺ HPV16- FarRed⁺ C33A cells (blue) as targets (T). Upper panels show frequencies of both cell lines relative to culture without effector cells (E:T=0). Lower panels show specific killing (black) calculated by the frequency of specific to unspecific target cells.

5 Discussion

This thesis aimed at identifying candidate epitopes for therapeutic HPV16 vaccine design, as well as providing a detailed performance evaluation of widely used HLA class I binding prediction methods. The first outlined research aim was attained by assessing the HLA class I binding affinity of 271 predicted HPV16 E6- and E7-derived peptides. Thereby, 69 out of 271 peptides were identified to be HLA ligands in the scope of this study. Combined with earlier results, 293 HPV16 E6- and E7-derived peptides were validated to be binders. The total dataset of 779 analyzed peptide-HLA combinations was used to evaluate the prediction performance of the 15 employed HLA class I binding prediction methods, which was the second aim of this thesis. Importantly, the evaluation revealed that prediction methods are not 100% precise, which confirms the continuous need for experimental validation of prediction results. The evaluation included detailed performance measures and novel individual threshold calculations for all combinations of analyzed predictors, HLA types and peptide lengths in order to provide recommendations for the best performance of available methods. The third aim of this thesis was accomplished when comparison of reference- and variant-derived peptides showed that many amino acid changes influenced the HLA binding affinities. To achieve the fourth aim, HLA A2- and A3-binding HPV16 E6- and E7-peptides were functionally characterized. Assessing the capacity to elicit IFN γ -secretion, 31 HLA ligands were identified to be immunogenic epitopes. Moreover, by investigating the potential to mediate specific killing of HPV16⁺ target cells, five A2-associated CTL epitopes were found.

The HPV16 E6/E7 dataset comprised 779 peptide binding affinity measurements for seven frequent HLA class alleles, which are representatives for HLA supertypes. In contrast, earlier datasets used for performance evaluation of prediction methods included several thousand peptide measurements. For example, Yu et al. investigated 1,230 HLA-A*02:01-restricted and 234 HLA-B*35:01-restricted 9-mer peptides with experimentally validated binding/non-binding (Yu et al., 2002). An even larger dataset of 48,828 peptide-binding affinity measurements for 48 MHC class I alleles of different species was used for benchmarking by Peters et al. (Peters et al., 2006). This dataset included 9- and 10-mer peptides related to 36 HLA alleles, ranging from 92 (HLA-A*30:02) up to 4,405 (HLA-A*02:01) analyzed peptides per HLA allele. However, these large datasets have been utilized to develop and train predictors. Therefore, Lin et al. used an independent dataset comprised of experimental binding affinities of 176 peptides for seven HLA alleles, six of which were analyzed in the scope of this thesis (Lin et al., 2008). All of the peptides were 9-mers derived from two antigens. A comprehensive dataset of 960 experimentally validated binders and nonbinders of seven HLA class I alleles, covering different antigens of HIV, influenza and cancer, was employed for prediction performance assessment by Gowthaman et al. (Gowthaman et al., 2010). However, as authors did not describe the detailed composition of this dataset, the peptide lengths and exact number of investigated peptides per HLA allele are unknown. Up-to-date performance evaluations can be accessed via the

IEDB as weekly benchmarks are performed on data that are newly entered into the database (Trolle et al., 2015). Nonetheless, such datasets are small in size and varying in HLA allele and peptide length coverage. In contrast to existing benchmark datasets, the HPV16 E6/E7 dataset is not focused on 9-mer peptides, but also included HLA class I binding experimental assessment of 8-, 10- and 11-mers. Thus, it represents a comprehensive independent dataset suitable for performance evaluation of HLA class I binding prediction methods. Its value for the field has been demonstrated by recent studies, which used our HPV16 E6/E7 data as validation datasets for the development of the predictors MHCflurry and MHCnuggets (O'Donnell et al., 2018; Shao et al.).

The binding affinity of HPV16 E6- and E7-driven HLA ligands was experimentally validated in cellular-based competitive binding assays as described by Kessler et al. (Kessler et al., 2004). Other experimental binding datasets were generated using cell free assays, e.g. competitive binding assays with purified MHC and radiolabeled peptide probes or reporter assays based on a conformation-dependent anti-HLA class I antibody (Sidney et al., 2001; Harndahl et al., 2009). The cellular assays represent a more natural setting. However, cells express several HLA types at the same time. Thus, test peptides with high affinity to the investigated and another free HLA molecule might not compete with the reference peptide and therefore may not be detected as binders. In this study, known high affinity binder or consensus peptides (i.e., peptides with the most beneficial amino acid at each position) were chosen as reference peptides. This resulted in IC_{50} -values in the μ M-range, which is in strong contrast to predicted IC_{50} -values in nM-range. Because of this difference between predicted and experimentally determined binding strength, predicted and experimental binding were only compared in a binary way, as binding or nonbinding.

Specific features of the HPV16 E6/E7 peptide binding data might have influenced the prediction performance assessment. First of all, sample sizes differ for sub-datasets. Especially peptides binding to the HLA types A1 and B7 and of 8- and 11-aa peptide length are underrepresented in this study and the respective results should be interpreted with caution. However, as described, previous datasets rarely included any 8- or 11-mer peptides, which makes our evaluation of these peptide lengths valuable, albeit being limited. Performance measures might have been overestimated as the dataset only contains a selection of all possible E6-/E7-derived peptides, concentrated at the top-range of predicted binding likelihoods. Moreover, the HPV16 E6/E7 dataset contains a few previously reported HLA binders, which likely are part of the training data of predictors. On the other hand, the performance assessment of prediction methods that are extensively trained on heterogeneous data might be impaired, as this study addressed only HLA binding. Especially predictors trained on MS data, which is naturally selected for processing, HLA binding affinity, peptide binding competition for HLA molecules and *bona fide* peptide presentation, would be affected (Creech et al., 2018). Additionally, cysteine-containing peptides are underrepresented in MS data due to technical reasons (Bassani-Sternberg et al., 2017; Abelin et al., 2017). As HPV16 E6 and E7 proteins are cysteine-rich, prediction method performance of MS-data trained algorithms might suffer from evaluation based on

this dataset. Furthermore, it has to be considered that only a single viral protein dataset was investigated. However, predictions are mostly based on HLA binding motifs that do not differentiate peptide sources. Although lacking a validation dataset of comparable size and composition, we were able to validate threshold-dependent performance measures by a bootstrapping method.

To date, many different algorithms for HLA class I binding prediction exist, and - with the advent of deep learning - new ones are constantly being developed (Han and Kim, 2017; Liu et al., 2019; Phloyphisut et al., 2019). This is driven by the advance of immunotherapies in personalized cancer treatment, where the prediction of suitable target epitopes remains a major challenge (Sahin and Türeci, 2018; Nogueira et al., 2018). The development of medical interventions based on HLA class I binding predictions requires reliable prediction methods. When comparing the different algorithms, we found the training-data dependent ANN methods superior to position-specific-scoring-matrix-based approaches, which is in line with previous studies (Gowthaman et al., 2010; Trolle et al., 2015; Lin et al., 2008; Yu et al., 2002; Peters et al., 2006). However, these studies did not include predictors trained on MS data, which were not outperforming other methods. Moreover, none of the prior evaluations examined prediction performance for individual peptide lengths due to the focus on 9-mer peptides.

The generally used decision thresholds of $IC_{50} \leq 50nM$ and $IC_{50} \leq 500nM$ were suggested based on a study published in 1994, focusing on validated 9- and 10-mer HLA-A2 ligands predicted by the SMM algorithm (Sette et al., 1994). Later, based on corresponding measurements for 9- and 10-mer peptides with IC_{50} values $>400nM$ in two experimental binding assay systems, the group recommended evaluating prediction performance by the ability to classify peptides into binders and nonbinders at a cutoff of 500 nM (Peters et al., 2006). These thresholds are now commonly applied across different predictors, HLA types and peptide lengths. However, practical examples imply that these thresholds might be too stringent for epitope prediction. Peptides eluted from HLA and identified by MS were found to have IC_{50} -values $>500nM$, as shown by Bassani-Sternberg and colleagues (Bassani-Sternberg et al., 2016). Duan et al. described that 8/10 investigated immunogenic neoepitopes were predicted with an affinity of $IC_{50} >500nM$ (Duan et al., 2014). Similar results were observed by Engels et al., who however additionally reported that only targeted high affinity binders led to tumor eradication, whereas targeting intermediate and low affinity binders resulted in relapse (Engels et al., 2013). Recently, MS data from our lab showed that also HPV16 E6- and E7-derived A2-restricted ligands of low affinity are presented on HPV16⁺ CaSki cells (Blatnik et al., 2018). Thus, applying the commonly used stringent decision thresholds is not generally suitable, which was demonstrated by a low sensitivity measured in our performance evaluation.

Low sensitivity is a drawback, especially when prediction methods are employed with the objective of identifying a high number of potential candidates. Thus, suitability of the commonly used thresholds is questionable. However, more tolerant low binding affinity thresholds ($IC_{50} \leq 5000nM$) were not generally improving prediction performance. For this reason, we defined optimal individual decision thresholds for each predictor, HLA type and peptide length. Earlier work by Paul et al. already

recommended HLA type dependent thresholds for 38 HLA-A and -B types (Paul et al., 2013). In line with our results, thresholds of IC₅₀ values >500nM were suggested for the majority (27 of 38) of alleles. For the remaining 11 types, including A2 and A11, thresholds below IC₅₀ 500nM were proposed. However, in contrast to the comprehensive recommendations given in this thesis, Paul and colleagues focused only on the SMM predictor and 9-mer peptides.

Our performance evaluation revealed that individual strengths of many approaches should be exploited and that the best methods for the HLA/peptide length to be investigated should be used. This is feasible using the web application MHCcombine, development of which was completed in the course of this thesis. MHCcombine facilitates using up to 12 different prediction algorithms. If offered by the algorithm, it allows querying individual prediction results for 8-11-mers for different HLA types combined in a spreadsheet. This is a decisive advantage over a previous method that provides the consensus output of several predictors and only allows 9-mer prediction (Trost et al., 2007).

The HPV16 E6/E7 dataset also included binding affinity measurements of HPV16 protein variants containing amino acid substitutions. Previous findings of a large whole-genome study of HPV16 identified more SNPs in E6 than in E7, which is generally conserved (Mirabello et al., 2017). Indeed, sequencing of E6 and E7 genes of HPV16⁺ cell lines in the cell bank of the research group revealed several SNPs resulting in amino acid substitutions, which mainly occurred in E6. In peptides, substitutions at anchor positions but also in the middle sequence drastically changed HLA binding properties. Based on known HLA binding motifs, effects are expected to be position-dependent (Falk et al., 1991; Kubo et al., 1994; Rammensee et al., 1999). Conventionally, certain amino acids at specific anchor positions are extremely important for HLA binding, whereas middle positions are more variable in their degree of influence. However, our results indicated that the role of the central peptide sequence in HLA binding might be more important than anticipated. Moreover, middle sequence amino acids are mainly involved in TCR interaction (Garboczi et al., 1996; Garcia et al., 1998; Calis et al., 2013). Hence, changed HLA binding affinity and chemical properties in middle residues of HPV16 E6/E7 variants might cause abolition of existing or creation of new target epitopes (Chowell et al., 2015). In this regard, HPV16 E6/E7 variant-derived epitopes are likely to be associated with a different T cell repertoire, which could be related to the reported differences in outcomes observed for infections with HPV16 variants (Zehbe et al., 1998; Tu et al., 2006; Xi et al., 2007; Zuna et al., 2011).

As outlined above, HLA binding and thus HLA presentation of a peptide is absolutely crucial for it to be a T cell epitope, however it is not sufficient for T cell recognition (Sidney et al., 2008a). The expression and processing of proteins is as important as effects of competition for HLA binding between peptides and the stability of the formed HLA complex. On the other side, the host T cell repertoire determines if a presented peptide can be recognized and if immune responses are induced. Further, the presented peptide needs to be different from endogenous peptides, as well as to be

presented in a stimulatory environment in order to elicit the immune responses desired for target cell killing. All these aspects are currently not heuristically considered by prediction methods. Thus, the functional characterization of identified HPV16 E6-/E7-derived HLA ligands was required in order to identify T cell epitopes.

To assess HPV16 E6- and E7-derived T cell epitopes, large numbers of PBMCs were required. As using buffy coats did not allow multiple blood drawings from the same donors, experiments were limited by the number of PBMCs isolated from a buffy coat. In order to perform consecutive experiments, it was necessary to work with varying numbers of frozen cell aliquots. However, handling of cells during freezing and thawing may influence cell viability or function (Owen et al., 2007) Also in our hands, analysis of one donor, who did not mediate target cell killing in cytotoxicity assays albeit demonstrating IFN γ -secretion in ELISpot assays, suggested that cytokine responses of frozen cells were impaired.

To determine the HLA type of donors, staining with HLA-specific conjugated antibodies and flow cytometry analysis were performed, as this method is suitable for work with low PBMC numbers. The limited availability of labeled HLA-type-specific antibodies without cross-reactivity narrowed the reliable characterization to three types: A2, A3 and B7. HLA type frequencies did not perfectly match with the reported distribution in the German population of 49.9%, 28.6% and 24.5%, respectively (González-Galarza et al., 2015), which is most likely due to the relatively small number of tested samples. Incomplete HLA-typing did not allow excluding cross-reactions associated with peptide presentation by uncharacterized HLA molecules. With complete HLA typing, e.g. by PCR of amplified HLA class I gene loci using HLA type-specific primers, cross-reactivity can be ruled out and more HLA types can be investigated in the future.

Earlier experiments with several voluntary HLA-typed blood donors showed no T cell reactivity against HPV16 E6-/E7-derived peptides (data not shown). As these donors were mostly young adults (age <30), this cohort had the chance of being prophylactically vaccinated against HPV16 and likelihood for exposure was low. Thus, we decided to work with buffy coat preparations of healthy blood donors. As blood donations were anonymous, HPV16 infection history was not known. To increase the chances for previous undetected or transient HPV16 infection, buffy coats from donors above the age of 40 were requested. Moreover, donors should preferably be female, because immune responses are expected to be stronger and more frequent in women (Bosch et al., 2013). However, the latter request was not always fulfilled. Although frequencies of HPV16 E6- and E7-specific T cells in the peripheral blood are known to be low, immunogenicity of target epitopes was detected in 18 of 24 analyzed donors (Youde et al., 2000; de Jong et al., 2005). This shows that our donor selection criteria are beneficial for detecting immune responses. Naturally, immune responses are varying by individual and thus, epitope immunogenicity is not likely to be the same for all donors. However, we identified immunogenic hotspot sequences in the E6 and E7 proteins, harboring epitopes detected in more than one donor. Based on shared epitope-specificity of measured memory T cell responses, peptide-

presentation on respective HLA class I molecules can be inferred. Such a conclusion is supported by the finding that the majority of HLA class I-presented viral peptides are immunogenic (Croft et al., 2019). Along the same lines, a targeted LC-MS strategy applied in our research group will additionally characterize the identified 293 HLA binders for natural translation, processing and *bona fide* HLA-presentation on HPV16⁺ target cells.

Despite donor-to-donor variations, five A2-restricted CTL epitopes derived from the HPV16 E7/7-21 hotspot region were found to mediate specific lysis of HPV16⁺ target cells. However, because of the observed differences in the capability of epitope-specific target cell killing, results need to be reproduced with cells from more donors. The used flow cytometry-based cytotoxicity assay was perfectly suited for this study, as it was designed to assess minute frequencies of CD8⁺ T cells' cytolytic function with 30 times higher sensitivity than the standard ⁵¹chromium-release assay (Stanke et al., 2010). This assay reduces the number of required input cells by co-incubation of effector cells together with only 1x10³ specific and 1x10³ unspecific target cells. For the purpose of this thesis, A2⁺ HPV16⁺ CaSki cells represented ideal specific target cells. The presentation of 11 HPV16 E6-/E7-derived epitopes on HLA-A*02:01 molecules on the surface of CaSki cells has already been proven by mass spectrometry (Blatnik et al., 2018). Out of the peptides tested for cytotoxicity induction in this thesis, the three peptides E7/11-19, E7/80-90 and E7/81-90 were reliably detected by MS on CaSki, and three other peptides (E7/11-20, E7/11-21 and E7/12-19) were found at the limit of detection. However, the peptides E7/12-20, E6/25-33 and the H85Q, L90V-variant-derived E6/83-90 peptide were not yet reliably detected to be displayed by HLA-A2 on CaSki cells. Thus, not all of the tested peptides were known to be presented on the specific target cells prior to cytotoxicity assays. In contrast, unspecific target cells should not present the specific epitopes, but should ideally be otherwise identical to the specific target cells. Currently, C33A is the only available HPV16⁻ cervical cancer cell line, and luckily it is HLA-A2⁺ (Yee et al., 1985). However, C33A cells grow faster than CaSki cells (doubling time 1.26 days vs. 3.2 days according to Cellosaurus) (Bairoch, 2018). In order to compensate different proliferation kinetics, specific killing was calculated based on the ratios between specific and unspecific target cells relative to co-cultures without effector cells. However, an unspecific decrease in C33A cell numbers was observed at high E:T ratios. It remains unclear if this is either an effect of unspecific killing or of undernutrition. Alternatively, HLA-matched peptide-pulsed B cells could be used as target cells in future assays. Autologous B cell lines can be generated from isolated PBMCs, but the approach would consume the anyway limited number of PBMCs and the process requires at least 14 days prior to starting the T cell line culture (von Bergwelt-Baildon et al., 2002; Liebig et al., 2009). Allogenic B-LCLs are easier to culture and would represent uniform controls over multiple donors, but as they are immortalized by EBV-transformation, unspecific killing by EBV-specific donor T cells might occur (Tosato and Cohen, 2007). However, the use of either B cells would allow investigating additional HLA types.

Previously, 89 different HPV16 E6-/E7-derived epitopes were reported in IEDB in context of the investigated HLA types. Overall, 102 peptide-HLA combinations were investigated, as some epitopes were associated with multiple HLA types. Of these, twelve combinations were not assessed in this study, due to negative binding predictions. Moreover, 25 reported HLA epitopes, of which 21 were described to be HLA ligands, did not demonstrate HLA binding in competition-based binding assays and were thus not functionally characterized. A special case is the peptide E7/78-86, which was characterized to be a binder in earlier experiments, but did not exhibit binding in later experiments. Hence, it was considered to be a nonbinder in the course of this thesis. Nonetheless it was detected to be presented on CaSki cells by targeted LC-MS (Blatnik et al., 2018) and showed immunogenicity in previous IFN γ -ELISpot assays. Thus, this peptide likely is a true epitope. Taken together, the majority of IEDB-registered HLA epitopes (64 of 102) was validated. More importantly, binding assays identified 229 HLA ligands that were not previously reported. Additionally, IFN γ -ELISpot assays revealed 13 A3 epitopes, of which only one was known to induce IFN γ -responses before. For A2, 18 IFN γ -inducing epitopes were identified, of which 8 were reported before, but 10 were novel. Also the cytotoxicity-mediating capacity of two out of five A2 epitopes has been demonstrated for the first time. Consequently, new promising target epitopes for HPV16 immunotherapy were identified in the course of this thesis.

The overall aim of the research group is to develop a therapeutic vaccine against HPV16 based on targeting the oncogenic proteins E6 and E7. The described HPV16 target identification approach pursued by the research group represents an approach often described as “reverse immunology” (Celis et al., 1994; Boon and van der Bruggen, 1996). First, potential HLA class I ligand peptides from a protein sequence are predicted *in silico* based on known binding motifs. Subsequently, candidate peptides are synthesized and actual HLA binding and induction of T cell immune responses is tested *in vitro* in order to identify targets for vaccine development. Our recently published achievements proved this strategy to be suitable for the detection of immunogenic HLA-A2-presented HPV16 E6 and E7 peptides (Blatnik et al., 2018). As an essential part of this approach, numerous E6- and E7-derived HLA ligands, A2- and A3-associated T-cell epitopes and five A2-restricted CTL epitopes were identified in the scope of this thesis. Thus, among all possible antigen-derived peptides, promising candidates for development of a therapeutic HPV16 vaccine were successfully characterized. Especially immunogenic peptides shared by multiple donors represent promising candidates for vaccine development. Moreover, the results obtained in this thesis imply that studying HPV16 E6-/E7-variants is clinically relevant. In HPV lesions, typically the HPV types but not the exact genomic variants are determined. Consequently, immunotherapy should ideally be focused on epitopes which are conserved between HPV16 variants. As outlined above, functional epitope-specific vaccination is dependent on the HLA types of patients. To avoid HLA-typing of patients, vaccine formulations can contain “promiscuous” epitopes, which are epitopes binding to multiple HLA types, or combine

epitopes for HLA supertypes. As described above, the selected alleles in this study have a high population frequency and are the representatives of supertypes. A vaccine combining epitopes for each of the five supertypes covered herein is expected to provide a population coverage of $\geq 95\%$ (Reche and Reinherz, 2007). Some of the candidate epitopes validated in the scope of this thesis were already analyzed in various vaccine formulations tested in a HLA-humanized A2.DR1 mouse model (Kruse et al., 2019). In this preclinical study, prophylactic and therapeutic epitope-specific vaccination provided survival benefits as well as anti-tumor effects in A2.DR1-mice challenged with PAP-A2 tumor cells. To further assess clinical relevance and therapeutic potential at a precancerous stage, the described epitope identification strategy will be extended to the investigation of clinical samples of CIN patients. In summary, this study describes the identification of a part of an epitome map for HPV16 E6- and E7-derived targets with clear translational potential. Knowledge on validated HLA-specific targets allows developing of various immunotherapies. The characterized HPV16 E6-/E7-derived epitopes can be employed for immunomonitoring, finding of epitope-specific TCRs for engineered T cell therapy, stimulation of isolated TILs for adoptive cell transfer and, obviously, for peptide-based vaccination approaches.

6 References

- Abbate, E.A., Berger, J.M., and Botchan, M.R. (2004). “The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2.” *Genes Dev.* 18, 1981–1996.
- Abelin, J.G., Keskin, D.B., Sarkizova, S., Hartigan, C.R., Zhang, W., Sidney, J., Stevens, J., Lane, W., Zhang, G.L., Eisenhaure, T.M., et al. (2017). “Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction.” *Immunity* 46, 315–326.
- Aggarwal, C., Cohen, R.B., Morrow, M.P., Kraynyak, K.A., Sylvester, A.J., Knoblock, D.M., Bauml, J.M., Weinstein, G.S., Lin, A., Boyer, J., et al. (2019). “Immunotherapy Targeting HPV16/18 Generates Potent Immune Responses in HPV-Associated Head and Neck Cancer.” *Clin. Cancer Res.* 25, 110–124.
- Ahmed, S., and Rai, K.R. (2003). “Interferon in the treatment of hairy-cell leukemia.” *Best Pract. Res. Clin. Haematol.* 16, 69–81.
- Andreatta, M., and Nielsen, M. (2016). “Gapped sequence alignment using artificial neural networks: application to the MHC class I system.” *Bioinformatics* 32, 511–517.
- Androphy, E.J., Hubbert, N.L., Schiller, J.T., and Lowy, D.R. (1987). “Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines.” *EMBO J.* 6, 989–992.
- Antonsson, A., Payne, E., Hengst, K., and McMillan, N.A.J. (2006). “The human papillomavirus type 16 E7 protein binds human interferon regulatory factor-9 via a novel PEST domain required for transformation.” *J. Interferon Cytokine Res.* 26, 455–461.
- Arbyn, M., Xu, L., Simoons, C., and Martin-Hirsch, P.P. (2018). “Prophylactic vaccination against human papillomaviruses to prevent cervical cancer and its precursors.” *Cochrane Database Syst. Rev.* 5, CD009069.
- Bagarazzi, M.L., Yan, J., Morrow, M.P., Shen, X., Parker, R.L., Lee, J.C., Giffear, M., Pankhong, P., Khan, A.S., Broderick, K.E., et al. (2012). “Immunotherapy against HPV16/18 generates potent TH1 and cytotoxic cellular immune responses.” *Sci. Transl. Med.* 4, 155ra138.
- Bairoch, A. (2018). “The Cellosaurus, a Cell-Line Knowledge Resource.” *J. Biomol. Tech.* 29, 25–38.
- Baldwin, P.J., van der Burg, S.H., Boswell, C.M., Offringa, R., Hickling, J.K., Dobson, J., Roberts, J.S.C., Latimer, J.A., Moseley, R.P., Coleman, N., et al. (2003). “Vaccinia-expressed human papillomavirus 16 and 18 e6 and e7 as a therapeutic vaccination for vulval and vaginal intraepithelial neoplasia.” *Clin. Cancer Res.* 9, 5205–5213.
- Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., et al. (2016). “Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry.” *Nat. Commun.* 7, 13404.
- Bassani-Sternberg, M., Chong, C., Guillaume, P., Solleder, M., Pak, H., Gannon, P.O., Kandalaf, L.E., Coukos, G., and Gfeller, D. (2017). “Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosteric regulating HLA specificity.” *PLoS Comput. Biol.* 13, e1005725.
- Bauer, M., Wagner, H., and Lipford, G.B. (2000). “HPV type 16 protein E7 HLA-A2 binding peptides are immunogenic but not processed and presented.” *Immunol. Lett.* 71, 55–59.
- von Bergwelt-Baildon, M.S., Vonderheide, R.H., Maecker, B., Hirano, N., Anderson, K.S., Butler, M.O., Xia, Z., Zeng, W.Y., Wucherpfennig, K.W., Nadler, L.M., et al. (2002). “Human primary and memory cytotoxic T lymphocyte responses are efficiently induced by means of CD40-activated B cells as antigen-presenting cells: potential for clinical application.” *Blood* 99, 3319–3325.

- Besser, M.J., Shapira-Frommer, R., Itzhaki, O., Treves, A.J., Zippel, D.B., Levy, D., Kubi, A., Shoshani, N., Zikich, D., Ohayon, Y., et al. (2013). “Adoptive transfer of tumor-infiltrating lymphocytes in patients with metastatic melanoma: intent-to-treat analysis and efficacy after failure to prior immunotherapies.” *Clin. Cancer Res.* 19, 4792–4800.
- Bhasin, M., and Raghava, G.P.S. (2004). “Analysis and prediction of affinity of TAP binding peptides using cascade SVM.” *Protein Sci.* 13, 596–607.
- Blatnik, R., Mohan, N., Bonsack, M., Falkenby, L.G., Hoppe, S., Josef, K., Steinbach, A., Becker, S., Nadler, W.M., Rucevic, M., et al. (2018). “A Targeted LC-MS Strategy for Low-Abundant HLA Class-I-Presented Peptide Detection Identifies Novel Human Papillomavirus T-Cell Epitopes.” *Proteomics* 18, e1700390.
- Boon, T., and van der Bruggen, P. (1996). “Human tumor antigens recognized by T lymphocytes.” *J. Exp. Med.* 183, 725–729.
- Borst, J., Ahrends, T., Bąbała, N., Melief, C.J.M., and Kastenmüller, W. (2018). “CD4+ T cell help in cancer immunology and immunotherapy.” *Nat. Rev. Immunol.* 18, 635–647.
- Bosch, F.X., Broker, T.R., Forman, D., Moscicki, A.-B., Gillison, M.L., Doorbar, J., Stern, P.L., Stanley, M., Arbyn, M., Poljak, M., et al. (2013). “Comprehensive control of human papillomavirus infections and related diseases.” *Vaccine* 31 Suppl 7, H1-31.
- Bourgault Villada, I., Moyal Barracco, M., Berville, S., Bafounta, M.L., Longvert, C., Prémel, V., Villefroy, P., Jullian, E., Clerici, T., Paniel, B., et al. (2010). “Human papillomavirus 16-specific T cell responses in classic HPV-related vulvar intra-epithelial neoplasia. Determination of strongly immunogenic regions from E6 and E7 proteins.” *Clin. Exp. Immunol.* 159, 45–56.
- Bradley, A.P. (1997). “The use of the area under the ROC curve in the evaluation of machine learning algorithms.” *Pattern Recognit.* 30, 1145–1159.
- Bravo, I.G., de Sanjosé, S., and Gottschling, M. (2010). “The clinical importance of understanding the evolution of papillomaviruses.” *Trends Microbiol.* 18, 432–438.
- Brown, D.R., Kitchin, D., Qadadri, B., Neptune, N., Batteiger, T., and Ermel, A. (2006). “The human papillomavirus type 11 E1–E4 protein is a transglutaminase 3 substrate and induces abnormalities of the cornified cell envelope.” *Virology* 345, 290–298.
- Brun, J.L., Dalstein, V., Leveque, J., Mathevet, P., Raulic, P., Baldauf, J.J., Scholl, S., Huynh, B., Douvier, S., Riethmuller, D., et al. (2011). “Regression of high-grade cervical intraepithelial neoplasia with TG4001 targeted immunotherapy.” *Am. J. Obstet. Gynecol.* 204, 169.e1-169.e8.
- Bruni, L., Diaz, M., Barrionuevo-Rosas, L., Herrero, R., Bray, F., Bosch, F.X., de Sanjosé, S., and Castellsagué, X. (2016). “Global estimates of human papillomavirus vaccination coverage by region and income level: a pooled analysis.” *Lancet. Glob. Heal.* 4, e453-63.
- Buck, C.B., Day, P.M., and Trus, B.L. (2013). “The papillomavirus major capsid protein L1.” *Virology* 445, 169–174.
- van der Burg, S.H., Arens, R., Ossendorp, F., van Hall, T., and Melief, C.J.M. (2016). “Vaccines for established cancer: overcoming the challenges posed by immune evasion.” *Nat. Rev. Cancer* 16, 219–233.
- Burgers, W.A., Blanchon, L., Pradhan, S., de Launoit, Y., Kouzarides, T., and Fuks, F. (2007). “Viral oncoproteins target the DNA methyltransferases.” *Oncogene* 26, 1650–1655.
- Burk, R.D., Harari, A., and Chen, Z. (2013). “Human papillomavirus genome variants.” *Virology* 445, 232–243.
- Calis, J.J.A., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Keşmir, C., and Peters, B. (2013). “Properties of MHC class I presented peptides that enhance immunogenicity.” *PLoS Comput. Biol.* 9, e1003266.

- Cameron, B.J., Gerry, A.B., Dukes, J., Harper, J. V., Kannan, V., Bianchi, F.C., Grand, F., Brewer, J.E., Gupta, M., Plesa, G., et al. (2013). "Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells." *Sci. Transl. Med.* 5, 197ra103.
- Campos, S.K. (2017). "Subcellular Trafficking of the Papillomavirus Genome during Initial Infection: The Remarkable Abilities of Minor Capsid Protein L2." *Viruses* 9.
- Carpino, L.A., and Han, G.Y. (1972). "9-Fluorenylmethoxycarbonyl amino-protecting group." *J. Org. Chem.* 37, 3404–3409.
- Castellsagué, X., Alemany, L., Quer, M., Halc, G., Quirós, B., Tous, S., Clavero, O., Alòs, L., Biegner, T., Szafarowski, T., et al. (2016). "HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients." *J. Natl. Cancer Inst.* 108, djv403.
- Celis, E., Tsai, V., Crimi, C., DeMars, R., Wentworth, P.A., Chesnut, R.W., Grey, H.M., Sette, A., and Serra, H.M. (1994). "Induction of anti-tumor cytotoxic T lymphocytes in normal humans using primary cultures and synthetic peptide epitopes." *Proc. Natl. Acad. Sci. U. S. A.* 91, 2105–2109.
- Chabeda, A., Yanez, R.J.R., Lamprecht, R., Meyers, A.E., Rybicki, E.P., and Hitzeroth, I.I. (2018). "Therapeutic vaccines for high-risk HPV-associated diseases." *Papillomavirus Res. (Amsterdam, Netherlands)* 5, 46–58.
- Chiang, C.L.-L., Coukos, G., and Kandalaft, L.E. (2015). "Whole Tumor Antigen Vaccines: Where Are We?" *Vaccines* 3, 344–372.
- Chiang, C.M., Ustav, M., Stenlund, A., Ho, T.F., Broker, T.R., and Chow, L.T. (1992). "Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins." *Proc. Natl. Acad. Sci. U. S. A.* 89, 5799–5803.
- Chow, L.T., Broker, T.R., and Steinberg, B.M. (2010). "The natural history of human papillomavirus infections of the mucosal epithelia." *APMIS* 118, 422–449.
- Chowell, D., Krishna, S., Becker, P.D., Cocita, C., Shu, J., Tan, X., Greenberg, P.D., Klavinskis, L.S., Blattman, J.N., and Anderson, K.S. (2015). "TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes." *Proc. Natl. Acad. Sci. U. S. A.* 112, E1754-62.
- Cicchini, L., Westrich, J.A., Xu, T., Vermeer, D.W., Berger, J.N., Clambey, E.T., Lee, D., Song, J.I., Lambert, P.F., Greer, R.O., et al. (2016). "Suppression of Antitumor Immune Responses by Human Papillomavirus through Epigenetic Downregulation of CXCL14." *MBio* 7, 1–13.
- Clifford, G.M., Tenet, V., Georges, D., Alemany, L., Pavón, M.A., Chen, Z., Yeager, M., Cullen, M., Boland, J.F., Bass, S., et al. (2019). "Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: Whole viral genome sequences from 7116 HPV16-positive women." *Papillomavirus Res. (Amsterdam, Netherlands)* 7, 67–74.
- Creech, A.L., Ting, Y.S., Goulding, S.P., Sauld, J.F.K., Barthelme, D., Rooney, M.S., Addona, T.A., and Abelin, J.G. (2018). "The Role of Mass Spectrometry and Proteogenomics in the Advancement of HLA Epitope Prediction." *Proteomics* 1700259, 1–10.
- 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., et al. (2019). "Most viral peptides displayed by class I MHC on infected cells are immunogenic." *Proc. Natl. Acad. Sci. U. S. A.* 116, 3112–3117.
- Daayana, S., Elkord, E., Winters, U., Pawlita, M., Roden, R., Stern, P.L., and Kitchener, H.C. (2010). "Phase II trial of imiquimod and HPV therapeutic vaccination in patients with vulval intraepithelial neoplasia." *Br. J. Cancer* 102, 1129–1136.
- Day, P.M., Roden, R.B., Lowy, D.R., and Schiller, J.T. (1998). "The papillomavirus minor capsid protein, L2, induces localization of the major capsid protein, L1, and the viral transcription/replication protein, E2, to PML oncogenic domains." *J. Virol.* 72, 142–150.

- Day, P.M., Lowy, D.R., and Schiller, J.T. (2003). "Papillomaviruses infect cells via a clathrin-dependent pathway." *Virology* 307, 1–11.
- Day, P.M., Weisberg, A.S., Thompson, C.D., Hughes, M.M., Pang, Y.Y., Lowy, D.R., and Schiller, J.T. (2019). "Human papillomavirus type 16 (HPV16) capsids mediate nuclear entry during infection." *J. Virol.*
- Dell, G., Wilkinson, K.W., Tranter, R., Parish, J., Leo Brady, R., and Gaston, K. (2003). "Comparison of the structure and DNA-binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus." *J. Mol. Biol.* 334, 979–991.
- Dendrou, C.A., Petersen, J., Rossjohn, J., and Fugger, L. (2018). "HLA variation and disease." *Nat. Rev. Immunol.* 18, 325–339.
- Deschuyteneer, M., Elouahabi, A., Plainchamp, D., Plisnier, M., Soete, D., Corazza, Y., Lockman, L., Giannini, S., and Deschamps, M. (2010). "Molecular and structural characterization of the L1 virus-like particles that are used as vaccine antigens in Cervarix™, the AS04-adjuvanted HPV-16 and -18 cervical cancer vaccine." *Hum. Vaccin.* 6, 407–419.
- Doorbar, J. (2013). "The E4 protein; structure, function and patterns of expression." *Virology* 445, 80–98.
- Doorbar, J., Ely, S., Sterling, J., McLean, C., and Crawford, L. (1991). "Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network." *Nature* 352, 824–827.
- Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., and Murakami, I. (2015). "Human papillomavirus molecular biology and disease association." *Rev. Med. Virol.* 25 Suppl 1, 2–23.
- Duan, F., Duitama, J., Al Seesi, S., Ayres, C.M., Corcelli, S.A., Pawashe, A.P., Blanchard, T., McMahan, D., Sidney, J., Sette, A., et al. (2014). "Genomic and bioinformatic profiling of mutational neopeptides reveals new rules to predict anticancer immunogenicity." *J. Exp. Med.* 211, 2231–2248.
- Dürst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H. (1983). "A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions." *Proc. Natl. Acad. Sci. U. S. A.* 80, 3812–3815.
- Egawa, N., Nakahara, T., Ohno, S.-I., Narisawa-Saito, M., Yugawa, T., Fujita, M., Yamato, K., Natori, Y., and Kiyono, T. (2012). "The E1 protein of human papillomavirus type 16 is dispensable for maintenance replication of the viral genome." *J. Virol.* 86, 3276–3283.
- Egawa, N., Egawa, K., Griffin, H., and Doorbar, J. (2015). "Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia." *Viruses* 7, 3863–3890.
- Eklund, C. (Karolinska I., and Dillner, J. (Karolinska I. (2019). "International Human Papillomavirus Reference Center" Accessed: 06/18/2019 via: "https://www.hpvcenter.se/human_reference_clones/."
- Engels, B., Engelhard, V.H., Sidney, J., Sette, A., Binder, D.C., Liu, R.B., Kranz, D.M., Meredith, S.C., Rowley, D.A., and Schreiber, H. (2013). "Relapse or eradication of cancer is predicted by peptide-major histocompatibility complex affinity." *Cancer Cell* 23, 516–526.
- Evans, M., Borysiewicz, L.K., Evans, A.S., Rowe, M., Jones, M., Gileadi, U., Cerundolo, V., and Man, S. (2001). "Antigen processing defects in cervical carcinomas limit the presentation of a CTL epitope from human papillomavirus 16 E6." *J. Immunol.* 167, 5420–5428.
- Falk, K., Rötzschke, O., Stevanović, S., Jung, G., and Rammensee, H.G. (1991). "Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules." *Nature* 351, 290–296.

- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012." *Int. J. Cancer* 136, E359-86.
- Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D.M., Piñeros, M., Znaor, A., and Bray, F. (2019). "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods." *Int. J. Cancer* 144, 1941–1953.
- Feuerstein, B., Berger, T.G., Maczek, C., Röder, C., Schreiner, D., Hirsch, U., Haendle, I., Leisgang, W., Glaser, A., Kuss, O., et al. (2000). "A method for the production of cryopreserved aliquots of antigen-preloaded, mature dendritic cells ready for clinical use." *J. Immunol. Methods* 245, 15–29.
- Finn, O.J., and Rammensee, H.-G. (2018). "Is It Possible to Develop Cancer Vaccines to Neoantigens, What Are the Major Challenges, and How Can These Be Overcome? Neoantigens: Nothing New in Spite of the Name." *Cold Spring Harb. Perspect. Biol.* 10.
- Folschweiller, N., Behre, U., Dionne, M., Durando, P., Esposito, S., Ferguson, L., Ferguson, M., Hillemanns, P., McNeil, S.A., Peters, K., et al. (2019). "Long-term Cross-reactivity Against Nonvaccine Human Papillomavirus Types 31 and 45 After 2- or 3-Dose Schedules of the AS04-Adjuvanted Human HPV-16/18 Vaccine." *J. Infect. Dis.* 219, 1799–1803.
- Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E., and Wiley, D.C. (1996). "Structure of the complex between human T-cell receptor, viral peptide and HLA-A2." *Nature* 384, 134–141.
- Garcia, K.C., Degano, M., Pease, L.R., Huang, M., Peterson, P.A., Teyton, L., and Wilson, I.A. (1998). "Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen." *Science* 279, 1166–1172.
- Garg, A.D., Coulie, P.G., Van den Eynde, B.J., and Agostinis, P. (2017). "Integrating Next-Generation Dendritic Cell Vaccines into the Current Cancer Immunotherapy Landscape." *Trends Immunol.* 38, 577–593.
- Ghosh, M., Di Marco, M., and Stevanović, S. (2019). "Identification of MHC Ligands and Establishing MHC Class I Peptide Motifs." *Methods Mol. Biol.* 1988, 137–147.
- Giguère, S., Drouin, A., Lacoste, A., Marchand, M., Corbeil, J., and Laviolette, F. (2013). "MHC-NP: predicting peptides naturally processed by the MHC." *J. Immunol. Methods* 400–401, 30–36.
- Gissmann, L., Pfister, H., and Zur Hausen, H. (1977). "Human papilloma viruses (HPV): characterization of four different isolates." *Virology* 76, 569–580.
- González-Galarza, F.F., Takeshita, L.Y.C., Santos, E.J.M., Kempson, F., Maia, M.H.T., da Silva, A.L.S., Teles e Silva, A.L., Ghattaoraya, G.S., Alfievic, A., Jones, A.R., et al. (2015). "Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations." *Nucleic Acids Res.* 43, D784-8.
- Gowthaman, U., Chodiseti, S.B., Parihar, P., and Agrewala, J.N. (2010). "Evaluation of different generic in silico methods for predicting HLA class I binding peptide vaccine candidates using a reverse approach." *Amino Acids* 39, 1333–1342.
- Grabowska, A.K., and Riemer, A.B. (2012). "The invisible enemy - how human papillomaviruses avoid recognition and clearance by the host immune system." *Open Virol. J.* 6, 249–256.
- Gulley, J.L. (2013). "Therapeutic vaccines: the ultimate personalized therapy?" *Hum. Vaccin. Immunother.* 9, 219–221.
- Hagensee, M.E., Yaegashi, N., and Galloway, D.A. (1993). "Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins." *J. Virol.* 67, 315–322.
- Van Hall, T., and Van der Burg, S.H. (2012). *Mechanisms of Peptide Vaccination in Mouse Models. Tolerance, Immunity, and Hyperreactivity.* (Elsevier Inc.).

- Han, Y., and Kim, D. (2017). “Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction.” *BMC Bioinformatics* 18, 585.
- Hanahan, D., and Weinberg, R.A. (2000). “The hallmarks of cancer.” *Cell* 100, 57–70.
- Hanahan, D., and Weinberg, R.A. (2011). “Hallmarks of cancer: the next generation.” *Cell* 144, 646–674.
- Hara, M., Matsueda, S., Tamura, M., Takedatsu, H., Tanaka, M., Kawano, K., Mochizuki, K., Kamura, T., Itoh, K., and Harada, M. (2005). “Identification of human papillomavirus 16-E6 protein-derived peptides with the potential to generate cytotoxic T-lymphocytes toward human leukocyte antigen-A24⁺ cervical cancer.” *Int. J. Oncol.* 27, 1371–1379.
- Harndahl, M., Justesen, S., Lamberth, K., Røder, G., Nielsen, M., and Buus, S. (2009). “Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays.” *J. Biomol. Screen.* 14, 173–180.
- Hasan, U.A., Zannetti, C., Parroche, P., Goutagny, N., Malfroy, M., Roblot, G., Carreira, C., Hussain, I., Müller, M., Taylor-Papadimitriou, J., et al. (2013). “The human papillomavirus type 16 E7 oncoprotein induces a transcriptional repressor complex on the Toll-like receptor 9 promoter.” *J. Exp. Med.* 210, 1369–1387.
- zur Hausen, H. (1996). “Papillomavirus infections-a major cause of human cancers.” *Biochim. Biophys. Acta* 1288, F55-78.
- zur Hausen, H. (2002). “Papillomaviruses and cancer: from basic studies to clinical application.” *Nat. Rev. Cancer* 2, 342–350.
- Hoffmann, R., Hirt, B., Bechtold, V., Beard, P., and Raj, K. (2006). “Different modes of human papillomavirus DNA replication during maintenance.” *J. Virol.* 80, 4431–4439.
- Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., and Nielsen, M. (2009). “NetMHCpan, a method for MHC class I binding prediction beyond humans.” *Immunogenetics* 61, 1–13.
- Horvath, C.A.J., Boulet, G.A. V, Renoux, V.M., Delvenne, P.O., and Bogers, J.-P.J. (2010). “Mechanisms of cell entry by human papillomaviruses: an overview.” *Virol. J.* 7, 11.
- Hughes, M.S., Yu, Y.Y.L., Dudley, M.E., Zheng, Z., Robbins, P.F., Li, Y., Wunderlich, J., Hawley, R.G., Moayeri, M., Rosenberg, S.A., et al. (2005). “Transfer of a TCR gene derived from a patient with a marked antitumor response conveys highly active T-cell effector functions.” *Hum. Gene Ther.* 16, 457–472.
- Huh, W.K., Joura, E.A., Giuliano, A.R., Iversen, O.-E., de Andrade, R.P., Ault, K.A., Bartholomew, D., Cestero, R.M., Fedrizzi, E.N., Hirschberg, A.L., et al. (2017). “Final efficacy, immunogenicity, and safety analyses of a nine-valent human papillomavirus vaccine in women aged 16-26 years: a randomised, double-blind trial.” *Lancet (London, England)* 390, 2143–2159.
- ICO/IARC HPV Information Centre (2019). “Human Papillomavirus and Related Diseases Report” Accessed: 06/18/2019 via: “<https://hpvcentre.net/statistics/reports/XWX.pdf>.”
- ICTV International Committee on Taxonomy of Viruses (2018). “Virus Taxonomy: 2018b Release” Accessed: 06/27/2019 via: “https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/121/papillomaviridae.”
- Johansson, C., Somberg, M., Li, X., Backström Winquist, E., Fay, J., Ryan, F., Pim, D., Banks, L., and Schwartz, S. (2012). “HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation.” *EMBO J.* 31, 3212–3227.
- de Jong, A., van der Hulst, J.M., Kenter, G.G., Drijfhout, J.W., Franken, K.L.M.C., Vermeij, P., Offringa, R., van der Burg, S.H., and Melief, C.J.M. (2005). “Rapid enrichment of human papillomavirus (HPV)-specific polyclonal T cell populations for adoptive immunotherapy of cervical cancer.” *Int. J. Cancer* 114, 274–282.

- Jonuleit, H., Kühn, U., Müller, G., Steinbrink, K., Paragnik, L., Schmitt, E., Knop, J., and Enk, A.H. (1997). "Pro-inflammatory cytokines and prostaglandins induce maturation of potent immunostimulatory dendritic cells under fetal calf serum-free conditions." *Eur. J. Immunol.* 27, 3135–3142.
- Jørgensen, K.W., Rasmussen, M., Buus, S., and Nielsen, M. (2014). "NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery." *Immunology* 141, 18–26.
- Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). "NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data." *J. Immunol.* 199, 3360–3368.
- Karbach, J., Neumann, A., Atmaca, A., Wahle, C., Brand, K., von Boehmer, L., Knuth, A., Bender, A., Ritter, G., Old, L.J., et al. (2011). "Efficient in vivo priming by vaccination with recombinant NY-ESO-1 protein and CpG in antigen naive prostate cancer patients." *Clin. Cancer Res.* 17, 861–870.
- Karosiene, E., Lundegaard, C., Lund, O., and Nielsen, M. (2012). "NetMHCcons: a consensus method for the major histocompatibility complex class I predictions." *Immunogenetics* 64, 177–186.
- Kaufmann, A.M., Stern, P.L., Rankin, E.M., Sommer, H., Nuessler, V., Schneider, A., Adams, M., Onon, T.S., Bauknecht, T., Wagner, U., et al. (2002). "Safety and immunogenicity of TA-HPV, a recombinant vaccinia virus expressing modified human papillomavirus (HPV)-16 and HPV-18 E6 and E7 genes, in women with progressive cervical cancer." *Clin. Cancer Res.* 8, 3676–3685.
- Kenter, G.G., Welters, M.J.P., Valentijn, A.R.P.M., Lowik, M.J.G., Berends-van der Meer, D.M.A., Vloon, A.P.G., Essahsah, F., Fathers, L.M., Offringa, R., Drijfhout, J.W., et al. (2009). "Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia." *N. Engl. J. Med.* 361, 1838–1847.
- Kessler, J.H., Mommaas, B., Mutis, T., Huijbers, I., Vissers, D., Benckhuijsen, W.E., Schreuder, G.M.T., Offringa, R., Goulmy, E., Melief, C.J.M., et al. (2003). "Competition-based cellular peptide binding assays for 13 prevalent HLA class I alleles using fluorescein-labeled synthetic peptides." *Hum. Immunol.* 64, 245–255.
- Kessler, J.H., Benckhuijsen, W.E., Mutis, T., Melief, C.J.M., van der Burg, S.H., and Drijfhout, J.W. (2004). "Competition-based cellular peptide binding assay for HLA class I." *Curr. Protoc. Immunol.* Chapter 18, Unit 18.12.
- Khallouf, H., Grabowska, A.K., and Riemer, A.B. (2014). "Therapeutic Vaccine Strategies against Human Papillomavirus." *Vaccines* 2, 422–462.
- Kim, Y., Sidney, J., Pinilla, C., Sette, A., and Peters, B. (2009). "Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior." *BMC Bioinformatics* 10, 394.
- Klaes, R., Friedrich, T., Spitkovsky, D., Ridder, R., Rudy, W., Petry, U., Dallenbach-Hellweg, G., Schmidt, D., and von Knebel Doeberitz, M. (2001). "Overexpression of p16(INK4A) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri." *Int. J. Cancer* 92, 276–284.
- Klingelutz, A.J., Foster, S.A., and McDougall, J.K. (1996). "Telomerase activation by the E6 gene product of human papillomavirus type 16." *Nature* 380, 79–82.
- Koşaloğlu, Z., Zörnig, I., Halama, N., Kaiser, I., Buchhalter, I., Grabe, N., Eils, R., Schlesner, M., Califano, A., and Jäger, D. (2016). "Identification of immunotherapeutic targets by genomic profiling of rectal NET metastases." *Oncoimmunology* 5, e1213931.
- Krishnamurthy, A., and Jimeno, A. (2018). "Bispecific antibodies for cancer therapy: A review." *Pharmacol. Ther.* 185, 122–134.

- Kruse, S., Büchler, M., Uhl, P., Sauter, M., Scherer, P., Lan, T.C.T., Zotnick, S., Klevenz, A., Yang, R., Rösl, F., et al. (2019). “Therapeutic vaccination using minimal HPV16 epitopes in a novel MHC-humanized murine HPV tumor model.” *Oncoimmunology* 8, e1524694.
- Kubo, R.T., Sette, A., Grey, H.M., Appella, E., Sakaguchi, K., Zhu, N.Z., Arnott, D., Sherman, N., Shabanowitz, J., and Michel, H. (1994). “Definition of specific peptide motifs for four major HLA-A alleles.” *J. Immunol.* 152, 3913–3924.
- Kuttler, C., Nussbaum, A.K., Dick, T.P., Rammensee, H.G., Schild, H., and Haderl, K.P. (2000). “An algorithm for the prediction of proteasomal cleavages.” *J. Mol. Biol.* 298, 417–429.
- Lacey, C.J.N., Lowndes, C.M., and Shah, K. V. (2006). “Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease.” *Vaccine* 24 Suppl 3, S3/35-41.
- Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., and Nielsen, M. (2007). “Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction.” *BMC Bioinformatics* 8, 424.
- Laurson, J., Khan, S., Chung, R., Cross, K., and Raj, K. (2010). “Epigenetic repression of E-cadherin by human papillomavirus 16 E7 protein.” *Carcinogenesis* 31, 918–926.
- Lenz, P., Thompson, C.D., Day, P.M., Bacot, S.M., Lowy, D.R., and Schiller, J.T. (2003). “Interaction of papillomavirus virus-like particles with human myeloid antigen-presenting cells.” *Clin. Immunol.* 106, 231–237.
- Lepique, A.P., Daghestanli, K.R.P., Cuccovia, I.M., and Villa, L.L. (2009). “HPV16 tumor associated macrophages suppress antitumor T cell responses.” *Clin. Cancer Res.* 15, 4391–4400.
- Liebig, T.M., Fiedler, A., Zoghi, S., Shimabukuro-Vornhagen, A., and von Bergwelt-Baildon, M.S. (2009). “Generation of human CD40-activated B cells.” *J. Vis. Exp.* 3–5.
- Lin, H.H., Ray, S., Tongchusak, S., Reinherz, E.L., and Brusica, V. (2008). “Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research.” *BMC Immunol.* 9, 8.
- Liu, G., Li, D., Li, Z., Qiu, S., Li, W., Chao, C.-C., Yang, N., Li, H., Cheng, Z., Song, X., et al. (2017). “PSSMHCpan: a novel PSSM-based software for predicting class I peptide-HLA binding affinity.” *Gigascience* 6, 1–11.
- Liu, Z., Cui, Y., Xiong, Z., Nasiri, A., Zhang, A., and Hu, J. (2019). “DeepSeqPan, a novel deep convolutional neural network model for pan-specific class I HLA-peptide binding affinity prediction.” *Sci. Rep.* 9, 794.
- Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., and Nielsen, M. (2008a). “NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11.” *Nucleic Acids Res.* 36, W509-12.
- Lundegaard, C., Lund, O., and Nielsen, M. (2008b). “Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers.” *Bioinformatics* 24, 1397–1398.
- Maciag, P.C., Radulovic, S., and Rothman, J. (2009). “The first clinical use of a live-attenuated *Listeria monocytogenes* vaccine: a Phase I safety study of Lm-LLO-E7 in patients with advanced carcinoma of the cervix.” *Vaccine* 27, 3975–3983.
- Maeda, E., Akahane, M., Kiryu, S., Kato, N., Yoshikawa, T., Hayashi, N., Aoki, S., Minami, M., Uozaki, H., Fukayama, M., et al. (2009). “Spectrum of Epstein-Barr virus-related diseases: a pictorial review.” *Jpn. J. Radiol.* 27, 4–19.
- de Martel, C., Plummer, M., Vignat, J., and Franceschi, S. (2017). “Worldwide burden of cancer attributable to HPV by site, country and HPV type.” *Int. J. Cancer* 141, 664–670.
- Maucort-Boulch, D., de Martel, C., Franceschi, S., and Plummer, M. (2018). “Fraction and incidence of liver cancer attributable to hepatitis B and C viruses worldwide.” *Int. J. Cancer* 142, 2471–2477.

- McBride, A.A., and Warburton, A. (2017). “The role of integration in oncogenic progression of HPV-associated cancers.” *PLOS Pathog.* 13, e1006211.
- McIntosh, P.B., Martin, S.R., Jackson, D.J., Khan, J., Isaacson, E.R., Calder, L., Raj, K., Griffin, H.M., Wang, Q., Laskey, P., et al. (2008). “Structural analysis reveals an amyloid form of the human papillomavirus type 16 E1–E4 protein and provides a molecular basis for its accumulation.” *J. Virol.* 82, 8196–8203.
- Mei, S., Li, F., Leier, A., Marquez-Lago, T.T., Giam, K., Croft, N.P., Akutsu, T., Smith, A.I., Li, J., Rossjohn, J., et al. (2019). “A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction.” *Brief. Bioinform.* 0, 1–17.
- Merrifield, R.B. (1963). “Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide.” *J. Am. Chem. Soc.* 85, 2149–2154.
- Mirabello, L., Yeager, M., Yu, K., Clifford, G.M., Xiao, Y., Zhu, B., Cullen, M., Boland, J.F., Wentzensen, N., Nelson, C.W., et al. (2017). “HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis.” *Cell* 170, 1164–1174.e6.
- Moody, C.A., and Laimins, L.A. (2010). “Human papillomavirus oncoproteins: pathways to transformation.” *Nat. Rev. Cancer* 10, 550–560.
- Morgan, R.A., Chinnasamy, N., Abate-Daga, D., Gros, A., Robbins, P.F., Zheng, Z., Dudley, M.E., Feldman, S.A., Yang, J.C., Sherry, R.M., et al. (2013). “Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy.” *J. Immunother.* 36, 133–151.
- Moutaftsi, M., Peters, B., Pasquetto, V., Tschärke, D.C., Sidney, J., Bui, H.-H., Grey, H., and Sette, A. (2006). “A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus.” *Nat. Biotechnol.* 24, 817–819.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). “Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.” *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1, 263–273.
- Murphy, K., and Weaver, C. (2017). *Janeway’s immunobiology* (New York: Garland Science/Taylor & Francis Group, LLC).
- Nardelli-Haeffliger, D., Dudda, J.C., and Romero, P. (2013). “Vaccination route matters for mucosal tumors.” *Sci. Transl. Med.* 5, 172fs4.
- Nielsen, M., and Andreatta, M. (2016). “NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets.” *Genome Med.* 8, 33.
- Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S.L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2003). “Reliable prediction of T-cell epitopes using neural networks with novel sequence representations.” *Protein Sci.* 12, 1007–1017.
- Nielsen, M., Lundegaard, C., Lund, O., and Keşmir, C. (2005). “The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage.” *Immunogenetics* 57, 33–41.
- Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., et al. (2007). “NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence.” *PLoS One* 2, e796.
- Nogueira, C., Kaufmann, J.K., Lam, H., and Flechtner, J.B. (2018). “Improving Cancer Immunotherapies through Empirical Neoantigen Selection.” *Trends in Cancer* 4, 97–100.
- O’Donnell, T.J., Rubinsteyn, A., Bonsack, M., Riemer, A.B., Laserson, U., and Hammerbacher, J. (2018). “MHCflurry: Open-Source Class I MHC Binding Affinity Prediction.” *Cell Syst.* 7, 129–132.e4.

- de Oliveira, C.M., Fregnani, J.H.T.G., and Villa, L.L. (2019). “HPV Vaccine: Updates and Highlights.” *Acta Cytol.* 63, 159–168.
- Ott, P.A., Hu, Z., Keskin, D.B., Shukla, S.A., Sun, J., Bozym, D.J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., et al. (2017). “An immunogenic personal neoantigen vaccine for patients with melanoma.” *Nature* 547, 217–221.
- Owen, R.E., Sinclair, E., Emu, B., Heitman, J.W., Hirschhorn, D.F., Epling, C.L., Tan, Q.X., Custer, B., Harris, J.M., Jacobson, M.A., et al. (2007). “Loss of T cell responses following long-term cryopreservation.” *J. Immunol. Methods* 326, 93–115.
- Paul, S., Weiskopf, D., Angelo, M.A., Sidney, J., Peters, B., and Sette, A. (2013). “HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity.” *J. Immunol.* 191, 5831–5839.
- Peters, B., and Sette, A. (2005). “Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method.” *BMC Bioinformatics* 6, 132.
- Peters, B., Bui, H.-H., Frankild, S., Nielson, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Fleri, W., et al. (2006). “A community resource benchmarking predictions of peptide binding to MHC-I molecules.” *PLoS Comput. Biol.* 2, e65.
- Phloyphisut, P., Pornputtpong, N., Sriswasdi, S., and Chuangsuwanich, E. (2019). “MHCSeqNet: a deep neural network model for universal MHC binding prediction.” *BMC Bioinformatics* 20, 270.
- Plummer, M., Franceschi, S., Vignat, J., Forman, D., and de Martel, C. (2015). “Global burden of gastric cancer attributable to *Helicobacter pylori*.” *Int. J. Cancer* 136, 487–490.
- Plummer, M., de Martel, C., Vignat, J., Ferlay, J., Bray, F., and Franceschi, S. (2016). “Global burden of cancers attributable to infections in 2012: a synthetic analysis.” *Lancet. Glob. Heal.* 4, e609-16.
- van Poelgeest, M.I.E., Welters, M.J.P., van Esch, E.M.G., Stynenbosch, L.F.M., Kerpershoek, G., van Persijn van Meerten, E.L., van den Hende, M., Löwik, M.J.G., Berends-van der Meer, D.M. a, Fathers, L.M., et al. (2013). “HPV16 synthetic long peptide (HPV16-SLP) vaccination therapy of patients with advanced or recurrent HPV16-induced gynecological carcinoma, a phase II trial.” *J. Transl. Med.* 11, 88.
- Vande Pol, S.B., and Klingelutz, A.J. (2013). “Papillomavirus E6 oncoproteins.” *Virology* 445, 115–137.
- Purcell, A.W., Ramarathinam, S.H., and Ternette, N. (2019). “Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics.” *Nat. Protoc.* 14, 1687–1707.
- Pyeon, D., Pearce, S.M., Lank, S.M., Ahlquist, P., and Lambert, P.F. (2009). “Establishment of human papillomavirus infection requires cell cycle progression.” *PLoS Pathog.* 5, e1000318.
- Rammensee, H.G. (1995). “Chemistry of peptides associated with MHC class I and class II molecules.” *Curr. Opin. Immunol.* 7, 85–96.
- Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanović, S. (1999). “SYFPEITHI: database for MHC ligands and peptide motifs.” *Immunogenetics* 50, 213–219.
- Rammensee, H.G., Falk, K., and Rötzschke, O. (1993). “Peptides naturally presented by MHC class I molecules.” *Annu. Rev. Immunol.* 11, 213–244.
- Reche, P.A., and Reinherz, E.L. (2007). “Definition of MHC supertypes through clustering of MHC peptide-binding repertoires.” *Methods Mol. Biol.* 409, 163–173.
- Reche, P.A., Glutting, J.-P., and Reinherz, E.L. (2002). “Prediction of MHC class I binding peptides using profile motifs.” *Hum. Immunol.* 63, 701–709.

- Ressing, M.E., van Driel, W.J., Brandt, R.M.P., Kenter, G.G., de Jong, J.H., Bauknecht, T., Fleuren, G.J., Hoogerhout, P., Offringa, R., Sette, A., et al. (2000). "Detection of T helper responses, but not of human papillomavirus-specific cytotoxic T lymphocyte responses, after peptide vaccination of patients with cervical carcinoma." *J. Immunother.* 23, 255–266.
- Riemer, A.B., Keskin, D.B., Zhang, G., Handley, M., Anderson, K.S., Brusica, V., Reinhold, B., and Reinherz, E.L. (2010). "A conserved E7-derived cytotoxic T lymphocyte epitope expressed on human papillomavirus 16-transformed HLA-A2+ epithelial cancers." *J. Biol. Chem.* 285, 29608–29622.
- Riley, R.S., June, C.H., Langer, R., and Mitchell, M.J. (2019). "Delivery technologies for cancer immunotherapy." *Nat. Rev. Drug Discov.* 18, 175–196.
- Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P., and Marsh, S.G.E. (2015). "The IPD and IMGT/HLA database: allele variant databases." Accessed: 06/22/2019 via: "<http://www.ncbi.nlm.nih.gov/pubmed/25414341>."
- Roman, A., and Munger, K. (2013). "The papillomavirus E7 proteins." *Virology* 445, 138–168.
- Rosenberg, S.A. (2014). "IL-2: the first effective immunotherapy for human cancer." *J. Immunol.* 192, 5451–5458.
- Sahin, U., and Türeci, Ö. (2018). "Personalized vaccines for cancer immunotherapy." *Science* 359, 1355–1360.
- Sahin, U., Derhovanesian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrörs, B., et al. (2017). "Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer." *Nature* 547, 222–226.
- Sandoval, F., Terme, M., Nizard, M., Badoual, C., Bureau, M.-F., Freyburger, L., Clement, O., Marcheteau, E., Gey, A., Fraisse, G., et al. (2013). "Mucosal imprinting of vaccine-induced CD8⁺ T cells is crucial to inhibit the growth of mucosal tumors." *Sci. Transl. Med.* 5, 172ra20.
- Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjosé, S., Fakhry, C., Monk, B.J., Stanley, M.A., and Franceschi, S. (2016). "Carcinogenic human papillomavirus infection." *Nat. Rev. Dis. Prim.* 2, 16086.
- Schoenborn, J.R., and Wilson, C.B. (2007). "Regulation of interferon-gamma during innate and adaptive immune responses." *Adv. Immunol.* 96, 41–101.
- Schroder, K., Hertzog, P.J., Ravasi, T., and Hume, D.A. (2004). "Interferon-gamma: an overview of signals, mechanisms and functions." *J. Leukoc. Biol.* 75, 163–189.
- Seedorf, K., Krämmmer, G., Dürst, M., Suhai, S., and Röwekamp, W.G. (1985). "Human papillomavirus type 16 DNA sequence." *Virology* 145, 181–185.
- Serrano, B., Brotons, M., Bosch, F.X., and Bruni, L. (2018). "Epidemiology and burden of HPV-related disease." *Best Pract. Res. Clin. Obstet. Gynaecol.* 47, 14–26.
- Sette, A., Vitiello, A., Reheman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., Oseroff, C., Yuan, L., Ruppert, J., et al. (1994). "The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes." *J. Immunol.* 153, 5586–5592.
- Shao, X.M., Bhattacharya, R., Huang, J., Sivakumar, A., Tokheim, C., Zheng, L., Kaminow, B., Omdahl, A., Bonsack, M., Riemer, A., et al. "High-throughput prediction of MHC Class I and Class II neoantigens with MHCnuggets." *submitted to Cancer Immunol. Res.* BioRxiv ID 154757.
- Sidney, J., Southwood, S., Oseroff, C., del Guercio, M.F., Sette, A., and Grey, H.M. (2001). "Measurement of MHC/peptide interactions by gel filtration." *Curr. Protoc. Immunol.* Chapter 18, Unit 18.3.
- Sidney, J., Peters, B., Frahm, N., Brander, C., and Sette, A. (2008a). "HLA class I supertypes: a revised and updated classification." *BMC Immunol.* 9, 1.

- Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A., and Peters, B. (2008b). "Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries." *Immunome Res.* 4, 2.
- Skwarczynski, M., and Toth, I. (2016). "Peptide-based synthetic vaccines." *Chem. Sci.* 7, 842–854.
- Smotkin, D., and Wettstein, F.O. (1986). "Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein." *Proc. Natl. Acad. Sci. U. S. A.* 83, 4680–4684.
- Stamova, S., Koristka, S., Keil, J., Arndt, C., Feldmann, A., Michalk, I., Bartsch, H., Bippes, C.C., Schmitz, M., Cartellieri, M., et al. (2012). "Cancer Immunotherapy by Retargeting of Immune Effector Cells via Recombinant Bispecific Antibody Constructs." *Antibodies* 1, 172–198.
- Stanke, J., Hoffmann, C., Erben, U., von Keyserling, H., Stevanovic, S., Cichon, G., Schneider, A., and Kaufmann, A.M. (2010). "A flow cytometry-based assay to assess minute frequencies of CD8+ T cells by their cytolytic function." *J. Immunol. Methods* 360, 56–65.
- Stanley, M.A. (2012). "Epithelial cell responses to infection with human papillomavirus." *Clin. Microbiol. Rev.* 25, 215–222.
- Stanley, M., Gissmann, L., and Nardelli-Haeffliger, D. (2008). "Immunobiology of human papillomavirus infection and vaccination - implications for second generation vaccines." *Vaccine* 26 Suppl 1, K62-7.
- Steinbach, A., and Riemer, A.B. (2018). "Immune evasion mechanisms of human papillomavirus: An update." *Int. J. Cancer* 142, 224–229.
- Steinbach, A., Winter, J., Reuschenbach, M., Blatnik, R., Klevenz, A., Bertrand, M., Hoppe, S., von Knebel Doeberitz, M., Grabowska, A.K., and Riemer, A.B. (2017). "ERAP1 overexpression in HPV-induced malignancies: A possible novel immune evasion mechanism." *Oncoimmunology* 6, e1336594.
- Steinman, R.M. (1991). "The dendritic cell system and its role in immunogenicity." *Annu. Rev. Immunol.* 9, 271–296.
- Steller, M.A., Gurski, K.J., Murakami, M., Daniel, R.W., Shah, K. V., Celis, E., Sette, A., Trimble, E.L., Park, R.C., and Marincola, F.M. (1998). "Cell-mediated immunological responses in cervical and vaginal cancer patients immunized with a lipidated epitope of human papillomavirus type 16 E7." *Clin. Cancer Res.* 4, 2103–2109.
- Stevanović, S., Helman, S.R., Wunderlich, J.R., Langhan, M.M., Doran, S.L., Kwong, M.L.M., Somerville, R.P.T., Klebanoff, C.A., Kammula, U.S., Sherry, R.M., et al. (2019). "A Phase II Study of Tumor-infiltrating Lymphocyte Therapy for Human Papillomavirus-associated Epithelial Cancers." *Clin. Cancer Res.* 25, 1486–1493.
- Sun, Y., Peng, S., Qiu, J., Miao, J., Yang, B., Jeang, J., Hung, C.-F., and Wu, T.-C. (2015). "Intravaginal HPV DNA vaccination with electroporation induces local CD8+ T-cell immune responses and antitumor effects against cervicovaginal tumors." *Gene Ther.* 22, 528–535.
- Suprynowicz, F.A., Krawczyk, E., Hebert, J.D., Sudarshan, S.R., Simic, V., Kamonjoh, C.M., and Schlegel, R. (2010). "The human papillomavirus type 16 E5 oncoprotein inhibits epidermal growth factor trafficking independently of endosome acidification." *J. Virol.* 84, 10619–10629.
- Tan, H.-X., Wheatley, A.K., Esterbauer, R., Jegaskanda, S., Glass, J.J., Masopust, D., De Rose, R., and Kent, S.J. (2018). "Induction of vaginal-resident HIV-specific CD8 T cells with mucosal prime-boost immunization." *Mucosal Immunol.* 11, 994–1007.
- Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M.M., Kloetzel, P.-M., Rammensee, H.-G., Schild, H., and Holzhütter, H.-G. (2005). "Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding." *Cell. Mol. Life Sci.* 62, 1025–1037.

- Tosato, G., and Cohen, J.I. (2007). "Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines." *Curr. Protoc. Immunol.* Chapter 7, Unit 7.22.
- Trimble, C.L., Morrow, M.P., Kraynyak, K.A., Shen, X., Dallas, M., Yan, J., Edwards, L., Parker, R.L., Denny, L., Giffear, M., et al. (2015). "Safety, efficacy, and immunogenicity of VGX-3100, a therapeutic synthetic DNA vaccine targeting human papillomavirus 16 and 18 E6 and E7 proteins for cervical intraepithelial neoplasia 2/3: a randomised, double-blind, placebo-controlled phase 2b trial." *Lancet (London, England)* 386, 2078–2088.
- Trolle, T., and Nielsen, M. (2014). "NetTepi: an integrated method for the prediction of T cell epitopes." *Immunogenetics* 66, 449–456.
- Trolle, T., Metushi, I.G., Greenbaum, J.A., Kim, Y., Sidney, J., Lund, O., Sette, A., Peters, B., and Nielsen, M. (2015). "Automated benchmarking of peptide-MHC class I binding predictions." *Bioinformatics* 31, 2174–2181.
- Trost, B., Bickis, M., and Kusalik, A. (2007). "Strength in numbers: achieving greater accuracy in MHC-I binding prediction by combining the results from multiple prediction tools." *Immunome Res.* 3, 5.
- Tu, J.J., Kuhn, L., Denny, L., Beattie, K.J., Lorincz, A., and Wright, T.C. (2006). "Molecular variants of human papillomavirus type 16 and risk for cervical neoplasia in South Africa." *Int. J. Gynecol. Cancer* 16, 736–742.
- Uhl, M., Aulwurm, S., Wischhusen, J., Weiler, M., Ma, J.Y., Almirez, R., Mangadu, R., Liu, Y.-W., Platten, M., Herrlinger, U., et al. (2004). "SD-208, a novel transforming growth factor beta receptor I kinase inhibitor, inhibits growth and invasiveness and enhances immunogenicity of murine and human glioma cells in vitro and in vivo." *Cancer Res.* 64, 7954–7961.
- Um, S., Rhyu, J., Kim, E., Jeon, K., Hwang, E.-S., and Park, J.-S. (2002). "Abrogation of IRF-1 response by high-risk HPV E7 protein in vivo." *Cancer Lett.* 179, 205–212.
- Vang, Y.S., and Xie, X. (2017). "HLA class I binding prediction via convolutional neural networks." *Bioinformatics* 33, 2658–2665.
- Veldman, T., Horikawa, I., Barrett, J.C., and Schlegel, R. (2001). "Transcriptional activation of the telomerase hTERT gene by human papillomavirus type 16 E6 oncoprotein." *J. Virol.* 75, 4467–4472.
- Wang, J.W., and Roden, R.B.S. (2013). "L2, the minor capsid protein of papillomavirus." *Virology* 445, 175–186.
- Werness, B.A., Levine, A.J., and Howley, P.M. (1990). "Association of human papillomavirus types 16 and 18 E6 proteins with p53." *Science* 248, 76–79.
- WHO (2019). "WHO-Cancer" Accessed: 06/18/2019 via: "<https://www.who.int/cancer/en/>."
- Woo, Y.L., Sterling, J., Damay, I., Coleman, N., Crawford, R., van der Burg, S.H., and Stanley, M. (2008). "Characterising the local immune responses in cervical intraepithelial neoplasia: a cross-sectional and longitudinal analysis." *BJOG* 115, 1616-21-2.
- Wu, S., Powers, S., Zhu, W., and Hannun, Y.A. (2016). "Substantial contribution of extrinsic risk factors to cancer development." *Nature* 529, 43–47.
- Xi, L.F., Koutsky, L.A., Hildesheim, A., Galloway, D.A., Wheeler, C.M., Winer, R.L., Ho, J., and Kiviat, N.B. (2007). "Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18." *Cancer Epidemiol. Biomarkers Prev.* 16, 4–10.
- Yan, W.-L., Shen, K.-Y., Tien, C.-Y., Chen, Y.-A., and Liu, S.-J. (2017). "Recent progress in GM-CSF-based cancer immunotherapy." *Immunotherapy* 9, 347–360.

- Yee, C., Krishnan-Hewlett, I., Baker, C.C., Schlegel, R., and Howley, P.M. (1985). "Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines." *Am. J. Pathol.* 119, 361–366.
- Yewdell, J.W., and Bennink, J.R. (1999). "Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses." *Annu. Rev. Immunol.* 17, 51–88.
- Yim, E., and Park, J. (2005). "The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis." *Cancer Res. Treat.* 37, 319–324.
- You, J. (2010). "Papillomavirus interaction with cellular chromatin." *Biochim. Biophys. Acta* 1799, 192–199.
- Youde, S.J., Dunbar, P.R., Evans, E.M., Fiander, A.N., Borysiewicz, L.K., Cerundolo, V., and 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 Res.* 60, 365–371.
- Yu, K., Petrovsky, N., Schönbach, C., Koh, J.Y.L., and Brusic, V. (2002). "Methods for prediction of peptide binding to MHC molecules: a comparative study." *Mol. Med.* 8, 137–148.
- Zehbe, I., Wilander, E., Delius, H., and Tommasino, M. (1998). "Human papillomavirus 16 E6 variants are more prevalent in invasive cervical carcinoma than the prototype." *Cancer Res.* 58, 829–833.
- Zhang, H., Lund, O., and Nielsen, M. (2009). "The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding." *Bioinformatics* 25, 1293–1299.
- Zuna, R.E., Tuller, E., Wentzensen, N., Mathews, C., Allen, R.A., Shanesmith, R., Dunn, S.T., Gold, M.A., Wang, S.S., Walker, J., et al. (2011). "HPV16 variant lineage, clinical stage, and survival in women with invasive cervical cancer." *Infect. Agent. Cancer* 6, 19.

7 Annex

Supplementary Table S1. HLA binding prediction scores and experimental binding affinity of HPV16 E6- and E7-derived peptides.

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
A1																	
E7/2-11	HGDPTLHEY	1.03±0.59	1142	250	679	414	267	257	2146	384	315	313	309	0.45	0.35	0.08	28
E6/80-88	ISEYRHYCY	1.20±0.96	33	55	56	108	81	67	1146	293	300	40	25	0.25	0.25	0.01	27
E6 D32E/30-39	IHEIILECVY	1.25±0.65	19613	21523	14364	12453	8016	13142	11604	2299	3082	13013	20469	4.90	4.70	11.00	24
E6 H85Y/80-88	ISEYRYCY	1.25±1.06	32	84	51	94	88	86	1062	319	323	31	29	0.25	0.25	0.02	27
E6 D32E/29-39	TIHEIILECVY	1.46±0.82	18897	8800	10976	10956	4236	6096	3308	2139	773	7199	6141	2.80	2.75	8.00	16
E7/14-23	DLQPETTDLY	1.62±0.91	3676	1275	1324	2246	2343	1728	4288	1041	941	194	1893	0.95	0.70	0.20	19
E6 A68G/67-77	YGVCDKCLKFY	1.81±1.35	11404	17179	6597	7392	14186	15625	12382	11097	6012	16400	17746	5.60	5.65	2.00	18
E6/68-77	AVCDKCLKFY	2.52±1.22	3210	3631	1547	1140	5683	4551	2123	1305	1753	6850	11927	1.30	1.05	1.00	18
E6/74-83	LKFYSKISEY	7.72±0.00	17014	9154	15773	14949	30939	16855	27876	1650	1270	12423	30978	3.00	2.75	4.00	16
E7 N29S/19-29	TTDLICYEQLS	10.45±10.39	14480	2391	7867	8697	1631	1978	2365	6717	75861	1449	2154	32.10	31.95	2.00	18
E6 E36Q/29-39	TIHDILQCVY	11.26±20.22	11046	8673	8854	8612	2834	4963	3100	1771	641	4615	7203	1.10	1.15	6.00	16
E7/18-25	ETTDLYCY	11.52±7.17	1749	322	4421	3575	211	262	2417	21858	40545	254	187	23.05	23.05	7.00	N/A
E6/92-99	GTTLQYQY	12.52±6.85	11702	2008	14781	15251	1216	1568	3933	3764	16632	1059	2537	11.25	11.20	6.00	N/A
E6/23-31	CTELQTTIH	13.09±15.44	6706	3087	2137	1037	2040	2510	1766	2486	3331	1528	4233	1.95	1.65	0.30	15
E6/79-88	KISEYRHYCY	15.66±10.16	331	2027	1230	1377	2366	2193	1968	765	867	1453	367	0.55	0.45	2.00	15
E6 H85Y/79-88	KISEYRYCY	16.68±15.01	203	1425	1069	1252	2455	1864	1747	607	799	1099	432	0.55	0.45	0.90	16
E6/82-91	EYRHYCYSLY	16.92±3.78	13541	4120	16338	19356	25392	10191	12249	1478	1193	13900	23159	2.00	1.70	3.00	17
E7/4-11	DPTLHEY	24.92±13.05	6601	2006	8641	7096	2190	2100	5211	26461	41299	1204	7031	23.60	23.55	6.00	N/A
E6/81-91	SEYRHYCYSLY	32.04±12.55	18653	18138	8914	21042	15291	16583	7054	525	2838	10236	13368	4.50	4.50	2.00	18
E6 A68G/68-77	GVCDKCLKFY	32.41±51.84	5164	4547	2474	1835	6808	5560	2752	1275	1981	8632	13079	1.40	1.15	2.00	17
E6 H85Y/82-91	EYRYCYSLY	35.16±11.67	9287	3982	12804	16601	25133	9973	12791	1544	1233	14285	18354	1.10	0.80	4.00	17
E6 H85Y/78-88	SKISEYRYCY	38.09±34.22	403	19758	958	2391	27963	23445	22696	13621	3556	13995	5216	2.95	2.90	4.00	17
E6/80-89	ISEYRHYCYS	39.31±27.97	2049	18380	4700	5104	15306	16855	13502	7457	8410	7803	994	4.45	4.20	0.80	13

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/19-28	TTDLICYEQL	41.65±38.75	3004	3204	3137	2318	1701	2340	2443	2226	2629	1250	3287	1.65	1.40	0.40	18
E6 H85Y/80-89	ISEYRYCYCS	43.40±9.46	1240	16575	4530	4699	14985	15795	11604	6927	7060	7415	1095	3.60	3.50	0.90	13
E6/78-88	SKISEYRHYCY	45.72±27.05	709	19881	1125	2744	28588	23828	25564	15639	3516	15008	5972	2.95	2.90	5.00	16
E6/81-88	SEYRHYCY	46.44±20.50	1014	19115	7833	12175	18241	18679	10191	11314	9439	15813	12167	6.55	6.55	71.00	N/A
E7/19-27	TTDLICYEQ	83.36±15.02	4336	5319	3126	1899	2317	3511	1453	3601	4504	1049	17540	2.40	2.15	0.50	17
E6/23-33	CTELQTTIHDI	nb	16586	3663	11743	11643	4426	4019	9246	54853	40459	2204	17921	24.10	24.05	2.00	15
E6/29-39	TIHDIILECVY	nb	17802	9128	8155	8316	3345	5530	3237	1982	671	6112	7234	2.40	2.30	6.00	16
E6/30-39	IHDIIILECVY	nb	13152	15699	8007	5279	1752	5239	5267	1591	1931	6432	20030	2.15	1.75	5.00	26
E6/31-39	HDIILECVY	nb	17445	17026	14334	16437	24096	20258	6331	3850	3842	16523	13338	5.70	5.55	13.00	17
E6/42-50	QQLLRREVY	nb	21851	19260	24239	26617	26226	22452	7776	20064	20585	19336	31692	20.00	18.50	3.00	17
E6/43-53	QLLRREYDFA	nb	26757	21133	34347	35934	32145	25983	31399	1123	514	21021	23626	9.35	8.85	48.00	2
E6/51-61	DFAFRDLCIVY	nb	11506	15873	12061	8603	10669	13000	9146	991	2042	14222	15999	2.40	2.40	12.00	16
E6/52-61	FAFRDLCIVY	nb	4526	5066	3342	1998	1736	2953	4197	2689	3032	1481	6285	1.80	1.55	4.00	16
E6/53-61	AFRDLCIVY	nb	19580	18827	25814	24198	21830	20258	3530	6675	6959	14491	16380	11.60	10.10	14.00	18
E6/54-61	FRDLCIVY	nb	13549	2913	21049	21874	2145	2497	6400	33544	27538	967	1397	16.90	16.85	2.00	N/A
E6/57-67	LCIVYRDGNPY	nb	15340	15200	13075	14874	21200	17985	9654	3624	465	7410	6043	1.55	1.40	5.00	16
E6/58-68	CIVYRDGNPYA	nb	27746	20711	26037	23974	27990	24087	46353	1876	1159	22102	15057	11.50	11.00	37.00	0
E6/59-67	IVYRDGNPY	nb	7734	11980	6340	5740	3695	6647	1438	5063	4792	6607	13311	2.60	2.30	2.00	15
E6/67-77	YAVCDKCLKFY	nb	6323	11458	1060	1403	5157	7692	7776	11754	9269	8405	1649	8.35	8.65	1.00	18
E6/69-77	VCDKCLKFY	nb	6457	5882	5057	4559	2287	3666	2443	1206	1001	2477	6886	0.95	0.65	0.03	26
E6/73-83	CLKFYSKISEY	nb	12283	11678	7773	9822	8283	9866	6683	1645	149	5032	12989	0.85	0.75	2.00	17
E6/75-83	KFYSKISEY	nb	18273	19888	20154	16829	21792	20814	4527	8738	6959	15877	30918	7.80	7.65	4.00	16
E6/76-86	FYSKISEYRHY	nb	7546	14923	11024	11374	15750	15291	14564	20759	5094	9714	8796	4.65	4.90	0.90	17
E6/77-86	YSKISEYRHY	nb	2088	434	2005	2071	5411	1526	9448	548	685	737	322	0.75	0.50	0.20	21
E6/78-86	SKISEYRHY	nb	27567	20630	35294	36816	40418	28795	23445	92559	83859	17555	22814	46.00	44.50	7.00	16
E6/83-91	YRHYCYSLY	nb	8111	17209	12767	14964	9459	12722	6331	2555	1839	8487	17722	1.25	0.95	0.70	17

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/88-95	YSLYGTTL	nb	22765	18459	21967	21545	13158	15625	15710	866	3250	19706	34277	2.95	2.95	18.00	N/A
E6/89-99	SLYGTTLQYY	nb	7370	8510	4838	7260	4101	5901	2193	4626	646	8079	7447	0.85	1.15	0.70	17
E6/90-99	LYGTTLQYY	nb	18888	7355	18184	18512	22188	12791	11113	1237	1999	14080	15189	4.10	3.90	4.00	15
E6/134-144	NIRGRWTGRCM	nb	34058	21649	38169	38473	33146	26840	43439	2601	9528	23134	33104	32.50	32.00	56.00	0
E6 A68G/60-69	VYRDGNPYGV	nb	24101	24096	31686	27659	33878	28640	50000	29960	23110	24136	33172	21.50	20.00	71.00	0
E6 E36Q/30-39	IHDILQCVY	nb	7212	14102	6692	4127	1109	3954	5155	1740	1918	5023	18897	1.40	1.10	3.00	26
E6 H85Y/76-85	FYSKISEYRY	nb	2954	1534	2244	1951	5926	3017	7609	707	591	4802	7497	0.70	0.45	2.00	17
E6 H85Y/76-86	FYSKISEYRY	nb	2231	12330	7691	6120	9562	10875	9654	11408	5496	9244	6775	4.95	5.25	0.50	17
E6 H85Y/77-85	YSKISEYRY	nb	471	282	670	545	761	464	2241	626	545	301	188	0.45	0.40	0.08	21
E6 H85Y/77-86	YSKISEYRY	nb	462	231	780	635	2109	696	6263	380	464	575	169	0.35	0.20	0.08	21
E6 H85Y/81-88	SEYRYICY	nb	647	18977	7419	11869	17519	18181	8120	44834	30335	14843	5752	17.55	17.55	74.00	N/A
E7/13-23	LDLQPETTDLY	nb	10702	17537	7768	9004	21942	19611	16228	3042	100	7967	2949	0.65	0.75	2.00	17
E7/15-23	LQPETTDLY	nb	7998	10499	5255	7312	7205	8711	2169	2645	3166	5389	16183	1.85	1.55	0.60	17
E7/15-25	LQPETTDLYCY	nb	16310	14564	12041	15648	9073	11479	7209	13589	2924	12110	5727	3.95	3.95	3.00	17
E7/16-25	QPETTDLYCY	nb	8074	5773	9370	7907	6441	6096	10642	498	546	10190	23714	0.70	0.40	0.60	25
E7/19-26	TTDLYCYE	nb	12345	3668	7559	8170	3349	3511	3100	11107	20415	2604	3600	13.30	13.25	1.00	N/A
E7/19-29	TTDLYCYEQLN	nb	16817	5300	11705	13612	4002	4626	2216	6702	38996	5595	3279	23.65	23.55	2.00	18
E7/42-52	AGQAEPDRAHY	nb	21457	17310	11592	24816	24890	20814	19088	8774	2819	15554	31073	5.75	5.70	6.00	16
E7/43-52	GQAEPDRAHY	nb	15534	12466	9864	15671	23790	17224	14099	4117	3348	13404	22907	3.40	3.35	2.00	16
E7/44-52	QAEPDRAHY	nb	4645	6096	3125	6669	5871	5998	4936	3728	2544	1302	3869	1.55	1.30	0.02	25
E7 N29S/28-38	LSDSSEEEDEI	nb	11966	4024	14067	13692	6246	5017	5933	16912	7329	1196	1551	7.20	7.15	2.00	15
A2																	
E7/83-93	LMGTLGIVCPI	0.38±0.38	248	558	215	262	79	211	121	206	1452	167	79	8.30	8.05	21.00	N/A
E7/82-91	LLMGTGIVC	1.03±0.40	325	916	462	355	1028	969	646	351	475	2426	478	2.75	2.45	6.00	15
E7/86-93	TLGIVCPI	1.27±1.54	980	183	2177	2198	393	269	212	174	167	578	34	1.00	0.80	5.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/11-19	YMLDLQPET	1.40±0.94	21	5	22	30	7	6	41	21	21	46	23	0.40	0.40	0.30	21
E7/85-93	GTLGIVCPI	1.96±1.93	203	155	75	107	121	137	176	387	427	309	139	4.40	4.40	7.00	21
E7/84-93	MGTGIVCPI	1.98±1.89	393	7314	296	322	10428	8711	3033	1294	1642	7561	1418	6.60	6.20	81.00	11
E7/11-20	YMLDLQPETT	2.19±2.71	176	44	187	237	25	33	184	170	165	280	102	1.45	1.20	0.50	19
E7/11-21	YMLDLQPETTD	2.37±1.67	4352	9395	9156	11401	2402	4727	503	940	889	6288	902	5.90	5.95	3.00	N/A
E6 D32E/28-38	TTIHEIILECV	2.47±1.71	3025	1499	4527	8614	994	1222	586	863	705	2555	520	4.95	5.05	6.00	N/A
E7/7-15	TLHEYMLDL	2.68±1.86	49	48	61	65	95	68	93	133	136	285	142	2.10	2.10	0.20	24
E7/77-87	RTLEDLLMGTL	2.69±2.94	3257	577	2582	3837	768	667	802	241	324	2830	1712	2.65	2.75	3.00	N/A
E6 R17I/9-17	FQDPQERPI	2.72±1.59	1166	1943	2258	2044	2713	2302	414	816	742	736	570	1.50	2.10	5.00	11
E7/82-90	LLMGTLGIV	3.30±2.58	25	19	17	16	20	20	12	25	26	65	122	0.50	0.50	2.00	29
E7/12-20	MLDLQPETT	3.60±0.73	1812	2403	2362	2157	2440	2417	561	1298	1162	3880	1038	7.90	7.90	5.00	16
E6 Q21D/18-28	KLPDLCTELQT	3.68±2.63	6743	1079	6485	11935	1007	1039	674	778	1489	1494	7691	9.45	9.25	0.90	N/A
E6/29-38	TIHDIILECV	3.69±2.17	276	303	316	320	145	210	269	445	450	971	179	2.65	2.35	2.00	23
E6/52-60	FAFRDLCIV	4.34±2.11	155	115	156	115	150	132	105	152	149	284	83	2.30	2.30	5.00	20
E6 D32E/29-38	TIHEIILECV	5.88±1.81	259	331	409	452	153	225	249	495	512	1428	150	2.85	2.55	2.00	24
E6 D32E, I34R/29-38	TIHEIRILECV	7.26±2.50	1219	12901	1587	1686	18908	15625	3686	3898	3745	2624	252	10.35	4.65	1.00	24
E6 Q21D/18-26	KLPDLCTEL	7.45±1.77	23	12	49	47	18	15	52	35	34	246	261	0.60	0.60	0.02	25
E7/11-18	YMLDLQPE	8.07±3.69	5042	2323	8103	7304	768	1333	234	822	2479	782	2057	7.85	8.00	29.00	N/A
E6/18-26	KLPQLCTEL	10.07±3.22	108	87	281	227	130	107	70	127	114	739	933	1.80	1.80	0.06	24
E7/80-90	EDLLMGTLGIV	14.20±2.47	971	18371	1511	1600	34740	25153	22696	2019	1161	14678	1134	7.35	7.05	67.00	N/A
E6 D32E/25-33	ELQTTIHEI	15.51±9.74	20154	29997	2930	26076	33645	31740	43439	401264	369922	2296	2381	64.00	6.60	0.20	22
E7/7-17	TLHEYMLDLQP	19.35±12.57	11451	9301	14462	20258	12314	10700	1926	283	90	13891	5243	2.95	1.60	10.00	N/A
E6 R17I/9-19	FQDPQERPIKL	20.75±15.64	8579	3176	6087	11214	740	1534	653	3501	12215	2232	3482	29.05	28.75	1.00	N/A
E6 H85Y/83-90	YRYCYSL	21.01±8.92	14920	6045	20213	23243	2298	3726	1926	19762	15639	2964	15411	23.10	22.65	39.00	N/A
E6 D32E, I34R/28-38	TTIHEIRILECV	29.14±16.90	10174	3865	12699	17971	3479	3666	953	11355	6136	5310	952	21.30	20.75	5.00	N/A
E7/76-86	IRTLEDLLMGT	31.85±22.45	8340	24652	5149	9448	39274	31229	50000	2946	3531	15750	15290	16.05	15.75	44.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/82-92	LLMGTLGIVCP	31.96±20.21	3240	5188	2994	3969	2069	3290	281	877	671	4453	340	4.75	4.85	5.00	N/A
E6 H85Y, L90V/83-90	YRYYCYSV	32.29±15.14	9128	2368	19446	22112	2302	2327	735	17492	7083	907	8357	13.50	13.50	33.00	N/A
E6/18-28	KLPQLCTELQT	34.50±19.55	10551	3392	8531	16616	4714	3997	855	1715	3311	4006	15762	16.05	15.25	3.00	N/A
E6 H85Y/84-93	RYYCYSLYGT	35.20±17.55	15989	4749	13986	11860	12796	7776	3849	972	1152	7614	18301	19.25	10.25	14.00	11
E6/28-38	TTIHDIILECV	35.72±4.66	3484	1444	3945	7569	909	1146	549	883	781	1683	636	5.40	5.45	6.00	N/A
E6/25-33	ELQTTIHDI	36.90±8.41	6830	7890	9733	9680	9381	8571	205	3072	3077	4896	12715	14.00	14.00	2.00	20
E7/66-74	RLCVQSTHV	37.18±10.65	558	781	719	571	765	776	36	213	208	730	661	2.80	2.80	0.20	20
E7/81-91	DLLMGTLGIVC	41.25±39.26	2747	9628	10688	5639	23175	14964	4107	2149	1066	8055	1175	7.65	7.75	22.00	N/A
E6/34-44	IIECVYCKQQL	45.15±8.59	8508	4362	4136	5306	813	1884	514	3429	9139	4866	3433	25.55	25.25	3.00	N/A
E6/79-87	KISEYRHYC	47.56±29.56	6936	6671	3524	2063	4164	5267	1534	1612	1390	3712	575	8.70	8.70	0.50	13
E7/81-90	DLLMGTLGIV	52.58±19.28	120	755	251	171	3220	1559	172	132	103	2104	225	0.90	0.85	7.00	25
E6 H85Y, L90V/81-90	SEYRYCYSL	52.62±28.64	6255	1041	7486	6926	3334	1864	704	206	175	1084	8890	2.75	1.40	13.00	16
E7/12-19	MLDLQPET	53.66±12.03	4073	1760	11632	9488	444	884	561	926	3992	4315	120	9.85	10.00	12.00	N/A
E6 H85Y/79-89	KISEYRYCYCS	59.11±35.98	25067	9962	18205	21152	2895	5382	2290	646	486	7743	1041	21.55	16.05	11.00	N/A
E7/86-94	TLGIVCPIC	62.31±18.61	4822	11770	5803	4733	8538	10027	1844	6433	5130	3341	777	18.00	18.00	23.00	11
E7/77-86	RTLEDLLMGT	72.72±17.36	2565	1045	2254	2033	3454	1905	1422	531	535	1887	1414	3.75	2.80	4.00	17
E6 H85Y/81-90	SEYRYCYSL	74.36±19.14	11629	3457	14031	12377	4922	4129	1844	656	575	3410	15416	5.70	4.15	16.00	16
E6/9-19	FQDPQERPRKL	nb	20790	9208	17751	23015	4827	6683	1347	7314	18319	6435	6353	47.00	41.50	2.00	N/A
E6/20-30	PQLCTELQTTI	nb	8717	4517	13293	14067	11283	7131	3380	237	271	13500	13251	2.80	2.45	15.00	N/A
E6/21-29	QLCTELQTT	nb	6636	13552	7781	6313	10642	11987	632	5278	5372	6238	3158	18.00	18.00	1.00	20
E6/21-30	QLCTELQTTI	nb	1394	1316	1091	1593	1060	1183	419	355	293	2846	1913	2.55	1.85	3.00	20
E6/24-34	TELQTTIHDI	nb	22327	10259	27527	29225	30114	17600	11858	640	997	18465	27720	22.00	16.50	66.00	N/A
E6/29-37	TIHDIILEC	nb	3935	6557	2853	2984	6472	6505	2470	3487	3192	5756	718	15.00	15.00	0.50	16
E6/29-39	TIHDIILECVY	nb	8437	23412	13566	16111	24586	23957	39837	1933	1264	18480	18935	8.55	8.25	28.00	N/A
E6/30-38	IHDIIILECV	nb	11618	15173	21610	21292	9987	12316	1710	7752	8056	11801	1766	17.00	17.00	12.00	16
E6/33-40	IIECVYC	nb	5793	1668	16281	13576	1958	1814	735	4941	1251	2605	707	5.20	5.30	17.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/33-41	IILECVYCK	nb	4246	4793	14731	17699	15200	8525	784	954	1245	3547	10477	5.50	5.50	2.00	19
E6/33-43	IILECVYCKQQ	nb	18479	10179	30770	33495	19263	14023	3380	354	298	12839	13234	14.90	8.90	6.00	N/A
E6/34-41	IILECVYCK	nb	13211	15745	23842	28087	23501	19192	5869	641	11072	12046	9997	18.45	17.85	19.00	N/A
E6/35-45	LECVYCKQQLL	nb	19399	21000	11673	18246	31770	25842	19296	28066	61647	15989	22006	57.00	51.50	33.00	N/A
E6/37-44	CVYCKQQL	nb	28998	12404	21594	22673	4644	7568	1051	15165	33901	13189	9010	44.00	42.50	24.00	N/A
E6/37-45	CVYCKQQLL	nb	3574	4148	2647	3027	2847	3435	442	1979	2033	5418	4447	12.00	12.00	0.50	16
E6/41-49	KQQLLRREV	nb	10474	9569	12030	12405	9835	9707	1568	4050	3237	11152	11329	15.00	15.00	3.00	14
E6/42-51	QQLLRREYD	nb	30968	28321	37816	38189	38363	32966	33505	23649	14206	22971	22316	48.50	41.00	45.00	4
E6/42-52	QQLLRREYDF	nb	22055	18764	27432	30679	18406	18579	20149	3517	3698	19049	27983	31.00	25.50	51.00	N/A
E6/43-52	QLLRREYDF	nb	11087	10566	12025	13898	8393	9397	1377	1300	1048	11557	10734	6.80	5.25	9.00	17
E6/43-53	QLLRREYDFA	nb	3519	2023	2327	5855	891	1347	255	1228	2523	799	1692	12.70	12.75	9.00	N/A
E6/44-53	LLRREYDFA	nb	2797	3637	1707	2360	1474	2315	784	1548	1260	2160	3118	5.85	4.90	38.00	16
E6/44-54	LLRREYDFAF	nb	13944	11801	13612	15236	8001	9759	3168	19150	8052	10422	10541	26.65	25.55	20.00	N/A
E6/50-59	YDFAFRDLICI	nb	7051	6139	8597	9357	14096	9296	7446	3663	20724	5783	11153	22.25	20.85	16.00	12
E6/50-60	YDFAFRDLCIV	nb	6606	8297	5664	9827	6861	7527	1637	830	713	8538	2371	5.25	5.10	13.00	N/A
E6/52-61	FAFRDLCIVY	nb	3192	6631	8256	6485	18170	10934	27279	1922	1856	6897	3568	7.85	6.85	25.00	6
E6/53-61	AFRDLCIVY	nb	24893	28282	34843	32066	33334	30727	50000	244023	258887	15569	24685	36.00	35.00	56.00	4
E6/59-68	IVYRDGNPYA	nb	1071	1159	587	858	758	938	503	543	568	2605	1751	3.25	2.85	2.00	12
E6/59-69	IVYRDGNPYAV	nb	3288	1197	518	2531	132	397	142	14628	15449	291	147	30.65	30.75	2.00	N/A
E6/60-69	VYRDGNPYAV	nb	15455	15777	19493	18526	21913	18579	7209	3190	2615	4585	4613	22.50	13.50	4.00	14
E6/61-69	YRDGNPYAV	nb	4312	3941	4991	7287	5938	4831	660	907	896	697	4654	5.90	5.90	0.80	16
E6/67-76	YAVCDKCLKF	nb	15272	20390	18813	18867	23085	21734	36534	8726	11181	14274	15813	30.00	20.95	9.00	9
E6/68-78	AVCDKCLKFYIS	nb	24098	12361	24857	29926	22102	16583	5267	791	794	7890	10062	22.25	16.75	8.00	N/A
E6/71-78	DKCLKFYIS	nb	39199	28441	45915	48134	45858	36141	50000	9243	132	23125	10243	31.60	30.10	95.00	N/A
E6/72-80	KCLKFYISKI	nb	12145	9867	16905	17601	16790	12860	3807	15397	14227	5453	1239	26.00	26.00	4.00	14
E6/74-81	LKFYISKIS	nb	40511	26199	45297	46367	41928	33144	50000	4751	183	22200	21960	35.80	34.30	94.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b		Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/77-87	YSKISEYRHYC	nb		24619	26570	16012	19293	30891	28640	50000	20473	3845	20768	18289	33.50	28.00	60.00	N/A
E6/79-89	KISEYRHYCYS	nb		25556	14150	20250	22368	5450	8759	2874	809	376	9022	2516	21.85	16.35	8.00	N/A
E6/81-90	SEYRHYCYSL	nb		14260	6383	15061	13026	6617	6505	2265	1140	945	5811	22900	10.25	7.85	12.00	16
E6/86-95	YCYSLYGTTL	nb		3602	1628	8328	9415	5072	2874	2693	1101	1440	4780	4277	6.25	5.25	3.00	15
E6/88-95	YSLYGTTL	nb		17993	5198	19224	20909	1016	2302	1017	23112	12001	2281	15868	20.80	20.50	48.00	N/A
E6/88-97	YSLYGTTLQ	nb		2507	15537	6125	4849	26117	20149	15045	4091	3610	4946	3926	10.80	9.85	34.00	6
E6/89-97	SLYGTTLQ	nb		4296	2395	7750	6626	5633	3666	482	2744	3267	4063	2451	2.40	2.10	2.00	20
E6/89-98	SLYGTTLQ	nb		8398	4206	14694	13854	8949	6129	1017	1640	1744	4484	3220	8.35	6.85	2.00	19
E6/89-99	SLYGTTLQ	nb		14538	4751	13567	18915	11581	7406	3686	1105	879	9041	15338	9.15	8.10	2.00	N/A
E6/93-103	TTLQYQNKPL	nb		17677	1268	9390	16075	3431	2077	1691	560	405	10436	15385	15.00	9.50	8.00	N/A
E6/94-103	TLEQQYQNKPL	nb		7527	6380	9945	8846	5238	5775	1146	2488	3439	8247	11421	10.75	9.35	16.00	18
E6/97-106	QQYQNKPLCDL	nb		9780	6150	8598	8779	4791	5412	1109	2182	1776	7262	5721	8.35	6.85	3.00	15
E6/97-107	QQYQNKPLCDLL	nb		14734	7209	7396	13660	2442	4197	1236	2473	2517	7882	7924	16.10	15.10	8.00	N/A
E6/98-108	QYQNKPLCDLLI	nb		23709	7711	31894	35466	27522	14564	19932	2433	1918	12473	16903	27.00	21.50	49.00	N/A
E6/99-108	YQNKPLCDLLI	nb		14169	11218	20471	22341	27418	17505	18280	5634	6346	10629	13515	18.00	16.35	28.00	12
E6/101-111	KPLCDLLIRCI	nb		9569	4519	24361	21676	19674	9448	8759	344	416	11670	17447	3.70	3.30	13.00	N/A
E6/102-111	PLCDLLIRCI	nb		2554	6081	6553	4404	5002	5500	531	2798	2585	6228	5845	8.80	7.85	6.00	20
E6/105-115	DLLIRCINCQK	nb		20836	23689	32321	33081	36110	29267	16583	2755	1374	15381	16105	22.50	17.00	26.00	N/A
E6/106-113	LLIRCINC	nb		9263	2451	18082	14592	2460	2456	802	533	235	5841	963	1.95	1.95	19.00	N/A
E6/106-114	LLIRCINCQ	nb		7168	8472	18562	17772	12566	10302	605	4607	5716	6195	1226	5.80	5.80	10.00	19
E6/106-115	LLIRCINCQK	nb		11949	13794	17020	16705	15713	14723	3134	2060	3087	8211	8589	11.95	10.40	8.00	14
E6/125-135	HLDKKQRFHNI	nb		13960	3145	13409	17238	1447	2134	266	6028	4168	2299	2414	20.15	19.05	7.00	N/A
E6/143-151	CMSCCRSSR	nb		28503	23210	18476	18386	25890	24482	19506	19881	16000	14190	8515	45.00	34.00	24.00	10
E6/143-152	CMSCCRSSRT	nb		16992	14740	13952	12198	7741	10700	2146	11424	10034	10536	16824	31.50	22.50	23.00	12
E6 D32E/25-35	ELQTTIHEIIL	nb		14053	4108	13078	17333	3974	4041	674	2826	1076	10684	7722	10.75	9.65	11.00	N/A
E6 D32E/29-37	TIHEIILEC	nb		3629	6115	3470	3635	6195	6162	2241	3912	3541	7047	575	15.00	15.00	0.50	17

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6 H85Y/79-87	KISEYRYYC	nb	2350	2754	1586	675	1329	1915	1062	734	669	1617	328	5.80	5.80	0.90	13
E6 H85Y, L90V/84-93	RYYCYSVYGT	nb	18722	6392	15049	13147	13233	9196	3933	1548	2130	9614	18721	24.50	16.00	21.00	11
E6 I34R/34-44	RLECVYCKQQL	nb	11255	4311	6180	7445	1266	2340	599	3275	9139	6057	5553	26.50	25.25	2.00	N/A
E6 L90V/80-90	ISEYRHYCYSV	nb	16198	9576	17417	22763	7510	8479	1926	2892	1374	2895	1854	17.00	12.20	42.00	N/A
E6 L90V/81-90	SEYRHYCYSV	nb	36945	32011	7819	38204	39377	35559	50000	211748	295447	2209	17162	80.50	1.85	9.00	16
E6 L90V/86-93	YCYSVYGT	nb	35996	24663	30514	41704	43190	32611	50000	3792	9467	6803	11141	38.00	38.40	26.00	N/A
E6 L90V/86-95	YCYSVYGTTL	nb	34593	34945	8457	39893	38687	36732	50000	180643	198372	4415	5541	74.50	6.35	3.00	15
E6 L90V/88-95	YSVYGTTL	nb	14823	8504	25590	25934	10648	9499	2969	38356	15930	6135	18257	23.10	29.00	64.00	N/A
E6 Q21D/21-30	DLCTELQTTI	nb	8440	7163	11452	13562	14092	10027	974	638	587	10315	4259	4.45	2.95	6.00	19
E6 R17I, Q21D/13-22	QERP ^I KLPDL	nb	32740	32349	37283	37703	36784	34610	50000	54679	38061	17464	25739	58.50	51.50	38.00	17
E6 R17I, Q21D/17-26	I ^K LPDLCTEL	nb	282	9456	354	225	17803	13000	2391	2327	3503	12549	5733	9.50	9.20	5.00	18
E6 R17T/9-19	FQDPQER ^P TKL	nb	32258	27800	9862	41501	43665	34798	50000	53973	14925	3131	6014	59.00	34.80	2.00	N/A
E6 R17T/13-22	QER ^P TKLPQL	nb	40296	33748	36315	43633	43533	38357	50000	481748	1731729	19319	28266	93.00	52.00	14.00	17
E6 R17T/17-26	TKLPQLCTEL	nb	1719	1170	4225	2086	1041	1103	696	1358	1160	15669	10494	5.55	16.20	6.00	16
E7/6-15	PTLHEYMLDL	nb	718	10537	3461	1675	24965	16228	5155	6149	4597	10994	4078	11.10	10.70	51.00	14
E7/7-16	TLHEYMLDLQ	nb	1879	8104	5804	3735	13194	10358	2874	4496	5052	8318	2363	12.25	11.35	12.00	16
E7/10-19	EYMLDLQPET	nb	293	22329	1594	1171	37220	28795	50000	15879	16768	19111	5311	19.50	19.20	72.00	5
E7/10-20	EYMLDLQPETT	nb	1548	21352	5367	6152	39252	28952	50000	6240	4752	19607	15102	18.40	18.75	62.00	N/A
E7/12-22	MLDLQPETDDL	nb	3950	2378	2557	4519	278	815	216	3974	4111	2691	2870	17.20	17.25	4.00	N/A
E7/14-22	DLQPETDDL	nb	17945	17374	17660	19296	19864	18579	514	6583	6182	12823	13825	19.00	13.00	2.00	23
E7/15-23	LQPETTDLY	nb	25371	21287	24832	24655	29823	25153	16764	40591	41602	15306	24304	38.00	38.00	26.00	6
E7/15-25	LQPETTDLYCY	nb	28443	14305	30165	34365	27128	19611	17505	13937	6052	17471	23640	43.00	37.50	11.00	N/A
E7/21-28	DLYCYEQL	nb	17632	7060	16955	17410	1650	3417	287	3980	3726	8137	15383	12.15	11.85	10.00	N/A
E7/27-35	QLNDSSEEE	nb	28773	25047	29820	27846	32684	28640	1691	24913	34923	14869	7730	36.00	35.00	2.00	14
E7/32-42	SEEEDEIDGPA	nb	39189	25923	42407	44656	40383	32260	45854	842	2202	21110	30162	53.00	48.00	43.00	N/A
E7/59-69	CKCDSTLR ^L CV	nb	18788	7516	34804	35739	32220	15541	7692	233	98	16497	10737	13.85	7.85	6.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/63-72	STLRRCVQST	nb	8271	8321	9472	10846	15416	11294	5382	4975	4107	8078	19931	11.85	10.35	26.00	18
E7/64-71	TLRLCVQS	nb	22847	14993	29625	31783	19456	17131	10993	11990	414	14923	5419	9.65	8.15	44.00	N/A
E7/64-72	TLRLCVQST	nb	7303	13963	5811	6676	9195	11294	2290	6721	6578	8243	10780	2.70	2.50	11.00	20
E7/64-74	TLRLCVQSTHV	nb	6275	3185	5424	9657	1511	2205	442	6858	4965	4643	4912	18.90	18.75	7.00	N/A
E7/65-74	LRLCVQSTHV	nb	1815	12227	2131	1971	21181	16141	3417	4548	5991	6258	9347	13.25	12.35	13.00	14
E7/66-76	RLCVQSTHVDI	nb	14791	1688	10274	14378	3207	2327	316	828	3971	5067	14924	20.15	19.15	5.00	N/A
E7/67-77	LCVQSTHVDIR	nb	34584	29802	36134	37864	43876	36141	50000	40474	10738	21318	23858	59.00	54.00	79.00	N/A
E7/69-76	VQSTHVDI	nb	31123	16006	31479	34281	7294	10816	1947	39340	99589	14870	11340	59.00	57.50	47.00	N/A
E7/71-79	STHVDIRTL	nb	13065	13272	21278	21370	21158	16764	3202	11205	9641	13081	13014	23.00	23.00	2.00	22
E7/73-82	HVDIRTLEDL	nb	12562	11704	14168	13627	8122	9759	2123	10042	9279	12415	12644	18.45	17.35	11.00	18
E7/73-83	HVDIRTLEDLL	nb	21136	8487	20053	26313	5721	6978	2216	1825	1004	11757	14871	21.00	15.50	41.00	N/A
E7/74-82	VDIRTLEDL	nb	25952	26391	33994	35515	32812	29425	5998	56811	43763	14061	17424	39.00	38.00	19.00	17
E7/75-83	DIRTLEDLL	nb	28506	26616	27806	31428	34841	30561	9654	71520	68251	16933	25804	45.00	43.00	49.00	18
E7/77-84	RTLEDLLM	nb	15876	8515	18336	19712	4251	5998	1844	20598	19152	2484	11543	26.00	25.50	34.00	N/A
E7/78-87	TLEDLLMGTL	nb	356	462	277	204	481	472	428	623	752	3978	3283	3.75	3.35	3.00	22
E7/81-89	DLLMGTLGI	nb	2529	3947	1728	1328	3076	3473	151	567	551	1247	1000	3.80	3.80	13.00	25
E7/82-89	LLMGTLGI	nb	212	123	329	184	14	42	48	706	740	40	201	3.20	3.00	6.00	N/A
E7/83-90	LMGTLGIV	nb	1106	911	3868	1875	269	495	114	574	468	1731	275	2.25	2.00	9.00	N/A
E7/83-91	LMGTLGIVC	nb	7548	13788	7937	6752	7538	10191	1290	5425	4625	5957	1409	17.00	17.00	38.00	11
E7/83-92	LMGTLGIVCP	nb	12140	13613	13463	13136	15166	14407	1249	6619	8760	10010	2069	18.10	16.60	32.00	16
E7/88-95	GIVCPICS	nb	31844	23041	34495	34852	23275	23067	5155	8353	464	16979	8633	17.85	16.35	72.00	N/A
E7/88-96	GIVCPICSQ	nb	24415	22439	29028	24846	30796	26265	2315	63450	88127	16340	10984	47.00	47.00	31.00	16
E7/78-86	TLEDLLMGT	nb*±3.01	842	1421	994	609	1898	1646	345	840	989	2140	2061	7.30	7.30	4.00	20
A3																	
E6/109-119	RCINCQKPLCP	1.71±0.22	35545	19965	40526	38523	42313	29109	50000	297	9243	20093	13227	56.00	53.00	17.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6 L90V/84-94	RHYCYSVYGTT	2.03±0.71	32699	10627	31415	37526	30380	17985	11479	691	2929	23034	22676	32.50	29.50	29.00	N/A
E6/106-115	LLIRCINCQK	2.06±0.97	172	90	80	79	128	107	2123	70	79	37	39	0.90	0.50	2.00	29
E6/72-80	KCLKFYSKI	3.80±0.80	28008	20793	30831	26463	26765	23572	19506	59753	69470	21373	23384	37.00	33.50	36.00	8
E7/89-97	IVCPICSQK	4.06±1.38	275	155	239	182	204	178	419	159	153	194	170	0.55	0.95	0.05	31
E6/89-99	SLYGTTLQYQY	6.68±2.49	234	955	132	163	131	354	1422	4685	2778	1272	199	6.65	6.60	0.50	N/A
E6/107-115	LIRCINCQK	7.03±3.61	375	593	146	162	227	366	586	489	465	204	406	1.10	1.50	0.80	24
E7/88-97	GIVCPICSQK	7.06±2.83	343	169	158	157	200	185	1864	108	100	215	95	1.00	0.55	0.40	24
E6/68-77	AVCDKCLKFY	9.68±2.59	932	1276	1365	991	1939	1576	4990	548	455	2047	458	2.15	1.50	0.70	20
E6/33-41	IILECVYCK	11.02±3.27	226	126	226	266	153	139	368	243	225	100	352	0.70	1.15	0.40	23
E6/75-84	KFYSKISEYR	11.40±0.71	319	329	1046	924	1119	608	1028	279	317	535	397	1.60	1.20	2.00	11
E6/59-67	IVYRDGNPY	11.58±1.76	975	1081	695	661	696	869	784	797	756	610	219	1.60	1.90	2.00	28
E6/93-101	TTLEQQYNK	15.55±5.05	493	248	658	1048	521	360	330	439	360	494	233	1.00	1.35	0.20	13
E6/92-101	GTTLEQQYNK	17.96±1.67	1292	2536	805	1092	1578	2000	2290	971	906	1026	185	3.50	2.80	3.00	11
E6/129-138	KQRFHNIRGR	23.03±4.59	1603	1119	1270	1769	2365	1628	2443	650	676	933	59	3.45	11.35	2.00	10
E6/125-133	HLDKKQRFH	29.58±13.63	3475	3086	8302	8408	7001	4651	9048	1881	1909	3505	191	2.70	2.90	2.00	15
E6/37-46	CVYCKQQLLR	36.33±19.78	113	232	320	378	248	240	428	248	211	56	285	1.15	0.95	0.40	21
E6/68-75	AVCDKCLK	38.20±10.85	6961	752	7695	5533	637	693	1551	21975	13611	220	3622	3.45	3.15	6.00	N/A
E6/75-83	KFYSKISEY	52.27±21.55	519	479	955	884	671	567	1655	572	665	1313	2433	1.45	1.80	0.30	17
E6/84-91	RHYCYSLY	64.46±20.70	10604	2148	12511	10536	1563	1834	2874	33879	42158	5011	5542	21.00	20.70	18.00	N/A
E6/8-18	MFQDPQERPRK	64.67±1.05	17041	12848	10200	9191	5057	8032	7287	11659	3705	2965	2296	13.10	12.70	6.00	N/A
E6/21-29	QLCTELQTT	nb	33763	24928	37823	37793	40224	31740	30727	164953	186982	20733	27714	54.50	51.50	17.00	16
E6/21-30	QLCTELQTTI	nb	21471	25775	26475	28167	33188	29267	48930	56114	44895	20494	22316	41.00	34.50	16.00	17
E6/23-33	CTELQTTIHDI	nb	35474	18831	35971	37597	36069	26124	47367	9742	291	24218	31137	32.10	29.10	70.00	N/A
E6/29-39	TIHDIILECVY	nb	8104	7833	7412	7794	3159	4990	5044	15089	9179	10239	702	24.90	24.20	23.00	N/A
E6/32-41	DIILECVYCK	nb	1829	6948	2341	1882	9298	8076	4990	1449	1564	6024	2167	5.15	13.05	20.00	22
E6/34-41	IILECVYCK	nb	6671	706	4308	3160	353	498	2290	20842	13517	890	1268	3.05	2.80	7.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/37-45	CVYCKQQLL	nb	11113	14359	11524	11532	11896	13071	2722	3291	2943	9444	9319	4.10	3.60	3.00	15
E6/37-47	CVYCKQQLLR	nb	2849	1030	2459	2238	348	599	503	526	1636	95	222	3.05	2.65	0.50	N/A
E6/38-46	VYCKQQLR	nb	11111	14467	15505	16175	17115	15710	3033	5667	5556	5567	14952	5.50	5.00	0.80	7
E6/42-50	QQLRREVY	nb	17322	16461	19322	18987	20352	18280	13213	19114	18698	10250	9654	15.50	14.15	42.00	18
E6/43-50	QLLRREVY	nb	21022	4694	23910	20930	4800	4727	11479	32729	43339	7983	10287	26.85	26.75	69.00	N/A
E6/43-51	QLLRREVYD	nb	29240	21987	36077	34485	31004	26124	11356	17273	14250	16006	9252	29.50	26.00	33.00	23
E6/43-52	QLLRREVYDF	nb	17677	24725	20195	19300	22371	23572	22452	12476	15639	14022	25863	29.00	24.50	17.00	21
E6/44-52	LLRREVYDF	nb	24721	21110	22277	21305	23841	22452	10993	48235	50327	15832	31322	31.50	27.50	73.00	20
E6/44-53	LLRREVYDFA	nb	21628	21822	20707	17155	23537	22573	25564	8167	7002	11632	9194	27.00	20.50	67.00	16
E6/48-55	EVYDFAFR	nb	19611	1119	15071	13702	1792	1415	561	31618	47850	1768	4497	33.40	33.30	17.00	N/A
E6/48-56	EVYDFAFRD	nb	31121	24018	34382	36902	37790	30069	2722	28272	22584	21154	15093	35.00	31.50	11.00	18
E6/48-57	EVYDFAFRDL	nb	22911	24229	23922	24652	26645	25426	8208	18368	22605	18611	21963	37.00	30.50	5.00	15
E6/52-61	FAFRDLCIVY	nb	6639	10871	7683	6734	6605	8479	5382	13276	12308	5604	5377	15.30	23.00	41.00	13
E6/52-62	FAFRDLCIVYR	nb	10465	3456	6145	5944	1780	2483	1249	970	855	2491	4055	1.90	1.10	29.00	N/A
E6/53-61	AFRDLCIVY	nb	7947	13978	9238	8918	9152	11294	5869	11333	10686	9428	10574	7.75	7.80	7.00	17
E6/53-62	AFRDLCIVYR	nb	5086	5136	3393	4587	6229	5651	3237	1556	1649	4400	7403	5.45	13.15	5.00	15
E6/58-67	CIVYRDGNPY	nb	3007	6794	2175	1467	9008	7860	9973	2698	2071	4835	5943	5.75	13.65	29.00	17
E6/59-68	IVYRDGNPYA	nb	7479	17009	8808	7085	7300	11113	4576	4593	3298	5575	519	7.80	15.50	13.00	19
E6/67-75	YAVCDKCLK	nb	8225	11823	6884	7155	9130	10414	2146	6100	5405	3822	2534	4.85	4.90	16.00	10
E6/68-76	AVCDKCLKF	nb	13832	19620	16946	18441	16787	18181	4242	18722	19943	7890	14977	12.25	11.70	0.30	22
E6/70-79	CDKCLKFYSK	nb	10655	10806	14495	15320	20437	14803	9866	1675	3035	4405	6634	8.15	15.50	14.00	13
E6/72-79	KCLKFYSK	nb	12372	528	17170	14702	998	727	2905	22435	13393	1162	3831	3.35	3.10	3.00	N/A
E6/73-83	CLKFYSKISEY	nb	1541	7090	1363	895	1720	3492	9866	1134	1659	2095	355	3.05	2.75	5.00	N/A
E6/74-83	LKFYSKISEY	nb	3754	11717	4158	3146	16994	14099	28485	4386	6842	9632	15399	11.15	19.00	18.00	9
E6/75-85	KFYSKISEYRH	nb	8574	5102	13864	12156	2831	3807	5267	2942	8929	8079	12451	24.40	23.70	7.00	N/A
E6/79-86	KISEYRHY	nb	12913	611	13744	12212	1275	884	3380	32880	35882	1746	7149	15.40	15.30	23.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/79-88	KISEYRHYCY	nb	573	430	626	421	573	495	2216	461	672	509	203	2.55	1.95	3.00	18
E6/80-88	ISEYRHYCY	nb	14211	16316	14320	15431	14386	15291	4288	10504	11690	12714	13604	10.45	9.90	38.00	10
E6/81-91	SEYRHYCYSLY	nb	8860	4532	13641	16141	3851	4174	5044	5066	5114	3824	8599	14.40	13.70	18.00	N/A
E6/84-93	RHYCYSLYGT	nb	20569	20078	27449	27461	24874	22330	8388	3027	958	18230	15439	18.95	12.45	18.00	8
E6/89-97	SLYGTTLQ	nb	4285	5751	1228	1059	4304	4963	933	6300	6963	2302	10187	5.40	5.55	0.40	20
E6/89-98	SLYGTTLQEQ	nb	5756	14450	2199	1801	10113	12052	5743	8651	11122	2883	8347	14.30	22.00	2.00	20
E6/91-101	YGTTLQEQYNK	nb	11681	18702	6768	8477	13689	15967	8950	10732	6820	9250	7260	19.55	18.90	25.00	N/A
E6/93-103	TTLQEQYNKPL	nb	27206	23949	28951	31980	28659	26265	8854	3554	1288	22232	28607	18.70	15.20	7.00	N/A
E6/94-101	TLEQQYNK	nb	10744	1343	10289	8604	1231	1290	3933	20746	13579	2690	1352	3.50	3.20	7.00	N/A
E6/105-115	DLLIRCINCQK	nb	2443	4239	2405	1146	3336	3766	3492	3018	218	1672	164	0.70	0.30	7.00	N/A
E6/106-114	LLIRCINCQ	nb	18801	22264	11564	8157	24479	23318	15710	87571	99038	12144	24530	31.00	27.00	56.00	16
E6/113-122	CQKPLCPEEK	nb	9042	6642	6630	6276	4342	5353	8032	3452	2001	7556	9345	6.00	13.60	3.00	15
E6/116-124	PLCPEEKQR	nb	34913	23558	36869	36295	35864	28952	9866	54873	43934	15986	25450	46.00	43.00	12.00	20
E6/122-129	KQRHLDKK	nb	12994	767	13836	14324	1853	1190	4626	21475	13768	978	771	4.20	4.10	9.00	N/A
E6/122-132	KQRHLDKKQRF	nb	33979	22040	34184	32190	30127	25703	42052	350	6321	18526	23787	45.50	42.50	10.00	N/A
E6/129-136	KQRFHNIR	nb	14665	2543	13352	14638	2352	2443	2635	31400	50105	7403	2866	35.20	35.00	49.00	N/A
E6/129-139	KQRFHNIRGRW	nb	15846	17671	20876	16748	19344	18478	15208	1908	4873	17770	11061	15.85	15.45	33.00	N/A
E6/133-142	HNIRGRWTGR	nb	919	17425	6983	5548	19060	18280	5743	8186	17147	5285	10897	16.25	15.60	20.00	11
E6/134-142	NIRGRWTGR	nb	3240	7258	8766	8302	7651	7487	1006	2974	3426	2543	1758	3.75	3.95	6.00	20
E6/136-146	RGRWTGRMSC	nb	32156	19609	31960	29346	30468	24349	22696	280	3410	20397	16936	32.50	29.50	35.00	N/A
E6/139-148	WTGRMSCCR	nb	12286	9643	10630	8888	11571	10584	1968	4065	3751	8770	13838	9.65	16.00	46.00	7
E6/142-151	RCMSCRSSR	nb	274	2109	229	339	6679	3746	3308	1397	1687	2146	2106	4.20	3.80	3.00	14
E6/143-151	CMSCRSSR	nb	712	373	666	891	608	477	667	660	435	698	665	1.20	1.50	3.00	10
E6/143-153	CMSCRSSRTR	nb	10382	5750	5407	4846	1196	2621	2216	4312	1373	1256	274	2.95	2.15	3.00	N/A
E6/144-151	MSCRSSR	nb	19626	2700	17132	13413	2101	2378	1438	34192	53812	4265	2270	40.40	40.30	32.00	N/A
E6/144-152	MSCRSSRT	nb	28749	21058	27339	24517	27263	23957	8032	21152	24592	15119	15448	31.50	28.00	62.00	5

(Continued)

Annex

Peptide ID ^a	Sequence ^b		Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/144-153	MSCCRSSRTR	nb		4818	2154	5367	4521	3957	2921	2077	989	721	1521	603	3.75	11.45	6.00	8
E6/144-154	MSCCRSSRTRR	nb		15035	7562	11875	11276	2469	4312	1805	4137	5079	2449	205	16.55	16.30	5.00	N/A
E6 L90V/89-99	SVYGTTLQYY	nb		721	1352	29778	347	244	573	1028	5763	4729	1982	738	12.15	12.10	0.50	N/A
E7/7-15	TLHEYMLDL	nb		15136	15410	18168	19458	18613	16946	3530	14137	12583	13479	18477	11.10	10.60	5.00	14
E7/12-20	MLDLQPETT	nb		32543	23566	34478	35580	37857	29745	32788	133463	126415	21275	30442	49.00	46.00	73.00	16
E7/14-23	DLQPETTDLY	nb		22015	15400	21016	21953	25320	19718	30396	12855	8039	15293	23937	28.50	22.00	28.00	19
E7/44-52	QAEPDRAHY	nb		27771	22103	31369	32322	33002	27132	15208	79499	75823	21057	26847	38.00	34.50	45.00	15
E7/49-57	RAHYNIVTF	nb		13237	16839	12923	11467	19464	18083	4152	33991	44238	9594	24294	16.85	16.25	20.00	16
E7/50-60	AHYNIVTFCK	nb		2918	2252	7882	10306	1675	1936	1947	332	586	943	220	1.05	0.65	3.00	N/A
E7/51-60	HYNIVTFCK	nb		3308	2689	4138	4050	4067	3308	2969	127	108	2199	1681	2.20	10.05	6.00	11
E7/52-60	YNIVTFCK	nb		7848	15180	6165	7522	7282	10527	2315	9558	10252	3253	2464	7.20	7.30	20.00	13
E7/53-60	NIVTFCK	nb		12936	3323	13003	10965	1813	2443	2607	21131	13517	1844	3885	3.65	3.55	13.00	N/A
E7/63-73	STLRVCVQSTH	nb		20788	2931	16029	18133	4771	3746	2722	1150	2896	6091	11072	17.00	13.50	7.00	N/A
E7/64-73	TLRVCVQSTH	nb		5914	4257	4625	5806	2514	3272	9146	3508	3624	4631	4644	8.30	16.00	15.00	25
E7/65-74	LRLVCVQSTHV	nb		19670	26030	25703	21694	40063	32435	50000	64131	67639	23725	31114	43.00	36.50	46.00	3
E7/66-73	RLCVQSTH	nb		15734	1369	10452	9517	1104	1229	4430	30898	50220	5586	739	35.45	35.25	7.00	N/A
E7/66-74	RLCVQSTHV	nb		9250	13538	15291	10478	13990	13723	5382	8637	8069	9182	10371	6.80	6.80	10.00	15
E7/66-75	RLCVQSTHVD	nb		18733	20298	30799	26869	29709	24614	16946	5809	3675	12125	10380	21.50	16.00	12.00	18
E7/73-81	HVDIRTLED	nb		29041	21958	34614	35870	30385	25842	7609	61145	46752	19553	17602	36.00	32.50	26.00	16
E7/78-87	TLEDLLMGTL	nb		23060	25539	30577	31871	31332	28332	25564	41983	39555	20850	29865	41.50	35.00	29.00	16
E7/81-89	DLLMGTLGI	nb		21838	21578	30568	31164	32692	26551	6063	63586	57916	20583	23491	29.50	25.50	58.00	17
E7/81-90	DLLMGTLGIV	nb		19532	22196	25829	23382	34987	27876	32788	17461	15893	19686	25085	30.50	24.50	70.00	13
E7/82-91	LLMGTLGIVC	nb		12569	17786	25636	25372	26190	21501	11356	1884	1973	6641	20295	7.35	13.55	34.00	19
E7/85-93	GTLGIVCPI	nb		16710	22145	14216	16237	16243	18985	2033	60864	61489	12525	15363	21.55	21.10	24.00	4
E7/87-97	LGIVCPICSQK	nb		4666	11010	3029	1894	3492	6195	6331	11986	2696	4029	1338	6.75	6.20	16.00	N/A
E7/89-98	IVCPICSQKP	nb		3036	26229	5633	3345	35410	30396	26408	55217	27114	14125	10856	21.10	29.00	16.00	15

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
A11																	
E6/106-115	LLIRCINCQK	1.36±0.30	351	147	200	197	119	132	1637	179	166	74	34	1.50	1.15	4.00	17
E7/88-97	GIVCPICSQK	2.22±0.84	185	79	113	109	87	83	2579	123	138	57	102	1.35	1.05	0.30	19
E6/52-62	FAFRDL CIVYR	2.67±1.28	6087	1546	1171	1597	158	495	704	1392	229	199	356	2.65	2.20	10.00	N/A
E7/89-97	IVCPICSQK	3.16±0.78	49	22	79	67	63	38	392	54	50	33	88	0.45	0.35	0.09	21
E6/93-101	TTLEQQYNK	4.71±1.74	24	18	18	28	20	19	167	30	25	17	17	0.30	0.25	0.01	21
E6/68-77	AVCDKCLKFY	4.80±0.55	220	312	420	534	400	354	3033	767	589	488	216	2.50	2.20	2.00	15
E6/33-41	IILECVYCK	4.98±1.33	43	22	53	36	32	27	241	83	80	30	25	0.55	0.45	0.30	18
E6/9-18	FQDPQERPRK	5.01±2.17	5528	3496	5224	5868	3275	3380	20590	1306	830	1263	3336	4.20	2.90	14.00	11
E6/107-115	LIRCINCQK	5.21±2.43	1089	2360	338	287	324	874	1158	697	836	399	1185	2.25	2.00	8.00	16
E6/92-101	GTTLEQQYNK	5.26±2.57	127	76	58	93	92	84	1333	119	65	29	20	0.80	0.75	0.40	23
E6/94-101	TLEQQYNK	6.03±2.84	3132	1354	7465	5308	349	685	3237	3425	15058	377	1230	11.05	10.65	22.00	N/A
E6/32-41	DIILECVYCK	6.80±2.41	448	483	924	496	1165	751	2216	452	315	612	189	2.00	1.65	13.00	18
E6/31-41	HDIILECVYCK	7.45±0.64	2169	6604	1213	971	6536	6575	4063	79	131	4300	89	1.90	1.65	10.00	N/A
E7/87-97	LGIVCPICSQK	7.64±2.77	1807	4075	1627	1398	889	1905	5933	1801	596	1572	820	3.80	3.75	26.00	N/A
E6/34-41	IILECVYCK	8.34±3.03	4751	642	4510	2359	241	395	2123	201	284	163	139	0.70	0.30	8.00	N/A
E6 D32E/31-41	HEIILECVYCK	8.70±4.08	2171	6972	1220	971	7181	7093	3726	159	1503	2828	418	7.70	7.45	12.00	N/A
E6 D32E/32-41	EIILECVYCK	8.79±1.87	459	261	616	398	437	339	1805	441	349	261	82	2.05	1.70	11.00	18
E6/68-75	AVCDKCLK	9.72±3.73	1413	25	3434	2359	56	38	819	422	3249	28	1027	2.95	2.90	7.00	N/A
E6/67-75	YAVCDKCLK	11.35±2.66	1462	2153	834	980	1371	1719	1485	935	850	305	5722	2.45	2.00	9.00	12
E6 D32E, I34R/31-41	HEIRILECVYCK	11.67±3.82	13242	5695	11653	11660	5677	5682	3380	305	1539	3796	1269	8.85	7.65	9.00	N/A
E6/142-151	RCMSCRSSLR	12.19±1.50	1928	11341	1956	1721	5277	7734	4626	5765	17492	10367	21148	12.65	11.65	4.00	9
E6/68-78	AVCDKCLKFYS	13.16±3.30	9183	1847	12748	11471	6654	3511	3033	69	1033	761	61	5.85	5.30	0.80	N/A
E7/52-60	YNIVTFCK	13.23±7.77	2722	2772	780	961	821	1510	1109	1253	1423	1366	1154	3.30	2.60	18.00	13
E6/105-115	DLLIRCINCQK	14.75±5.33	4278	2803	5361	3140	2240	2510	4478	1966	1148	3142	372	6.05	5.65	30.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/53-60	NIVTFCCK	15.96±4.18	8169	166	7719	5011	157	162	2391	2888	3039	177	629	3.15	2.75	6.00	N/A
E6/66-75	PYAVCDKCLK	15.99±2.11	8296	18387	7230	7038	23080	20702	31399	9546	5766	6694	10152	8.60	7.15	48.00	14
E6 I34R/34-41	RLECVYCK	17.52±4.56	4437	277	2509	1319	199	235	1453	322	307	117	265	0.75	0.35	4.00	N/A
E6/59-67	IVYRDGNPY	19.84±4.91	576	407	541	480	369	386	704	483	449	218	156	1.45	1.35	11.00	10
E6/65-75	NPYAVCDKCLK	22.27±17.81	19279	3338	12853	18020	11107	6096	3726	510	131	5561	5143	4.85	4.30	6.00	N/A
E7 S63F/56-66	TFCCKCDFTLR	24.31±10.02	19753	2790	17962	21979	7353	4527	6683	423	661	1348	6041	7.45	6.90	16.00	N/A
E6/48-55	EVYDFAFR	27.75±5.14	9342	94	6461	4961	107	100	419	1464	556	146	266	0.85	0.45	10.00	N/A
E7/50-60	AHYNIVTFCCK	28.96±12.39	4515	1025	3048	5909	874	948	1844	266	113	1799	3427	1.95	1.55	2.00	N/A
E6/37-46	CVYCKQQLLR	36.49±6.55	506	291	357	349	113	181	561	648	297	55	50	1.90	1.55	0.50	24
E6/67-77	YAVCDKCLKFY	40.88±5.62	2635	21532	10968	7709	8055	13142	11479	1834	1159	7953	13620	6.00	5.70	62.00	N/A
E6/91-101	YGTTLQYYNK	47.05±8.49	963	9590	540	957	3197	5530	6331	1873	1374	4319	6339	6.25	6.05	10.00	N/A
E6/68-76	AVCDKCLKF	48.82±11.21	9085	13066	10508	12744	11306	12183	3134	4000	3633	3531	6299	5.55	4.65	2.00	19
E6/53-62	AFRDLCIVYR	53.65±13.73	6053	5149	2782	3801	3766	4406	8208	2715	2923	2623	5953	6.15	4.85	8.00	18
E6/79-88	KISEYRHYCY	55.30±14.40	264	222	716	554	463	320	2011	698	642	97	106	2.60	2.30	21.00	6
E6/69-79	VCDKCLKFYSK	58.43±5.66	16577	4488	16473	20499	2959	3646	11113	2901	11275	2927	4468	28.35	27.75	3.00	N/A
E6/80-88	ISEYRHYCY	62.36±12.84	4912	6762	7399	7568	6917	6829	2417	3945	4210	1090	1075	5.70	5.00	52.00	10
E6/139-148	WTGRCSCCR	82.74±12.11	15228	5497	5392	4431	1326	2693	1263	3372	2398	3412	2002	8.00	6.75	27.00	18
E6/8-18	MFQDPQERPRK	nb	11309	3363	9296	10202	1758	2443	7287	1847	17	1280	5135	2.05	1.05	3.00	N/A
E6/18-25	KLPQLCTE	nb	41574	23226	38757	38182	36610	29109	50000	143764	296319	16633	23568	80.50	80.00	66.00	N/A
E6/23-33	CTELQTTIHDI	nb	35051	20321	35188	35926	29034	24349	33505	1847	1164	22799	24855	27.75	25.25	44.00	N/A
E6/27-36	QTTIHDIILE	nb	25662	23215	17923	15821	28102	25564	26695	27659	21971	5541	15293	29.00	24.50	19.00	17
E6/28-36	TTIHDIILE	nb	9741	6489	6683	5708	12956	9146	3726	2845	1553	2152	9126	3.70	2.80	3.00	18
E6/28-37	TTIHDIILEC	nb	15893	23057	21278	20096	17149	19825	4936	3731	5406	13870	14442	10.95	10.05	6.00	17
E6/29-37	TIHDIILEC	nb	17265	18314	24875	27745	22615	20368	4626	5483	6044	14935	15950	11.30	10.85	8.00	14
E6/29-39	TIHDIILECVY	nb	4812	5405	2673	3895	1041	2365	6400	4926	35490	2175	560	39.85	39.45	48.00	N/A
E6/32-42	DIILECVYCKQ	nb	12219	7740	25455	17383	40060	17696	50000	1642	3237	15002	19725	13.60	12.45	46.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/33-42	IILECVYCKQ	nb	2187	24375	3232	1575	19168	21617	12517	27532	52101	3679	7104	19.65	18.65	32.00	7
E6/37-44	CVYCKQQL	nb	39605	20488	31864	31728	18718	19611	8297	4108	696	14035	17919	23.90	23.40	55.00	N/A
E6/37-47	CVYCKQQLLR	nb	4832	1260	2457	2404	182	479	612	78	232	94	114	2.65	2.25	0.40	N/A
E6/38-46	VYCKQQLLR	nb	6611	8382	9908	8471	8972	8664	2265	2643	2015	4227	15378	4.10	3.35	10.00	14
E6/47-55	REVDFAFR	nb	13510	17314	10984	12365	10367	13429	3344	5943	8323	2239	15158	10.55	10.15	27.00	9
E6/48-56	EVYDFAFRD	nb	22421	15320	29166	33872	30898	21852	1290	3048	3567	10001	14611	17.45	13.95	3.00	14
E6/52-61	FAFRDLCIVY	nb	6696	4063	5787	4891	2976	3473	2782	3061	2365	2112	1561	5.80	4.45	59.00	6
E6/53-63	AFRDLCIVYRD	nb	27978	12086	26498	26640	38835	21734	50000	416	397	23325	24906	14.50	11.50	69.00	N/A
E6/54-61	FRDLCIVY	nb	36720	22894	37300	37093	28083	25426	50000	234234	69304	22575	27541	48.50	48.00	31.00	N/A
E6/58-67	CIVYRDGNPY	nb	4176	5581	1954	1550	5378	5471	14099	2863	2280	1202	1713	5.60	4.40	61.00	6
E6/62-72	RDGNPYAVCDK	nb	21484	7901	22986	23426	15751	11113	7287	405	216	2351	7852	6.15	5.75	8.00	N/A
E6/63-73	DGNPYAVCDKC	nb	36277	26843	40097	40539	46011	35177	50000	6569	24496	22645	33805	60.50	58.00	94.00	N/A
E6/64-72	GNPYAVCDK	nb	12578	18284	9064	5427	9189	12930	6400	7397	7300	3811	8955	9.60	8.90	32.00	14
E6/69-77	VCDKCLKFY	nb	30699	22591	27152	27718	21307	21971	32788	130326	176319	19430	25705	54.00	51.00	42.00	0
E6/70-77	CDKCLKFY	nb	38208	21756	36761	37951	30934	25983	50000	10535	636	19012	29119	19.35	18.85	52.00	N/A
E6/70-79	CDKCLKFYSK	nb	5288	12019	6894	8274	10721	11356	6978	1086	4829	1058	9526	7.90	6.65	22.00	11
E6/71-79	DKCLKFYSK	nb	18144	19383	27730	25329	28657	23572	6400	16789	18980	5625	12559	22.00	19.00	14.00	11
E6/72-79	KCLKFYSK	nb	6484	1728	15934	15542	1122	1392	3272	12096	11214	1689	9415	9.05	8.65	5.00	N/A
E6/73-83	CLKFYSKISEY	nb	19570	9932	8921	7377	7205	8479	27279	3323	10817	7637	5332	28.70	28.15	61.00	N/A
E6/74-84	LKFYSKISEYR	nb	15476	12298	12678	11848	10622	11417	6063	1642	496	13037	11586	5.35	4.40	60.00	N/A
E6/75-83	KFYSKISEY	nb	4210	4948	4132	4278	4103	4502	4107	2594	2373	4973	18754	4.35	3.65	8.00	3
E6/75-84	KFYSKISEYR	nb	2367	584	1633	1605	1462	928	1551	625	622	1137	6674	3.60	2.55	5.00	9
E6/75-85	KFYSKISEYRH	nb	25752	8713	19655	18416	9791	9246	7692	161	1671	14699	27147	18.00	15.00	10.00	N/A
E6/76-84	FYSKISEYR	nb	11508	12364	7323	7460	10358	11294	1453	7781	7367	7599	11446	9.10	8.35	56.00	8
E6/77-84	YSKISEYR	nb	19079	3096	13960	10650	1468	2134	2100	15405	5981	5074	14433	6.95	6.75	63.00	N/A
E6/77-85	YSKISEYRH	nb	19254	16361	21121	23417	12001	14023	4335	17260	15146	13898	20138	21.00	17.50	87.00	10

(Continued)

Annex

Peptide ID ^a	Sequence ^b		Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/78-88	SKISEYRHYCY	nb		3460	19566	12020	7956	23662	21501	37740	1554	4899	8211	17267	16.80	16.45	17.00	N/A
E6/79-86	KISEYRHY	nb		15027	2286	19199	17932	1679	1957	5621	67398	72738	1523	11312	35.75	35.40	22.00	N/A
E6/79-87	KISEYRHYC	nb		11648	14766	19173	20197	14166	14407	2782	2911	2827	10281	2609	5.55	4.80	17.00	6
E6/79-89	KISEYRHYCYS	nb		11597	11997	14596	13338	5419	8076	4936	699	7501	4377	247	22.00	20.95	19.00	N/A
E6/81-89	SEYRHYCYS	nb		25883	19890	20097	23398	27662	23445	4478	10940	18548	7243	27180	29.50	26.00	40.00	3
E6/81-91	SEYRHYCYSLY	nb		20957	6186	11631	17142	4371	5211	6829	7736	11752	4964	24634	30.25	30.05	35.00	N/A
E6/84-91	RHYCYSLY	nb		26264	10994	18175	20679	3180	5901	7054	39233	18355	11030	28424	16.50	16.40	17.00	N/A
E6/85-95	HYCYSLYGTTL	nb		37470	25551	32541	36003	32211	28640	41599	45761	62244	22095	30317	73.00	70.50	86.00	N/A
E6/88-97	YSLYGTTLQ	nb		16898	21458	5992	4900	15327	18181	9973	15305	19048	5990	17462	16.35	15.85	36.00	16
E6/89-97	SLYGTTLQ	nb		5669	13070	2519	1994	9063	10875	1805	5001	5168	910	20130	6.90	6.15	2.00	15
E6/89-99	SLYGTTLQYY	nb		1707	1606	508	751	579	963	2843	2246	8572	1394	2182	22.35	22.45	4.00	N/A
E6/92-102	GTTLEQQYNKP	nb		4733	21741	9681	6918	37343	28485	43912	783	3674	14847	16813	13.85	13.45	17.00	N/A
E6/92-99	GTTLEQQY	nb		29129	9316	28051	28460	6620	7860	10875	45777	9263	4644	9252	14.00	13.50	61.00	N/A
E6/93-102	TTLQYYNKP	nb		710	25384	1893	1706	30885	27876	12653	34028	37831	14394	16619	16.20	15.85	6.00	11
E6/93-103	TTLQYYNKPL	nb		7924	21009	14605	16816	16392	18579	6611	1598	2085	16752	22578	9.45	8.95	9.00	N/A
E6/101-109	KPLCDLLIR	nb		16969	19683	15103	15434	15356	17411	2524	2302	2240	8793	21212	8.05	7.60	33.00	13
E6/108-115	IRCINCQK	nb		18779	18875	21088	16016	17465	18181	20814	15763	1559	8481	16023	3.15	3.00	19.00	N/A
E6/113-122	CQKPLCPEEK	nb		4576	2973	4838	5449	1259	1936	11356	2387	1202	3343	1925	4.55	3.35	6.00	13
E6/115-122	KPLCPEEK	nb		28490	12327	32109	29853	12098	12249	5743	1034	459	8723	16931	5.25	5.20	53.00	N/A
E6/116-124	PLCPEEKQR	nb		32709	23219	37177	36830	32555	27427	13649	97281	81339	16598	29756	51.50	48.50	41.00	15
E6/118-128	CPEEKQRHLDK	nb		36268	18581	35259	36150	27538	22696	26408	1206	306	13099	17701	27.55	25.05	23.00	N/A
E6/122-129	KQRHLDDK	nb		11498	4979	18901	23658	3343	4085	13357	3861	4072	2586	2227	4.15	3.75	25.00	N/A
E6/129-136	KQRFHNIR	nb		25255	10718	19893	21443	4510	6978	9346	7721	31100	5788	8921	22.55	22.45	32.00	N/A
E6/129-138	KQRFHNIRGR	nb		12019	14503	4445	4250	6159	9448	12930	15200	12046	5409	10893	12.25	10.55	41.00	9
E6/134-142	NIRGRWTGR	nb		20180	20213	17811	14494	13249	16316	3646	18622	21004	5891	18208	24.00	20.50	47.00	15
E6/140-148	TGRMSSCCR	nb		18365	18486	14196	11665	12020	14883	4576	26244	32307	7303	13808	26.00	22.50	61.00	8

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/143-151	CMSCCRSSR	nb	1843	2401	1773	1592	512	1109	1290	2183	1366	488	1272	3.20	2.55	10.00	14
E6/143-153	CMSCCRSSRTR	nb	12933	13841	9396	8620	978	3666	4676	9874	11671	2050	6312	27.80	26.60	12.00	N/A
E6/144-151	MSCCRSSR	nb	13501	327	11910	8485	461	388	1196	19528	7191	745	637	7.55	7.30	13.00	N/A
E6/144-152	MSCCRSSRT	nb	23954	18779	27596	22982	25499	21852	7692	21283	20198	11061	9725	27.50	24.00	59.00	10
E6/144-153	MSCCRSSRTR	nb	1644	1031	2309	1534	1231	1127	2123	853	738	904	1212	3.45	2.75	18.00	18
E6/144-154	MSCCRSSRTRR	nb	9728	1378	5072	5471	653	948	1534	79	499	1091	1068	3.85	3.30	4.00	N/A
E6/145-153	SCCRSSRTR	nb	26662	19672	24146	18884	24619	22090	8388	29925	37265	15486	27913	35.50	32.00	0.80	11
E6 D32E/23-33	CTELQTTIHEI	nb	32014	16635	32406	32670	25839	20702	24218	1895	711	22160	24837	20.80	17.80	26.00	N/A
E6 D32E/28-36	TTIHEIILE	nb	6961	5687	6279	5394	13543	8759	4152	2825	1568	1936	8587	3.60	2.80	3.00	18
E6 L90V/89-97	SVYGTTLLEQ	nb	2254	5675	734	643	3693	4576	776	1538	1582	222	12523	3.55	2.85	0.20	19
E6 L90V/89-99	SVYGTTLLEQQY	nb	369	354	100	181	143	226	1222	691	2291	373	222	9.20	9.05	0.70	N/A
E7/7-15	TLHEYMLDL	nb	22420	18281	22887	23312	21434	19825	7366	22858	26080	14281	25502	27.50	24.00	54.00	10
E7/18-25	ETTDLYCY	nb	27010	7943	27712	28057	9094	8479	3380	23749	45789	3114	12560	30.30	30.20	18.00	N/A
E7/18-26	ETTDLYCYE	nb	23907	19260	16298	16808	30867	24349	4727	23444	18806	7639	19569	27.00	23.50	19.00	12
E7/19-27	TTDLYCYEQ	nb	10164	15002	11982	10078	16057	15541	4335	12707	12031	2173	19452	10.60	9.80	7.00	11
E7/19-29	TTDLYCYEQLN	nb	29490	9840	32337	35659	29561	17038	21971	1098	1104	10812	9480	19.10	16.10	39.00	N/A
E7/31-40	SSEEEDEIDG	nb	42990	29440	45399	46389	45071	36337	50000	109353	85872	21626	33015	69.00	66.00	33.00	17
E7/36-43	DEIDGPAG	nb	46217	23914	47383	48003	47892	33869	50000	158362	41282	27793	34959	74.00	73.50	48.00	N/A
E7/41-50	PAGQAEPDRA	nb	41497	30039	44969	46355	46427	37333	50000	820030	482892	23063	32517	79.50	76.50	91.00	1
E7/42-50	AGQAEPDRA	nb	32898	21926	40701	41723	43042	30727	50000	310477	299433	19488	31857	62.00	59.00	49.00	5
E7/51-60	HYNIVTFCK	nb	1042	1761	1214	1213	1013	1340	1728	961	1007	986	2066	3.30	2.85	14.00	11
E7/55-63	VTFCKCDS	nb	20535	18556	18465	15433	17410	17888	2315	13123	13593	6336	1326	21.50	18.00	41.00	11
E7/56-66	TFCKCDSTLR	nb	22831	8579	20910	22934	9971	9246	8854	671	675	2175	10407	13.15	9.15	13.00	N/A
E7/63-71	STLRVCVQS	nb	11216	8885	8399	7312	8634	8759	3066	3750	3866	2592	8124	6.05	5.25	1.00	18
E7/63-73	STLRVCVQSTH	nb	21889	449	8508	11210	979	663	3237	2757	1812	3677	12740	13.50	12.20	5.00	N/A
E7/66-75	RLCVQSTHVD	nb	31503	27490	38874	36666	38339	32435	26124	89708	165520	16440	26199	51.00	47.50	46.00	8

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/68-77	CVQSTHVDIR	nb	3271	3233	2116	2449	1566	2253	2011	8163	5700	1292	1393	8.30	7.15	10.00	20
E7/69-77	VQSTHVDIR	nb	3616	4855	5840	6053	4921	4883	2905	2183	2199	1799	2502	4.15	3.45	9.00	10
E7/71-80	STHVDIRTLE	nb	17397	19635	10690	8181	20654	20149	24218	24936	18529	2351	11709	15.90	15.35	6.00	16
E7/77-84	RTLEDLLM	nb	28534	8942	25150	25073	4760	6540	3308	28884	63644	8568	11670	37.50	37.00	15.00	N/A
E7/77-86	RTLEDLLMGT	nb	17957	20735	11548	10615	11108	15208	3454	7275	4919	5428	12450	13.00	10.60	5.00	11
E7/85-93	GTLGIVCPI	nb	11139	13790	4369	5656	7782	10358	1236	4831	5988	6941	15101	7.90	7.10	2.00	16
E7/88-95	GIVCPICS	nb	37711	18661	36952	36530	22815	20590	19296	1861	2192	12178	21861	19.35	18.85	38.00	N/A
E7/88-98	GIVCPICSQKP	nb	7427	24720	9779	5978	36023	29907	50000	1519	19683	19797	18277	33.45	32.95	24.00	N/A
E7/89-98	IVCPICSQKP	nb	2205	27040	4961	2826	33583	30069	22943	198989	216197	16509	16847	31.15	30.15	39.00	11
A24																	
E6/87-94	CYSLYGTT	0.08±0.00	18030	11113	27793	26920	8751	9866	4676	12171	8495	2537	5050	10.55	11.60	33.00	N/A
E6 L90V/86-95	YCYSVYGTTL	0.08±0.11	188	5938	298	305	12415	8571	4936	1488	1913	4369	5506	2.30	2.00	3.00	10
E6/87-95	CYSLYGTTL	0.10±0.02	107	113	205	189	286	180	477	655	578	79	58	0.75	0.60	0.06	20
E6/85-95	HYCYSLYGTTL	0.10±0.03	1871	496	783	649	503	501	903	2075	6502	358	1107	11.30	13.45	0.80	N/A
E6/49-57	VYDFAFRDL	0.10±0.04	1877	769	726	967	596	678	1146	1306	1030	884	105	1.65	1.55	0.20	23
E7 S63F/56-65	TFCKCDFTL	0.11±0.17	1907	1649	2338	1735	1390	1518	1947	1398	2102	691	1316	2.80	2.15	2.00	18
E7/49-57	RAHYNIVTF	0.15±0.07	763	1699	1104	910	334	751	1518	2014	1647	770	589	1.80	1.45	0.80	10
E6 L90V/87-95	CYSVYGTTL	0.24±0.04	142	120	184	206	304	191	482	643	586	85	79	0.85	0.60	0.07	20
E6/82-90	EYRHYCYSL	0.32±0.10	856	807	532	573	352	534	419	853	999	205	327	1.30	0.95	0.04	19
E6/49-59	VYDFAFRDLCI	0.40±0.29	2481	189	2074	2621	365	263	1392	7909	3927	648	75	5.40	7.45	3.00	N/A
E6/66-76	PYAVCDKCLKF	0.48±0.21	1977	236	1844	3418	386	301	893	670	2729	83	34	3.05	5.20	0.50	N/A
E6/98-107	QYNKPLCDLL	0.54±0.24	2235	1042	536	614	517	731	1469	820	1096	329	324	1.85	1.20	0.30	23
E6/98-108	QYNKPLCDLLI	0.56±0.08	2769	449	1162	1660	151	260	913	7693	10967	135	94	19.90	21.95	1.00	N/A
E7/56-65	TFCKCDSTL	0.58±0.55	9077	5078	8383	5243	4108	4576	3726	3473	4940	3332	1741	4.40	3.55	3.00	17
E6 L90V/90-99	VYGTTLQY	0.68±0.05	6615	7372	8381	8526	3113	4779	3033	1492	1895	1630	785	2.70	1.95	0.50	12

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/18-26	KLPQLCTEL	0.83±0.07	11101	13344	6747	6113	8995	10934	2315	8279	8844	1551	9767	8.40	7.65	0.40	13
E6/60-69	VYRDGNPYAV	0.95±0.39	3260	2003	2111	1934	1597	1795	1347	825	920	402	790	1.70	1.05	0.80	13
E7 S63F/56-63	TFCKCD F	1.03±0.41	18965	7402	16431	11201	2218	4041	1277	5772	7519	2719	6732	10.20	11.10	14.00	N/A
E6/67-76	YAVCDKCLKF	1.09±0.30	12915	9223	7693	7763	7291	8208	4727	4706	3004	4306	1737	4.60	4.20	4.00	12
E7/48-57	DRAHYNIVTF	1.32±0.35	6106	11562	12097	11145	18181	14564	12382	3645	3355	2429	983	3.55	2.85	4.00	11
E6/26-34	LQTTIHDII	1.42±0.23	11192	10542	12843	14956	14228	12249	5621	5807	6586	7615	5027	6.40	5.65	2.00	12
E6/82-91	EYRHYCYSLY	1.59±0.41	6231	3583	7218	7320	5263	4335	4527	1081	1330	703	6922	2.15	1.45	2.00	11
E6 D32E/26-34	LQTTI HE II	1.65±0.37	7820	5961	11329	12345	10228	7776	5382	4073	4504	5914	4255	4.60	4.30	1.00	13
E6/66-74	PYAVCDKCL	1.97±0.77	9731	3772	7066	5470	5884	4702	3272	3373	2470	1091	235	3.15	2.80	0.80	20
E7/50-57	AHYNIVTF	2.41±0.76	12214	4182	10388	7996	1684	2664	2874	53495	140650	535	331	42.65	44.60	14.00	N/A
E6/98-106	QYNKPLCDL	2.97±0.67	5603	4509	3321	2279	1869	2905	1017	2183	2475	774	1138	3.05	2.80	0.09	21
E6/51-59	DFAFRDLCI	3.27±0.99	5300	2280	8359	6895	7359	4107	3202	1941	1363	2322	858	2.05	1.80	6.00	17
E6/88-95	YSLYGTTL	3.36±1.09	5320	17026	10801	9147	7313	11113	4936	86559	120542	7199	1116	40.85	43.10	22.00	N/A
E6/76-83	FYSKISEY	3.36±2.28	17138	7035	12705	12310	1131	2827	3607	10973	18161	2205	1033	16.85	18.10	2.00	N/A
E6/38-45	VYCKQQLL	3.51±1.64	1910	2178	3319	2262	267	759	1158	14700	3992	52	135	6.20	8.45	4.00	N/A
E6/44-54	LLRREYDFAF	3.84±0.77	23190	7208	14267	18252	3850	5267	3646	1556	1387	2028	491	5.85	5.65	19.00	N/A
E6/49-58	VYDFAFRDLC	4.30±0.37	7810	14614	5873	6027	7541	10471	4626	582	721	3262	9637	1.65	0.85	6.00	14
E6/48-57	EVYDFAFRDL	4.75±0.58	10889	26900	7301	7779	26976	26985	20814	25981	33863	10585	11189	14.70	13.90	24.00	12
E6/76-85	FYSKISEYRH	4.76±0.60	4654	9583	19540	17447	6083	7609	5211	818	360	2558	1544	1.10	0.45	4.00	11
E6/90-99	LYGTTLEQQY	5.63±3.15	10663	9201	11003	9536	3570	5743	3530	1674	2316	1942	1818	3.30	2.50	1.00	10
E7/47-57	PDRAHYNIVTF	5.75±4.30	20603	17668	20625	25910	18064	17888	18083	1500	6111	7038	3584	14.35	14.10	9.00	N/A
E7/51-59	HYNIVTFCC	6.06±2.70	3701	6548	7451	5298	6494	6505	579	586	1098	883	2588	1.85	1.60	0.60	12
E6/87-96	CYSLYGTTL	6.87±1.89	1125	7385	5861	3660	15918	10816	5044	415	319	1460	771	0.80	0.40	4.00	12
E6/72-80	KCLKFYSKI	7.50±2.33	2342	2669	17531	14133	3487	3050	1655	1371	1998	699	767	2.50	2.30	0.40	16
E6/35-44	LECVYCKQQL	8.13±0.64	37771	33156	37066	38822	36027	34610	49462	32185	35214	10865	22228	38.00	36.00	30.00	10
E6 L90V/82-90	EYRHYCYS V	8.64±0.72	3061	2457	1399	1604	1622	2000	963	1853	2386	518	1884	2.90	2.70	0.40	9

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/76-86	FYSKISEYRHY	10.60±1.96	14076	2765	8763	11705	1544	2066	3891	9018	6517	2915	1074	12.55	13.45	0.70	N/A
E6/66-75	PYAVCDKCLK	10.93±1.40	26543	17628	27683	25169	25705	21385	20368	4494	2442	6886	2263	13.75	11.75	11.00	12
E6/50-59	YDFAFRDLCI	11.01±0.86	2957	17550	4658	2878	6759	10875	14252	3620	3025	12352	1705	3.35	2.70	30.00	10
E6/81-90	SEYRHYCYSL	11.98±1.32	825	14977	744	480	8304	11173	11858	5354	8684	4777	7562	5.35	5.00	24.00	10
E7/51-58	HYNIVTFC	16.51±1.47	23030	7863	21161	22350	7787	7818	2290	1325	5677	4943	6470	9.65	9.60	11.00	N/A
E6/38-47	VYCKQQLRR	22.28±2.19	13917	10568	28045	25203	16490	13142	11356	3553	1311	3204	1079	3.65	3.15	1.00	13
E6/47-54	REVDYFAF	23.27±1.19	23692	7169	23030	18654	9647	8297	3766	5074	3035	1850	9089	7.45	7.40	45.00	N/A
E6/131-139	RFHNIRGRW	23.93±3.24	2159	2788	1704	1124	621	1311	1121	2104	1716	354	812	2.25	2.05	0.50	6
E6 A68G/60-69	VYRDGNPYGV	24.65±2.54	4789	2905	3668	3729	2450	2664	1602	1376	1544	679	810	2.30	1.65	1.00	13
E6/128-136	KKQRFHNIR	28.64±3.18	35997	29829	38575	36419	37407	33324	50000	28181	31378	18840	6674	35.50	34.00	45.00	2
E7 L28F/21-28	DLYCYEQF	29.23±2.92	24390	14260	20680	23229	7056	10027	5099	23513	30001	2811	202	25.10	25.05	26.00	N/A
E6/127-134	DKKQRFHN	31.27±15.48	46950	40157	48607	48874	48881	44390	50000	33675	34288	25047	6880	67.00	66.50	75.00	N/A
E6 L90V/90-97	VYGTTLQ	31.98±16.05	25903	7583	32850	34571	10002	8711	5044	1123	59	3436	6321	3.90	3.85	7.00	N/A
E6/75-85	KFYSKISEYRH	32.00±4.31	19160	22292	34292	32363	19059	20590	9866	4075	8671	8006	7593	18.65	18.40	3.00	N/A
E6 L90V/87-96	CYSVYGTTL	32.71±2.44	1650	8585	6458	4246	16616	11922	5099	394	323	1570	974	0.90	0.40	5.00	12
E7/67-76	LCVQSTHVDI	34.74±11.56	21249	25625	32845	34208	26981	26265	10302	36783	42729	9468	11136	25.00	25.50	33.00	11
E6/86-96	YCYSLYGTTL	36.66±4.08	2314	30606	9766	7611	39763	34987	24218	5599	7535	16140	14997	13.85	15.95	55.00	N/A
E6 L90V/88-95	YSVYGTTL	37.88±0.91	7092	20039	11347	11242	14992	17317	9654	95567	32972	8065	1382	23.85	26.10	14.00	N/A
E6/75-83	KFYSKISEY	41.70±23.44	14961	14099	5644	4220	4664	8120	3344	11040	11006	4294	28390	10.70	10.05	0.60	9
E6/38-46	VYCKQQLLR	46.40±5.14	6745	4118	20124	17707	15602	7989	4626	1524	1724	1776	1457	2.30	2.05	0.70	15
E6/128-135	KKQRFHNI	46.84±2.41	24112	13730	29623	30642	9434	11417	14723	12368	26190	3941	1140	23.50	23.45	63.00	N/A
E6/127-135	DKKQRFHNI	55.95±3.54	25336	24382	38730	39079	38876	30893	41152	13645	11057	8260	3169	18.00	16.00	3.00	11
E6/125-135	HLDKKQRFHNI	60.80±1.43	35115	16930	30329	35515	17357	17131	24218	19639	15598	10190	6917	42.00	40.50	15.00	N/A
E6 L90V/81-91	SEYRHYCYSVY	63.32±8.71	18432	28728	22981	18169	30610	29745	50000	9957	13617	12712	32388	26.40	26.15	39.00	N/A
E7/69-76	VQSTHVDI	84.61±2.96	29711	15729	33856	33737	15104	15374	8032	46592	10427	7668	14279	16.50	16.00	64.00	N/A
E6/11-19	DPQERPRKL	nb	40129	34851	44372	45928	45152	39837	50000	105187	62897	18732	21547	51.50	50.00	3.00	16

(Continued)

Annex

Peptide ID ^a	Sequence ^b		Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/36-46	ECVYCKQQLR	nb		31380	39352	37342	38120	43553	41375	50000	9730	12025	25524	16348	33.00	31.50	45.00	N/A
E6/42-52	QQLLRREVDYDF	nb		18490	3737	13841	13219	3235	3473	4242	3035	10425	2629	25332	21.90	21.65	9.00	N/A
E6/43-52	QLLRREVDYDF	nb		6035	7442	9138	8434	3289	4936	3237	1856	2629	1310	1172	3.15	2.45	11.00	13
E6/44-52	LLRREVDYDF	nb		4528	7209	5067	3268	4035	5382	2123	3281	3464	1643	5880	3.80	3.55	3.00	10
E6/45-54	LRREVDYDFAF	nb		9033	12546	13888	15282	15576	13947	5267	1873	1895	3396	6883	2.80	1.95	8.00	12
E6/47-57	REVDYDFAFRDLD	nb		14574	25290	6586	8447	25531	25426	30069	12221	11352	11116	16109	21.75	22.45	18.00	N/A
E6/48-58	EVDYDFAFRDLC	nb		22725	39872	25416	24586	39680	39837	37740	15385	41028	20922	22169	47.75	47.55	49.00	N/A
E6/50-60	YDFAFRDLCIV	nb		16421	31557	26268	21789	19151	24482	40271	18928	13185	18867	20628	24.75	24.95	33.00	N/A
E6/51-60	DFAFRDLCIV	nb		20731	19913	20286	19690	18672	19296	19932	40239	38170	7286	12721	23.50	24.50	44.00	6
E6/51-61	DFAFRDLCIVY	nb		33255	18144	32031	33778	24872	21154	25842	949	10048	13383	31992	32.00	30.50	14.00	N/A
E6/59-68	IVYRDGNPYA	nb		27618	36911	29113	30399	37469	37132	44390	10015	7811	14404	27624	17.10	15.10	14.00	0
E6/62-69	RDGNPYAV	nb		38038	18403	38619	39114	19560	18985	25564	587	109	10139	9244	12.70	12.20	33.00	N/A
E6/64-74	GNPYAVCDKCL	nb		34978	35991	27741	26546	34797	35367	33144	15564	14292	16939	30659	39.50	38.00	52.00	N/A
E6/65-74	NPYAVCDKCL	nb		18381	36136	15643	10468	36642	36337	50000	166207	127562	19232	31871	35.90	35.65	37.00	10
E6/73-83	CLKFYISKISEY	nb		37525	35096	23225	24856	26618	30561	50000	11223	15779	26915	35882	45.00	43.50	31.00	N/A
E6/76-84	FYSKISEYR	nb		7664	3686	9578	9771	4156	3912	4626	2094	1980	2517	3041	2.60	2.30	0.50	13
E6/80-90	ISEYRHYCYSL	nb		3113	13390	2355	2834	12137	12722	11987	11861	22703	8899	2261	33.90	35.95	21.00	N/A
E6/81-91	SEYRHYCYSLY	nb		9612	24433	16610	10564	22661	23572	50000	10306	10141	10111	29162	18.55	20.45	34.00	N/A
E6/85-94	HYCYSLYGTT	nb		9118	13239	13727	13921	13999	13575	3849	7172	16662	4160	8822	8.90	8.05	6.00	10
E6/86-95	YCYSLYGTTL	nb		135	5738	342	269	13312	8759	5099	1218	1458	4045	6021	1.80	1.60	3.00	10
E6/90-97	LYGTTLEQ	nb		29647	10217	35137	35254	11432	10816	5869	1251	154	3972	4449	6.50	6.00	9.00	N/A
E6/90-98	LYGTTLEQQ	nb		14067	10130	26818	28990	13417	11604	3237	3373	2126	3472	4616	4.20	3.60	0.90	11
E6/93-101	TITLEQQYNK	nb		40786	28244	33754	39059	39153	33144	12249	72604	61324	22380	24601	53.00	51.50	27.00	5
E6/96-106	EQQYNKPLCDL	nb		30404	28837	23578	24508	29712	29267	31740	10522	30625	18047	28302	50.00	48.50	39.00	N/A
E6/97-106	QQYNKPLCDL	nb		14362	25657	11155	8432	22418	23957	24218	38340	51135	14481	12744	20.95	20.45	14.00	10
E6/109-117	RCINCQKPL	nb		23544	22414	37656	37135	26694	24482	8571	22027	24470	10266	5407	23.00	21.00	9.00	12

(Continued)

Annex

Peptide ID ^a	Sequence ^b		Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/124-132	RHLDKKQRF	nb		2766	1862	9996	5098	10835	4502	1222	1510	1389	624	1086	2.05	1.85	0.08	14
E6/126-135	LDKKQRFHNI	nb		24255	26354	33721	39332	22184	24218	31740	25683	15802	10760	2755	18.50	17.50	19.00	11
E6/128-137	KKQRFHNIRG	nb		39485	40557	40205	40321	44254	42509	50000	101076	98564	18274	21785	55.00	53.00	61.00	1
E6/128-138	KKQRFHNIRGR	nb		44708	37019	42683	42717	41857	39409	50000	51063	35406	25756	15922	78.00	76.50	59.00	N/A
E6/138-146	RWTGRCMSC	nb		4276	14865	11769	8183	12458	13649	1710	2004	4061	1252	3796	4.25	4.00	4.00	1
E6/138-147	RWTGRCMSCC	nb		9956	15827	17843	12646	16822	16316	6683	1131	1674	2328	5274	2.65	1.80	8.00	0
E6 L90V/81-90	SEYRHYCYSV	nb		3325	25053	2608	1674	13278	18181	27279	21660	36284	8284	13788	15.20	14.55	62.00	0
E6 L90V/82-91	EYRHYCYSVY	nb		10241	4655	11271	11142	9522	6683	4936	961	1224	1111	6591	2.30	1.45	2.00	9
E6 L90V/85-93	HYCYSVYGT	nb		18142	18075	14351	13021	19694	18883	1319	5861	8084	5787	12623	10.50	10.00	4.00	11
E6 L90V/85-94	HYCYSVYGT	nb		8440	11811	13983	14147	14401	13071	3646	5870	13053	3960	10863	7.85	7.05	7.00	10
E6 L90V/85-95	HYCYSVYGTTL	nb		2450	459	796	668	601	526	865	1672	4376	428	1546	6.90	8.95	2.00	N/A
E6 L90V/87-94	CYSVYGT	nb		18438	12605	27337	26634	8493	10358	4676	13222	10645	2701	4869	12.60	13.60	26.00	N/A
E6 L90V/90-98	VYGTTL	nb		10684	9604	24869	28209	11957	10758	2782	2786	1869	3000	5310	2.75	2.20	0.50	13
E7/10-18	EYMLDLQPE	nb		13459	10614	25762	24906	23465	15795	2607	3334	3852	2628	2045	5.45	4.65	6.00	10
E7/10-19	EYMLDLQPET	nb		21243	15043	25810	27656	19675	17224	5682	9720	6248	4745	2955	12.95	13.45	6.00	10
E7/10-20	EYMLDLQPETT	nb		26570	4960	29579	33132	12872	7989	5621	18540	9206	2314	1013	25.00	23.50	3.00	N/A
E7/22-31	LYCYEQLNDS	nb		26308	21157	30784	33079	26731	23828	10527	12754	4883	7872	12899	15.00	13.00	13.00	10
E7/24-32	CYEQLNDSS	nb		25011	28364	31257	31068	25891	27132	6469	23821	17932	7807	12342	21.50	19.50	5.00	10
E7/24-33	CYEQLNDSSE	nb		27772	28748	33039	33222	30776	29745	12117	1930	2476	10511	4374	14.80	12.80	13.00	10
E7/51-60	HYNIVTFCK	nb		16645	11946	12502	11612	13070	12517	5099	4137	3773	8950	6282	6.70	6.20	5.00	11
E7/56-66	TFCKCDSTLR	nb		29079	19569	32333	26034	27384	23067	26408	4700	6384	16551	8039	21.50	20.00	14.00	N/A
E7/61-69	CDSTLR	nb		35315	34576	34181	37420	29048	31740	50000	90565	58834	17019	24742	41.50	40.00	63.00	0
E7/67-77	LCVQSTHVDIR	nb		37951	38141	42730	44115	42413	40271	50000	10644	5944	23082	14498	29.50	28.00	71.00	N/A
E7/74-82	VDIRTLEDL	nb		26183	25706	32842	31431	23061	24349	19296	25642	22575	14396	16245	23.50	21.50	3.00	15
E7/74-83	VDIRTLEDLL	nb		23644	28546	31902	30708	17687	22452	34238	19171	23000	16245	21666	21.00	20.00	12.00	16
E7/77-87	RTLEDLLMGTL	nb		26563	7563	19336	23834	6693	7093	4288	20707	17461	7132	16783	36.50	35.00	4.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/78-87	TLEDLLMGTL	nb	28965	29750	28756	30400	22660	25983	15710	41846	49172	11968	11965	31.50	29.50	8.00	12
E7/82-91	LLMGTLGIVC	nb	31101	35423	38468	41830	37725	36534	16054	3645	8886	8536	16133	20.50	18.50	20.00	0
E7/83-90	LMGTLGIV	nb	36257	26354	37668	38662	26576	26551	17131	3276	572	9288	20393	13.10	12.60	44.00	N/A
E7 L28F/19-28	TTDL ^Y CYEQ ^F	nb	4301	9903	5931	5744	3968	6263	5044	3211	4322	2204	2344	4.00	3.35	5.00	12
E7 L28F/20-28	TDLYCYEQ ^F	nb	3242	4913	11619	10716	3811	4335	1710	1585	1533	2109	1817	2.15	1.95	0.80	14
E7 L28F/22-30	LYCYEQ ^F ND	nb	8080	16105	23406	25552	25674	20258	1319	1694	1106	1737	1691	1.90	1.60	5.00	11
E7 L28F/22-31	LYCYEQ ^F NDS	nb	18526	17546	25488	29358	22678	19932	6903	7900	3418	6067	10241	7.25	7.00	9.00	11
E7 L28F/24-32	CYEQ ^F NDSS	nb	23616	27069	28994	27366	22685	24748	5933	18662	14984	6992	10442	19.00	17.00	5.00	10
E7 N29S/24-33	CYEQ ^L SDSSE	nb	26823	27718	32607	33222	32326	29907	11858	1639	1999	8763	3987	14.00	12.00	14.00	10
E7 N29S/24-34	CYEQ ^L SDSSEE	nb	34616	23322	38610	39947	32372	27576	10993	4888	6139	12394	3008	26.00	24.50	14.00	N/A
B7																	
E6/107-117	LIRCINCQKPL	0.15±0.06	10602	750	4560	3065	236	421	1568	26888	49182	1945	299	31.40	31.25	16.00	N/A
E6/15-22	RPRKLPQL	0.20±0.06	35	45	15	44	5	15	117	401	286	8	18	0.65	0.60	0.02	N/A
E6/148-158	RSSRTRRETQL	0.32±0.16	13278	3233	9899	8042	1635	2302	4779	15016	49069	3295	344	31.90	31.25	0.50	N/A
E6/136-144	RGRWTGRCM	0.35±0.10	167	682	737	544	723	700	462	236	294	88	106	0.60	0.60	0.70	10
E6/15-23	RPRKLPQLC	0.41±0.28	286	131	71	100	310	202	309	425	583	57	85	1.00	1.00	0.20	14
E6/15-24	RPRKLPQLCT	0.50±0.18	170	80	414	496	79	79	586	199	638	62	65	1.50	1.35	0.50	22
E6/19-28	LPQLCTELQT	0.64±0.52	3061	5373	5644	4632	3231	4174	1585	14277	24936	2064	2558	24.25	24.75	8.00	18
E6/15-25	RPRKLPQLCTE	0.68±0.25	3632	1025	4260	5682	447	678	573	3528	88	305	1170	1.10	1.30	0.80	N/A
E6 R17T/15-25	RPT ^K L ^P QLCTE	0.69±0.37	16556	14182	17526	14473	4408	7946	1766	4442	88	2877	5994	2.75	2.10	5.00	N/A
E7/5-12	TPTLHEYM	0.71±0.48	6867	509	5515	6604	859	660	605	162	42	1465	1945	0.55	0.45	5.00	N/A
E6/151-158	RTRRETQL	1.14±0.48	4307	2161	2410	3896	358	884	2340	27902	16596	302	435	17.90	17.85	7.00	N/A
E6/19-26	LPQLCTEL	1.30±0.43	591	192	405	427	46	94	234	721	333	285	127	0.75	0.70	0.60	N/A
E6/11-19	DPQERPRKL	3.94±1.82	6673	5180	8586	6431	3666	4335	236	1084	922	4448	9055	1.50	1.50	0.40	20
E7/5-13	TPTLHEYML	4.50±3.30	522	921	482	505	487	671	372	566	394	118	214	1.70	1.70	0.90	20

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/49-57	RAHYNIVTF	4.65±2.28	2595	2136	1074	1801	1100	1534	1453	1133	1256	1032	2408	3.80	3.80	2.00	9
E6/142-151	RCMSCCRSSR	4.72±5.40	26769	19179	23568	25546	27683	22943	23445	666	650	8606	11803	15.25	12.25	5.00	3
E6/118-126	CPEEKQRHL	5.90±3.90	2464	3283	1529	893	723	1543	275	811	692	1394	496	2.50	2.50	0.30	20
E6/136-146	RGRWTGRMSC	13.95±14.34	20222	7396	20480	19735	7505	7446	8479	12461	10985	2587	201	13.50	13.05	5.00	N/A
E6/95-103	LEQQYNKPL	14.86±2.04	16354	17963	14910	13789	22381	20040	2874	8244	11612	6260	14323	16.00	16.00	16.00	11
E6/101-111	KPLCDLLIRCI	16.12±10.66	12510	2000	7330	6870	1253	1585	667	4483	3818	801	8533	4.65	4.20	7.00	N/A
E7/46-55	EPDRAHYNIV	18.36±13.23	1084	4462	2782	2352	3279	3828	1362	1308	687	6278	2912	1.75	1.45	6.00	18
E6/59-69	IVYRDGNPYAV	20.28±10.18	29570	5251	26764	26980	4553	4910	9048	4566	16704	7480	1782	24.50	23.00	12.00	N/A
E7/83-93	LMGTLGIVCPI	21.78±29.30	34058	18547	29391	30687	15999	17224	18478	11791	1516	14798	19969	18.05	16.55	70.00	N/A
E7/82-89	LLMGTLGI	22.70±11.69	21900	16001	24323	29523	11224	13429	5044	3196	279	9659	23137	3.65	3.50	80.00	N/A
E6/101-108	KPLCDLLI	35.23±12.60	8853	642	5418	6518	1846	1091	743	240	167	1815	4125	0.85	0.70	14.00	N/A
E6/134-144	NIRGRWTGRCM	54.83±5.62	12158	1451	10711	12005	533	879	1422	41645	213211	1904	486	49.20	48.75	15.00	N/A
E6/13-22	QERPRKLPQL	nb	11032	17593	13332	13747	16500	17038	6469	5870	19401	10893	20635	21.85	21.75	3.00	14
E6/17-26	RKLPQLCTEL	nb	17333	12454	20344	18396	8347	10246	10642	1415	475	9299	14022	4.95	4.55	4.00	13
E6/19-27	LPQLCTELQ	nb	8417	17659	16362	11909	15276	16405	1017	2939	1836	5938	2141	4.70	4.70	6.00	11
E6/19-29	LPQLCTELQTT	nb	11800	4094	12515	11771	3142	3587	1568	14144	10347	4420	9518	10.65	10.25	5.00	N/A
E6/37-44	CVYCKQQL	nb	29637	16733	17574	24776	8492	11922	6063	19572	29041	9462	11733	31.50	30.50	25.00	N/A
E6/40-49	CKQQLLRREV	nb	33314	23223	29492	29879	26691	24882	10758	2580	360	10096	6283	24.30	21.30	22.00	7
E6/44-54	LLRREYDFAF	nb	26385	10221	25001	22487	4236	6575	4242	3904	18828	2863	5648	23.50	22.00	26.00	N/A
E6/52-60	FAFRDLCIV	nb	19885	21410	17280	19560	21764	21617	4883	75711	70289	15649	20914	36.00	36.00	57.00	6
E6/65-74	NPYAVCDKCL	nb	3555	3032	2858	2834	548	1290	646	1681	1773	1510	4046	4.10	4.60	2.00	21
E6/65-75	NPYAVCDKCLK	nb	19960	23132	23422	20023	18483	20702	6978	1814	286	9321	23033	4.05	3.55	14.00	N/A
E6/78-86	SKISEYRHY	nb	32669	24243	42055	40921	40546	31399	50000	1432716	1293868	25507	39741	73.00	67.00	66.00	0
E6/81-90	SEYRHYCYSL	nb	23378	13200	13972	16945	16285	14643	12791	2054	1922	11602	20969	14.55	11.55	8.00	10
E6/86-95	YCYSLYGTTL	nb	20702	11320	15841	18704	9160	10191	3380	888	619	4907	17037	10.70	7.20	3.00	11
E6/88-95	YSLYGTTL	nb	23550	11859	12209	15487	3030	5998	2635	58836	13032	8042	14825	19.15	19.00	13.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/98-107	QYNKPLCDLL	nb	35724	27858	33348	36165	33600	30561	30396	34884	31320	23056	31855	55.00	52.00	20.00	11
E6/109-117	RCINCQKPL	nb	10882	13630	10021	5830	4578	7903	3380	4844	6308	3949	3583	12.00	12.00	9.00	12
E6/118-127	CPEEKQRHLD	nb	16926	11329	24314	16454	31074	18781	4527	1232	313	11978	11639	4.40	4.00	6.00	11
E6/136-145	RGRWTGRCMS	nb	4482	13789	14621	14415	15608	14643	14407	623	430	836	915	1.70	2.15	21.00	4
E6/149-158	SSRTRRETQL	nb	2017	4523	2816	2090	1226	2352	3202	1492	1667	2539	9124	3.95	4.45	0.70	13
E6 L90V/86-95	YCYSVYGTTL	nb	20272	9222	15299	18450	9897	9551	3134	767	616	4962	18781	10.70	7.20	2.00	11
E6 R17T/11-19	DPQERPTKL	nb	7596	6571	6523	5653	3913	5072	281	1578	1346	4980	9871	4.00	4.00	0.20	21
E7/6-13	PTLHEYML	nb	15643	25868	22197	17191	40592	32435	50000	104145	128236	22647	25118	38.85	38.60	94.00	N/A
E7/16-25	QPETTDLYCY	nb	28507	19458	26394	26188	23556	21385	12791	18777	18443	18331	34949	35.50	32.50	19.00	10
E7/16-26	QPETTDLYCYE	nb	36819	25550	38389	37789	34278	29585	6683	4738	1122	19944	32526	24.20	23.20	26.00	N/A
E7/39-49	DGPAGQAEPDR	nb	37208	25009	44391	38803	47632	34610	50000	31229	13798	28303	35912	35.50	34.50	80.00	N/A
E7/40-48	GPAGQAEPD	nb	10408	19898	22759	18549	23090	21501	1109	7164	16822	5409	2912	19.00	19.00	5.00	14
E7/46-53	EPDRAHYN	nb	38616	25321	37677	40246	32286	28640	4936	1785	4246	19453	29277	33.50	32.50	83.00	N/A
E7/46-56	EPDRAHYNIVT	nb	15185	14656	22448	19907	9877	12052	2443	639	474	6608	12837	2.60	1.85	6.00	N/A
E7/78-85	TLEDLLMG	nb	39378	30118	43834	45979	42804	35946	50000	10131	1084	26009	36709	31.20	30.20	98.00	N/A
B15																	
E6/73-83	CLKFYSKISEY	0.50±0.28	3334	339	1149	978	113	195	1905	N/A	N/A	100	47	0.70	0.10	0.40	N/A
E6 L90V/81-91	SEYRHYCYSVY	2.87±0.84	4699	91	1130	1137	113	102	1209	N/A	N/A	508	96	0.80	0.10	2.00	N/A
E6 L90V/83-91	YRHYCYSVY	3.56±1.20	1307	1005	1237	1221	938	974	1407	877	859	900	2968	5.00	5.00	3.00	N/A
E6/81-91	SEYRHYCYSLY	3.73±2.09	11011	372	2054	1945	272	318	1710	N/A	N/A	1402	390	1.90	1.00	3.00	N/A
E6/74-83	LKFYSKISEY	3.83±2.91	670	289	360	329	926	517	3417	354	105	464	261	3.85	3.15	2.00	N/A
E6 L90V/84-91	RHYCYSVY	3.91±1.26	3245	761	2200	3892	731	747	1673	N/A	N/A	2824	4821	0.80	0.10	16.00	N/A
E7/43-52	GQAEPDRAHY	4.07±1.50	19	14	27	39	32	21	1551	20	3	17	19	0.30	0.20	0.02	N/A
E6/68-77	AVCDKCLKFY	4.10±2.35	1317	495	2645	3065	1447	851	9866	69	28	179	67	2.30	1.10	1.00	N/A
E6/79-88	KISEYRHYCY	4.67±0.95	533	165	674	722	683	336	5382	164	89	188	35	3.50	2.80	3.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/15-23	LQPETTDLY	4.68±1.28	115	73	308	233	476	188	1097	132	114	86	52	0.80	0.80	0.09	N/A
E6/84-91	RHYCYSLY	4.85±1.96	9120	2654	4021	6469	1359	1905	2315	N/A	N/A	6001	7308	1.00	0.10	29.00	N/A
E6/52-61	FAFRDLCIVY	5.07±2.18	72	145	54	30	70	100	1196	35	29	85	37	1.25	1.10	0.80	N/A
E6/41-50	KQQLLRREVY	5.80±2.97	76	24	75	178	160	62	2123	5	4	30	20	0.40	0.25	0.20	N/A
E6/83-91	YRHYCYSLY	6.00±1.26	4254	5652	2396	2227	2650	3870	2123	3380	3436	2232	5112	15.00	15.00	4.00	N/A
E6/53-61	AFRDLCIVY	6.13±2.15	460	278	1057	1031	744	454	2722	477	426	424	481	2.80	2.80	0.20	N/A
E6/59-67	IVYRDGNPY	6.22±1.77	61	64	32	46	46	54	776	104	96	53	30	0.70	0.70	0.20	N/A
E7/15-25	LQPETTDLYCY	7.30±2.00	1905	81	2382	2535	473	196	3001	N/A	N/A	107	20	0.60	0.10	0.80	N/A
E6 L90V/89-99	SVYGTTLQYQY	7.51±2.18	3382	133	319	640	140	137	2391	N/A	N/A	256	11	0.70	0.10	0.06	N/A
E7/82-90	LLMGTLGIV	7.54±1.87	2323	1172	937	861	2089	1568	2905	1449	1556	535	181	8.00	8.00	22.00	N/A
E6/58-67	CIVYRDGNPY	7.98±2.78	110	362	114	126	212	278	2169	76	54	41	76	2.05	1.85	5.00	N/A
E6/42-50	QQLLRREVY	8.12±3.38	194	144	207	300	343	221	1209	207	166	231	494	1.10	1.10	0.07	N/A
E6/89-99	SLYGTTLQYQY	8.18±4.40	1741	68	230	338	79	73	1620	N/A	N/A	117	10	0.50	0.10	0.04	N/A
E6/57-67	LCIVYRDGNPY	10.49±0.99	983	1474	301	495	1370	1422	2365	N/A	N/A	340	849	0.20	0.10	7.00	N/A
E6/82-91	EYRHYCYSLY	10.57±7.23	9585	1259	7669	7003	10190	3587	10875	1651	1518	5190	16655	18.50	16.10	14.00	N/A
E6/18-26	KLPQLCTEL	11.13±2.74	5950	8263	4086	3979	6874	7568	4831	5533	5260	3836	4431	15.00	15.00	0.40	N/A
E6/44-54	LLRREVDFAF	11.22±6.36	1705	46	399	382	36	41	1362	N/A	N/A	68	5	0.50	0.10	5.00	N/A
E6/75-83	KFYSKISEY	11.36±9.44	1026	900	986	1178	944	923	1362	881	813	600	218	4.80	4.80	0.07	N/A
E6/45-54	LRREVDFAF	12.52±5.40	6140	3350	4932	5790	5318	4219	4430	86	49	1691	10904	3.65	1.70	24.00	N/A
E6/44-52	LLRREVDYDF	12.88±2.32	282	170	94	88	159	164	1158	218	180	73	82	1.20	1.20	0.70	N/A
E6/97-106	QQYNKPLCDL	14.38±8.21	1346	478	1621	1829	2584	1109	3807	316	139	600	1783	5.00	3.75	2.00	N/A
E7/49-57	RAHYNIVTF	15.72±15.79	82	87	62	65	67	77	423	166	184	60	36	0.50	0.50	0.20	N/A
E6/67-77	YAVCDKCLKFY	19.64±11.44	11119	438	3384	2994	709	558	4335	N/A	N/A	1314	69	2.00	1.10	7.00	N/A
E7/7-15	TLHEYMLDL	19.98±4.76	5412	7918	4781	3961	6941	7406	2722	4121	4634	3188	11909	14.00	14.00	2.00	N/A
E6/42-52	QQLLRREVDYDF	20.84±1.26	6524	1211	3005	4778	1250	1229	2443	N/A	N/A	1396	899	0.90	0.10	6.00	N/A
E6/60-67	VYRDGNPY	21.93±13.93	2754	1126	3245	3951	2808	1776	5933	N/A	N/A	813	9058	0.60	0.10	6.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/42-52	AGQAEPDRAHY	23.28±4.09	1407	712	1362	1405	3431	1559	11987	N/A	N/A	523	49	0.30	0.10	2.00	N/A
E6/26-34	LQTTIHDII	24.11±6.46	6475	11974	3183	3202	7508	9499	3344	4841	3837	3508	5950	16.00	16.00	52.00	N/A
E6/129-139	KQRFHNIRGRW	28.75±12.42	12284	1308	6520	10845	2365	1756	9346	N/A	N/A	3209	274	3.00	2.20	3.00	N/A
E7/43-51	GQAEPDRAH	28.79±16.56	1527	1933	887	1672	1814	1864	3001	2910	2296	628	120	12.00	12.00	0.30	N/A
E6/77-86	YSKISEYRHY	29.36±13.83	48	55	236	175	217	109	3686	130	182	50	42	4.65	4.55	2.00	N/A
E6/76-86	FYSKISEYRHY	32.33±8.76	5410	3255	6033	5921	3230	3254	7131	N/A	N/A	6512	4016	0.80	0.10	11.00	N/A
E6/40-50	CKQQLLRREVY	33.39±18.26	2925	10652	2678	3766	3731	6297	11113	N/A	N/A	11115	1934	0.60	0.10	20.00	N/A
E6/43-50	QLLRREVY	36.81±17.10	4022	583	4959	7374	515	549	5933	N/A	N/A	334	5919	0.80	0.10	10.00	N/A
E6/31-39	HDIILECVY	37.48±12.61	1976	3786	4656	2733	2787	3254	1277	1294	1125	1365	4676	5.10	5.10	6.00	N/A
E6/29-39	TIHDIILECVY	37.75±17.06	2106	151	1127	1073	208	177	3308	N/A	N/A	596	12	0.60	0.10	0.70	N/A
E6/59-68	IVYRDGNPYA	38.45±10.14	935	14388	545	872	8218	10875	29425	1526	987	5431	6597	13.80	13.10	19.00	N/A
E7/15-24	LQPETTDLYC	38.45±10.85	2365	3924	6028	4566	17186	8208	22696	74	22	2586	6154	2.55	0.95	10.00	N/A
E6/122-132	KQRHLDKKQRF	38.74±15.84	3955	92	1482	4350	368	184	1766	N/A	N/A	264	55	0.70	0.10	0.20	N/A
E7/82-89	LLMGTLGI	42.43±21.41	14125	410	6789	6293	954	625	3933	N/A	N/A	465	686	3.60	2.80	82.00	N/A
E6/136-144	RGRWTGRCM	43.73±9.05	2545	5050	3340	4008	6047	5530	1926	1841	1614	1820	284	8.20	8.20	3.00	N/A
E6/67-76	YAVCDKCLKF	46.75±21.63	1324	2551	721	342	950	1559	2169	1174	960	590	822	14.30	13.10	6.00	N/A
E7/43-53	GQAEPDRAHYN	46.88±22.26	2848	15787	2698	2978	18728	17131	16946	N/A	N/A	8454	4269	0.60	0.10	5.00	N/A
E6/68-76	AVCDKCLKF	47.83±8.39	2323	3332	1426	1434	2357	2812	2193	1452	1040	763	2484	5.80	5.80	0.50	N/A
E6/81-88	SEYRHYCY	48.91±15.52	8495	497	7695	6850	664	576	2241	N/A	N/A	1628	633	1.00	0.10	5.00	N/A
E6/30-39	IHDIILECVY	54.36±24.59	4433	4551	6136	4466	17359	8854	26985	25	3	9240	13570	2.00	0.25	42.00	N/A
E6/134-144	NIRGRWTGRCM	56.25±15.87	25100	6712	17382	16677	5392	5998	6756	N/A	N/A	4269	1716	25.00	20.00	45.00	N/A
E6/59-69	IVYRDGNPYAV	58.67±16.94	13626	5894	5189	7816	4571	5182	14252	N/A	N/A	6853	126	4.30	3.50	14.00	N/A
E6/78-86	SKISEYRHY	58.91±4.96	2652	2875	1679	1436	1778	2265	3001	1758	1256	3237	5011	2.70	2.50	3.00	N/A
E6/79-86	KISEYRHY	68.31±17.99	3431	116	3403	4276	248	171	5682	N/A	N/A	213	80	0.80	0.10	17.00	N/A
E6/113-121	CQKPLCPEE	70.23±6.13	11536	15412	13486	16285	13831	14564	8297	3917	3999	7714	7521	16.00	16.00	10.00	N/A
E6/78-88	SKISEYRHYCY	75.26±44.92	11407	6192	8779	9065	2957	4288	5682	N/A	N/A	6824	2359	2.20	1.30	9.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/43-52	QLLRREYVDF	75.66±18.15	2564	2179	1583	1215	2771	2456	6129	147	87	331	3135	4.40	2.80	12.00	N/A
E6/76-83	FYSKISEY	77.10±23.81	7638	4276	2380	2370	2405	3219	3646	N/A	N/A	2415	8619	0.90	0.10	0.70	N/A
E7/67-76	LCVQSTHVDI	79.92±23.62	12925	15141	16211	13894	21041	17888	15208	1822	580	9607	28889	14.75	13.35	59.00	N/A
E7/82-91	LLMGTLGIVC	84.41±38.55	1730	3877	6468	7102	2838	3326	7946	37	45	247	408	3.05	1.55	10.00	N/A
E6/20-29	PQLCTELQTT	nb	34905	12222	34093	33438	34643	20590	50000	641	165	12046	14487	43.15	54.15	23.00	N/A
E6/21-30	QLCTELQTTI	nb	11462	15403	9069	9800	5611	9296	10414	729	994	2414	872	17.15	15.05	5.00	N/A
E6/25-34	ELQTTIHDII	nb	17096	13342	15214	14309	21356	16855	21040	426	2781	7744	25711	39.50	30.00	77.00	N/A
E6/26-35	LQTTIHDIIL	nb	3065	1470	1433	1308	3942	2404	3933	43	31	1052	105	2.80	1.15	33.00	N/A
E6/32-39	DIILECVY	nb	18539	5131	15902	16568	2804	3807	4527	N/A	N/A	6207	19659	7.20	6.70	35.00	N/A
E6/33-42	IILECVYCKQ	nb	28329	21095	31088	31827	30029	25153	50000	7738	1350	10988	6441	45.00	36.50	61.00	N/A
E6/35-44	LECVYCKQQL	nb	18025	16905	16756	16640	17026	16946	6331	2956	1420	3833	20951	35.00	25.00	47.00	N/A
E6/37-44	CVYCKQQL	nb	30864	14668	21054	26148	9278	11667	6829	N/A	N/A	3813	8169	33.00	31.00	45.00	N/A
E6/42-51	QQLLRREYVD	nb	2592	22405	4021	5027	34471	27876	50000	323	110	13999	26753	4.80	3.20	44.00	N/A
E6/47-54	REYDFAF	nb	4471	177	8117	7894	360	253	759	N/A	N/A	320	2285	0.80	0.10	71.00	N/A
E6/51-60	DFAFRDLCIV	nb	28432	22902	20253	19142	34491	28179	50000	5710	569	18609	27215	40.00	31.50	96.00	N/A
E6/52-60	FAFRDLCIV	nb	11153	10769	4422	4062	8949	9812	4288	16864	14521	4740	8652	31.00	31.00	49.00	N/A
E6/73-82	CLKFYSKISE	nb	20676	14658	20294	23745	18325	16405	33505	113	66	10128	857	24.10	14.10	27.00	N/A
E6/81-90	SEYRHYCYSL	nb	4629	7856	5064	4283	3731	5412	2874	1126	231	3966	938	7.90	6.10	24.00	N/A
E6/88-95	YSLYGTTL	nb	15086	1579	9888	13124	834	1152	2265	N/A	N/A	739	3989	4.40	4.30	16.00	N/A
E6/89-97	SLYGTTLEQ	nb	4880	5667	8402	7925	9831	7487	10081	2606	1816	3680	502	8.80	8.80	7.00	N/A
E6/90-99	LYGTTLEQQY	nb	10883	5292	8193	8952	8539	6719	11987	926	517	2786	16393	12.70	10.50	9.00	N/A
E6/95-103	LEQQYNKPL	nb	3725	4599	11395	10517	8319	6162	1926	2958	2780	2764	5060	13.00	13.00	21.00	N/A
E6/97-107	QQYNKPLCDLL	nb	16933	3872	6080	8662	2481	3100	3726	N/A	N/A	3028	492	7.70	7.10	4.00	N/A
E6/98-107	QYNKPLCDLL	nb	28147	25392	24378	24646	27133	26265	34610	60345	197811	17064	30368	77.50	69.00	38.00	N/A
E6/106-114	LLIRCINCQ	nb	7809	9423	17114	15786	13060	11113	13947	3451	2168	4894	1091	3.30	3.00	41.00	N/A
E6/107-115	LIRCINCQK	nb	15500	16852	18955	20384	17580	17131	18280	29850	29242	6841	15362	29.00	29.00	19.00	N/A

(Continued)

Annex

Peptide ID ^a	Sequence ^b		Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E6/107-117	LIRCINCQKPL	nb		8294	1034	6136	4241	1126	1079	5044	N/A	N/A	1327	112	1.00	0.10	53.00	N/A
E6/113-122	CQKPLCPEEK	nb		21798	11881	18226	22047	15451	13502	25842	701	1085	6583	16051	37.00	27.00	6.00	N/A
E6/122-129	KQRHLDDK	nb		29632	18220	28410	29028	17987	18181	29425	N/A	N/A	9377	28606	29.00	31.00	26.00	N/A
E6/129-137	KQRFHNIRG	nb		2107	3728	4659	5308	6175	4804	9551	2443	2188	3899	936	10.00	10.00	8.00	N/A
E6/143-151	CMSCCRSSR	nb		15117	16483	16100	17865	16801	16583	7287	10964	10162	6857	19560	25.00	18.00	29.00	N/A
E6/144-151	MSCCRSSR	nb		33102	14502	31263	27740	20793	17411	16405	N/A	N/A	9089	9978	39.00	37.00	65.00	N/A
E6/144-152	MSCCRSRT	nb		20597	13773	24486	21090	19495	16405	9448	12301	11967	9844	12417	28.00	22.00	62.00	N/A
E6/148-158	RSSRTRRETQL	nb		28366	6721	24116	25513	10521	8433	7527	N/A	N/A	5674	170	32.00	27.00	16.00	N/A
E6/149-158	SSRTRRETQL	nb		11299	15989	17521	16278	10058	12722	6683	9029	6184	2627	10123	31.50	29.55	12.00	N/A
E6/150-158	SRTRRETQL	nb		35474	19929	34055	38393	35085	26408	19718	339586	344350	22124	35163	79.00	78.00	16.00	N/A
E6/151-158	RTRRETQL	nb		24588	7125	19529	20971	4177	5471	6469	N/A	N/A	3000	8987	19.00	31.00	36.00	N/A
E6 L90V/82-91	EYRHVCYSVY	nb		5809	445	5150	4957	5567	1576	7692	246	236	2409	10811	8.05	6.10	12.00	N/A
E6 L90V/88-95	YSVYGTTL	nb		10415	351	7446	9535	386	368	1602	N/A	N/A	383	2123	1.20	0.30	13.00	N/A
E7/2-11	HGDPTLHEY	nb		9071	2358	4080	3324	14783	5901	13502	508	189	4794	14936	7.00	4.65	9.00	N/A
E7/11-20	YMLDLQPETT	nb		16163	5248	13971	15539	7457	6263	17696	214	232	2715	5685	23.50	13.00	8.00	N/A
E7/12-21	MLDLQPETTD	nb		37581	25456	34771	36368	36930	30727	50000	922	1898	21095	35014	61.50	68.00	50.00	N/A
E7/14-23	DLQPETTDLY	nb		1131	618	2981	1986	4723	1710	8388	45	46	3901	11524	2.45	1.60	2.00	N/A
E7/27-35	QLNDSSEEE	nb		17376	13421	31284	32349	28656	19611	11858	7161	7889	13514	16923	31.00	23.00	4.00	N/A
E7/42-51	AGQAEPDRAH	nb		1583	17824	2036	3483	30012	23192	50000	4588	4190	3360	4007	26.10	24.60	24.00	N/A
E7/44-52	QAEPDRAHY	nb		3474	4410	6088	5004	6219	5239	10191	3605	3319	2522	1219	14.00	14.00	2.00	N/A
E7/45-52	AEPDRAHY	nb		8989	1571	17889	16044	7200	3362	7860	N/A	N/A	1797	3135	1.00	0.10	16.00	N/A
E7/48-57	DRAHYNIVTF	nb		831	3130	1024	846	15827	7054	6829	706	252	5063	19192	7.30	6.60	31.00	N/A
E7/49-58	RAHYNIVTFC	nb		1857	20441	2398	2053	16485	18280	18083	4223	2853	8021	9063	22.65	21.10	33.00	N/A
E7/63-72	STLRVCVQST	nb		21261	22774	13381	14105	21151	21852	50000	7025	1090	10169	16313	36.50	26.50	54.00	N/A
E7/64-73	TLRVCVQSTH	nb		6302	6671	3603	3357	2810	4335	13649	884	3114	1126	202	24.10	22.10	3.00	N/A
E7/66-73	RLCVQSTH	nb		17595	3111	13985	15172	3972	3511	10302	N/A	N/A	1230	263	6.40	5.90	53.00	N/A

(Continued)

Peptide ID ^a	Sequence ^b	Exp. IC ₅₀ ± SD ^c	NetMHC 4.0 ^d	NetMHC 3.4 ^d	NetMHCpan 4.0 ^d	NetMHCpan 3.0 ^d	NetMHCpan 2.8 ^d	NetMHCcons 1.1 ^d	PickPocket 1.1 ^d	IEDB SMMPMBEC ^d	IEDB SMM ^d	MHCflurry 1.2 ^d	MHCnuggets 2.0 ^d	IEDB recommended ^e	IEDB consensus ^e	MixMHCpred2.0.2 ^e	SYFPEITHI ^f
E7/66-74	RLCVQSTHV	nb	1399	622	3977	3307	2498	1249	5325	1066	944	992	632	2.30	1.50	4.00	N/A
E7/66-75	RLCVQSTHVD	nb	12566	16735	18570	19160	18972	17792	25017	148	134	5693	1677	7.95	6.90	21.00	N/A
E7/69-79	VQSTHVDIRTL	nb	11578	627	4918	7355	1586	995	2241	N/A	N/A	871	17	2.40	1.50	8.00	N/A
E7/77-84	RTLEDLLM	nb	26819	5700	15109	15406	2892	4063	5621	N/A	N/A	1999	6238	23.00	31.00	85.00	N/A
E7/78-87	TLEDLLMGTL	nb	21717	9843	14779	13333	16478	12722	14407	606	2953	4318	18454	44.50	34.50	11.00	N/A
E7/81-90	DLLMGTLGIV	nb	14331	7013	9292	7669	27380	13797	37740	239	223	13806	21420	18.00	10.40	86.00	N/A
E7/83-91	LMGTLGIVC	nb	2284	4750	9196	10941	3830	4265	4676	849	1549	1163	4264	8.00	8.00	23.00	N/A
E7/83-93	LMGTLGIVCPI	nb	5377	139	5684	3805	432	245	2905	N/A	N/A	300	17	0.80	0.10	75.00	N/A
E7/85-93	GTLGIVCPI	nb	6259	10035	3917	3304	6199	7903	4382	4886	5284	4512	1968	19.00	19.00	75.00	N/A
E7/86-93	TLGIVCPI	nb	18481	7105	18271	20397	9723	8342	10642	N/A	N/A	3543	8859	7.10	6.60	84.00	N/A
E7/86-95	TLGIVCPICS	nb	30590	20044	28227	29407	33263	25842	50000	2720	967	14463	17789	46.00	38.50	82.00	N/A
E7/87-96	LGIVCPICSQ	nb	4074	19216	20330	17865	23359	21269	40709	995	178	2725	2511	6.25	4.50	78.00	N/A
E7/88-96	GIVCPICSQ	nb	6329	10914	18113	15377	15978	13213	19932	8491	5321	6001	669	18.00	18.00	34.00	N/A
E7/89-98	IVCPICSQKP	nb	29504	25580	31503	36517	31417	28332	50000	8583	408	10301	2896	40.00	31.50	43.00	N/A

a: Peptide ID contains the protein (E6/E7), amino acid changes compared to the reference sequence (e.g. L90V), and region within the amino acid sequence

b: Peptide sequence, amino acid changes are highlighted in red

c: Experimental binding capacity expressed IC₅₀ ± SD

d: Prediction score as putative IC₅₀ value (the lower, the better)

e: Prediction score as percentile rank (the lower, the better)

f: SYFPEITHI-specific prediction score in arbitrary units (the higher, the better)

N/A: Prediction was not available

nb: nonbinder

*: Recent A2 binding assays for E7/78-86 showed no binding affinity. However, prior binding assays resulted in binding affinity for A2.

(Continued)

Supplementary Table S2. HPV16 E6/E7-derived peptides and their reported binding affinities

Peptide ID	Sequence	A1 ^a	A2 ^a	A3 ^a	A11 ^a	A24 ^a	B7 ^a	B15 ^a	IEDB ID ^b	Exp. binding affinity ^c	Functionality ^d
E6/1-9	MHQKRTAMF					ligand			41674	A24: -	
E6/7-15	AMFQDPQER		ligand	ligand	ligand				3089	A2: -; A3: -; A11: -	
E6/9-18	FQDPQERPRK				ligand				111289	A11: 5.01±2.17	
E6/15-22	QERPRKLPQL						ligand		(Bourgault Villada et al., 2010)	B7: nb	
E6/15-22	RPRKLPQL						ligand		(Bourgault Villada et al., 2010)	B7: 0.20±0.06	
E6/15-24	RPRKLPQLCT						IFN γ		911841	B7: 0.50±0.18	
E6/18-26	KLPQLCTEL		ligand, IFN γ						32085	A2: 10.07±3.22	
E6/19-26	LPQLCTEL						ligand		(Bourgault Villada et al., 2010)	B7: 1.30±0.43	
E6/21-29	QLCTELQTT		ligand						111674	A2: nb	
E6/21-30	QLCTELQTTI		ligand							A2: nb	
E6/25-33	ELQTTIHDI		ligand						111250	A2: 36.90±8.41	A2: IFN γ
E6/26-34	LQTTIHDI		ligand, cytolysis			ligand			39002	A2: -; A24: 1.42±0.23	
E6/29-37	TIHDIILEC		ligand						110720	A2: nb	
E6/29-38	TIHDIILECV		ligand, IFN γ , cytolysis						64320	A2: 3.69±2.17	A2: IFN γ
E6/30-38	IHDIILECV		ligand						111421	A2: nb	
E6/32-41	DIILECVYCK				ligand				111199	A11: 6.80±2.41	A11: IFN γ
E6/33-41	IILECVYCK			ligand, IFN γ	ligand				26568	A3: 11.02±3.27; A11: 4.98±1.33	
E6/37-46	CVYCKQQLLR				IFN γ				911713	A11: 36.49±6.55	A11: IFN γ
E6/42-50	QQLLRREVY			ligand	ligand				52063	A3: nb; A11: -	
E6/44-52	LLRREVYDF					ligand			37754	A24: nb	
E6/49-57	VYDFAFRDL					ligand, IFN γ , cytolysis			71988	A24: 0.10±0.04	
E6/52-60	FAFRDLCIV		ligand, IFN γ , cytolysis				ligand		15173	A2: 4.34±2.11; B7: nb	A2: IFN γ

(Continued)

Annex

Peptide ID	Sequence	A1 ^a	A2 ^a	A3 ^a	A11 ^a	A24 ^a	B7 ^a	B15 ^a	IEDB ID ^b	Exp. binding affinity ^c	Functionality ^d
E6/54-61	FRDLCIVY	ligand		ligand	ligand				(Bourgault Villada et al., 2010)	A1: nb; A3: - ; A11: nb	
E6/59-67	IVYRDGNPY			ligand	ligand				29519	A3: 11.58±1.76; A11: 19.84±4.91	A3: IFN γ ; A11: IFN γ
E6/65-74	NPYAVCDKCL						IFN γ		911816	B7: nb	
E6/66-74	PYAVCDKCL					ligand, IFN γ			50064	A24: 1.97±0.77	
E6/68-75	AVCDKCLK				ligand				111129	A11: 9.72±3.73	A11: IFN γ
E6/68-77	AVCDKCLKFY				IFN γ				911710	A11: 4.80±0.55	A11: IFN γ
E6/69-77	VCDKCLKFY	cytolysis							67833	A1: nb	
E6/75-83	KFYKISEY			ligand					30892	A3: 52.27±21.55	A3: IFN γ
E6/77-86	YSKISEYRHY	ligand							111973	A1: nb	
E6/79-87	KISEYRHYC		ligand, IFN γ , cytolysis						111456	A2: 47.56±29.56	
E6 L90V/80-88	CYSVYGTTL				ligand				111184	A11: 62.36±12.84	
E6/80-88	ISEYRHYCY	ligand, IFN γ			ligand				28484	A1: 1.20±0.96; A11: 62.36±12.84	
E6/82-90	EYRHYCYSL					ligand, IFN γ			110846	A24: 0.32±0.10	
E6/87-95	CYSLYGTTL					ligand, IFN γ			7439	A24: 0.10±0.04	A24: IFN γ
E6/89-97	SLYGTTLQ			ligand					59598	A3: nb	
E6/97-106	QQYNKPLCDL		ligand, IFN γ , cytolysis						604613	A2: nb	
E6/92-101	GTTLEQQYNK				ligand				111384	A11: 5.26±2.57	A11: IFN γ
E6/93-101	TTLEQQYNK			ligand	ligand				66689	A3: 15.55±5.05; A11: 4.71±1.74	A3: IFN γ ; A11: IFN γ
E6/98-106	QYNKPLCDL					ligand, IFN γ			111698	A24: 2.97±0.67	
E6/98-107	QYNKPLCDLL					IFN γ		(Hara et al., 2005)		A24: 0.54±0.24	A24: IFN γ
E6/102-110	PLCDLLIRC		ligand						111630	A2: nb	

(Continued)

Annex

Peptide ID	Sequence	A1 ^a	A2 ^a	A3 ^a	A11 ^a	A24 ^a	B7 ^a	B15 ^a	IEDB ID ^b	Exp. binding affinity ^c	Functionality ^d
E6/106-115	LLIRCINCQK			ligand	ligand, IFN γ , cytolysis				37421	A3: 2.06 \pm 0.97; A11: 1.36 \pm 0.30	A3: IFN γ ; A11: IFN γ
E6/118-126	CPEEKQRHL						IFN γ		110197	B7: 5.90 \pm 3.90	
E6/125-133	HLDKKQRFH			ligand					24182	A3: 29.58 \pm 13.63	
E6/131-139	RFHNIRGRW					ligand			53711	A24: 23.93 \pm 3.24	
E7/2-11	HGDTPTLHEY	ligand							111395	A1: 1.03 \pm 0.59	
E7/5-13	TPTLHEYML						IFN γ		911868	B7: 4.50 \pm 3.30	
E7/7-15	TLHEYMLDL		ligand, IFN γ , cytolysis		ligand				64830	A2: 2.68 \pm 1.86; A11: nb	
E7/7-17	TLHEYMLDLQP		ligand						768668	A2: 19.35 \pm 12.57	A2: IFN γ
E7/10-19	EYMLDLQPET		ligand						768427	A2: nb	
E7/11-18	YMLDLQPE		ligand						768725	A2: 8.07 \pm 3.69	A2: IFN γ
E7/11-19	YMLDLQPET		ligand, IFN γ , cytolysis						75074	A2: 1.40 \pm 0.94	A2: IFN γ ; cytolysis
E7/11-20	YMLDLQPETT		ligand, IFN γ , cytolysis						75075	A2: 2.19 \pm 2.71	A2: IFN γ ; cytolysis
E7/11-21	YMLDLQPETTD		ligand						768729	A2: 2.37 \pm 1.67	A2: IFN γ ; cytolysis
E7/12-19	MLDLQPET		ligand						768546	A2: 53.66 \pm 12.03	A2: IFN γ ; cytolysis
E7/12-20	MLDLQPETT		ligand, IFN γ , cytolysis						41919	A2: 3.60 \pm 0.73	A2: IFN γ ; cytolysis
E7/14-22	DLQPETDL		ligand					(Bauer et al., 2000)		A2: nb	
E7/15-23	LQPETDLY						IFN γ		911800	B15: 4.68 \pm 1.28	
E7/18-25	ETTDLYCY	ligand							110220	A1: 11.52 \pm 7.17	
E7/19-27	TTDLYCYEQ	ligand							66569	A1: 83.36 \pm 15.02	
E7/37-45	EIDGPAGQA	ligand							12414	A1: -	
E7/44-52	QAEPDRAHY	ligand							50240	A1: nb	
E7/49-57	RAHYNIVTF					ligand			53112	A24: 0.15 \pm 0.07	
E7/51-60	HYNIVFCCK			IFN γ					164805	A3: nb	

(Continued)

Annex

Peptide ID	Sequence	A1 ^a	A2 ^a	A3 ^a	A11 ^a	A24 ^a	B7 ^a	B15 ^a	IEDB ID ^b	Exp. binding affinity ^c	Functionality ^d
E7/61-69	CDSTLRRCV					IFN γ , cytolysis			168256	A24: nb	
E7/66-74	RLCVQSTHV		ligand						54496	A2: 37.18 \pm 10.65	A2: IFN γ
E7/67-76	LCVQSTHVDI					IFN γ , cytolysis			169070	A24: 34.74 \pm 11.56	A24: IFN γ
E7/73-81	HVDIRTLED	ligand							25028	A1: -	
E7/73-82	HVDIRTLEDL		ligand						(Bauer et al., 2000)	A2: nb	
E7/75-83	DIRTLEDLL		ligand						(Bauer et al., 2000)	A2: nb	
E7/76-86	IRTLEDLLMGT		ligand						113033	A2: 31.85 \pm 22.45	
E7/77-86	RTLEDLLMGT		ligand, IFN γ , cytolysis						139234	A2: 72.72 \pm 17.36	A2: IFN γ
E7/77-87	RTLEDLLMGTL		ligand						768628	A2: 2.69 \pm 2.94	A2: IFN γ
E7/78-86	TLEDLLMGT		ligand						111833	A2: nb*	A2: IFN γ
E7/78-87	TLEDLLMGTL		ligand							A2: nb	
E7/80-90	EDLLMGTLGIV		ligand						768404	A2: 14.20 \pm 2.47	A2: IFN γ
E7/81-90	DLLMGTLGIV		ligand						139028	A2: 52.58 \pm 19.28	A2: IFN γ
E7/81-91	DLLMGTLGIVC		cytolysis						688709	A2: 41.25 \pm 39.26	A2: IFN γ
E7/82-90	LLMGTLGIV		ligand, IFN γ , cytolysis						37573	A2: 3.30 \pm 2.58	
E7/83-93	LMGTLGIVCPI					ligand, IFN γ			165037	A24: -	
E7/85-93	GTLGIVCPI		ligand, IFN γ , cytolysis						22738	A2: 1.96 \pm 1.93	A2: IFN γ
E7/86-93	TLGIVCPI		ligand, IFN γ , cytolysis						64818	A2: 1.27 \pm 1.54	A2: IFN γ
E7/86-94	TLGIVCPIC		ligand, cytolysis						64819	A2: 62.31 \pm 18.61	
E7/87-93	LGIVCPI		ligand						36171	A2: -	
E7/87-95	LGIVCPICS		ligand						111480	A2: -	

(Continued)

Annex

Peptide ID	Sequence	A1 ^a	A2 ^a	A3 ^a	A11 ^a	A24 ^a	B7 ^a	B15 ^a	IEDB ID ^b	Exp. binding affinity ^c	Functionality ^d
E7/88-97	GIVCPICSQK				ligand, IFN γ , cytolysis				20471	A11: 2.22 \pm 0.84	A11: IFN γ
E7/89-97	IVCPICSQK			ligand	ligand, IFN γ				29222	A3: 4.06 \pm 1.38; A11: 3.16 \pm 0.78	A3: IFN γ ; A11: IFN γ

a: HLA-restricted epitopes reported to exhibit HLA binding, induce IFN γ -responses (IFN γ) or mediate cytotoxicity (cytolysis)

b: Epitope identifier in IEDB; if epitope is not entered in IEDB specific reference is given

c: Experimental binding affinity is expressed as IC₅₀-value \pm Standard deviation in μ M; nb: nonbinder; -: not tested

d: Functionality assessed as IFN γ -responses in ELISpot assays (IFN γ) or cytotoxicity assays (cytolysis)

(Continued)

Supplementary Table S3. HLA-type and peptide length-dependent threshold recommendations and performance indicators for prediction methods

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
HLA A1	8-mers (n=8)			9-mers (n=15)			10-mers (n=23)			11-mers (n=24)			Pooled lengths (n=70)		
NetMHC 4.0	0.75	11702	0.75	0.91	6706	0.73	0.65	3676	0.70	0.63	709	1.00	0.73	6706	0.76±0.11
		(80.00)	1.00		(57.14)	1.00		(72.73)	0.62		(100.00)	0.29		(64.29)	0.64±0.14
NetMHC 3.4	0.75	2008	1.00	0.95	5319	0.91	0.57	4547	0.70	0.48	2391	1.00	0.67	5319	0.77±0.09
		(100.00)	0.75		(80.00)	1.00		(75.00)	0.69		(100.00)	0.14		(62.96)	0.60±0.12
NetMHCpan 4.0	0.63	4421	1.00	0.93	3126	0.82	0.62	1547	0.90	0.78	6597	0.88	0.72	6597	0.73±0.10
		(100.00)	0.25		(66.67)	1.00		(83.33)	0.38		(60.00)	0.43		(60.71)	0.63±0.11
NetMHCpan 3.0	0.75	7096	1.00	0.95	1899	0.91	0.59	1835	0.90	0.71	8697	0.71	0.72	5105	0.81±0.09
		(100.00)	0.50		(80.00)	1.00		(83.33)	0.38		(50.00)	0.71		(66.67)	0.56±0.14
NetMHCpan 2.8	0.69	2190	0.75	0.93	2317	0.82	0.44	267	1.00	0.56	4236	0.82	0.63	2835	0.83±0.09
		(75.00)	0.75		(66.67)	1.00		(100.00)	0.08		(50.00)	0.43		(66.67)	0.50±0.15
NetMHCcons 1.1	0.75	2100	1.00	0.95	3511	0.91	0.50	257	1.00	0.52	4963	0.88	0.66	4552	0.75±0.09
		(100.00)	0.75		(80.00)	1.00		(100.00)	0.08		(50.00)	0.29		(58.33)	0.51±0.16
PickPocket 1.1	0.69	5211	0.75	0.95	1766	0.91	0.62	4288	0.90	0.57	3308	0.82	0.64	4289	0.77±0.08
		(75.00)	0.75		(80.00)	1.00		(87.50)	0.54		(50.00)	0.43		(61.54)	0.58±0.14
IEDB SMMPMBEC	0.56	3764	0.75	0.86	2486	0.82	0.46	384	0.90	0.52	525	1.00	0.60	608	0.93±0.06
		(50.00)	0.25		(60.00)	0.75		(50.00)	0.08		(100.00)	0.14		(62.50)	0.18±0.10
IEDB SMM	0.38	16632	0.75	0.75	323	1.00	0.48	315	1.00	0.43	641	0.76	0.52	315	0.95±0.04
		(66.67)	0.50		(100.00)	0.50		(100.00)	0.08		(20.00)	0.14		(50.00)	0.06±0.07
MHCflurry 1.2	0.69	1204	0.75	0.93	1528	0.82	0.53	313	1.00	0.51	4615	0.88	0.65	1529	0.83±0.09
		(75.00)	0.75		(66.67)	1.00		(100.00)	0.15		(50.00)	0.29		(65.00)	0.45±0.14
MHCnuggets 2.0	0.56	2537	0.75	0.80	4233	0.82	0.62	3287	0.80	0.60	2154	0.88	0.67	5217	0.77±0.09
		(66.67)	0.50		(60.00)	0.75		(77.78)	0.54		(33.33)	0.14		(58.33)	0.50±0.13
IEDB recommended	0.38	11.25	0.75	0.77	0.25	1.00	0.52	0.55	0.90	0.55	1.10	0.82	0.59	0.55	0.96±0.05
		(66.67)	0.50		(100.00)	0.50		(75.00)	0.23		(25.00)	0.14		(71.43)	0.17±0.10
IEDB consensus	0.38	11.20	0.75	0.77	0.25	1.00	0.49	0.35	0.90	0.55	1.15	0.82	0.59	0.35	0.98±0.03
		(66.67)	0.50		(100.00)	0.50		(50.00)	0.08		(25.00)	0.14		(75.00)	0.13±0.09
MixMHCpred 2.0.2	0.44	6.00	0.50	0.86	0.50	0.73	0.60	0.08	0.90	0.45	2.00	0.47	0.56	0.05	0.95±0.05
		(50.00)	0.50		(57.14)	1.00		(50.00)	0.08		(25.00)	0.43		(50.00)	0.06±0.06
HLA A2	8-mers (n=21)			9-mers (n=39)			10-mers (n=46)			11-mers (n=50)			Pooled lengths (n=156)		
NetMHC 4.0	0.83	9128	0.81	0.86	1812	0.96	0.78	2565	0.69	0.83	11451	0.71	0.79	6937	0.70±0.06
		(57.14)	0.80		(90.91)	0.71		(42.11)	0.73		(60.00)	0.94		(50.75)	0.73±0.08

(Continued)

Annex

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
NetMHC 3.4	0.81	6045	0.69	0.83	2403	0.92	0.86	4749	0.80	0.68	4362	0.74	0.79	6045	0.72±0.06
		(50.00)	1.00		(83.33)	0.71		(56.25)	0.82		(50.00)	0.56		(52.31)	0.74±0.09
NetMHCpan 4.0	0.76	11632	0.88	0.90	5803	0.72	0.80	2254	0.80	0.80	10688	0.68	0.80	8103	0.70±0.07
		(60.00)	0.60		(65.00)	0.93		(53.33)	0.73		(54.17)	0.81		(50.00)	0.74±0.08
NetMHCpan 3.0	0.74	9488	0.88	0.87	4733	0.76	0.81	2033	0.80	0.81	11935	0.79	0.78	9488	0.67±0.07
		(60.00)	0.60		(66.67)	0.86		(53.33)	0.73		(63.16)	0.75		(48.53)	0.71±0.10
NetMHCpan 2.8	0.80	2302	0.69	0.85	4164	0.84	0.81	4922	0.83	0.72	4714	0.71	0.79	4923	0.74±0.06
		(50.00)	1.00		(73.33)	0.79		(57.14)	0.73		(54.55)	0.75		(55.38)	0.78±0.09
NetMHCcons 1.1	0.81	2327	0.75	0.85	5267	0.76	0.83	8711	0.69	0.72	5382	0.71	0.80	5383	0.74±0.06
		(50.00)	0.80		(64.71)	0.79		(47.62)	0.91		(54.55)	0.75		(56.25)	0.79±0.09
PickPocket 1.1	0.76	735	0.75	0.83	561	0.80	0.76	704	0.86	0.68	953	0.76	0.76	954	0.74±0.05
		(50.00)	0.80		(68.75)	0.79		(54.55)	0.55		(57.89)	0.69		(53.33)	0.71±0.10
IEDB SMMPMBEC	0.60	926	0.75	0.84	3072	0.72	0.91	1294	0.83	0.65	283	0.94	0.78	1715	0.70±0.06
		(42.86)	0.60		(63.16)	0.86		(62.50)	0.91		(60.00)	0.19		(50.75)	0.74±0.09
IEDB SMM	0.51	167	0.94	0.85	3077	0.76	0.90	1642	0.74	0.61	90	1.00	0.73	1277	0.68±0.07
		(50.00)	0.20		(66.67)	0.86		(52.63)	0.91		(100.00)	0.06		(46.03)	0.64±0.10
MHCflurry 1.2	0.84	4315	0.69	0.92	4896	0.76	0.89	3410	0.89	0.74	7743	0.71	0.83	5310	0.74±0.06
		(50.00)	1.00		(70.00)	1.00		(69.23)	0.82		(54.55)	0.75		(56.72)	0.82±0.08
MHCnuggets 2.0	0.73	8357	0.69	0.88	1038	0.84	0.81	1418	1.00	0.84	7691	0.68	0.82	3482	0.73±0.06
		(44.44)	0.80		(75.00)	0.86		(100.00)	0.73		(56.00)	0.88		(55.38)	0.77±0.10
IEDB recommended	0.70	1.00	1.00	0.80	8.70	0.68	0.85	6.60	0.80	0.76	16.05	0.71	0.77	10.35	0.71±0.06
		(100.00)	0.20		(57.89)	0.79		(56.25)	0.82		(54.55)	0.75		(51.52)	0.73±0.09
IEDB consensus	0.70	0.80	1.00	0.85	6.60	0.72	0.89	6.20	0.77	0.75	9.25	0.79	0.77	8.98	0.70±0.06
		(100.00)	0.20		(58.82)	0.71		(55.56)	0.91		(58.82)	0.63		(49.25)	0.70±0.10
MixMHCpred 2.0.2	0.65	12.00	0.81	0.74	0.50	0.88	0.68	2.00	0.94	0.66	6.00	0.74	0.68	5.00	0.72±0.06
		(40.00)	0.40		(70.00)	0.50		(66.67)	0.36		(52.63)	0.63		(46.43)	0.56±0.10
HLA A3	8-mers (n=13)			9-mers (n=34)			10-mers (n=35)			11-mers (n=23)			Pooled lengths (n=105)		
NetMHC 4.0	0.91	10604	0.91	0.89	3475	0.92	0.94	1603	0.89	0.34	234	1.00	0.78	6962	0.73±0.07
		(66.67)	1.00		(77.78)	0.88		(70.00)	1.00		(100.00)	0.25		(41.03)	0.78±0.14
NetMHC 3.4	0.55	752	0.73	0.91	3086	0.96	0.97	2536	0.89	0.45	955	1.00	0.84	3086	0.78±0.06
		(25.00)	0.50		(87.50)	0.88		(70.00)	1.00		(100.00)	0.25		(48.57)	0.79±0.14
NetMHCpan 4.0	0.82	12511	0.73	0.88	8302	0.85	0.95	1365	0.93	0.42	132	1.00	0.78	7696	0.68±0.07
		(40.00)	1.00		(63.64)	0.88		(77.78)	1.00		(100.00)	0.25		(35.71)	0.71±0.15

(Continued)

Annex

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
NetMHCpan 3.0	0.82	10536	0.73	0.88	8408	0.77	0.95	1769	0.89	0.39	163	1.00	0.78	5534	0.77±0.06
		(40.00)	1.00		(53.85)	0.88		(70.00)	1.00		(100.00)	0.25		(44.12)	0.72±0.14
NetMHcpan 2.8	0.73	637	0.91	0.90	7001	0.92	0.98	2365	0.96	0.36	131	1.00	0.79	2366	0.79±0.07
		(50.00)	0.50		(77.78)	0.88		(87.50)	1.00		(100.00)	0.25		(47.06)	0.75±0.15
NetMHCcons 1.1	0.64	693	0.91	0.90	4651	0.96	0.98	2000	0.96	0.39	354	1.00	0.82	2000	0.85±0.06
		(50.00)	0.50		(87.50)	0.88		(87.50)	1.00		(100.00)	0.25		(57.14)	0.77±0.14
PickPocket 1.1	0.68	1551	0.82	0.78	1655	0.88	0.92	4990	0.71	0.36	1422	0.84	0.73	2874	0.72±0.06
		(33.33)	0.50		(66.67)	0.75		(46.67)	1.00		(25.00)	0.25		(39.47)	0.70±0.15
IEDB SMMPMBEC	0.36	21975	0.64	0.90	1881	0.96	0.96	971	0.93	0.54	691	0.79	0.80	1881	0.82±0.06
		(20.00)	0.50		(87.50)	0.88		(77.78)	1.00		(33.33)	0.50		(51.61)	0.77±0.15
IEDB SMM	0.55	13611	0.64	0.88	1909	0.96	0.96	906	0.89	0.36	2778	0.53	0.79	2930	0.76±0.06
		(20.00)	0.50		(87.50)	0.88		(70.00)	1.00		(10.00)	0.25		(43.24)	0.78±0.13
MHCflurry 1.2	0.64	220	1.00	0.85	3505	0.85	0.97	2047	0.93	0.41	1272	0.84	0.79	3506	0.72±0.06
		(100.00)	0.50		(63.64)	0.88		(77.78)	1.00		(25.00)	0.25		(41.46)	0.82±0.13
MHCnuggets 2.0	0.32	3622	0.45	0.90	2433	0.92	0.98	458	0.96	0.46	199	0.95	0.81	3623	0.73±0.07
		(14.29)	0.50		(77.78)	0.88		(87.50)	1.00		(50.00)	0.25		(40.48)	0.82±0.14
IEDB recommended	0.64	3.45	0.82	0.88	2.70	0.96	0.98	3.50	0.93	0.33	6.65	0.68	0.77	3.50	0.86±0.06
		(33.33)	0.50		(87.50)	0.88		(77.78)	1.00		(14.29)	0.25		(55.56)	0.71±0.14
IEDB consensus	0.64	3.15	0.82	0.88	2.90	0.96	0.98	11.35	0.89	0.32	6.60	0.63	0.77	11.35	0.68±0.07
		(33.33)	0.50		(87.50)	0.88		(70.00)	1.00		(12.50)	0.25		(38.10)	0.79±0.13
MixMHCpred 2.0.2	0.64	6.00	0.91	0.87	2.00	0.88	0.99	3.00	0.86	0.54	0.50	0.89	0.82	6.00	0.74±0.07
		(50.00)	0.50		(70.00)	0.88		(63.64)	0.57		(33.33)	0.25		(43.59)	0.80±0.13
HLA A11		8-mers (n=24)			9-mers (n=40)			10-mers (n=33)			11-mers (n=40)			Pooled lengths (n=137)	
NetMHC 4.0	0.98	9342	0.94	0.95	9085	0.77	0.82	1928	0.85	0.76	9183	0.70	0.85	9342	0.72±0.07
		(85.71)	1.00		(56.25)	1.00		(75.00)	0.85		(52.94)	0.69		(57.14)	0.89±0.08
NetMHC 3.4	0.98	1354	0.94	0.94	13066	0.71	0.87	5497	0.70	0.73	6972	0.70	0.86	6972	0.72±0.06
		(85.71)	1.00		(50.00)	1.00		(64.71)	0.77		(57.89)	0.85		(58.06)	0.87±0.08
NetMHCpan 4.0	1.00	7719	1.00	0.92	7399	0.71	0.83	2782	0.70	0.71	5361	0.81	0.85	7720	0.72±0.07
		(100.00)	1.00		(47.06)	0.89		(62.50)	0.69		(58.33)	0.54		(55.74)	0.83±0.09
NetMHCpan 3.0	1.00	5308	1.00	0.90	980	0.97	0.82	1721	0.70	0.71	7709	0.70	0.84	7568	0.68±0.06
		(100.00)	1.00		(87.50)	0.78		(60.00)	0.92		(50.00)	0.62		(52.31)	0.84±0.08

(Continued)

Annex

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
NetMHcpan 2.8	1.00	349	1.00	0.94	6917	0.87	0.86	5277	0.70	0.69	158	0.96	0.85	7182	0.69±0.07
		(100.00)	1.00		(66.67)	0.89		(66.67)	0.85		(50.00)	0.08		(53.73)	0.87±0.07
NetMHCcons 1.1	0.98	685	0.94	0.94	6829	0.87	0.86	4406	0.70	0.72	7093	0.67	0.86	7734	0.71±0.07
		(85.71)	1.00		(66.67)	0.89		(64.71)	0.77		(57.14)	0.92		(57.58)	0.93±0.07
PickPocket 1.1	0.94	3237	0.89	0.90	2417	0.74	0.75	4626	0.70	0.75	6331	0.70	0.77	3726	0.67±0.07
		(75.00)	1.00		(50.00)	0.89		(62.50)	0.77		(55.56)	0.77		(49.18)	0.72±0.10
IEDB SMMPMBEC	0.95	3425	0.89	0.92	1253	1.00	0.86	2715	0.75	0.66	510	0.81	0.84	1966	0.74±0.06
		(75.00)	1.00		(100.00)	0.78		(66.67)	0.85		(58.33)	0.54		(55.17)	0.77±0.09
IEDB SMM	0.81	3249	0.72	0.92	3633	0.68	0.86	2923	0.70	0.71	229	0.93	0.82	1539	0.80±0.06
		(50.00)	0.83		(44.44)	0.89		(64.71)	0.69		(66.67)	0.31		(62.00)	0.76±0.11
MHCflurry 1.2	1.00	377	1.00	0.94	3531	0.71	0.81	1263	0.75	0.72	1799	0.81	0.84	3412	0.67±0.06
		(100.00)	1.00		(50.00)	1.00		(64.29)	0.85		(50.00)	0.38		(50.77)	0.80±0.09
MHCnuggets 2.0	0.98	1230	0.94	0.97	6299	0.87	0.83	5953	0.70	0.81	6339	0.67	0.90	6339	0.78±0.06
		(85.71)	1.00		(69.23)	1.00		(64.71)	0.77		(57.14)	0.92		(64.41)	0.94±0.06
IEDB recommended	0.94	11.05	0.67	0.94	5.70	0.74	0.87	6.15	0.70	0.83	8.85	0.81	0.87	8.60	0.67±0.06
		(50.00)	1.00		(52.94)	1.00		(62.50)	0.77		(70.59)	0.92		(55.88)	0.91±0.06
IEDB consensus	0.94	10.65	0.67	0.94	5.00	0.74	0.87	4.85	0.70	0.82	7.65	0.81	0.87	7.65	0.68±0.06
		(50.00)	1.00		(52.94)	1.00		(62.50)	0.46		(70.59)	0.92		(55.88)	0.94±0.06
MixMHCpred 2.0.2	0.91	10.00	0.94	0.71	2.00	0.84	0.75	4.00	1.00	0.66	3.00	0.89	0.72	10.50	0.67±0.06
		(83.33)	0.83		(44.44)	0.44		(100.00)	0.62		(50.00)	0.23		(45.61)	0.64±0.11
HLA A24		8-mers (n=18)			9-mers (n=36)			10-mers (n=46)			11-mers (n=29)			Pooled lengths (n=129)	
NetMHC 4.0	0.79	25903	0.75	0.76	7820	0.71	0.76	9077	0.75	0.78	20603	0.67	0.73	12915	0.70±0.09
		(92.31)	0.86		(72.22)	0.68		(71.43)	0.77		(60.00)	0.82		(69.35)	0.65±0.07
NetMHC 3.4	0.71	11113	0.75	0.81	6548	0.82	0.78	14977	0.67	0.78	17668	0.72	0.76	11562	0.75±0.08
		(88.89)	0.57		(82.35)	0.74		(68.00)	0.77		(61.54)	0.73		(73.33)	0.67±0.08
NetMHcpan 4.0	0.86	33856	0.75	0.82	12843	0.71	0.77	12097	0.75	0.74	22981	0.67	0.74	14268	0.71±0.09
		(92.86)	0.93		(75.00)	0.79		(73.91)	0.77		(60.00)	0.82		(69.84)	0.65±0.08
NetMHcpan 3.0	0.86	34571	0.75	0.81	12345	0.71	0.76	11145	0.71	0.72	18252	0.72	0.74	12346	0.73±0.07
		(92.86)	0.93		(73.68)	0.74		(70.83)	0.64		(61.54)	0.73		(71.67)	0.66±0.08
NetMHcpan 2.8	0.79	10002	0.75	0.81	10228	0.82	0.76	12415	0.88	0.77	19059	0.72	0.78	12415	0.78±0.08
		(91.67)	0.79		(83.33)	0.79		(82.35)	0.82		(64.29)	0.82		(76.67)	0.70±0.07

(Continued)

Annex

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
NetMHCcons 1.1	0.79	10027	1.00	0.81	10934	0.71	0.79	14564	0.67	0.78	20590	0.72	0.78	13142	0.68±0.08
		(100.00)	0.64		(76.19)	0.84		(69.23)	0.64		(64.29)	0.82		(72.86)	0.78±0.07
PickPocket 1.1	0.77	5099	0.75	0.69	1655	0.82	0.71	5211	0.71	0.79	24218	0.67	0.71	5100	0.67±0.09
		(90.91)	0.71		(76.92)	0.53		(66.67)	0.55		(62.50)	0.91		(67.69)	0.67±0.07
IEDB SMMPMBEC	0.18	1123	0.75	0.72	2183	0.71	0.74	1674	0.83	0.75	9957	0.67	0.64	1674	0.84±0.07
		(50.00)	0.07		(70.59)	0.63		(75.00)	0.77		(62.50)	0.91		(70.59)	0.37±0.07
IEDB SMM	0.18	59	1.00	0.71	1724	0.82	0.76	3355	0.71	0.75	8671	0.78	0.64	3025	0.75±0.07
		(100.00)	0.07		(75.00)	0.47		(70.83)	0.73		(66.67)	0.73		(66.67)	0.49±0.08
MHCflurry 1.2	0.71	3941	0.75	0.77	2322	0.71	0.75	4369	0.67	0.82	10190	0.67	0.76	3942	0.71±0.08
		(90.00)	0.64		(73.68)	0.74		(66.67)	0.73		(60.00)	0.82		(70.49)	0.63±0.07
MHCnuggets 2.0	0.73	1382	1.00	0.80	3169	0.71	0.81	5506	0.71	0.81	14997	0.67	0.80	5050	0.70±0.09
		(100.00)	0.50		(73.68)	0.74		(69.57)	0.82		(62.50)	0.91		(70.77)	0.71±0.08
IEDB recommended	0.36	6.20	1.00	0.74	3.15	0.71	0.78	5.35	0.75	0.85	19.90	0.89	0.70	6.40	0.77±0.07
		(100.00)	0.14		(72.22)	0.68		(75.00)	0.82		(81.82)	0.82		(71.70)	0.58±0.09
IEDB consensus	0.34	3.85	1.00	0.73	2.80	0.71	0.77	5.00	0.75	0.84	21.95	0.78	0.69	6.53	0.74±0.08
		(100.00)	0.07		(72.22)	0.68		(75.00)	0.73		(69.23)	0.82		(67.92)	0.55±0.09
MixMHCpred 2.0.2	0.55	22.00	0.75	0.79	2.00	0.71	0.75	6.00	0.75	0.76	9.00	0.78	0.68	4.00	0.77±0.07
		(88.89)	0.57		(76.19)	0.84		(72.73)	0.77		(63.64)	0.64		(72.22)	0.59±0.09
HLA B7	8-mers (n=11)			9-mers (n=13)			10-mers (n=16)			11-mers (n=15)			Pooled lengths (n=55)		
NetMHC 4.0	0.97	21900	0.80	0.90	6673	1.00	0.79	3061	0.92	0.70	16556	0.67	0.77	12895	0.70±0.12
		(85.71)	1.00		(100.00)	0.86		(75.00)	0.75		(75.00)	0.67		(66.67)	0.69±0.13
NetMHC 3.4	0.97	16001	0.80	0.93	5180	1.00	0.75	5373	0.83	0.87	14182	0.67	0.87	7396	0.87±0.10
		(85.71)	1.00		(100.00)	0.86		(60.00)	0.75		(80.00)	0.89		(84.00)	0.81±0.11
NetMHCpan 4.0	0.90	5515	1.00	0.93	14910	0.67	0.79	5644	0.83	0.81	20480	0.83	0.81	10711	0.86±0.09
		(100.00)	0.83		(77.78)	1.00		(60.00)	0.75		(87.50)	0.78		(82.61)	0.73±0.12
NetMHCpan 3.0	0.90	6604	1.00	0.88	6431	0.67	0.75	4632	0.83	0.80	19735	0.83	0.79	14473	0.72±0.11
		(100.00)	0.83		(75.00)	0.86		(60.00)	0.75		(87.50)	0.78		(72.41)	0.82±0.10
NetMHCpan 2.8	0.93	1846	1.00	0.90	3666	1.00	0.71	3279	0.83	0.83	7505	0.67	0.86	7506	0.76±0.10
		(100.00)	0.83		(100.00)	0.86		(60.00)	0.75		(80.00)	0.89		(75.86)	0.84±0.11
NetMHCcons 1.1	0.93	1091	1.00	0.93	4335	1.00	0.71	4174	0.83	0.85	7946	0.67	0.87	7946	0.76±0.10
		(100.00)	0.83		(100.00)	0.86		(60.00)	0.75		(80.00)	0.89		(75.86)	0.85±0.09

(Continued)

Annex

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
PickPocket 1.1	0.93	2340	1.00	0.79	462	0.83	0.73	1585	0.92	0.65	1766	0.83	0.79	2874	0.77±0.11
		(100.00)	0.83		(83.33)	0.71		(75.00)	0.75		(83.33)	0.56		(74.07)	0.76±0.12
IEDB SMMPMBEC	0.87	3196	0.80	0.90	1133	1.00	0.69	1308	0.67	0.35	3528	0.67	0.65	1309	0.82±0.11
		(83.33)	0.83		(100.00)	0.86		(42.86)	0.75		(33.33)	0.11		(72.22)	0.50±0.15
IEDB SMM	0.90	333	1.00	0.93	1256	1.00	0.40	687	0.50	0.39	88	1.00	0.64	923	0.72±0.11
		(100.00)	0.83		(100.00)	0.86		(33.33)	0.50		(100.00)	0.22		(65.22)	0.55±0.12
MHCflurry 1.2	0.93	1815	1.00	0.88	4448	0.83	0.75	2064	0.83	0.83	3295	0.83	0.85	4448	0.81±0.13
		(100.00)	0.83		(85.71)	0.86		(50.00)	0.75		(87.50)	0.78		(76.92)	0.77±0.11
MHCnuggets 2.0	0.93	4125	1.00	0.81	2408	0.83	0.85	2912	0.92	0.91	8533	0.83	0.87	9056	0.75±0.10
		(100.00)	0.83		(83.33)	0.71		(75.00)	0.50		(88.89)	0.89		(75.86)	0.84±0.10
IEDB recommended	1.00	17.90	1.00	0.93	3.80	1.00	0.63	1.75	0.92	0.44	1.10	1.00	0.74	4.65	0.77±0.13
		(100.00)	1.00		(100.00)	0.86		(66.67)	0.50		(100.00)	0.11		(69.57)	0.63±0.13
IEDB consensus	1.00	17.85	1.00	0.93	3.80	1.00	0.63	1.45	1.00	0.46	1.30	1.00	0.74	4.20	0.82±0.10
		(100.00)	1.00		(100.00)	0.86		(100.00)	0.25		(100.00)	0.11		(76.19)	0.61±0.14
MixMHCpred 2.0.2	0.90	14.00	0.80	0.79	2.00	0.83	0.58	0.50	1.00	0.70	12.00	0.67	0.71	0.60	0.85±0.10
		(83.33)	0.83		(85.71)	0.86		(100.00)	0.75		(75.00)	0.67		(87.50)	0.43±0.14
HLA B15		8-mers (n=20)			9-mers (n=34)			10-mers (n=48)			11-mers (n=25)			Pooled lengths (n=127)	
NetMHC 4.0	0.93	14125	0.75	0.83	4254	0.67	0.89	6140	0.70	0.82	6524	0.80	0.83	7638	0.68±0.09
		(72.73)	1.00		(75.00)	0.79		(64.00)	0.89		(93.33)	0.70		(72.60)	0.82±0.06
NetMHC 3.4	0.80	2654	0.67	0.82	5652	0.67	0.89	4551	0.83	0.60	438	0.80	0.82	5050	0.69±0.08
		(63.64)	0.88		(75.00)	0.79		(76.19)	0.94		(90.91)	0.50		(73.61)	0.81±0.08
NetMHCpan 4.0	0.99	7695	0.92	0.94	4781	0.73	0.86	7669	0.67	0.85	5189	0.80	0.89	7230	0.72±0.09
		(88.89)	1.00		(81.82)	0.95		(62.96)	0.94		(94.12)	0.80		(75.64)	0.91±0.05
NetMHCpan 3.0	1.00	7374	1.00	0.94	4008	0.87	0.87	7102	0.67	0.79	3766	1.00	0.88	7596	0.72±0.08
		(100.00)	1.00		(90.00)	0.95		(62.96)	0.83		(100.00)	0.65		(76.62)	0.91±0.05
NetMHcpan 2.8	0.80	2808	0.67	0.90	7508	0.67	0.87	10190	0.70	0.62	709	0.80	0.85	6047	0.70±0.07
		(66.67)	1.00		(78.26)	0.95		(62.50)	0.94		(90.91)	0.50		(74.32)	0.85±0.06
NetMHCcons 1.1	0.82	3219	0.75	0.87	7406	0.67	0.89	10875	0.67	0.61	558	0.80	0.84	5183	0.72±0.08
		(72.73)	1.00		(76.19)	0.84		(62.96)	0.78		(90.91)	0.50		(74.65)	0.82±0.07
PickPocket 1.1	0.71	3933	0.75	0.94	4831	0.73	0.83	10875	0.73	0.53	2443	0.80	0.81	6756	0.70±0.08
		(62.50)	0.63		(81.82)	0.95		(63.64)	0.78		(90.00)	0.45		(72.86)	0.78±0.07

(Continued)

Annex

Predictor ^a	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Threshold ^b (PPV [%])	Specificity Sensitivity	A _{ROC}	Val. Threshold ^c (PPV [%])	Specificity Sensitivity
IEDB SMMPMBEC	-	-	-	0.82	3380	0.67	0.79	354	0.73	-	-	-	0.72	478	<i>0.80±0.10</i>
		(-)	-		(75.00)	0.79		(63.64)	0.78		(-)	-		(68.97)	<i>0.54±0.11</i>
IEDB SMM	-	-	-	0.84	2296	0.67	0.82	182	0.77	-	-	-	0.71	184	<i>0.83±0.08</i>
		(-)	-		(73.68)	0.74		(66.67)	0.78		(-)	-		(73.08)	<i>0.52±0.12</i>
MHCflurry 1.2	0.67	813	0.75	0.88	3836	0.67	0.86	2586	0.87	0.55	596	0.80	0.80	2586	0.73±0.08
		(57.14)	0.50		(78.26)	0.95		(77.78)	0.72		(90.91)	0.50		(74.60)	0.73±0.08
MHCnuggets 2.0	0.65	686	0.92	0.74	494	1.00	0.78	6154	0.67	0.42	12	1.00	0.76	1934	0.71±0.08
		(75.00)	0.38		(100.00)	0.53		(56.52)	0.94		(100.00)	0.20		(69.49)	0.65±0.08
IEDB recommended	0.91	3.60	0.75	0.81	12.00	0.67	0.89	14.75	0.67	0.81	0.90	0.80	0.85	5.80	0.82±0.08
		(72.73)	1.00		(73.68)	0.74		(62.96)	0.78		(93.33)	0.70		(81.25)	0.80±0.07
IEDB consensus	0.91	2.80	0.75	0.81	12.00	0.67	0.87	4.55	0.87	0.81	0.10	0.60	0.84	5.80	0.78±0.07
		(72.73)	1.00		(73.68)	0.74		(77.78)	0.83		(87.50)	0.70		(78.79)	0.80±0.07
MixMHCpred 2.0.2	0.81	29.00	0.67	0.88	10.00	0.67	0.80	19.00	0.67	0.83	7.00	0.80	0.84	14.00	0.71±0.09
		(63.64)	0.88		(77.27)	0.89		(60.00)	0.00		(93.75)	0.75		(74.65)	0.81±0.07

a: Units of thresholds depend on predictor output. IEDB recommended, IEDB consensus and MixMHCpred 2.0.2 express results as percentile rank, all others as IC₅₀ in nM.

b: Single length analysis was performed applying criteria-based thresholds.

c: Pooled peptide length analysis was performed applying bootstrapping-validated thresholds. Specificity and sensitivity are expressed mean ± SD of 100 runs.

A_{ROC}: area under receiver operating characteristic curve, maximum = 1

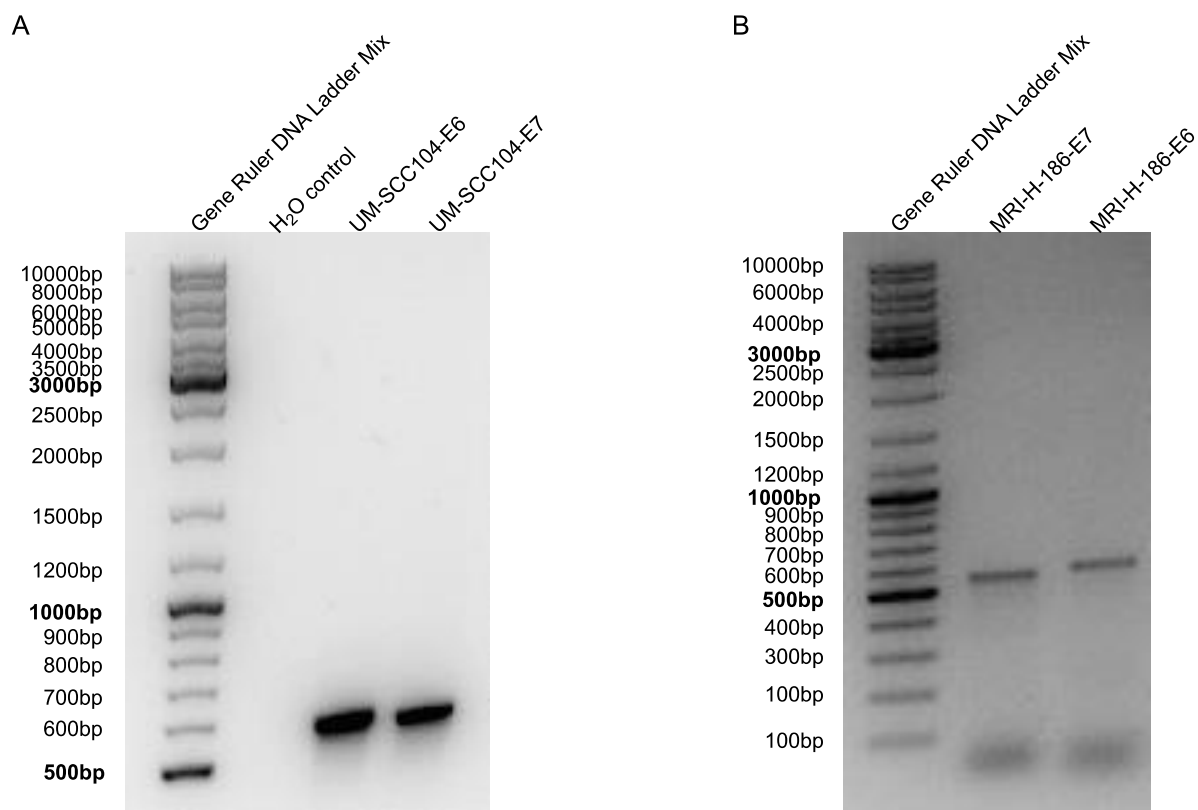
PPV [%]: positive predictive value expressed in %

Grey cell: No possible threshold met the criteria for optimal thresholds. Analysis was performed for the threshold resulting in lowest FPR and highest TPR.

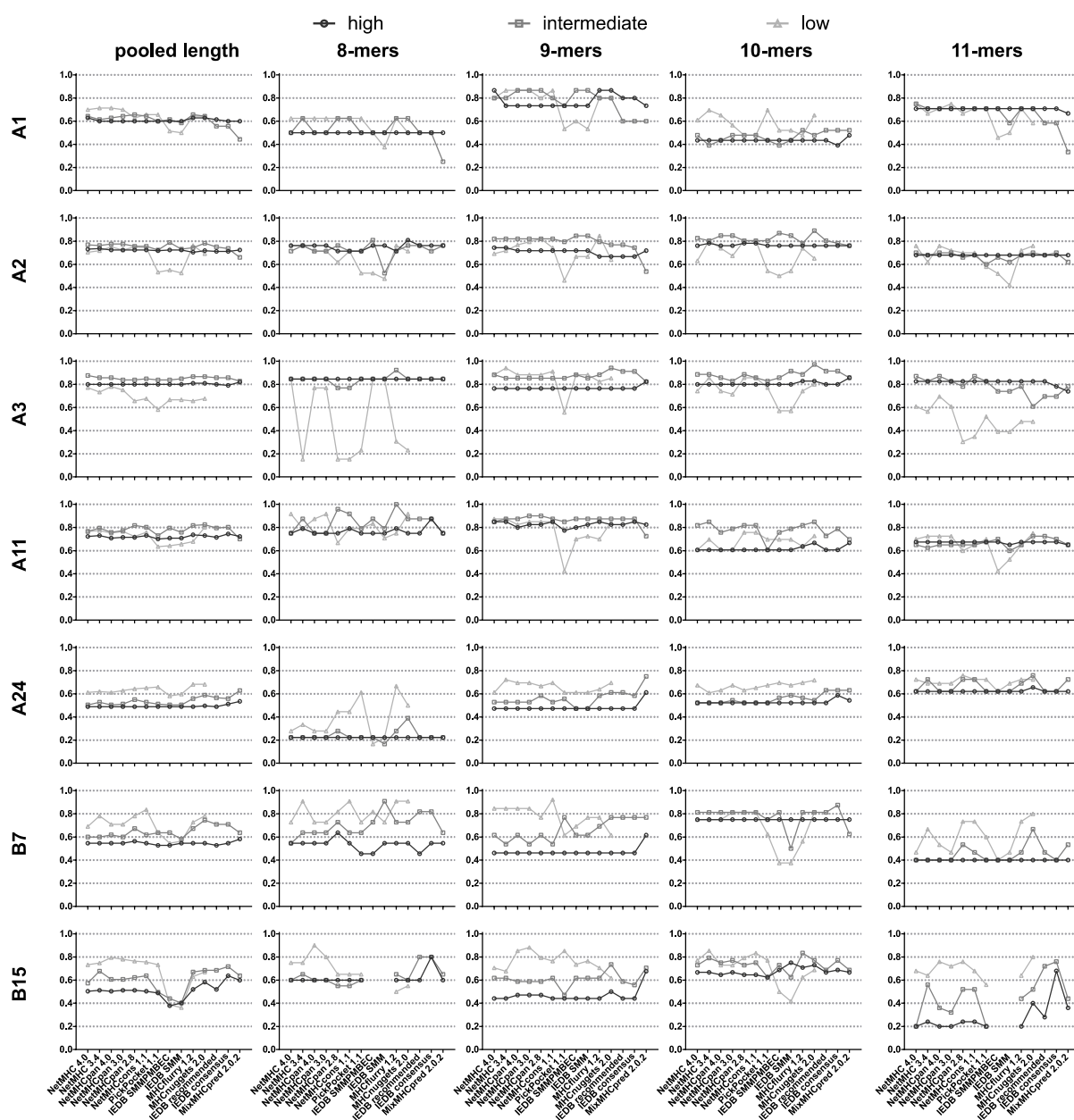
- : Prediction was not available.

Italic font: Numbers of peptides differ.

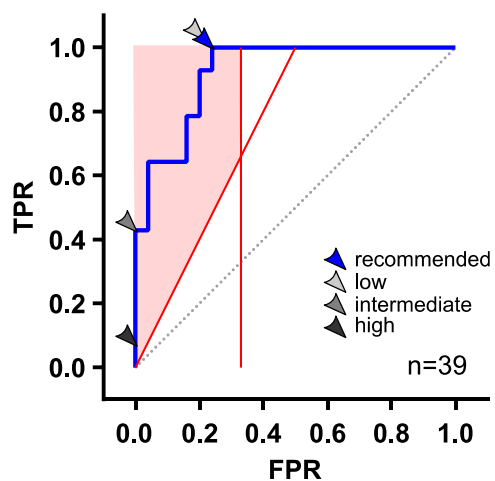
(Continued)



Supplementary Figure S1. Agarose gel electrophoresis of E6- and E7-specific PCR products amplified from genomic DNA of HPV16-positive cell lines. Amplification of E6 and E7 genes by PCR was verified by electrophoresis of a 1% agarose gel. From left to right first lanes show the marker (GeneRuler ladder mix with a range of 100-10000bp) followed by (A) a negative control with water, the UM-SCC104 PCR products of E6 and E7 or (B) the MRI-H-186 PCR products for E7 and E6. Expected amplicon sizes are 524bp for E6 and 505pb for E7.



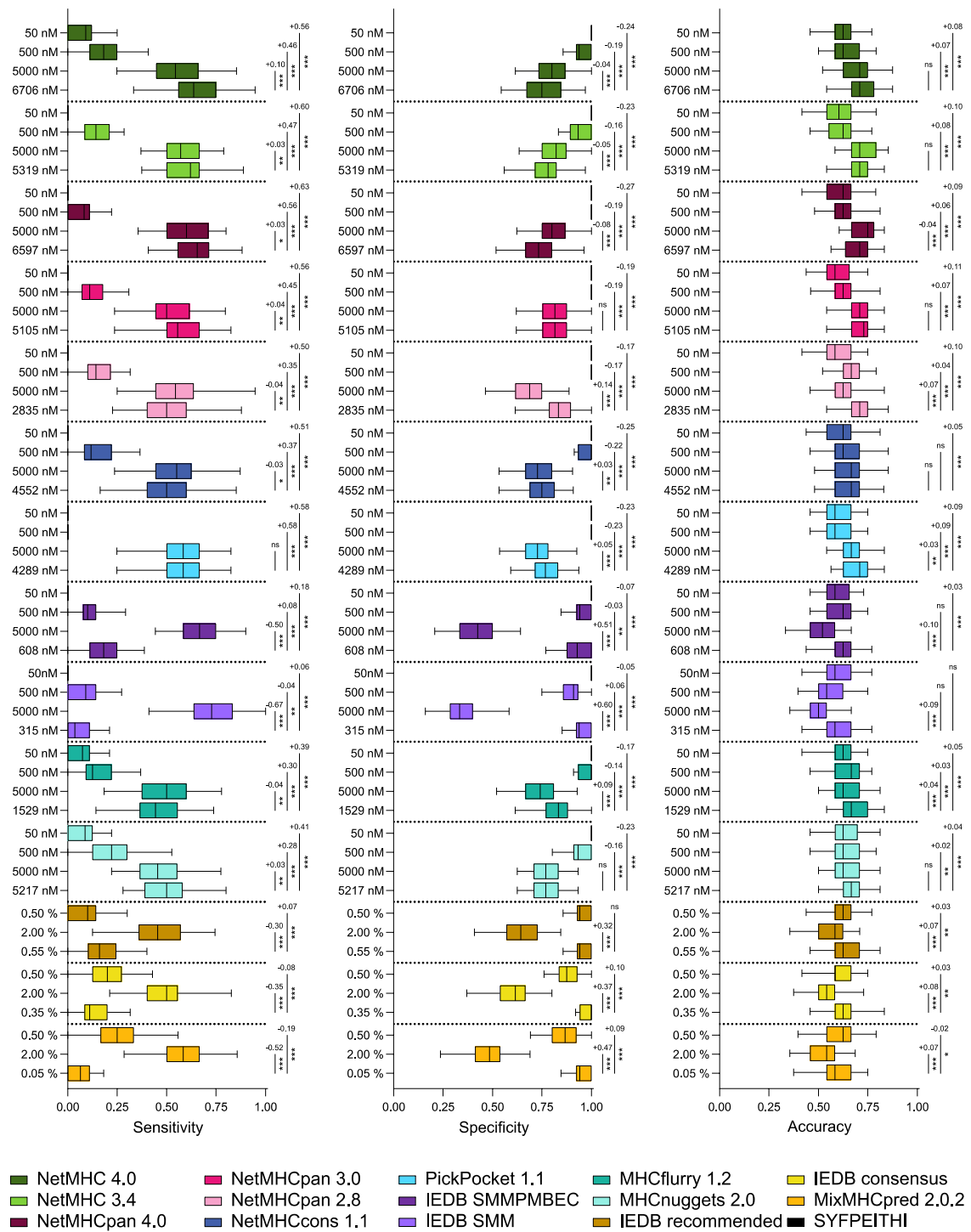
Supplementary Figure S2. Threshold-dependent accuracy of predictors for different HLA types and peptide lengths. Accuracy of predictors is shown in line graphs for HLA types and pooled and single 8-, 9-, 10-, and 11-mer peptides. Symbols in different tones of grey represent the accuracy of a predictor by applying different decision thresholds: high ($IC_{50} \leq 50nM$ or percentile rank ≤ 0.5), intermediate (inter, $IC_{50} \leq 500nM$ or percentile rank ≤ 2) or low ($IC_{50} \leq 5000nM$) binding likelihood.



Supplementary Figure S3. Illustration of criteria defined in this study for optimal decision thresholds. Optimal decision threshold were defined by resulting in a maximal FPR of 0.33 and a TPR which is minimal twice as high as the FPR, depicted by red lines in the ROC graph. Within the pink area resulting from these two criteria, the threshold resulting in the highest possible TPR and lowest possible FPR was selected for recommendation (blue arrow). For comparison, high (dark grey, $IC_{50} \leq 50\text{nM}$ or percentile rank ≤ 0.5), intermediate (medium grey, $IC_{50} \leq 500\text{nM}$ or percentile rank ≤ 2) and low (light grey, $IC_{50} \leq 5000\text{nM}$) binding affinity thresholds are also indicated by arrows.

A

A1

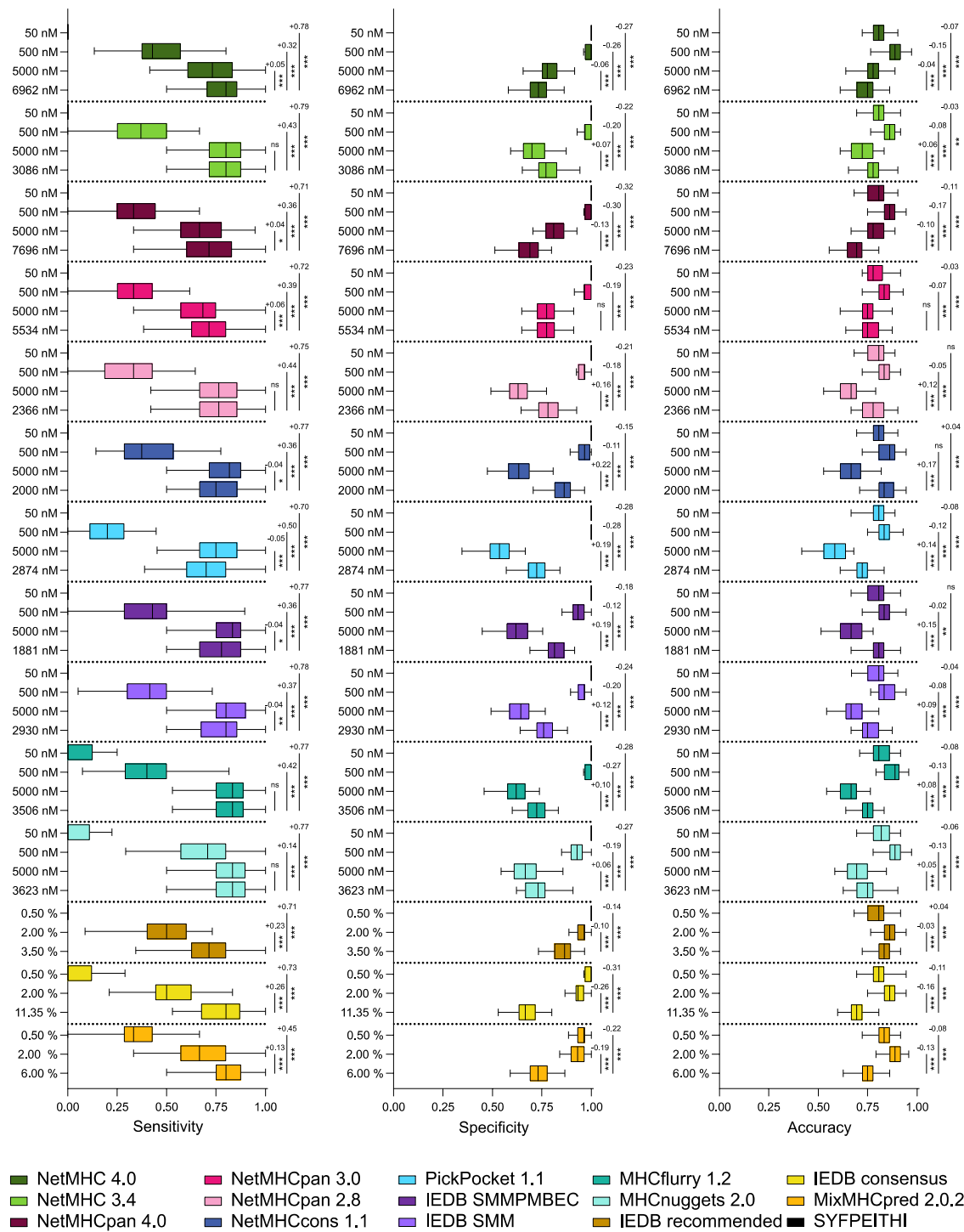


Supplementary Figure S4. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended or general thresholds.

(Figure continues, see Figure legend on page 174)

B

A3

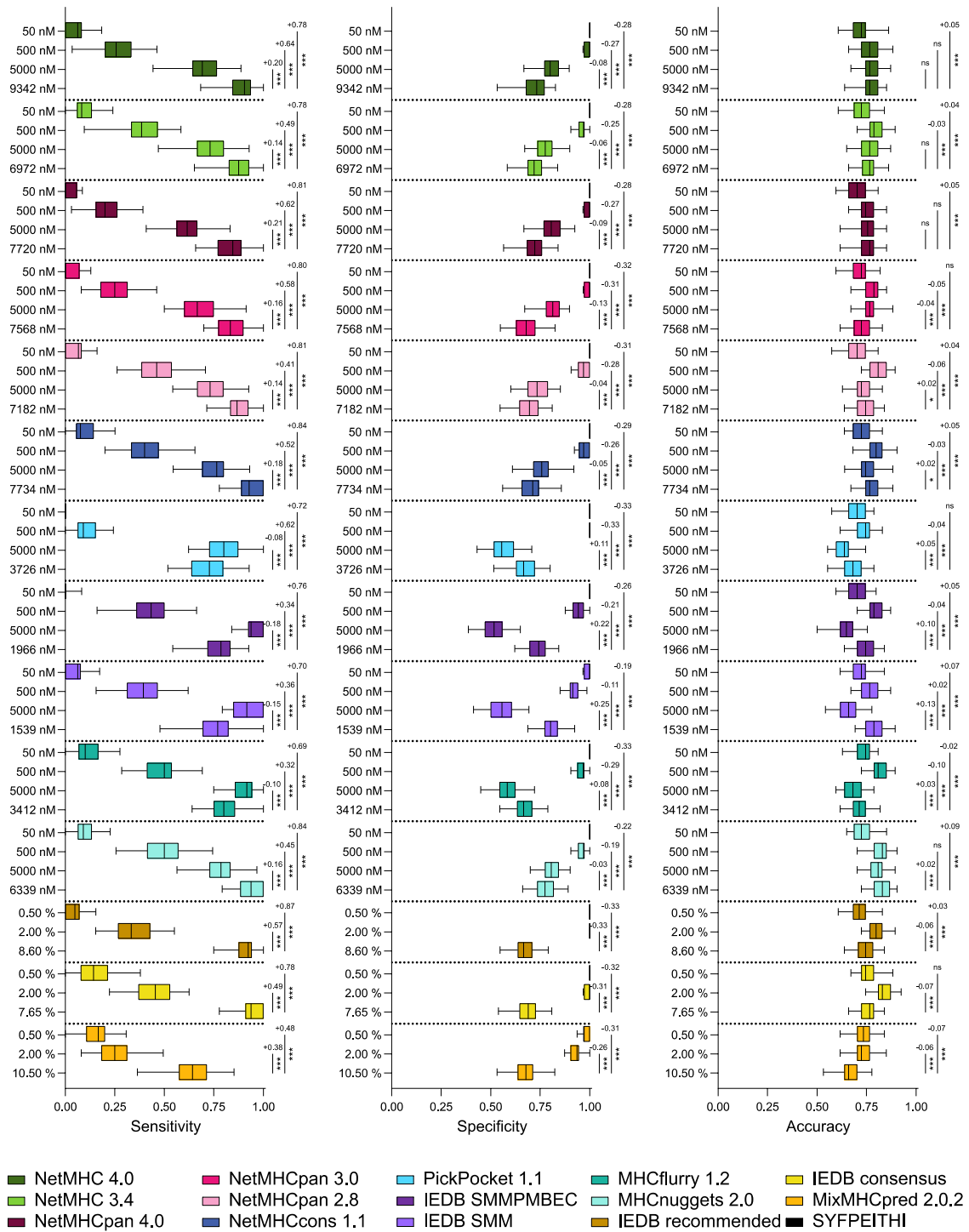


Supplementary Figure S4. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended or general thresholds.

(Figure continues, see Figure legend on page 174)

C

A11

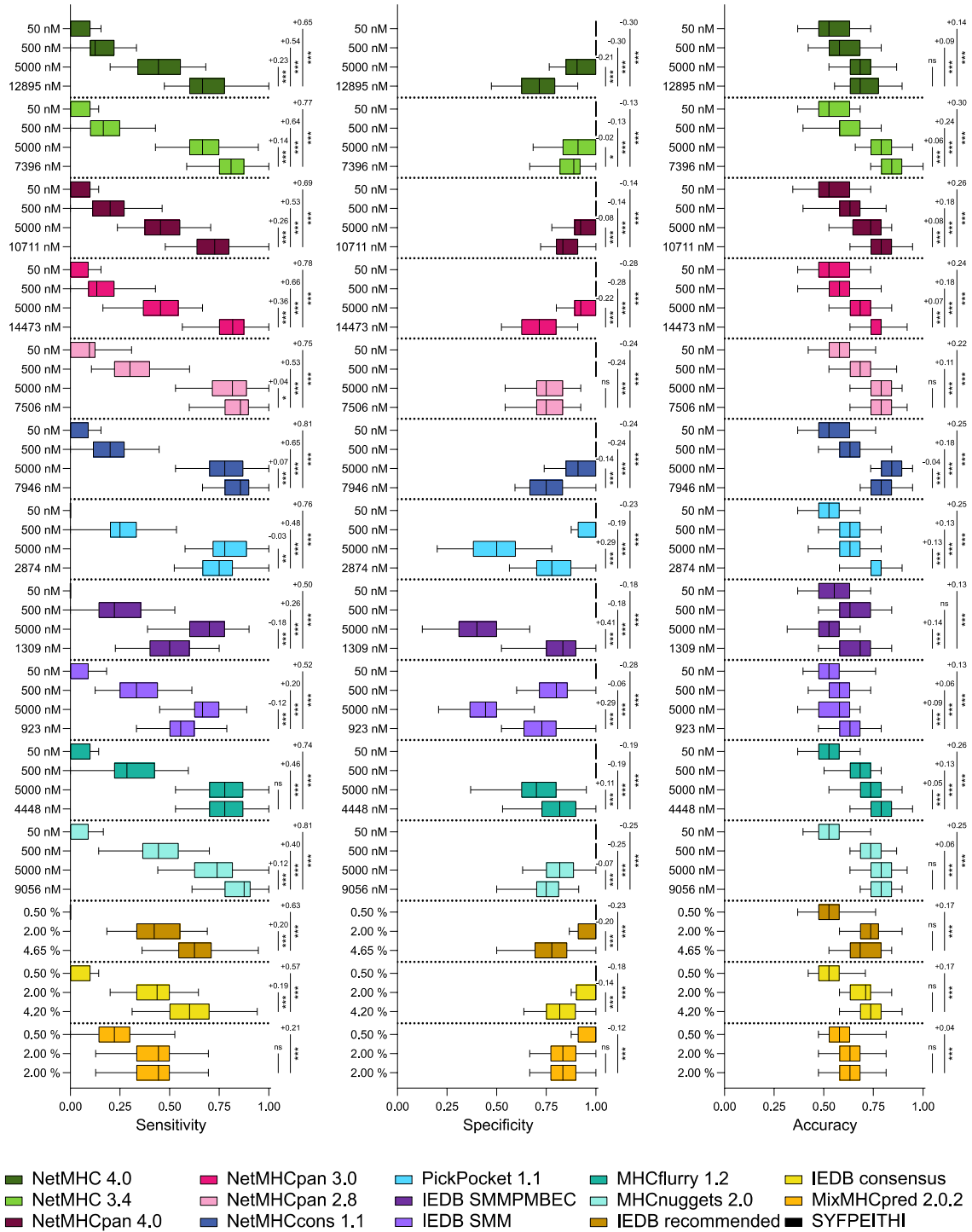


Supplementary Figure S4. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended or general thresholds.

(Figure continues, see Figure legend on page 174)

D

B7

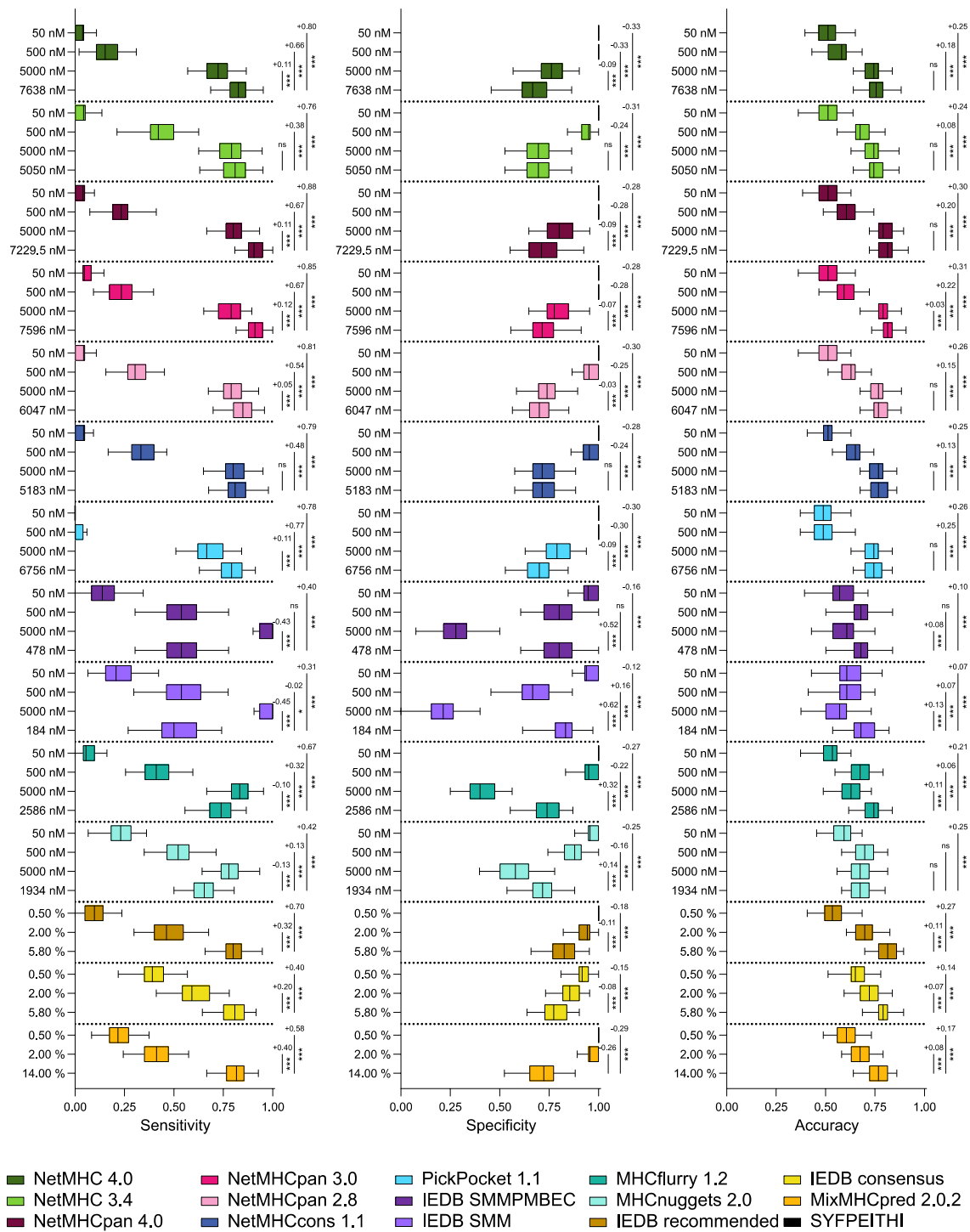


Supplementary Figure S4. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended or general thresholds.

(Figure continues, see Figure legend on page 174)

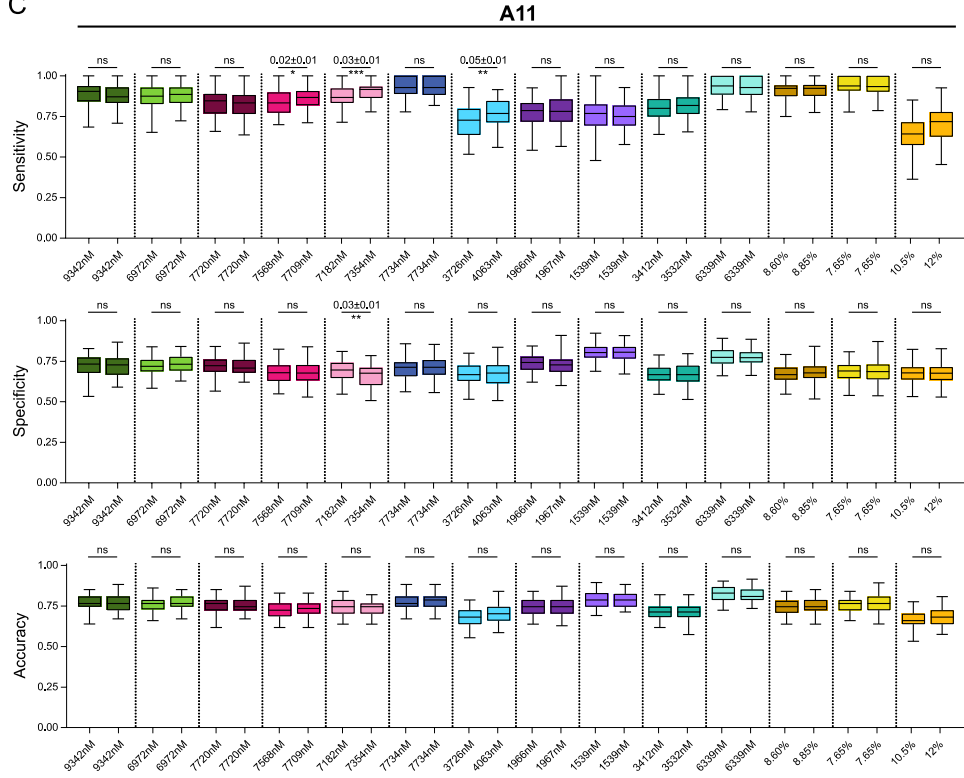
E

B15

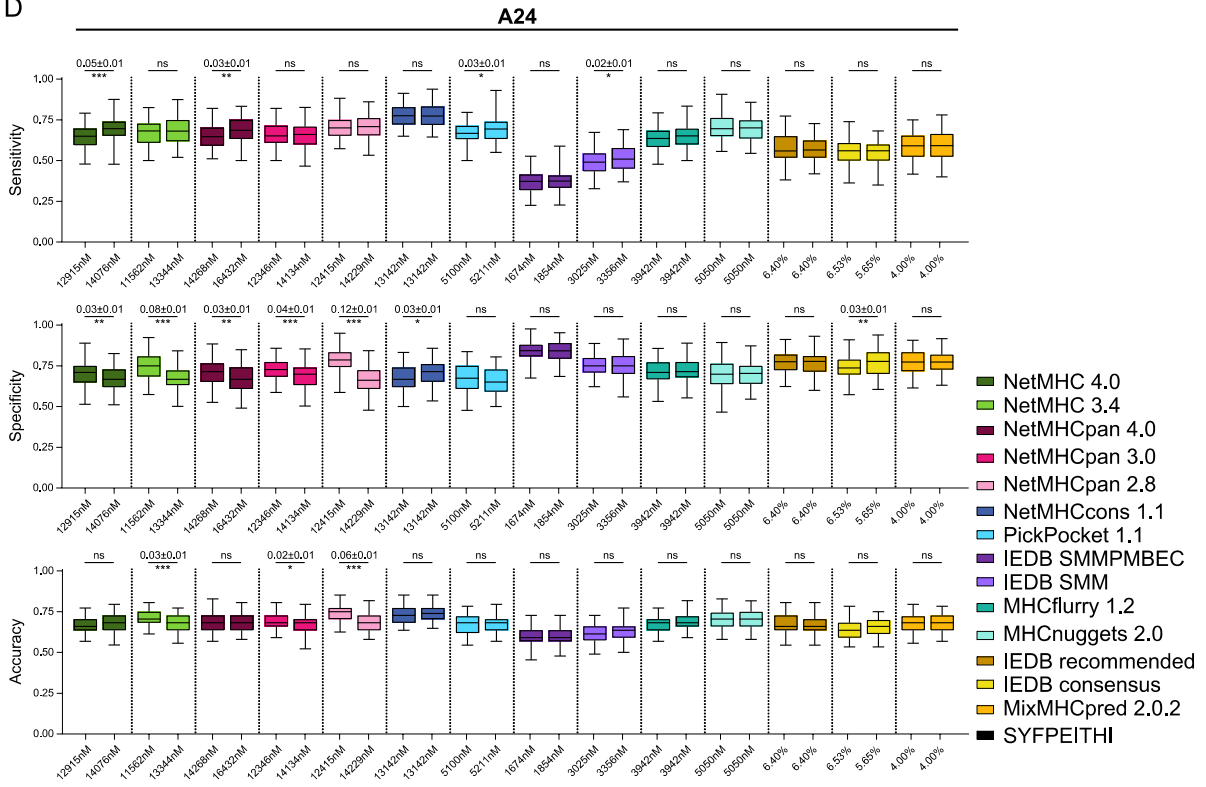


Supplementary Figure S4. Bootstrapping-based comparison of sensitivity, specificity and accuracy of predictors by applying individually recommended or general thresholds. Results are shown for (A) HLA A1, (B) HLA A3, (C) HLA A11, (D) HLA B7 and (E) HLA B15. Recommended thresholds were calculated and validated by bootstrapping as described. In a second bootstrapping sensitivity, specificity and accuracy of predictors applying recommended thresholds were compared to general thresholds for predicting high ($IC_{50} \leq 50\text{nM}$ or percentile rank (%) ≤ 0.5), intermediate (inter, $IC_{50} \leq 500\text{nM}$ or percentile rank (%) ≤ 2) or low ($IC_{50} \leq 5000\text{nM}$) binding likelihood. Box plots and whiskers show bootstrapping quartiles and the 95% interval of data, respectively. Significant differences of means were determined by one-way ANOVA followed by Dunnett multiple comparisons test (significance, $p < 0.05$). (***) $p < 0.001$, (**) $p < 0.01$, (*) $p < 0.05$, (ns) not significant

C



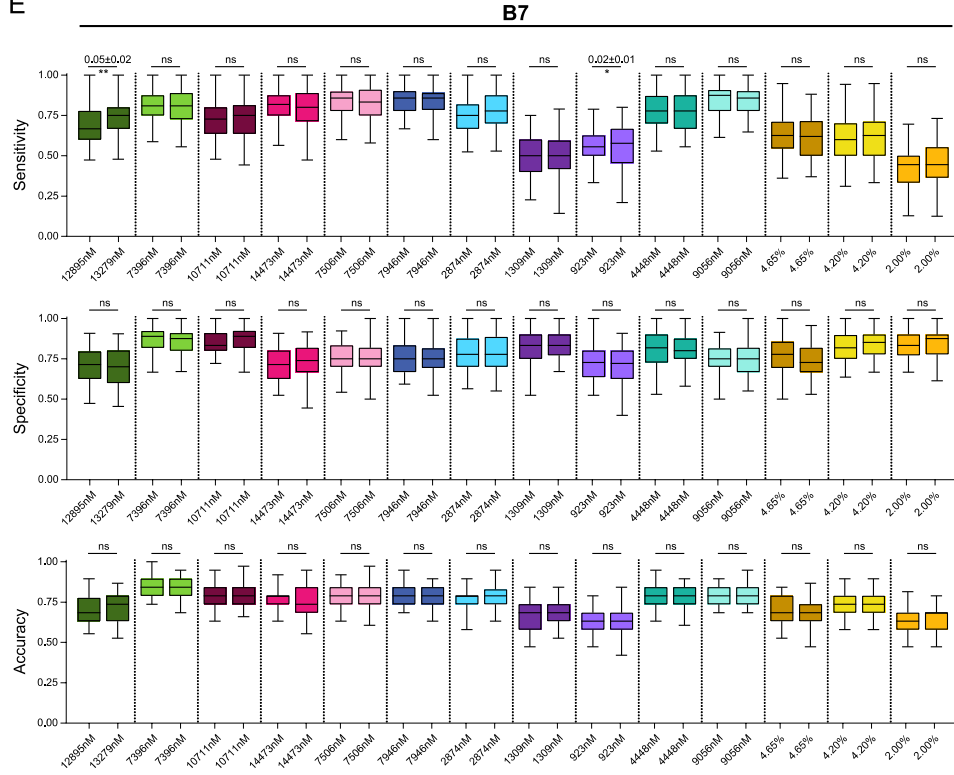
D



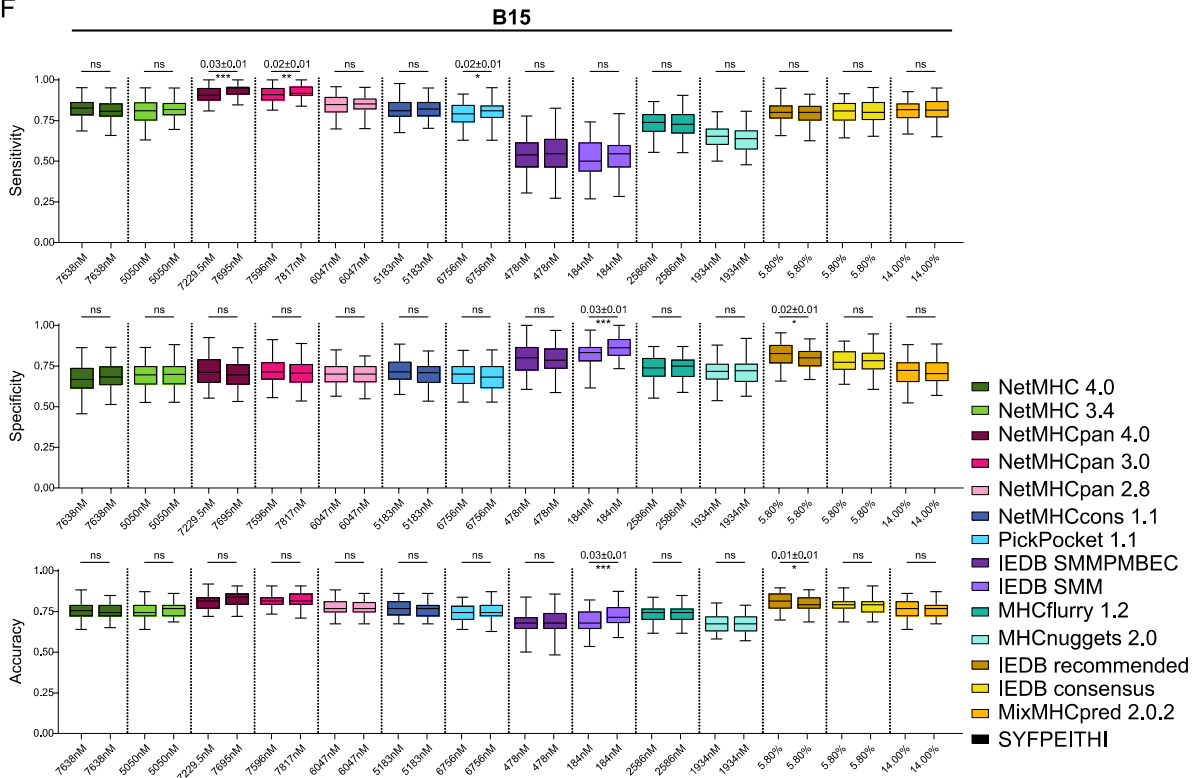
Supplementary Figure S5. Comparison of predictor performance applying criteria-based thresholds and recommended thresholds.

(Figure continues, see Figure legend on page 177)

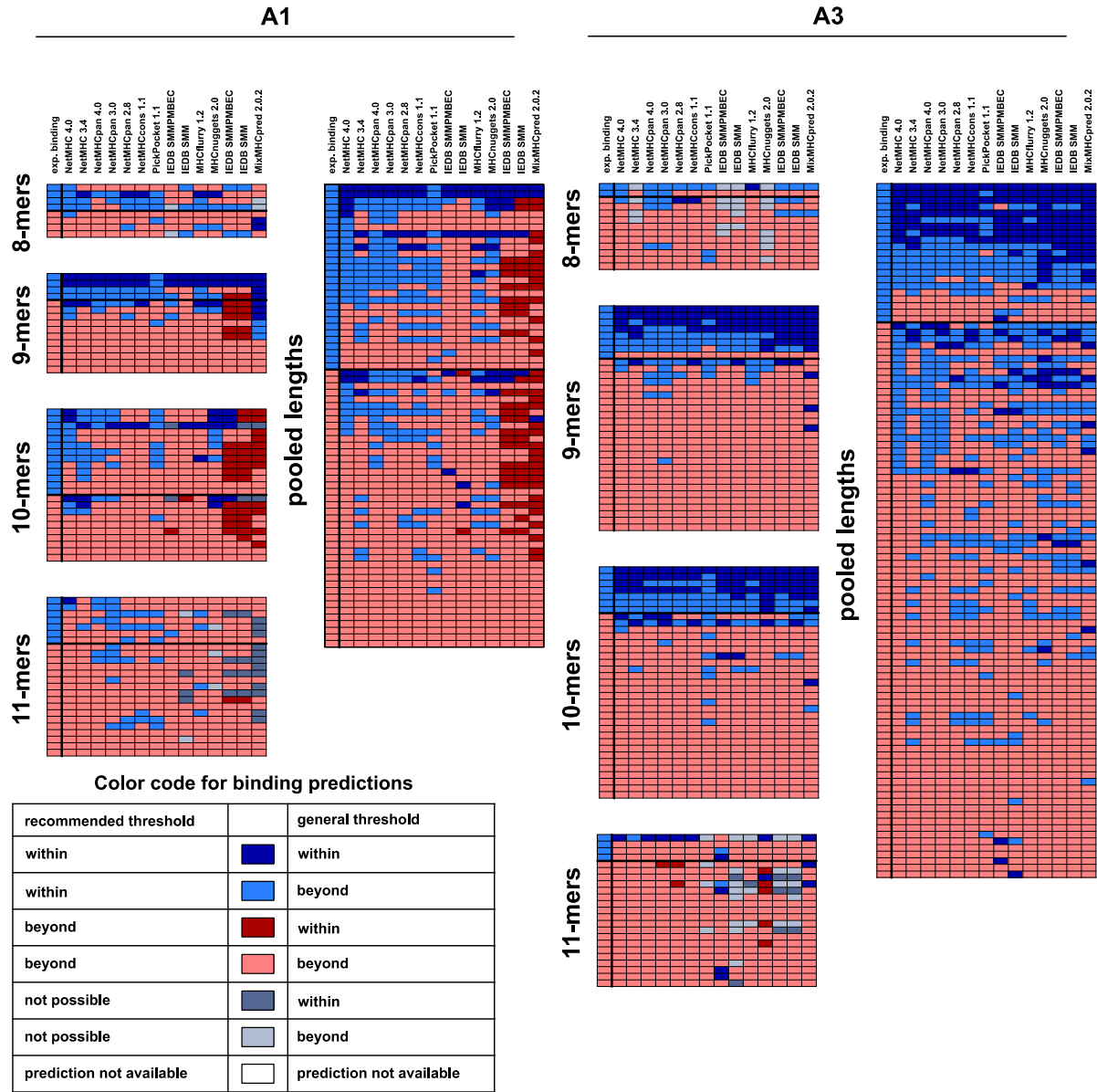
E



F

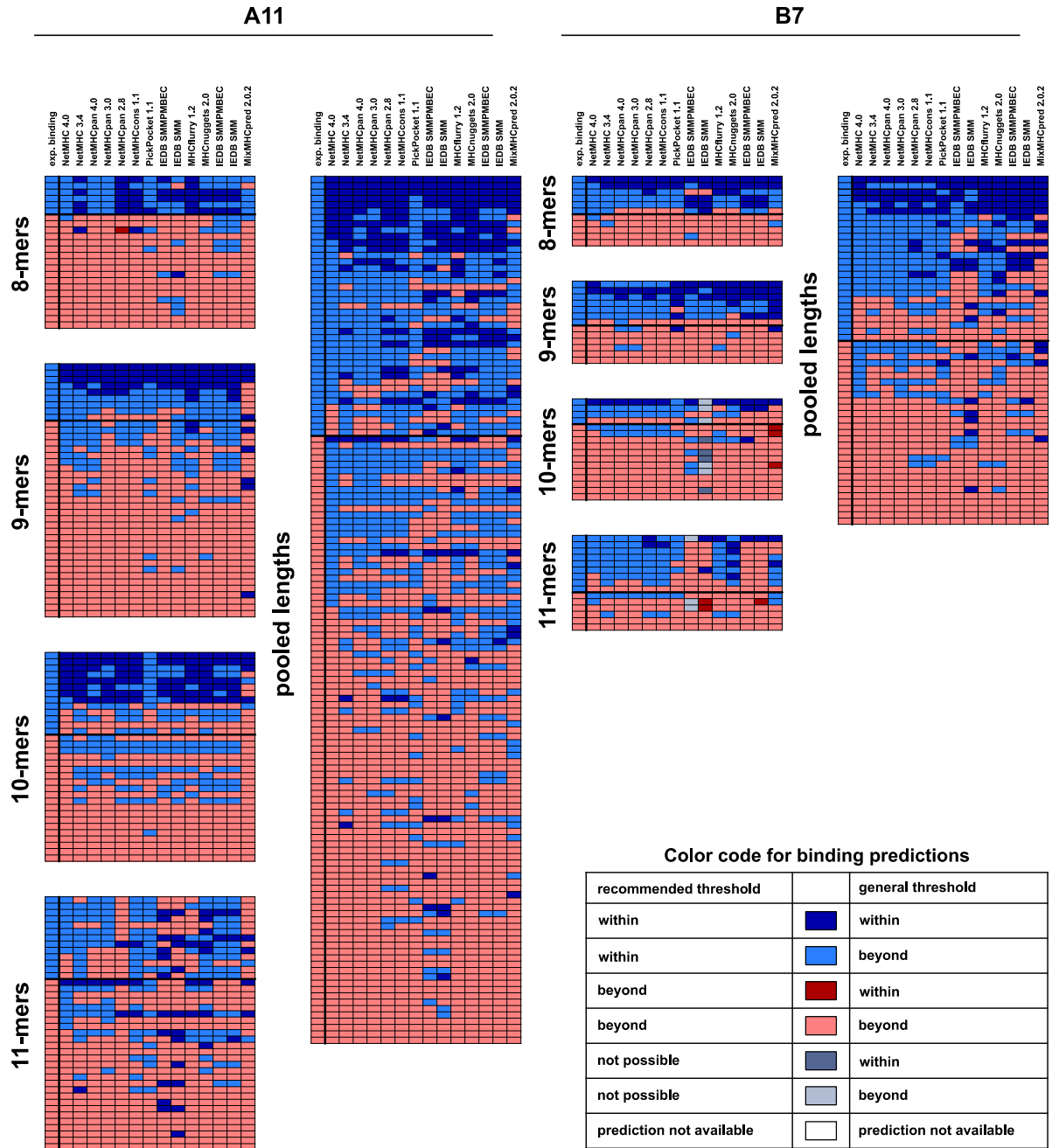


Supplementary Figure S5. Comparison of predictor performance applying criteria-based thresholds and recommended thresholds. Results are shown for HLA (A) A1, (B) A3, (C) A11, (D) A24, (E) B7 and (F) B15. Recommended thresholds (left) were calculated by bootstrapping whereas criteria-based thresholds (right) were calculated by applying the defined optimal threshold criteria to the respective peptide set as described. In a second bootstrapping the two thresholds were applied and the confidence intervals of sensitivity, specificity and accuracy were calculated for 100 samplings. Box plots and whiskers show bootstrapping quartiles and the 95% interval of data, respectively. Significant differences of means were determined using Student's t test (significance, $p < 0.05$). (***) $p < 0.001$, (**) $p < 0.01$, (*) $p < 0.05$, (ns) not significant.



Supplementary Figure S6. Classification of HLA binding prediction of HPV16 E6/E7 peptides to HLA A1, A3, A11, B7 and B15 according to application of different thresholds.

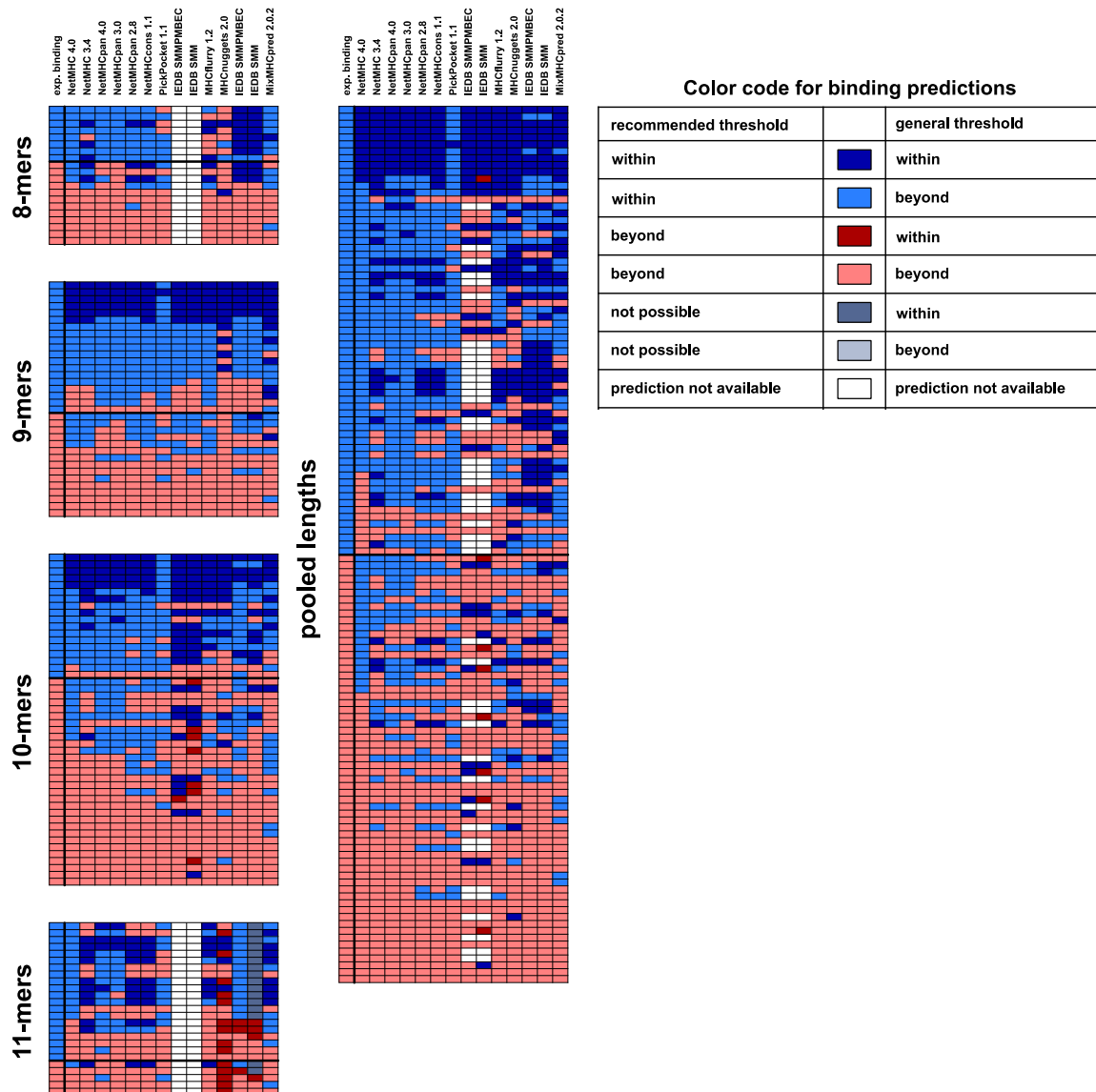
(Figure continues, see Figure legend on page 180)



Supplementary Figure S6. Classification of HLA binding prediction of HPV16 E6/E7 peptides to HLA A1, A3, A11, B7 and B15 according to application of different thresholds.

(Figure continues, see Figure legend on page 180)

B15



Supplementary Figure S6. Classification of HLA binding prediction of HPV16 E6/E7 peptides to HLA A1, A3, A11, B7 and B15 according to application of different thresholds. HLA-ligands derived from HPV16 E6/E7 were validated by experimental assessment (first column) and categorized into binders (blue) and nonbinders (red). Following columns indicate the predicted binding likelihood of peptides classified by different thresholds: predicted within (blue) or beyond (red) individual criteria-based (single peptide lengths) or recommended threshold (pooled lengths) or predicted within general threshold of $IC_{50} \leq 500nM$ or percentile rank ≤ 2 (dark shade of blue, red or grey). Grey fields indicate if calculation of a recommended threshold was not possible. In this case the binding likelihood of a binder resulting in lowest FPR and highest TPR was applied as threshold. If a prediction was not available, the field was left blank.

Supplementary Table S4. Table of HPV16⁺ cell lines with amino acid changes compared to the reference sequence.

Cell line ID	aa changes in E6	aa changes in E7	Nucleotide changes within E6 ORF	Nucleotide changes within E7 ORF
UM-SCC-104	L90V		T350G (L90V)	
CaSki	R17T; L90V		A131G (R17T), T350G (L90V)	
SCC090	R17T; L90V		A131G (R17T), T350G (L90V)	
SCC152	R17T; L90V		A131G (R17T), T350G (L90V)	
866	R17I; Q21D; H85Y	N29S	T109C, G132T (R170I), C143G (Q21D in combi), G145T (Q21D in combi), T256A, T286A, A289G, C335T (H85Y), A403G	A647G (N29S), C765T, T789C, T795G
UM-SCC-47	R17I; Q21D; E36Q; A68G; H85Y		T109C, G132T (R17I), C143G (Q21D in combi), G145T (Q21D in combi), G188C (E36Q), C285G (A68G), T286A, A289G, C335T (H85Y), A403G	A647G (N29S), T789C, T795G
SCC154	Q21D; H85Y; L90V		G145T (Q21D), T286A, A288G, C335T (H85Y), T350G (L90V), A532G	T732C, T789C, T795G
SNU-1299	Q21D; H85Y; L90V		G145T (Q21D), T286A, A289G, C335T (H85Y), T350G (L90V), A532G	T732C, T789C, T795G
SiHa	L90V; E120D	L28F	T350G (L90V), A442C (E120D)	A645C (L28F)
C66#3	L90V		C256T, T350G (L90V)	
C66#7	L90V		C265T, T350G (L90V)	
MRI-H-196	L90V		T350G (L90V)	
W12 20861	L90V		C256T, T350G (L90V)	
W12 20863	L90V		C256T, T350G (L90V)	
SNU-1000	D32E; I34R	N29S	T178G (D32E), T183G (I34R)	A647G (N29S), T846C
SNU-1005	D32E	N29S	T178G (D32E)	A647G (N29S), T846C
SNU-17	D32E	N29S	T178G (D32E)	A647G (N29S), T828C, T846C
SNU-703	D32E	N29S	T178G (D32E)	A647G (N29S), T846C
SNU-902	D32E	N29S; S63F	T178G (D32E)	A647G (N29S), C749T (S63F), T846C
UD-SCC2		H51N		C712A (H51N)

E6 and E7 sequences of HPV16⁺ cell lines in our cell bank were compared to the reference sequence to determine presence of amino acid (aa) changes. Nucleotide changes, which did not result in silent mutation, were amended with the respective aa change. The E6 and E7 sequences of the cell lines MRI-H-186, 879, 915, 93VU147T, FK16A, Goerke, HPK 1A and Marqu are identical to the reference sequence and thus not listed in the table.

Supplementary Table S5. Sequence alignment of amplified HPV16 E6 and E7 sequences with reference sequences.

MRI-H-186 E6 sequence vs. E6 reference sequence			
Query 1	TTTTATGCACCAAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTT	60	
Ref 79	TTTTATGCACCAAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTT	138	
Query 61	ACCACAGTTATGCACAGAGCTGCAAACAACATAACATGATATAATATTAGAATGTGTGTA	120	
Ref 139	ACCACAGTTATGCACAGAGCTGCAAACAACATAACATGATATAATATTAGAATGTGTGTA	198	
Query 121	CTGCAAGCAACAGTTACTGCGACGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCAT	180	
Ref 199	CTGCAAGCAACAGTTACTGCGACGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCAT	258	
Query 181	AGTATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTATTCTAA	240	
Ref 259	AGTATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTATTCTAA	318	
Query 241	AATTAGTGAGTATAGACATTATTGTTATAGTTTGTATGGAACAACATTAGAACAGCAATA	300	
Ref 319	AATTAGTGAGTATAGACATTATTGTTATAGTTTGTATGGAACAACATTAGAACAGCAATA	378	
Query 301	CAACAAACCGTTGTGTGATTTGTTAATTAGGTGTATTAAGTGTCAAAAGCCACTGTGTCC	360	
Ref 379	CAACAAACCGTTGTGTGATTTGTTAATTAGGTGTATTAAGTGTCAAAAGCCACTGTGTCC	438	
Query 361	TGAAGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAATATAAGGGGTCGGTG	420	
Ref 439	TGAAGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAATATAAGGGGTCGGTG	498	
Query 421	GACCGGTCGATGTATGTC	438	
Ref 499	GACCGGTCGATGTATGTC	516	
Query 439	TTGTTGCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCATGGAGATAC	497	
Ref 517	TTGTTGCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCATGGAGATAC	575	
MRI-H-186 E7 sequence vs. E7 reference sequence			
Query 4	AGAAACCCAGCTGTAATCATGCATGGAGATACACCTACATTGCATGAATATATGTTAGAT	63	
Ref 544	AGAAACCCAGCTGTAATCATGCATGGAGATACACCTACATTGCATGAATATATGTTAGAT	603	
Query 64	TTGCAACCAGAGACAACCTGATCTCTACTGTTATGAGCAATTAATGACAGCTCAGAGGAG	123	
Ref 604	TTGCAACCAGAGACAACCTGATCTCTACTGTTATGAGCAATTAATGACAGCTCAGAGGAG	663	
Query 124	GAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCATTACAATATT	183	
Ref 664	GAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCATTACAATATT	723	
Query 184	GTAACCTTTTGTGCAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTA	243	
Ref 724	GTAACCTTTTGTGCAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTA	783	
Query 244	GACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGT	303	
Ref 784	GACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGT	843	
Query 304	TCTCAGAAACCATAATCTACCATGGCTGATCCTGCAGGTACCAATGGGGAAGAGGGTACG	363	
Ref 844	TCTCAGAAACCATAATCTACCATGGCTGATCCTGCAGGTACCAATGGGGAAGAGGGTACG	903	

UM-SCC104 E6 sequence vs. E6 reference sequence				
Query	8	ATGCACCAAAGAGAAGCTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTACC	67	
Ref	83	ATGCACCAAAGAGAAGCTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTACC	141	
Query	68	ACAGTTATGCACAGAGCTGCAAACAACACTATACATGATATAATATTAGAATGTGTGTACTG	127	
Ref	142	ACAGTTATGCACAGAGCTGCAAACAACACTATACATGATATAATATTAGAATGTGTGTACTG	203	
Query	128	CAAGCAACAGTTACTGCGACGTGAGGTATATGACTTTTGCTTTTCGGGATTTATGCATAGT	187	
Ref	202	CAAGCAACAGTTACTGCGACGTGAGGTATATGACTTTTGCTTTTCGGGATTTATGCATAGT	261	
Query	188	ATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTATTCTAAAAAT	247	
Ref	262	ATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTATTCTAAAAAT	321	
Query	248	TAGTGAGTATAGACATTATTGTTATAGT G TGTATGGAACAACATTAGAACAGCAATACAA	307	
Ref	322	TAGTGAGTATAGACATTATTGTTATAGT T TGTATGGAACAACATTAGAACAGCAATACAA	381	
Query	308	CAAACCGTTGTGTGATTTGTTAATTAGGTGTATTAAGTGTCAAAGCCACTGTGTCTCTGA	367	
Ref	382	CAAACCGTTGTGTGATTTGTTAATTAGGTGTATTAAGTGTCAAAGCCACTGTGTCTCTGA	441	
Query	368	AGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAATATAAGGGGTCGGTGGAC	427	
Ref	442	AGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAATATAAGGGGTCGGTGGAC	501	
Query	428	CGGTCGATGTATGTCTTGTTCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATC	487	
Ref	502	CGGTCGATGTATGTCTTGTTCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATC	561	
UM-SCC104 E7 sequence vs. E7 reference sequence				
Query	8	<u>GATGTATGTCTTGTTCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCA</u>	67	
Ref	507	<u>GATGTATGTCTTGTTCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCA</u>	566	
Query	68	TGGAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATCT	127	
Ref	567	TGGAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATCT	626	
Query	128	CTACTGTTATGAGCAATTAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGC	187	
Ref	627	CTACTGTTATGAGCAATTAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGC	686	
Query	188	TGGACAAGCAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTTCAGAGTGTGA	247	
Ref	687	TGGACAAGCAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTTCAGAGTGTGA	746	
Query	248	CTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTTCGTACTTTGGAAGACCT	307	
Ref	747	CTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTTCGTACTTTGGAAGACCT	806	
Query	308	GTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAATCTACCAT	367	
Ref	807	GTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAATCTACCAT	866	

grey: sequence beyond ORF

underlined: overlap between E6 and E7 ORFs

bold red: sequence mismatch

NCBI reference sequence NC_001526.4

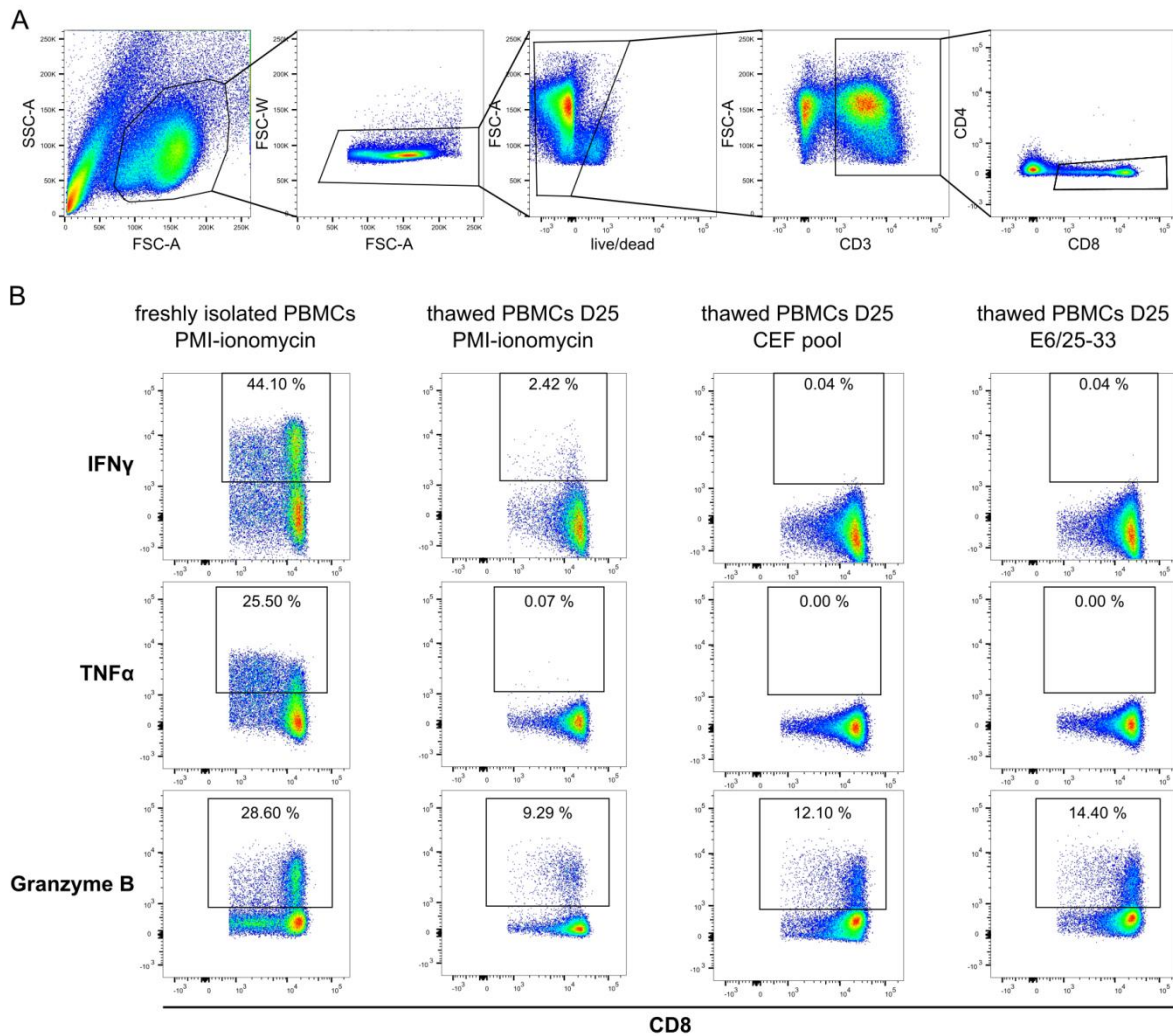
Supplementary Table S6. List of donors with positive IFN γ -responses in ELISpots assays

Donor	Epitope	SI	SFU/1x10 ⁶ cells	A1	A2	A3	A11	A24	B7	B15
BC4	E6/37-46	3.56	201.67			x	x			
	E6/67-75	3.56	202.00				x			
	E6/59-67	3.60	204.00			x	x			x
	E7/77-86	3.88	220.00		x					
	E6/48-55	4.32	245.00				x			
	E7/78-86	4.35	246.67		x					
	E6/68-75	4.44	251.67			x	x			
	E7/89-97	4.88	276.67			x	x			
	E6/52-60	5.21	295.00		x					
	E6/34-44	5.44	308.33		x					
	E7/88-97	5.91	335.00			x	x			
	E6/34-41	11.18	633.33				x			
	E6/68-77	11.44	648.33		x	x	x			x
	E7/87-97	14.62	828.33				x			
	E6/93-101	14.65	830.00			x	x			
	E7/50-60	18.79	1065.00				x			
	E6/106-115	23.32	1321.67			x	x			
E7/53-60	27.35	1550.00				x				
E6/52-62	27.59	1563.33				x				
E7/52-60	34.12	1933.33				x				
BC7	E7/67-76	16.63	221.67					x		x
	E6/107-115	19.25	256.67			x	x			
BC9	E6/134-144	6.75	225.00						x	x
	E6/85-95	6.77	225.56					x		
	E6/98-107	7.47	248.89					x		
	E6/87-95	18.40	613.33					x		
	E6/49-57	19.43	647.78					x		
BC14	E6/53-61	2.01	315.71							x
	E6/97-106	2.06	323.57							x
	E7/66-74	2.35	370.00		x					
	E7/81-91	2.35	370.00		x					
	E7/82-91	2.37	373.57		x					x
	E6/32-41	2.40	378.00				x			
	E6/94-101	2.46	386.67				x			
	E7/80-90	2.51	395.33		x					
	E6/69-79	2.73	429.00				x			
	E7/12-19	2.77	435.38		x					
	E6/68-77	2.78	436.67		x	x	x			x
	E6/73-83	2.81	442.86							x
	E6/41-50	2.82	443.33							x
	E6/68-76	2.83	446.00				x			x
	E6/44-54	2.85	448.00					x	x	x
	E6/139-148	3.00	472.00				x			
	E7/89-97	3.26	512.86			x	x			
	E7/88-97	3.63	570.67			x	x			
	E7/12-20	3.65	574.00		x					
	E6/57-67	4.04	635.00							x
E6/89-99	4.15	653.00			x	x			x	
E7/87-97	4.62	726.67				x				
E7/82-89	7.67	1206.00						x	x	
D01	E7/11-21	13.29	349.62		x					
D05	E7/11-20	3.34	529.69		x					
D06	E7/81-90	2.25	833.33		x					
	E7/80-90	10.28	548.33		x					
	E7/12-19	183.00	1830.00		x					
	E7/11-19	1348.38	17978.33		x					
D12	E7/11-19	58.92	1963.81		x					
	E7/12-19	64.50	716.64		x					
D18	E7/88-97	3.02	372.37			x	x			
D21	E7/11-19	1003.65	7232.97		x					
D24	E6/75-83	2.03	333.35			x		x		x
	E6/106-115	2.35	230.01			x	x			
	E6/129-138	2.54	960.05			x				
	E6/107-115	2.59	553.36			x	x			
	E6/68-77	6.20	385.57		x	x	x			x
D25	E7/11-18	6.55	218.33		x					
	E6/25-33	12.00	240.00		x					
	E6 H85Y, L90V/81-90	12.42	248.33		x					
	E6 Q21D/18-26	15.17	303.33		x					
D30	E6/109-119	1.99	243.85			x				
D32	E7/77-87	7.88	315.00		x					
D33	E6 D32E/29-38	2.30	242.95		x					

Annex

Donor	Epitope	SI	SFU/1x10 ⁶ cells	A1	A2	A3	A11	A24	B7	B15
	E7/11-20	11.86	312.63		x					
D35	E6/68-77	2.56	269.22	x		x	x			x
	E6/72-80	2.63	565.36			x		x		
	E6/8-18	2.10	699.98			x				
D37	E6/68-77	2.34	352.62	x		x	x			x
	E6/93-101	2.54	373.67			x	x			
	E6 L90V/84-94	2.88	242.10			x				
	E6/107-115	2.11	364.72			x	x			
D43	E6/106-115	2.17	382.36			x	x			
	E6/68-77	2.20	741.20	x		x	x			x
	E6/107-115	1.98	216.67			x	x			
D44	E6/93-101	2.13	233.34			x	x			
	E6/106-115	2.40	297.63			x	x			
	E6/72-80	5.41	1159.55			x		x		
	E7/85-93	2.13	971.72		x					
	E7/77-87	2.51	514.44		x					
	E7/11-21	2.74	221.50		x					
	E6/37-46	2.84	1150.35			x	x			
D45	E6/18-28	3.13	1388.51		x					
	E7/80-90	3.53	554.93		x					
	E7/7-17	4.04	807.39		x					
	E7/11-19	4.11	352.49		x					
	E6/29-38	4.20	200.06		x					
	E6 D32E/29-38	4.38	583.51		x					
	E6/68-75	19.25	366.78			x	x			
	E7/77-87	2.21	773.33		x					
D46	E6 H85Y/81 90	2.40	208.33		x					
	E6/29-38	3.34	813.33		x					
	H85Y, L90V/81-90	4.09	831.67		x					
	E7/86-93	6.97	395.00		x					

X: marks the HLA alleles that can bind the epitope



Supplementary Figure S7. Flow cytometry analysis of intracellular cytokine production of freshly and thawed isolated PBMCs from healthy donors. Production of cytokines was analyzed by intracellular staining and flow cytometry measurement. (A) Gating strategy for analyzed CD8⁺ T cells is shown. (B) Presence of intracellular IFN γ , TNF α and granzyme B was assessed for (left to right): freshly isolated PBMCs stimulated with PMI/Ionomycin, thawed PBMCs of HLA-A2⁺ donor D25 stimulated with PMI/Ionomycin, CEF peptide pool, or the E6/25-33 epitope (which showed induction of IFN γ -response in this donor in a previous ELISpot assay). Of note: granzyme B is stored intracellularly and present in all conditions.

