Analysis of the Immunopeptidomes of Ependymomas and Oropharyngeal Squamous Cell Carcinomas by Mass Spectrometry

Analyse der Immunpeptidome von Ependymomen und Oropharyngealen Plattenepithelkarzinomen mittels Massenspektrometrie

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

M. Sc. Lena Mühlenbruch

aus Lohr am Main

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	02.11.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Stefan Stevanović
2. Berichterstatter:	Prof. Dr. Oliver Planz

Contents

Summary

Presentation of antigens by human leukocyte antigen (HLA) molecules on cell surfaces and their recognition through T cells is essential for the initiation of an adaptive immune response. The analysis of the entirety of HLA-presented antigens, referred to as immunopeptidome or HLA ligandome, plays an important role for the development of immunotherapeutic methods to treat cancer patients. The present work addresses the immunopeptidomic analyses in two different tumor entities.

Ependymomas are a heterogenous tumor entity of the central nervous system and a long-term disease control is often challenging. Oropharyngeal squamous cell carcinomas (OPSCCs) belong to the head and neck tumors. Besides smoking and alcohol consumption, infections with oncogenic human papilloma viruses (HPV) are an important risk factor for the development of OPSCCs. Due to a rising number of HPV infections, the incidence rate of OPSCCs is increasing.

HLA ligands were isolated from tumor tissue samples by immunoaffinity chromatography and were subsequently analyzed through liquid chromatography followed by tandem mass spectrometry (MS/MS). This method was applied to 22 ependymomas and 40 OPSCCs. In addition, the immunopeptidomic data of five tonsil tissue samples were collected and served as benign reference for the OPSCC cohort.

The MS-based analysis enabled the detection of 6,801 HLA class I and 3,855 HLA class II ligands in the ependymomas as well as of 25,228 HLA class I and 15,203 HLA class II-ligands in the OPSCCs. By detailed investigation of the immunopeptidomic datasets, potential tumor-associated antigens (TAAs) presented by different HLA allotypes were identified. This included 8 ependymoma TAAs of which the most frequent peptides derived from the source proteins CCDC180 and DNAH6. 15 OPSCC TAAs were identified, of which the most frequent peptides derived from the source proteins UBD and PKP1. Additionally, HPV-associated differences between the HLA ligandomes as well as between the RNA transcriptomes of HPV^+ and HPV^- OPSCCs were revealed. These differences partly concerned the selected TAAs indicating a benefit for the consideration of the HPV status during immunotherapeutic target selection.

The suitability of the selected TAAs as immunotherapeutic targets must be further evaluated in future studies. Still, the results illustrate that both tumor entities provide opportunities for immunotherapeutic approaches such as personalized or semi-personalized peptide vaccinations. Hence, by mapping the immunopeptidomes, a major step towards the development of immunotherapeutic strategies for the treatment of ependymoma and OPSCC patients was taken.

Zusammenfassung

Die Präsentation von Antigenen durch humane Leukozytenantigene (HLA) auf Zelloberflächen und deren Erkennung durch T-Zellen ist essenziell für die Initiierung von adaptiven Immunantworten. Die Analyse der Gesamtheit der HLA-präsentierten Antigene, bezeichnet als Immunpeptidom oder HLA-Ligandom, spielt eine bedeutende Rolle für die Entwicklung immuntherapeutischer Methoden zur Behandlung von Krebspatienten. Die vorliegende Studie thematisiert die immunpeptidomischen Analysen in zwei verschiedenen Tumorentitäten.

Bei Ependymomen handelt es sich um eine heterogene Tumorentität des zentralen Nervensystems. Eine langfristige Kontrolle dieser Erkrankung ist häufig schwierig. Oropharyngeale Plattenepithelkarzinome (OPSCCs) gehören zu den Kopf-Hals-Tumoren. Neben Rauchen und Alkoholkonsum, sind Infektionen mit onkogenen humanen Papillomaviren (HPV) ein bedeutender Risikofaktor für die Entwicklung von OPSCCs. Aufgrund eines Anstiegs an HPV-Infektionen, steigt die Inzidenzrate von OPSCCs.

HLA-Liganden wurden aus Tumorgewebeproben durch Immunaffinitätschromatographie isoliert und anschließend mithilfe von Flüssigkeitschromatographie gefolgt von Tandem-Massenspektrometrie (MS/MS) analysiert. Dieses Verfahren wurde auf 22 Ependymome sowie 40 OPSCCs angewendet. Zusätzlich wurden immunpeptidomische Daten von fünf Tonsillengewebeproben erhoben, die als benigne Referenz für die OPSCC-Kohorte dienten.

Die MS-basierte Analyse ermöglichte die Detektion von 6.801 HLA-Klasse-I- und 3.855 HLA-Klasse-II-Liganden in den Ependymomen sowie von 25.228 HLA-Klasse-I- und 15.203 HLA-Klasse-II-Liganden in den OPSCCs. Durch detaillierte Untersuchung der immunpeptidomischen Datensätze wurden potenzielle tumor-assoziierte Antigene (TAAs), die von verschiedenen HLA-Allotypen präsentiert wurden, identifiziert. Dies beinhaltete 8 Ependymom-TAAs, von denen die häufigsten aus den Quellproteinen CCDC180 und DNAH6 stammten. 15 OPSCC-TAAs wurden identifiziert, von denen die häufigsten aus den Quellproteinen UBD und DSC3 stammten. Außerdem wurden Unterschiede zwischen den HLA-Ligandomen sowie zwischen den RNA-Transkriptomen von HPV^{+} und HPV^{-} OPSCCs offengelegt. Diese Unterschiede betrafen zum Teil die ausgewählten TAAs und weisen darauf hin, dass die Beachtung des HPV Status bei der Auswahl immuntherapeutischer Ziele von Nutzen sein kann.

Die Eignung der ausgewählten TAAs als immuntherapeutische Ziele muss in zukünftigen Studien weiter überprüft werden. Dennoch verdeutlichen die Ergebnisse, dass beide Tumorentitäten Möglichkeiten für immuntherapeutische Ansätze, wie personalisierte oder semi-personalisierte Peptid-vakzinierungen, bieten. Mit der Charakterisierung der Immunpeptidome wurde daher ein wichtiger Schritt in Richtung der Entwicklung von immuntherapeutischen Strategien für die Behandlung von Ependymom- und OPSCC-Patienten vollbracht.

1 Introduction

1.1 The immune system

The immune system of an organism has the crucial functions to differentiate between self and non-self patterns, initiate a response against non-self and spare self subjects. An immune response is a complex functional interaction between several cellular and humoral elements as defense system against recognized foreign substances, which can derive from microbes, viruses or tissue damage [1]. The human immune system is subdivided into two main sections, the innate immunity and the adaptive immunity **(Figure 1)** [2]. These two branches are defined according to their reaction speed and specificity of the immune response after exposure to a non-self subject. The evolutionary oldest type of immunity is the innate immunity. It is quite universal and, for most organisms, the survival with only an innate immune system is possible. The adaptive immunity evolved later around 450 million years ago as additional defense mechanism solely in jawed vertebrates [3, 4].

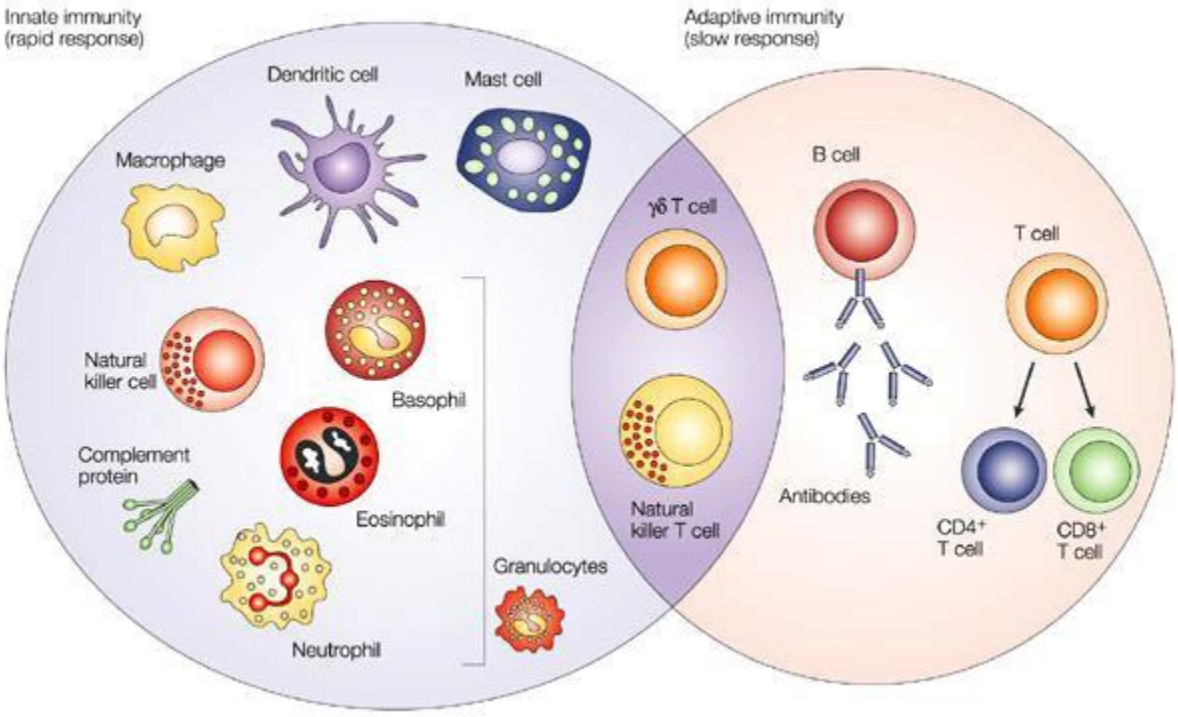


Figure 1: Cellular compartments of the innate and the adaptive immunity. Innate immune responses occur very fast within minutes or hours. The innate immune system is capable of recognizing pathogenic patterns and consists of dendritic cells, macrophages, mast cells, natural killer (NK) cells and granulocytes. γδ T cells and NK cells are the intermediate element between innate and adaptive immunity displaying lymphocyte characteristics without antigen specificity. Adaptive immune responses are more specific reactions that take days or weeks. Main players of the adaptive immune system are B and T cells. [2]

The innate immune system

The main characteristic of the innate immunity is the very rapid but rather unspecific reaction. The innate immune system consists of cellular and humoral elements, both including afferent (sensing) and efferent (effector) segments [3].

Receptors of the innate immune system were already described by Janeway (1989) [5]. These germline-encoded microbial sensors are today referred to as pattern recognition receptors (PRRs) and are expressed by dendritic cells (DCs), macrophages and neutrophilic granulocytes. They recognize invariant pathogenic molecules termed pathogen-associated molecular patterns (PAMPs). In addition, damage-associated molecular patterns (DAMPs) can induce PRR recognition and innate immune response activation in case of cell stress, which entails tissue damage or cell death [6]. Toll-like receptors (TLRs), nucleotide-binding oligomerization domain (NOD) -like receptors (NLRs), retinoic acid-inducible gene 1 (RIG-1) -like receptors (RLRs) and C-type lectin receptors (CLRs) constitute the four main PRR families [7]. PRRs can detect an extensive variety of PAMPs. For example, TLRs are able to sense lipids, lipoproteins, lipopolysaccharides, lipoteichoic acids, flagellin, glycolipids, β-glucans or nucleic acids derived from bacteria, viruses, parasites or fungi [8, 9]. Recognition of PAMPs through TLRs induces the activation of intracellular signaling pathways which leads to the release of inflammatory cytokines such as the tumor necrosis factor (TNF) α, type I interferon (IFN) or the interleukins IL-6, IL-1β or IL-12. Additionally, TLR signaling can upregulate co-stimulatory molecules on the surfaces of DCs, which is a crucial step for the activation of the adaptive immune system [10].

Further important components of the innate immune system are granulocytes, which are polymorphonuclear phagocytes and comprise neutrophils, eosinophils and basophils. They are capable of destructing pathogenic microbes and influence the inflammatory milieu. Furthermore, mast cells are important mediators of allergic responses [3]. γδ T cells and natural killer (NK) cells represent intermediate elements between innate and adaptive immunity due to their classification as lymphocytes without antigen specificity. γδ T cells express T cell receptors (TCRs) with adaptive characteristics on their surfaces and react *via* rapid, innate-like effector mechanisms during the immune responses [11]. NK cells react to pathogen recognition as highly cytotoxic immune effector cells by non-specifically killing their targets [12]. Moreover, the complement system displays an important non-cellular (humoral) component of the innate immune system [13].

The adaptive immune system

Macrophages, B cells and especially DCs serve as professional antigen presenting cells (APCs) that present peptides derived from digested pathogenic proteins by MHC (major histocompatibility complex) molecules on their cell surfaces. Thereby, they represent a link between the innate and the adaptive immunity [4]. The adaptive immune system consists of T lymphocytes, arising in the thymus,

and antibody-producing B lymphocytes, maturing in the bone marrow. These two cell types act as the effectors for cellular immune responses of the adaptive immune system [14]. Naïve T cells are divided into two classes, $CD8^+$ and $CD4^+$ T cells. Priming through interaction with an APC in the lymph nodes activates T cells and leads to their differentiation and clonal expansion. This activation requires three signals. First, the T cell receptors (TCRs) on the $CD4^+$ or $CD8^+$ T cell surfaces bind to MHC class I or class II molecules, respectively, which present peptides on the APC surfaces. This interaction is stabilized by either the CD4 or CD8 co-receptor. Second, a co-stimulatory signal is necessary for enhanced T cell survival and proliferation. This requires an interaction between surface molecules like CD80/86 on the APC and CD28 on the T cell surface or CD40 on the APC and CD40L on the T cell surface. The third signal is transduced *via* the release of cytokines, such as IL-4, IL-6, IL-12 or TGF-β by the APC and defines the definite T cell fate [15, 16]. $CD8^+$ T lymphocytes differentiate into cytotoxic T cells upon priming. They are capable of killing cells infected with intracellular pathogens and play an important role for the immune defense against viral infections [17]. $CD4^+$ T cells can differentiate into regulatory T (T_{reg}) cells or several kinds of T helper (T_H) cells. T_H cells are subdivided into different lineages depending on their characteristic expression of cytokines, transcription factors or homing receptors and the emerging immunological functions. They comprise the two main groups T_H1 and T_H2 cells as well as T_H9, T_H17, T_H22 and T follicular helper (T_{FH}) cells. T_H cells are essential for the direction and initiation of B cell responses resulting in the production of high-affinity antibodies [18]. These antibodies lead either to neutralization and opsonization of pathogens or to activation of the complement system against pathogens as part of the innate immune response [19]. In contrast to $CD8^+$ T and T_H cells, T_{reg} cells have regulatory functions and control immune responses in order to prevent autoimmune reactions [20]. Clonal expansion of T cells, contraction of the T cell population and immunological memory formation are the three phases of a T cell immune response in reaction to an acute infection [17]. Furthermore, the immunological memory is a crucial part of the adaptive immunity. B cells as well as $CD4^+$ and $CD8^+$ T cells can differentiate into long-lived memory cell populations, which allow more rapid and effective immune responses against previously exposed antigens. Recent publications showed that the development of an immunogenic memory is a very complex process and that both innate and antigen-inexperienced adaptive cells are involved in immunogenic memory responses in addition to memory cells [19, 21, 22].

1.2 The HLA system

Peptide presentation by MHC molecules on cell surfaces is essential for the initiation and regulation of an adaptive immune response. The MHC gene region exists in all jawed vertebrates where it encodes crucial components of the adaptive immune system [23]. The human MHC molecules are referred to

as human leukocyte antigen (HLA) molecules and display an enormous diversity of alleles in a variety of combinations. HLA molecules are members of the immunoglobulin gene family and are encoded on the short arm of chromosome 6 (6p21.3) in a 4,000 kb long gene region within the human genome [24, 25]. According to the current release of the IPD-IMGT/HLA Database (release 3.40.0, April 20th, 2020), the HLA gene region contains more than 25,000 allele sequences for 45 genes [26]. This includes 17.099 HLA class I and 6.695 HLA class II alleles as well as other encoded proteins with immunological functions. Among the latter are the hereditary hemochromatosis protein (HFE), the MHC class I polypeptide-related sequence A and B (MIC-A and MIC-B) as well as the antigen peptide transporters 1 and 2 (TAP1 and TAP2). The occurrence of different gene variants within a population is termed polymorphism. The HLA gene region is the most polymorphic one within the human genome. Polygeny describes the circumstance that several genes are involved in one phenotypic expression. The high diversity of the HLA gene region can be ascribed to the combination of polygeny and polymorphism [26–28].

HLA class I genes are subdivided into the classical HLA-A, -B and -C genes and the non-classical HLA-E, -F and -G genes. HLA class I molecules are expressed on the surfaces of all nucleated cells where they usually present antigen-derived peptides of intracellular origin to CD8$^+$ T cells. They are heterodimers assembled in the endoplasmic reticulum (ER) of a 43 kDa heavy α-chain and a non-covalently linked invariant 12 kDa β_2-microglobulin (β_2m) domain, which is encoded on the human chromosome 15 **(Figure 2A)**. The heavy chain consists of three α-domains. The α1- and α2-domains constitute the peptide binding cleft whereas the α3-domain serves as anchor into the cell membrane [29]. The HLA class I binding cleft is separated into six pockets (A-F) where antigenic peptides are bound and presented to T cells [24, 25]. Due to the closed characteristic of the binding cleft, the length of HLA class I-presented peptides is limited. They can have a length between 8 and 12 amino acids (AAs), among which 9mers prevail [30, 31]. The HLA binding clefts are very polymorphic, which is why different HLA allotypes prefer very restricted peptide motifs that they can bind and present to TCRs. Especially the second AA of a peptide interacting with pocket B and its carboxy (C) -terminal AA interacting with pocket F are the determinant anchor positions **(Figure 2B)** [32].

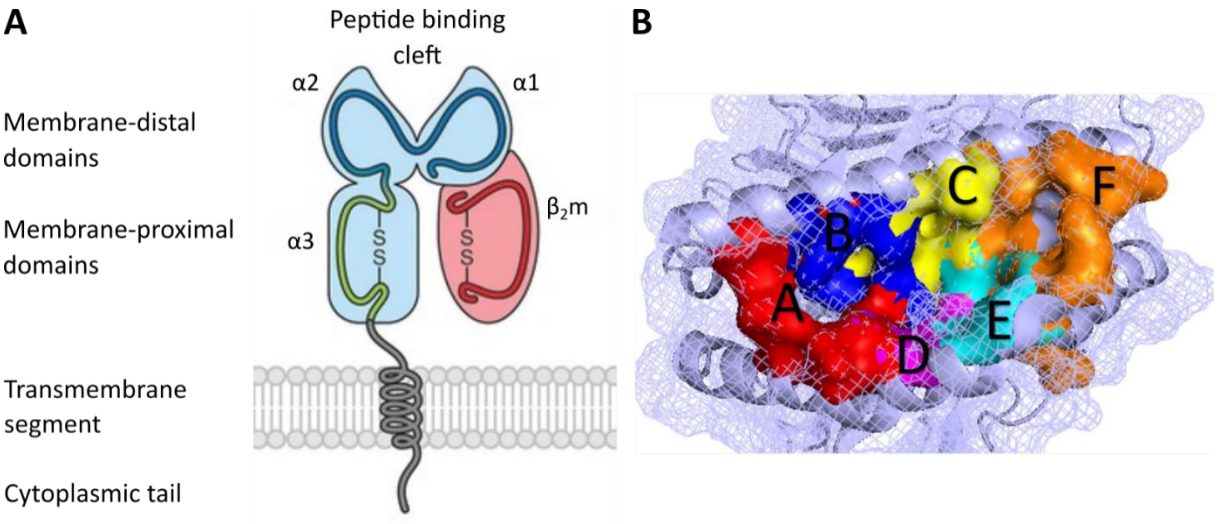


Figure 2: Structure of the HLA class I molecule. (A) HLA class I molecules are heterodimers comprised of a heavy α-chain and a non-covalently linked invariant 12 kDa β_2m domain. The α1- and α2-domains form the peptide binding cleft and the α3-domain serves as anchor into the cell membrane. Modified from [33]. (B) The HLA class I binding cleft is closed, which limits length of presented peptides to 8-12 AAs. It consists of six pockets (A-F) to which peptides derived from intracellular antigens are bound and presented to T cells. Modified from [34]. β_2m – β_2-microglobulin; AA – amino acid.

The HLA class II gene region consists of the HLA-DR, -DQ, -DP, -DM, and -DO genes. HLA class II molecules are expressed solely on the surfaces of professional APCs such as DCs, macrophages and B cells where they present peptides of extracellular origin to the TCRs of $CD4^+$ T cells. Their heterodimeric structure is composed of two homogenous polypeptides: one α-chain consisting of an α1- and an α2-domain and one β-chain consisting of a β1- and a β2-domain **(Figure 3A)**. The anchoring into the cell membrane is mediated *via* the transmembrane segments of the α2- and β2- domains. The α1- and β1-domains constitute the HLA class II binding cleft [33]. In contrast to HLA class I molecules, HLA class II molecules have an open binding cleft, which allows binding of peptides with a length between 8 and 25 AAs [24, 31]. The HLA class II binding cleft comprises several pockets, of which the five pockets P1, P4, P6, P7, P9 are responsible for peptide binding **(Figure 3B)**. Compared to that of HLA class I molecules, the HLA class II binding clef It is less constrained. This allows more promiscuous binding of peptides to different HLA class II allotypes [35, 36]. Peptide motif preferences of various HLA class I or class II allotypes are collected and accessible in databases such as SYFPEITHI [37] or NetMHC [38].

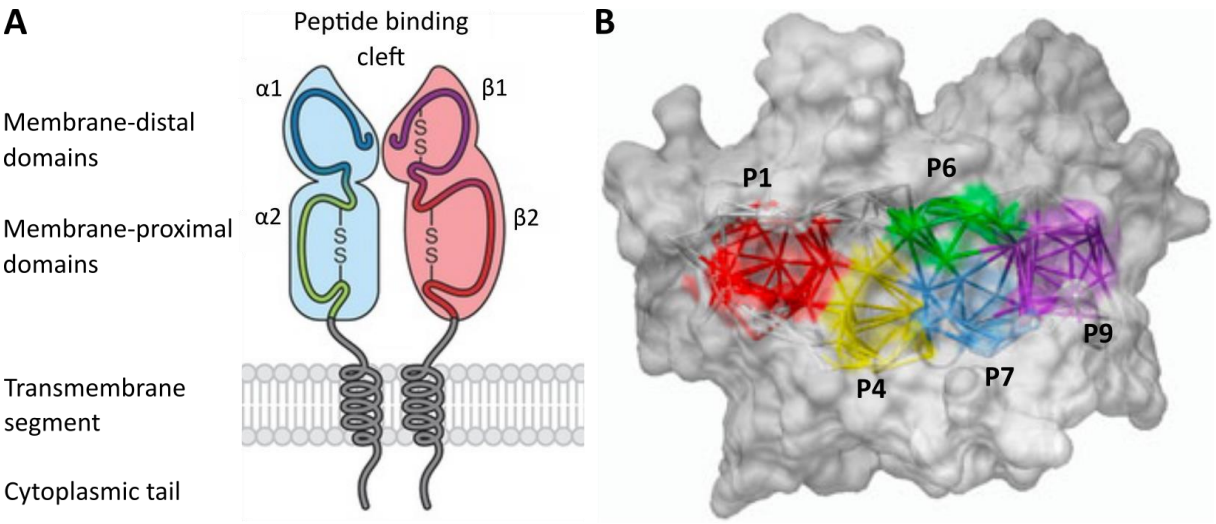


Figure 3: Structure of the HLA class II molecule. (A) The HLA class II heterodimer consists of an α- and a β-chain. The α1- and β1-domains form the peptide binding cleft whereas the α2- and β2-chains are anchored into the plasma membrane. Modified from [33]. (B) HLA class II molecules have an open binding cleft allowing presentation of longer peptides (8-25 AAs) than HLA class I molecules. They exhibit five pockets for peptide binding (P1, P4, P6, P7, P9). Modified from [36]. AA – amino acid.

1.2.1 HLA class I antigen processing

HLA class I molecules are expressed on the surfaces of all nucleated cells and platelets and present self or non-self peptides derived from intracellular cytosolic proteins. Their binding capacity is preferably restricted to 9mer peptides exhibiting hydrophobic or charged residues in the C-terminus and the second AA position [39]. These peptides are produced through a proteasome-mediated protein degradation mechanism termed antigen processing **(Figure 4)** [30].

HLA class I-presented peptides are generated from intracellular proteins by the 26S proteasome [40]. A ubiquitin modification code determines target proteins for the proteasome [41]. The 26S proteasome is a catalytic multi-protein complex, which consists of a 20S core barrel entailing protease activity and two 19S caps [40]. The C-terminus of a peptide is determined by the proteasome whereas the amino (N) -terminus can remain elongated after proteasomal protein degradation [42]. Subsequently, the encounter between HLA molecules and peptides occurs within the ER. The generated peptides diffuse through the cytosol to the TAP, which is located in the ER membrane and translocates the peptides into the ER lumen [43]. The N-terminal trimming of the peptides is performed by the ER-aminopeptidases ERAP1 and ERAP2 [44]. The peptide-loading complex (PLC) is compiled of the HLA class I molecule, the TAP as well as the chaperones tapasin, ERp57 and calreticulin [45]. These chaperones facilitate folding and stabilizing of the HLA class I molecules within the ER. Efficient loading of the peptide onto the HLA class I molecules is enabled by the PLC [46]. For degradation or another round of TAP translocation and possible HLA association, peptides and HLA molecules that are not conjoined in the ER are recycled and return to the cytosol *via* the ER-associated protein degradation

(ERAD) pathway [47]. After release of the peptide-HLA class I complexes from the ER, they are translocated through the Golgi to the plasma membrane, where they present loaded peptides to $CD8^+$ T cells [30].

HLA class I-presented antigens commonly derive from the degradation of intracellular proteins. A large proportion (30-70%) of newly synthesized proteins is degraded before they can develop cellular functions. These defective ribosomal proteins (DRiPs) serve as important source for HLA-presented peptides [48, 49]. DRiPs originate from defective transcription or translation processes and allow rapid presentation of antigens such as viral peptides [50, 51]. A rather uncommon source for HLA class I-presented peptides are exogenous antigens. *Via* cross presentation, extracellular antigens can be loaded onto HLA class I molecules and presented on cell surfaces for $CD8^+$ T cell activation. This mechanism appears in specialized subsets of DCs [52].

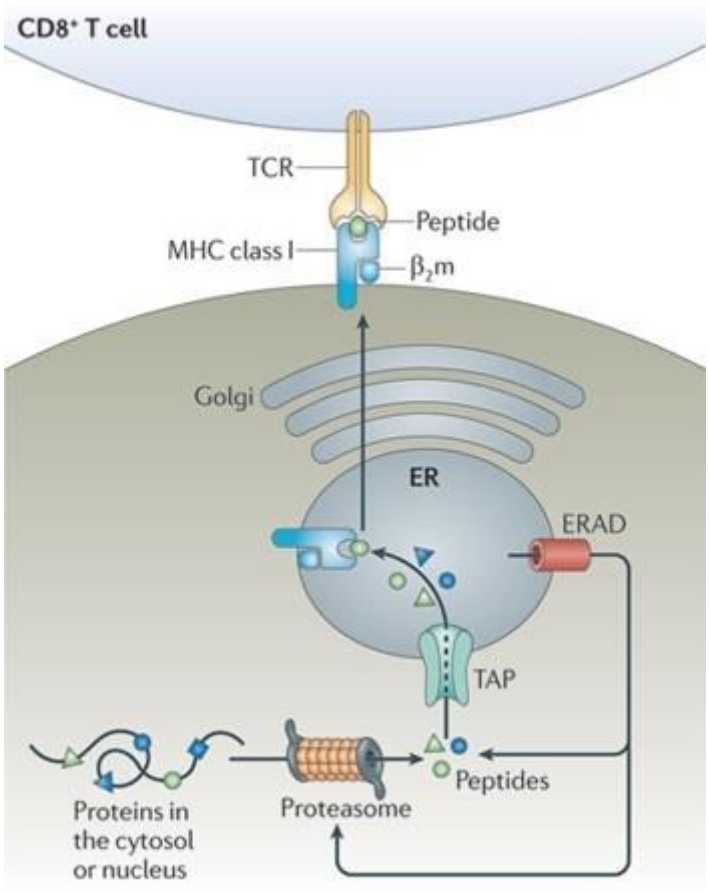


Figure 4: The HLA class I antigen presentation pathway. Proteins of intracellular origin are degraded by the 26S proteasome and translocated from the cytosol into the ER *via* TAP. Here, peptides are trimmed to their final length and loaded into the peptide binding cleft of HLA class I molecules. Peptide-HLA class I complexes are transported through the Golgi apparatus to the plasma membrane where they present their peptide cargo to the TCRs of $CD8^+$ T cells. HLA molecules and peptides that are not assembled in the ER are recycled through the ERAD pathway. TAP – transporter associated with antigen processing; ER – endoplasmic reticulum; MHC – major histo-compatibility complex; TCR – T cell receptor; ERAD – ER-associated protein degradation. [30]

1.2.2 HLA class II antigen processing

Professional APCs, mainly including DCs, macrophages and B cells, express HLA class II molecules on their cell surfaces. HLA class II molecule expression is primarily regulated by the MHC class II

transactivator (CIITA) [53]. Due to differences between the structures and cargo sources of HLA class I and HLA class II molecules, the processing mechanism of HLA class II-presented antigens differs compared to that of HLA class I-presented antigens **(Figure 5)** [30].

In the ER, the α- and β-chains of HLA class II are assembled and the HLA class II molecules are coupled to the invariant chain (Ii). In this Ii-HLA class II complex, the MHC class II-associated Ii peptide (CLIP) region of Ii blocks the HLA class II binding cleft and prevents peptide binding in the ER. Whereas single HLA class II chains remain within the ER, only HLA class II heterodimers can exit the ER. Directed by the di-leucine of Ii, the Ii-HLA class II complex translocates to the MHC class II compartment (MIIC) of the late endosome [54]. In the MIIC, Ii is degraded by the proteases cathepsin S and cathepsin L, leaving the CLIP fragment attached to the HLA class II binding cleft [55]. The HLA class II chaperone HLA-DM accelerates the substitution of CLIP with a peptide cargo derived from the endosomal degradation pathway [56, 57]. Influencing factors like cholesterol, cytosolic pH, kinases and GTPases regulate the release of peptide-HLA class II complexes from the MIIC. After the transport to the plasma membrane, the HLA class II molecules present their loaded peptides to $CD4^+$ T cells by an interaction with the TCR [30].

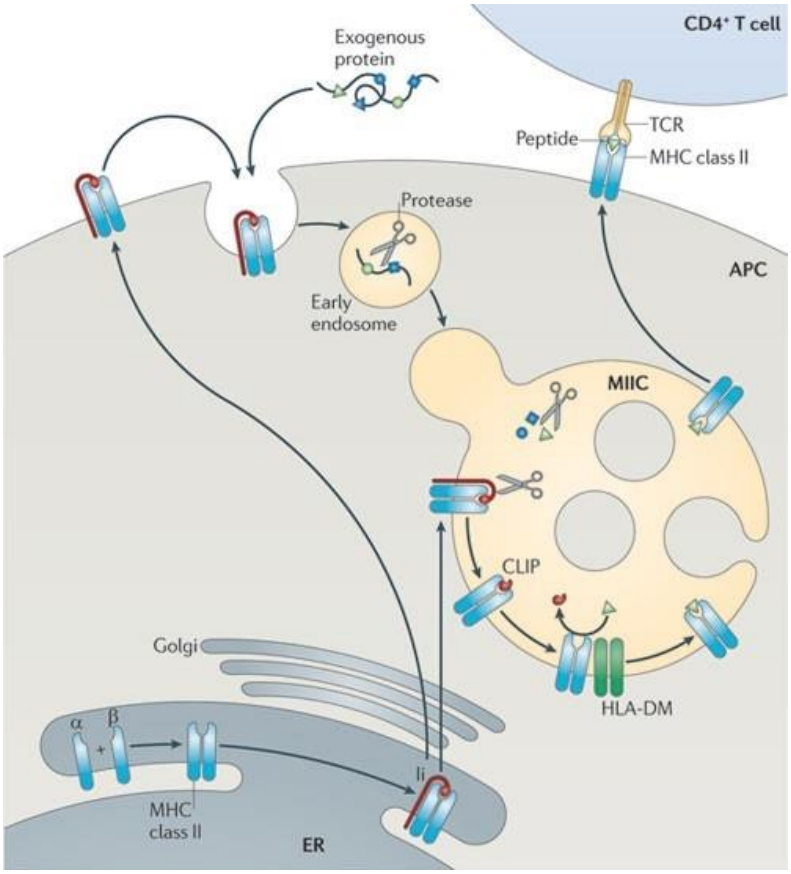


Figure 5: The HLA class II antigen presentation pathway. HLA class II expression is restricted to professional APCs. HLA class II chains and Ii are assembled in the ER. After transport of the Ii-HLA class II complex to MIIC, Ii is degraded leaving the CLIP fragment bound to the HLA binding cleft. With help of the chaperone HLA-DM, CLIP is exchanged for a peptide cargo. The peptide-HLA class II complex is translocated to the plasma membrane where interactions with TCRs of $CD4^+$ T cells take place. APC – antigen presenting cell; ER – endoplasmic reticulum; MHC – major histocompatibility complex; Ii – invariant chain; MIIC – MHC class II compartment; CLIP – MHC class II-associated Ii peptide; HLA – human leukocyte antigens; TCR – T cell receptor; ERAD – ER-associated protein degradation. [30]

1.3 Cancer immunity

Natural immune reactions play an essential role for the defense against cancer. Tumor immunosurveillance includes the identification and destruction of tumor cells by innate and adaptive immune reactions to suppress cancer development. Especially the elimination of cancer cells through T cell responses is of importance [58]. Cellular and genetic alterations occur during the transformation of a normal into a cancer cell. These changes have impact on the antigenic landscape of the cells. The presentation of tumor-specific antigens (TSAs) and tumor-associated antigens (TAAs) by HLA molecules on the tumor cell surfaces enables T cells to recognize cancer cells [59]. TSAs are antigens derived from proteins that are unique to tumor cells. This applies for example to neoantigens. Somatic mutations in protein-coding DNA sequences are the source of neoantigens that contain a mutated AA sequence [60]. In contrast, TAAs derive from unmutated proteins that are differentially HLA-presented on tumor cells compared to healthy tissue. They include cancer-testis antigens (CTAs), viral, overexpressed or differentiation antigens [61]. Among TAAs, CTAs are immunogenic proteins with a characteristic expression that is restricted to the human germ line and several distinct cancer entities [62]. Van der Bruggen *et al.* (1991) identified the first CTA, MAGE-A1 (melanoma-associated antigen 1) [63]. Several hundred additional CTAs have been revealed since [64]. Viral TAAs occur in tumors caused by a viral infection. 20% of the worldwide cancer burden can be ascribed to infectious agents. Proteins of oncogenic viruses can serve as source for antigens in infected tumor cells [65].Overexpressed TAAs derive from proteins that occur in benign cells but are overexpressed in tumor cells [59]. Differentiation TAAs derive from tissue-related proteins and were initially discovered in melanoma [66]. The release of antigens from cancer cells including TSAs and TAAs trigger the cancer-immunity cycle, which includes the recognition of a tumor by T cells and the initiation of an adaptive response **(Figure 6)**. The resulting destruction of cancer cells self-propagates the immune responses by an amplification of the antigen release [67].

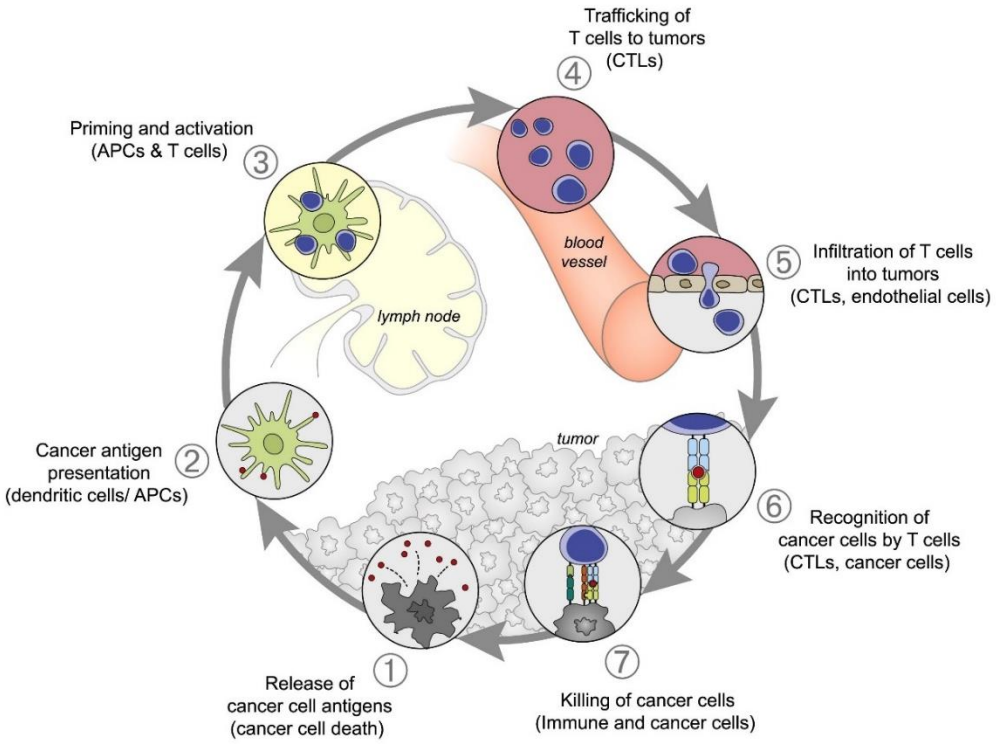


Figure 6: Cycle of cancer immunity. Cancer cells harbor antigens that can be used by APCs to prime T cells. Subsequent trafficking and tumor infiltration by T cells lead to the recognition and killing of tumor cells by the adaptive immune system. Destruction of cancer cells initiates further antigen release. The cancer-immunity cycle has self-propagating characteristics amplifying the immune reaction as well as regulatory features limiting the response. APC – antigen presenting cell; CTL – cytotoxic T lymphocyte. [67]

Avoiding the recognition and elimination of tumor cells through the immune system has been defined as one important hallmark of cancer [68]. Tumors possess a wide range of strategies to escape from immune recognition or to even use the immune system for cancer progression [69]. To avoid a recognition, tumor cells can interfere with the antigen processing machinery and HLA antigen presentation pathway of T cells [70, 71]. Another immunosuppressive mechanism is the capability of tumors to recruit T_{reg} cells that infiltrate the tumor tissue and have immunoregulatory functions [72, 73]. Additionally, many tumors are able to induce the expression of immunosuppressive cytokines such as transforming growth factor β (TGF-β) either in the tumor cells themselves or in cells of the tumor environment [74]. Besides immunosuppression, tumors are capable of inducing tolerance and immune deviation by the expression of inhibitory molecules like programmed death-ligand 1 (PD-L1). This results in anergy and deletion of cancer reactive immune cells [75]. Another strategy of tumor cells to escape from immune responses is to trigger cell death in cancer-specific T cells *via* apoptotic pathways [76].

The natural immune system is capable of recognizing tumor cells and inducing an anti-cancer response. However, most cancers develop various strategies to escape from the host immune responses. Restoration of the immune response against cancer is the main issue of immunotherapy development.

The capability and mechanisms of tumors to avoid immune reactions are suggested to have an impact on patient outcome and the effectiveness of immunotherapeutic strategies [77].

1.4 Cancer immunotherapy

Immunotherapy for cancer aims to activate the immune system of patients against tumor cells with the eventual goal to achieve an immunologic memory. Currently, there are three main approaches of T cell-based tumor immunotherapy: immune checkpoint therapy, adoptive cell transfer and cancer vaccination [78].

The application of immune checkpoint inhibitors (ICIs) is a promising approach in the research field of immunotherapy. Tumors are able to escape from immunosurveillance through the activation of co-inhibitory signaling pathways suppressing the elimination of tumor cells through the immune system. Disruption of this downregulation and promotion of anti-tumor immune responses can be achieved using ICIs [79]. The human monoclonal IgG1 antibody Ipilimumab was one of the earliest ICIs and was first approved in melanoma [80]. The cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) belongs to the immune checkpoint molecules and inhibits T cell responses by the downregulation of T cell activation pathways [81]. Ipilimumab blocks CTLA-4 to facilitate anti-cancer immune responses [82, 83]. The programmed cell death protein 1 (PD-1) is another target of ICIs. This receptor is expressed on the surfaces of activated T cells. Binding of the ligands PD-L1 or PD-L2 to PD-1 leads to a downregulation of T cell activity resulting in tumor immune evasion [81]. Pembrolizumab, a humanized monoclonal IgG4 antibody, and nivolumab, a human monoclonal IgG4 antibody, target PD-1 and interrupt PD-1-mediated signal transduction to restore T cell activity. Application of nivolumab and pembrolizumab have proven success as cancer therapy for example in non-small cell lung carcinoma (NSCLC) patients [84, 85].

The immunotherapeutic approach of adoptive cell transfer utilizes functionally modified T lymphocytes for the treatment of cancer patients. This implies either the selection of autologous tumor-infiltrating lymphocytes (TILs) or genetically engineering autologous T cells with TCRs or chimeric antigen receptors (CAR). Thereby, specific T cells are created that are able to recognize and attack tumor cells [86]. Therapy with CAR T cells demonstrated success in advanced leukemia patients [87]. However, undesirable on-target/off-tumor side effects were revealed in clinical studies and the selection of suitable target antigens remains as key challenge for developing CAR T cell therapy for solid tumors [88].

Cancer vaccination is another promising immunotherapeutic approach that aims to induce specific recognition and elimination of tumor cells by T cell immunity. During the development of a cancer vaccine, four important aspects must be precisely elaborated: the vaccine formulation, the selection

of tumor antigens, the usage of suitable immune adjuvants and of delivery vehicles. The combination of these components is necessary to assure an effective and specific action of the vaccine and to induce an activation of naïve T cells by APCs [78]. There are different approaches regarding the vaccine formulation and forms of target delivery. One possibility is the usage of nucleic acids. This includes DNA or RNA directly encoding the targets [89, 90]. Another option are cellular cancer vaccines, which are based on antigen-loaded or genetically engineered tumor cells, tumor cell lysates or DCs [91, 92]. Additionally, cancer vaccines on the basis of synthetic peptides of defined targets have proven successful for the induction of T cell reactivity against tumor tissue [93]. The major challenge for the development of cancer immunotherapeutic approaches is to avoid the occurrence of autoimmunity side effects. Therefore, an accurate selection of suitable target antigens is of substantial importance. Antigens that are differentially expressed and HLA-presented on tumor cells compared to normal cells should preferably be selected. TSAs and TAAs display promising targets for the use in cancer immunotherapeutic vaccination [94]. Another important aspect to improve the effect of a cancer vaccine is the usage of an effective adjuvant. An adjuvant must assure the antigen protection from degradation, the antigen uptake by APCs and the induction of APC activation [95]. For peptide-based vaccines, a commonly used adjuvant is an oil-in-water emulsion with Montanide ISA™51 (incomplete Freund's adjuvant analogue) [96]. Recently, XS15, a synthetic TLR1/2 ligand, was suggested as suitable adjuvant for peptide-based cancer vaccination [97]. To further assure the effectiveness of a cancer vaccine, a protecting delivery system is necessary. Delivery vehicles consist of emulsions, liposomes or virosomes. Often, the adjuvant agent simultaneously serves as delivery vehicle, which is for instance the case with Montanide [98].

Cancer immunotherapy has a vast impact on the evolvement of cancer therapy and is a rapidly changing research area. Novel approaches become more and more specialized. Personalized or semi-personalized immunotherapies have the goal to treat individual patients or patients of a small group with adjusted treatment methods based on unique genetic and molecular characteristics. This will enable the development of more precise and effective treatment options with less side-effects for cancer patients [99, 100].

1.5 Immunopeptidomics

The selection of suitable target antigens is essential for the development of effective cancer immunotherapeutic approaches. The entirety of HLA-presented antigens, termed the immunopeptidome or HLA ligandome, serves as basis for the search for target antigens. The analysis and characterization of the immunopeptidome is referred to as immunopeptidomics [101].

One possibility to identify HLA-presented antigens is their *in silico* prediction based on sequencing data. This approach is often used for the detection of neoantigens. By whole exome or transcriptome sequencing, somatic mutations of a tumor are identified. On this basis, potential immunogenic HLA epitopes are predicted according to HLA molecule binding restrictions for the usage in cancer vaccines [102]. However, the correlation between the immunopeptidome and the expression profile in a tissue is limited [103]. To access the collectivity of pathophysiologically relevant antigens naturally HLA-presented on tumor cell surfaces, mass spectrometric (MS) analysis of the immunopeptidome is still the method of choice. Mapping of the HLA ligandome includes isolation of HLA-presented peptides by immunoaffinity chromatography and analysis of isolated peptides *via* MS-based strategies. Subsequently, the resulting datasets comprised of the immunopeptidomes of the respective cancer tissues serve as basis for the precise selection of immunogenic targets [104]. The MS-based analysis method of the HLA ligandome is a very slow and work-intensive approach taking several days for target identification [105]. The applicability of sample processing, MS analysis and data interpretation tools are rapidly improving. Recently, a high-throughput method was presented enabling simultaneous processing of a high number of immunopeptidomic cell and tissue samples for MS analysis within a few hours [106].

Presently, both *in silico* prediction- and MS-based immunopeptidomic methods turned out to be essential tools for developing novel immunotherapeutic options to treat cancer patients. A combinatory treatment, which applied a vaccination with MS-analyzed unmutated TAAs followed by a vaccination with predicted neoantigens, resulted in the successful induction of immune responses in glioblastoma patients in the phase I trial GAPVAC-101 [107]. Continuous progression in this research field will further facilitate, improve and expand the application of immunopeptidomics.

1.6 Aim of the study

In the present work, MS-based immunopeptidomic analysis methods are applied to tumor tissues with the aim to characterize the HLA ligandomes of two different tumor entities: ependymomas and oropharyngeal squamous cell carcinomas (OPSCCs).

Ependymomas are rare gliomas that can occur within different anatomic compartments of the central nervous system (CNS) of children and adults. They are a heterogeneous group of tumors and can be classified into different tumor subgroups according to histopathological and molecular characteristics as well as different locations within the CNS [108]. This impedes the disease management and therapy development. Despite quite high overall survival (OS) rates of ependymoma patients, a long-term control of the disease is rarely achieved and relapse remains a challenging issue [109]. The analysis of

the ependymoma immunopeptidome aims to reveal potential target antigens for the application in immunotherapeutic approaches of ependymoma treatment.

OPSCCs belong to the head and neck squamous cell carcinomas (HNSCCs) and arise from the mucosal lining of the oropharynx. Risk factors of this tumor entity are smoking and alcohol consumption. Additionally, infections with high-risk types of the human papillomavirus (HPV), especially type HPV16, are important risk factors for the development of OPSCCs. Over the past decades, the number of HPV infections increased, which is accompanied by a rising incidence of OPSCC cases [110]. Similar to the ependymomas, the OPSCC immunopeptidome is analyzed with the aim to identify potential immunotherapeutic target antigens. Furthermore, multi-dimensional data integration is expected to improve data density. Whole exome sequencing (WES) data are used to search for neoantigens within the HLA ligandomes. The analysis of transcriptomic data aims to reveal potential correlations between protein expression and antigen presentation. Since OPSCC tumors display distinct clinical and biological features depending on their HPV status [111], another focus of this study is on the comparative analysis of HPV^+ and HPV^- OPSCCs on the immunopeptidomic and the transcriptomic level. This aims to unravel potential HPV-associated characteristics and to reveal a potential necessity for the consideration of the HPV status during the development of novel immunotherapeutic approaches.

The present work contains two main chapters that separately address ependymomas (see chapter 2) and OPSCCs (see chapter 3).

1.7 References

[1] Parkin J and Cohen B. An overview of the immune system. *The Lancet* 357(9270): 1777–1789 (2001).

[2] Dranoff G. Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* 4(1): 11–22 (2004).

[3] Beutler B. Innate immunity: an overview. *Mol Immunol* 40(12): 845–859 (2004).

[4] Yatim KM and Lakkis FG. A brief journey through the immune system. *Clin J Am Soc Nephrol* 10(7): 1274–1281 (2015).

[5] Janeway CA. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol* 54 Pt 1: 1–13 (1989).

[6] Matzinger P. Tolerance, danger, and the extended family. *Annu Rev Immunol* 12: 991–1045 (1994).

[7] Walsh D, McCarthy J, O'Driscoll C, and Melgar S. Pattern recognition receptors--molecular orchestrators of inflammation in inflammatory bowel disease. *Cytokine Growth Factor Rev* 24(2): 91–104 (2013).

[8] Iwasaki A and Medzhitov R. Control of adaptive immunity by the innate immune system. *Nat Immunol* 16(4): 343–353 (2015).

[9] Kawai T and Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* 11(5): 373–384 (2010).

[10] Kawai T and Akira S. TLR signaling. *Semin Immunol* 19(1): 24–32 (2007).

[11] Vantourout P and Hayday A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat Rev Immunol* 13(2): 88–100 (2013).

[12] Farag SS and Caligiuri MA. Human natural killer cell development and biology. *Blood Rev* 20(3): 123–137 (2006).

[13] Medzhitov R and Janeway C. Innate immune recognition: mechanisms and pathways. *Immunol Rev* 173: 89–97 (2000).

[14] Bonilla FA and Oettgen HC. Adaptive immunity. *J Allergy Clin Immunol* 125(2 Suppl 2): S33-40 (2010).

[15] Delves PJ and Roitt IM. The immune system. First of two parts. *N Engl J Med* 343(1): 37–49 (2000).

[16] Delves PJ and Roitt IM. The immune system. Second of two parts. *N Engl J Med* 343(2): 108–117 (2000).

[17] Kaech SM and Cui W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol* 12(11): 749–761 (2012).

[18] Crotty S. Follicular helper CD4 T cells (TFH). *Annu Rev Immunol* 29: 621–663 (2011).

[19] Eibel H, Kraus H, Sic H, Kienzler A-K, and Rizzi M. B cell biology: an overview. *Curr Allergy Asthma Rep* 14(5): p. 434 (2014).

[20] Cosmi L, Maggi L, Santarlasci V, Liotta F, and Annunziato F. T helper cells plasticity in inflammation. *Cytometry A* 85(1): 36–42 (2014).

[21] Jameson SC and Masopust D. Diversity in T cell memory: an embarrassment of riches. *Immunity* 31(6): 859–871 (2009).

[22] Kirman JR, Quinn KM, and Seder RA. Immunological memory. *Immunol Cell Biol* 97(7): 615–616 (2019).

[23] Trowsdale J and Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* 14: 301–323 (2013).

[24] Goldberg AC and Rizzo LV. MHC structure and function – antigen presentation. Part 1. *Einstein (Sao Paulo)* 13(1): 153–156 (2015).

[25] Goldberg AC and Rizzo LV. MHC structure and function - antigen presentation. Part 2. *Einstein (Sao Paulo)* 13(1): 157–162 (2015).

[26] Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, and Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res* 48(D1): D948-D955 (2020).

[27] Robinson J, Guethlein LA, Cereb N, Yang SY, Norman PJ, Marsh SGE, and Parham P. Distinguishing functional polymorphism from random variation in the sequences of 10,000 HLA-A, -B and -C alleles. *PLoS Genet* 13(6): e1006862 (2017).

[28] Klein J, Satta Y, O'hUigin C, and Takahata N. The molecular descent of the major histocompatibility complex. *Annu Rev Immunol* 11: 269–295 (1993).

[29] Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, and Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329(6139): 506–512 (1987).

[30] Neefjes J, Jongsma MLM, Paul P, and Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11(12): 823–836 (2011).

[31] Rammensee H-G. Chemistry of peptides associated with MHC class I and class II molecules. *Curr Opin Immunol* 7(1): 85–96 (1995).

[32] Rammensee HG, Falk K, and Rötzschke O. Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol* 11: 213–244 (1993).

[33] Schumacher F-R, Delamarre L, Jhunjhunwala S, Modrusan Z, Phung QT, Elias JE, and Lill JR. Building proteomic tool boxes to monitor MHC class I and class II peptides. *Proteomics* 17(1-2) (2017).

[34] Ramsbottom KA, Carr DF, Jones AR, and Rigden DJ. Critical assessment of approaches for molecular docking to elucidate associations of HLA alleles with adverse drug reactions. *Mol Immunol* 101: 488–499 (2018).

[35] Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, and Wiley DC. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368(6468): 215–221 (1994).

[36] Yeturu K, Utriainen T, Kemp GJL, and Chandra N. An automated framework for understanding structural variations in the binding grooves of MHC class II molecules. *BMC Bioinformatics* 11 Suppl 1: S55 (2010).

[37] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, and Stevanović S. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics* 50(3-4): 213–219 (1999).

[38] Andreatta M and Nielsen M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* 32(4): 511–517 (2016).

[39] van Endert, PM, Riganelli D, Greco G, Fleischhauer K, Sidney J, Sette A, and Bach JF. The peptide-binding motif for the human transporter associated with antigen processing. *J Exp Med* 182(6): 1883–1895 (1995).

[40] Bard JAM, Goodall EA, Greene ER, Jonsson E, Dong KC, and Martin A. Structure and Function of the 26S Proteasome. *Annu Rev Biochem* 87: 697–724 (2018).

[41] Komander D and Rape M. The ubiquitin code. *Annu Rev Biochem* 81: 203–229 (2012).

[42] Cascio P, Hilton C, Kisselev AF, Rock KL, and Goldberg AL. 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *EMBO J* 20(10): 2357–2366 (2001).

[43] Reits E, Griekspoor A, Neijssen J, Groothuis T, Jalink K, van Veelen P, Janssen H, Calafat J, Drijfhout JW, and Neefjes J. Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. *Immunity* 18(1): 97–108 (2003).

[44] Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, Greer F, Schomburg L, Fruci D, Niedermann G, and van Endert PM. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol* 6(7): 689–697 (2005).

[45] Wearsch PA and Cresswell P. The quality control of MHC class I peptide loading. *Curr Opin Cell Biol* 20(6): 624–631 (2008).

[46] Schölz C and Tampé R. The peptide-loading complex--antigen translocation and MHC class I loading. *Biol Chem* 390(8): 783–794 (2009).

[47] Hughes EA, Hammond C, and Cresswell P. Misfolded major histocompatibility complex class I heavy chains are translocated into the cytoplasm and degraded by the proteasome. *Proc Natl Acad Sci U S A* 94(5): 1896–1901 (1997).

[48] Reits EA, Vos JC, Grommé M, and Neefjes J. The major substrates for TAP in vivo are derived from newly synthesized proteins. *Nature* 404(6779): 774–778 (2000).

[49] Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW, and Bennink JR. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* 404(6779): 770–774 (2000).

[50] Khan S, Giuli R de, Schmidtke G, Bruns M, Buchmeier M, van den Broek M, and Groettrup M. Cutting edge: neosynthesis is required for the presentation of a T cell epitope from a long-lived viral protein. *J Immunol* 167(9): 4801–4804 (2001).

[51] Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, and Cheung VG. Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333(6038): 53–58 (2011).

[52] Li B and Hu L. Cross-presentation of Exogenous Antigens. *Transfus Clin Biol* 26(4): 346–351 (2019).

[53] Choi NM, Majumder P, and Boss JM. Regulation of major histocompatibility complex class II genes. *Curr Opin Immunol* 23(1): 81–87 (2011).

[54] Busch R, Doebele RC, Patil NS, Pashine A, and Mellins ED. Accessory molecules for MHC class II peptide loading. *Curr Opin Immunol* 12(1): 99–106 (2000).

[55] Hsing LC and Rudensky AY. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol Rev* 207: 229–241 (2005).

[56] Landsverk OJB, Bakke O, and Gregers TF. MHC II and the endocytic pathway: regulation by invariant chain. *Scand J Immunol* 70(3): 184–193 (2009).

[57] Kropshofer H, Vogt AB, Moldenhauer G, Hammer J, Blum JS, and Hämmerling GJ. Editing of the HLA-DR-peptide repertoire by HLA-DM. *EMBO J* 15(22): 6144–6154 (1996).

[58] Vesely MD, Kershaw MH, Schreiber RD, and Smyth MJ. Natural innate and adaptive immunity to cancer. *Annu Rev Immunol* 29: 235–271 (2011).

[59] Seremet T, Brasseur F, and Coulie PG. Tumor-specific antigens and immunologic adjuvants in cancer immunotherapy. *Cancer J* 17(5): 325–330 (2011).

[60] Schumacher TN, Scheper W, and Kvistborg P. Cancer Neoantigens. *Annu Rev Immunol* 37: 173–200 (2019).

[61] Janelle V, Rulleau C, Del Testa S, Carli C, and Delisle J-S. T-Cell Immunotherapies Targeting Histocompatibility and Tumor Antigens in Hematological Malignancies. *Front Immunol* 11: p. 276 (2020).

[62] Fratta E, Coral S, Covre A, Parisi G, Colizzi F, Danielli R, Nicolay HJM, Sigalotti L, and Maio M. The biology of cancer testis antigens: Putative function, regulation and therapeutic potential. *Mol Oncol* 5(2): 164–182 (2011).

[63] van der Bruggen P, Traversari C, Chomez P, Lurquin C, Plaen E de, van den Eynde B, Knuth A, and Boon T. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* 254(5038): 1643–1647 (1991).

[64] Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-T, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJG, Old LJ, Camargo AA, and Vasconcelos ATR. CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37(Database issue): D816-9 (2009).

[65] zur Hausen H. The search for infectious causes of human cancers: where and why. *Virology* 392(1): 1–10 (2009).

[66] Anichini A, Maccalli C, Mortarini R, Salvi S, Mazzocchi A, Squarcina P, Herlyn M, and Parmiani G. Melanoma cells and normal melanocytes share antigens recognized by HLA-A2-restricted cytotoxic T cell clones from melanoma patients. *J Exp Med* 177(4): 989–998 (1993).

[67] Chen DS and Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* 39(1): 1–10 (2013).

[68] Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 144(5): 646–674 (2011).

[69] Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, Lichtor T, Decker WK, Whelan RL, Kumara HMCS, Signori E, Honoki K, Georgakilas AG, Amin A, Helferich WG, Boosani CS, Guha G, Ciriolo MR, Chen S, Mohammed SI, Azmi AS, Keith WN, Bilsland A, Bhakta D, Halicka D, Fujii H, Aquilano K, Ashraf SS, Nowsheen S, Yang X, Choi BK, and Kwon BS. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin Cancer Biol* 35 Suppl: S185-S198 (2015).

[70] Garrido F, Ruiz-Cabello F, Cabrera T, Pérez-Villar JJ, López-Botet M, Duggan-Keen M, and Stern PL. Implications for immunosurveillance of altered HLA class I phenotypes in human tumours. *Immunol Today* 18(2): 89–95 (1997).

[71] Hicklin DJ, Marincola FM, and Ferrone S. HLA class I antigen downregulation in human cancers: T-cell immunotherapy revives an old story. *Mol Med Today* 5(4): 178–186 (1999).

[72] Gasparoto TH, Souza Malaspina TS de, Benevides L, Melo EJF de, Costa MRSN, Damante JH, Ikoma MRV, Garlet GP, Cavassani KA, da Silva JS, and Campanelli AP. Patients with oral squamous cell carcinoma are characterized by increased frequency of suppressive regulatory T cells in the blood and tumor microenvironment. *Cancer Immunol Immunother* 59(6): 819–828 (2010).

[73] Lee I, Wang L, Wells AD, Dorf ME, Ozkaynak E, and Hancock WW. Recruitment of Foxp3+ T regulatory cells mediating allograft tolerance depends on the CCR4 chemokine receptor. *J Exp Med* 201(7): 1037–1044 (2005).

[74] Elliott RL and Blobe GC. Role of transforming growth factor Beta in human cancer. *J Clin Oncol* 23(9): 2078–2093 (2005).

[75] Driessens G, Kline J, and Gajewski TF. Costimulatory and coinhibitory receptors in anti-tumor immunity. *Immunol Rev* 229(1): 126–144 (2009).

[76] Lauritzsen GF, Hofgaard PO, Schenck K, and Bogen B. Clonal deletion of thymocytes as a tumor escape mechanism. *Int J Cancer* 78(2): 216–222 (1998).

[77] Beatty GL and Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin Cancer Res* 21(4): 687–692 (2015).

[78] da Silva JL, Dos Santos ALS, Nunes NCC, Moraes Lino da Silva F de, Ferreira CGM, and Melo AC de. Cancer immunotherapy: the art of targeting the tumor immune microenvironment. *Cancer Chemother Pharmacol* 84(2): 227–240 (2019).

[79] Darvin P, Toor SM, Sasidharan Nair V, and Elkord E. Immune checkpoint inhibitors: Recent progress and potential biomarkers. *Exp Mol Med* 50(12): 1–11 (2018).

[80] Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, Gonzalez R, Robert C, Schadendorf D, Hassel JC, Akerley W, van den Eertwegh AJM, Lutzky J, Lorigan P, Vaubel JM, Linette GP, Hogg D, Ottensmeier CH, Lebbé C, Peschel C, Quirt I, Clark JI, Wolchok JD, Weber JS, Tian J, Yellin MJ, Nichol GM, Hoos A, and Urba WJ. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 363(8): 711–723 (2010).

[81] Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 12(4): 252–264 (2012).

[82] O'Day SJ, Hamid O, and Urba WJ. Targeting cytotoxic T-lymphocyte antigen-4 (CTLA-4): a novel strategy for the treatment of melanoma and other malignancies. *Cancer* 110(12): 2614–2627 (2007).

[83] Fong L and Small EJ. Anti-cytotoxic T-lymphocyte antigen-4 antibody: the first in an emerging class of immunomodulatory antibodies for cancer treatment. *J Clin Oncol* 26(32): 5275–5283 (2008).

[84] Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L, Carcereny E, Ahn M-J, Felip E, Lee J-S, Hellmann MD, Hamid O, Goldman JW, Soria J-C, Dolled-Filhart M, Rutledge RZ, Zhang J, Lunceford JK, Rangwala R, Lubiniecki GM, Roach C, Emancipator K, and Gandhi L. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* 372(21): 2018–2028 (2015).

[85] Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhäufl M, Arrieta O, Burgio MA, Fayette J, Lena H, Poddubskaya E, Gerber DE, Gettinger SN, Rudin CM, Rizvi N, Crinò L, Blumenschein GR, Antonia SJ, Dorange C, Harbison CT, Graf Finckenstein F, and Brahmer JR. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 373(17): 1627–1639 (2015).

[86] Yeku O, Li X, and Brentjens RJ. Adoptive T-Cell Therapy for Solid Tumors. *Am Soc Clin Oncol Educ Book* 37: 193–204 (2017).

[87] Kalos M, Levine BL, Porter DL, Katz S, Grupp SA, Bagg A, and June CH. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci Transl Med* 3(95): 95ra73 (2011).

[88] Morgan RA, Yang JC, Kitano M, Dudley ME, Laurencot CM, and Rosenberg SA. Case report of a serious adverse event following the administration of T cells transduced with a chimeric antigen receptor recognizing ERBB2. *Mol Ther* 18(4): 843–851 (2010).

[89] Yang B, Jeang J, Yang A, Wu TC, and Hung C-F. DNA vaccine for cancer immunotherapy. *Hum Vaccin Immunother* 10(11): 3153–3164 (2014).

[90] Pardi N, Hogan MJ, Porter FW, and Weissman D. mRNA vaccines - a new era in vaccinology. *Nat Rev Drug Discov* 17(4): 261–279 (2018).

[91] Santos PM and Butterfield LH. Dendritic Cell-Based Cancer Vaccines. *J Immunol* 200(2): 443–449 (2018).

[92] Kozłowska A, Mackiewicz J, and Mackiewicz A. Therapeutic gene modified cell based cancer vaccines. *Gene* 525(2): 200–207 (2013).

[93] Calvo Tardón M, Allard M, Dutoit V, Dietrich P-Y, and Walker PR. Peptides as cancer vaccines. *Curr Opin Pharmacol* 47: 20–26 (2019).

[94] Buonaguro L, Petrizzo A, Tornesello ML, and Buonaguro FM. Translating tumor antigens into cancer vaccines. *Clin Vaccine Immunol* 18(1): 23–34 (2011).

[95] Gouttefangeas C and Rammensee H-G. Personalized cancer vaccines: adjuvants are important, too. *Cancer Immunol Immunother* 67(12): 1911–1918 (2018).

[96] Baumgaertner P, Jandus C, Rivals J-P, Derré L, Lövgren T, Baitsch L, Guillaume P, Luescher IF, Berthod G, Matter M, Rufer N, Michielin O, and Speiser DE. Vaccination-induced functional competence of circulating human tumor-specific CD8 T-cells. *Int J Cancer* 130(11): 2607–2617 (2012).

[97] Rammensee H-G, Wiesmüller K-H, Chandran PA, Zelba H, Rusch E, Gouttefangeas C, Kowalewski DJ, Di Marco M, Haen SP, Walz JS, Gloria YC, Bödder J, Schertel J-M, Tunger A, Müller L, Kießler M, Wehner R, Schmitz M, Jakobi M, Schneiderhan-Marra N, Klein R, Laske K, Artzner K, Backert L, Schuster H, Schwenck J, Weber ANR, Pichler BJ, Kneilling M, La Fougère C, Forchhammer S, Metzler G, Bauer J, Weide B, Schippert W, Stevanović S, and Löffler MW. A new synthetic toll-like receptor 1/2 ligand is an efficient adjuvant for peptide vaccination in a human volunteer. *J Immunother Cancer* 7(1): p. 307 (2019).

[98] He X, Abrams SI, and Lovell JF. Peptide Delivery Systems for Cancer Vaccines. *Adv. Therap.* 1(5): p. 1800060 (2018).

[99] Mandal R and Chan TA. Personalized Oncology Meets Immunology: The Path toward Precision Immunotherapy. *Cancer Discov* 6(7): 703–713 (2016).

[100] Löffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, Trautwein N, Hilke FJ, Zinser RS, Mühlenbruch L, Kowalewski DJ, Schuster H, Sturm M, Matthes J, Riess O, Czemmel S, Nahnsen S, Königsrainer I, Thiel K, Nadalin S, Beckert S, Bösmüller H, Fend F, Velic A, Maček B, Haen SP, Buonaguro L, Kohlbacher O, Stevanović S, Königsrainer A, and Rammensee H-G. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med* 11(1): p. 28 (2019).

[101] Ternette N and Purcell AW. Immunopeptidomics Special Issue. *Proteomics* 18(12): e1800145 (2018).

[102] Guo Y, Lei K, and Tang L. Neoantigen Vaccine Delivery for Personalized Anticancer Immunotherapy. *Front Immunol* 9: p. 1499 (2018).

[103] Weinzierl AO, Lemmel C, Schoor O, Müller M, Krüger T, Wernet D, Hennenlotter J, Stenzl A, Klingel K, Rammensee H-G, and Stevanovic S. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol Cell Proteomics* 6(1): 102–113 (2007).

[104] Freudenmann LK, Marcu A, and Stevanović S. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. *Immunology* 154(3): 331–345 (2018).

[105] Purcell AW, Ramarathinam SH, and Ternette N. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat Protoc* 14(6): 1687–1707 (2019).

[106] Marino F, Chong C, Michaux J, and Bassani-Sternberg M. High-Throughput, Fast, and Sensitive Immunopeptidomics Sample Processing for Mass Spectrometry. *Methods Mol Biol* 1913: 67–79 (2019).

[107] Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, Platten M, Tabatabai G, Dutoit V, van der Burg SH, Thor Straten P, Martínez-Ricarte F, Ponsati B, Okada H, Lassen U, Admon A, Ottensmeier CH, Ulges A, Kreiter S, Deimling A von, Skardelly M, Migliorini D, Kroep JR, Idorn M, Rodon J, Piró J, Poulsen HS, Shraibman B, McCann K, Mendrzyk R, Löwer M, Stieglbauer M, Britten CM, Capper D, Welters MJP, Sahuquillo J, Kiesel K, Derhovanessian E, Rusch E, Bunse L, Song C, Heesch S, Wagner C, Kemmer-Brück A, Ludwig J, Castle JC, Schoor O, Tadmor AD, Green E, Fritsche J, Meyer M, Pawlowski N, Dorner S, Hoffgaard F, Rössler B, Maurer D, Weinschenk T, Reinhardt C, Huber C, Rammensee H-G, Singh-Jasuja H, Sahin U, Dietrich P-Y, and Wick W. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* 565(7738): 240–245 (2019).

[108] Wu J, Armstrong TS, and Gilbert MR. Biology and management of ependymomas. *Neuro Oncol* 18(7): 902–913 (2016).

[109] Thorp N and Gandola L. Management of Ependymoma in Children, Adolescents and Young Adults. *Clin Oncol (R Coll Radiol)* 31(3): 162–170 (2019).

[110] Budu VA, Decuseară T, Balica NC, Mogoantă CA, Rădulescu LM, Chirilă M, Maniu AA, Mistra DM, Muşat GC, Oprişcan IC, and Georgescu MG. The role of HPV infection in oropharyngeal cancer. *Rom J Morphol Embryol* 60(3): 769–773 (2019).

[111] Parvathaneni U, Lavertu P, Gibson MK, and Glastonbury CM. Advances in Diagnosis and Multidisciplinary Management of Oropharyngeal Squamous Cell Carcinoma: State of the Art. *Radiographics* 39(7): 2055–2068 (2019).

2 The ependymoma HLA ligandome

2.1 Introduction

Ependymomas are rare neuroepithelial tumors occurring within the central nervous system (CNS). These primary malignancies originate from radial glia cells (RGs) and arise from the wall of the ventricular system and the central canal of the spinal cord along the entire craniospinal axis [1, 2]. Ependymomas are among the rarest types of gliomas constituting 3.3% of all malignant primary brain and other CNS tumors. Malignant ependymomas occur throughout all age groups with different average annual age-adjusted incidence rates (IRs) per 100,000 individuals. The IRs amount to 0.27 patients in children (age 0-14 years), 0.20 patients in adolescents and young adults (age 15-39 years) and 0.28 patients in adults (age ≥ 40 years). With an IR of 0.28 patients per 100,000 individuals, more men than women (IR = 0.23) are affected by ependymal malignancies [3]. Ependymomas are a heterogeneous tumor entity and can be classified into several tumor subgroups [4]. Therefore, numerous symptoms can emerge prior to ependymoma diagnosis. This wide range of ependymoma symptoms includes headache, visual problems, nausea/vomiting, weakness, numbness or tingling of limbs and (radiating) back pain [5].

Stem cells of ependymomas

RGs generate most CNS neurons and glia cells, which makes them the principal neuroepithelial progenitor cells of the CNS [6]. The neural stem cell pool is increased by symmetric division of RGs resulting in two identical daughter stem cells. By dividing asymmetrically, each RG produces one stem cell and one cell designated for differentiation [7]. Furthermore, RGs are suggested to act as stem cells of ependymomas [2]. A conjuncture of several mechanisms is assumed to lead to the conversion of RGs into cancer stem cells (CSCs) **(Figure 7)** [8]. This includes aberrant RG division induced by a deviating and prolonged expression of specific transcription factors such as the Homeobox protein EMX2 [2]. Disrupted adherens junction complexes can have further impact on the RG fate [9]. Additionally, dysregulation of cell signaling pathways involved in proliferation and self-renewal of neural stem cells is suggested to play a crucial role in the transformation of RGs into CSCs [10, 11]. An increased self-renewal and an enhanced mutational rate in these aberrant RGs might constitute the initiation of ependymoma development [8].

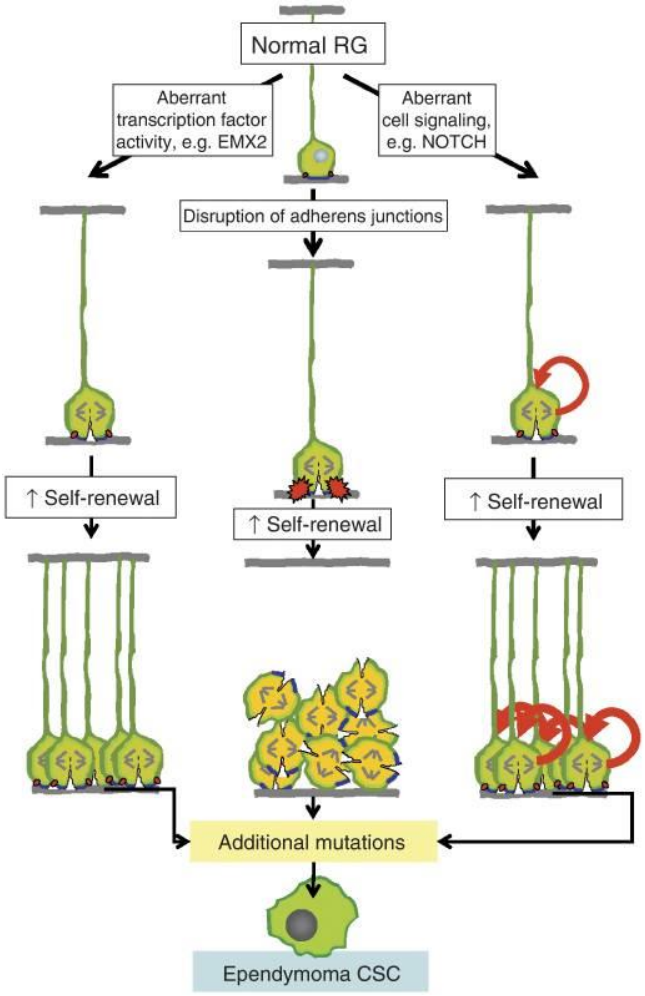


Figure 7: RGs are considered as stem cells of ependymomas. A combination of aberrant RG division induced by a differential and prolonged expression of specific transcription factors, disrupted adherens junctions and a dysregulation of cell signaling pathways are suggested to induce the transformation of RGs into CSCs. Additionally, self-renewal and further mutations of aberrant RGs contribute to the development of ependymomas. RG – radial glia cell; CSC – cancer stem cell. [8]

Ependymoma classification

Ependymomas are a very heterogeneous tumor entity. The ependymal tumors can be subdivided into several groups according to clinical, histopathological and molecular criteria.

The World Health Organization (WHO) differentiates on the histopathological level between subependymoma, myxopapillary ependymoma, ependymoma and anaplastic ependymoma **(Table 1)** [1]. Based on their level of anaplasia, which describes the progressing loss of normal morphological cell characteristics, ependymal tumors are categorized into the WHO grades I, II and III [4]. Subependymomas and myxopapillary ependymomas are classified as WHO grade I tumors. Subependymomas are regarded as own tumor entity and are histologically discernable due to their hypocellularity and their commonly occurring microcystic change. Myxopapillary ependymomas are characterized by pseudo-papillary structures, which are cuboidal ependymal cells radially arranged

around a myxoid stroma. Their tumor cells are surrounded by mucoid material and microcysts. Further, collagen balls (eosinophilic structures with reticulin) occur as side effect of this subgroup. The subgroup termed ependymoma is categorized as WHO grade II tumors and comprises three subtypes: Papillary ependymomas have a central vascular core surrounded by cylindrical cells; clear cell ependymomas are regularly arranged cells with a clear cytoplasm and inconspicuous perivascular rosettes; tanycytic ependymomas consist of elongated spindle shaped cells arranged in fascicles and prominent pseudorosettes. Anaplastic ependymomas are categorized as WHO grade III tumors. Their main characteristics are a hypercellular appearance and hyperchromatic, pleomorphic nuclei [1].

Table 1: Different ependymoma subgroups as categorized by WHO grading. Subependymoma, myxopapillary ependymoma, ependymoma and anaplastic ependymoma are the main subgroups defined according to histopathological characteristics. [1, 4]

Ependymoma subtype		Histopathological characteristics	WHO grade
Subependymoma		- Hypocellular - Microcystic change	I
Myxopapillary ependymoma		- Pseudo-papillary structures - Surrounded by mucoid material - Mucin-rich microcysts - Collagen balls	I
Ependymoma	- **Papillary**	- Central vascular core - Cylindrical cells	II
	- **Clear cell**	- Clear cytoplasm - Perivascular rosettes	
	- **Tanycytic**	- Spindle shaped cells in fascicles - Pseudorosettes	
Anaplastic ependymoma		- Hypercellularity - Hyperchromatic and pleomorphic nuclei - Abundant mitosis - Pseudopalisading necrosis - Microvascular proliferation - Ependymal rosettes absent	III

In clinical application, histopathological diagnosis and grading of ependymomas according to WHO classification criteria is often challenging and the results are sometimes inconclusive. There are several further classification attempts for ependymal tumors to develop a more prognostic and reproducible grading.

Ependymomas can be discriminated according to three different locations within the main anatomic compartments of the CNS: spinal, infratentorial (posterior fossa) and supratentorial [12]. Myxopapillary ependymomas are mainly located in the lower portion of the spinal cord [13, 14], whereas subependymomas are often found within the ventricular wall [15]. Clear cell, tanycytic and papillary ependymomas are described to occur in the upper spinal cord or intracranially [16]. Anaplastic ependymomas are commonly found intracranially [17]. There is evidence that the anatomic site of the tumor within the CNS has an impact on the patient´s outcome [18]. Therefore, Raghunathan *et al.* (2013) suggested to include tumor localization as important factor into prospective classification schemes [18].

On the genetic level, the RELA (v-rel avian reticuloendotheliosis viral oncogene homolog A) fusion-positive ependymoma has been described as molecular tumor subtype. It is defined as either WHO grade II or III and often occurs in supratentorial location. Tumors of this subgroup are characterized by an oncogenic gene fusion between RELA and the uncharacterized gene C11orf95. RELA is the principal effector of canonical NF-κB signaling [19, 20]. Members of the nuclear factor-κB (NF-κB) transcription factor family are central mediators of inflammatory processes. The NF-κB signaling pathway is active in most human cancers and activates survival genes in tumor cells and inflammatory genes within the tumor microenvironment [21]. In RELA fusion-positive ependymomas, the described gene fusion leads to an aberrantly elevated NF-κB signaling [19, 20].

Johnson *et al.* (2010) categorized ependymomas into different subgroups by analyzing the tumor gene expression profiles. Thereby, they defined nine different ependymal tumor subgroups based on focal gene deletions and amplifications [22]. Pajtler *et al.* (2015) used DNA methylation profiling of ependymal tumors as another novel approach to discriminate between different subgroups. They reported nine distinct epigenetic subgroups of ependymomas, three in each of the possible anatomical compartments of the CNS (spinal, intracranial, supratentorial) [23]. The molecular subgroups described by Pajtler *et al.* (2015) display genetically, epigenetically, transcriptionally, demographically and clinically distinct characteristics **(Figure 8)**. They partly overlap with the nine previously defined subgroups by Johnson *et al.* (2010) based on gene expression profiling [22].

Numerous attempts have been made in the past to grade ependymomas and distinguish between distinct subtypes of this tumor entity. Tumor localization combined with data received from molecular analyses are considered as currently most precise ependymoma classification strategy, which is continuously optimized and extended.

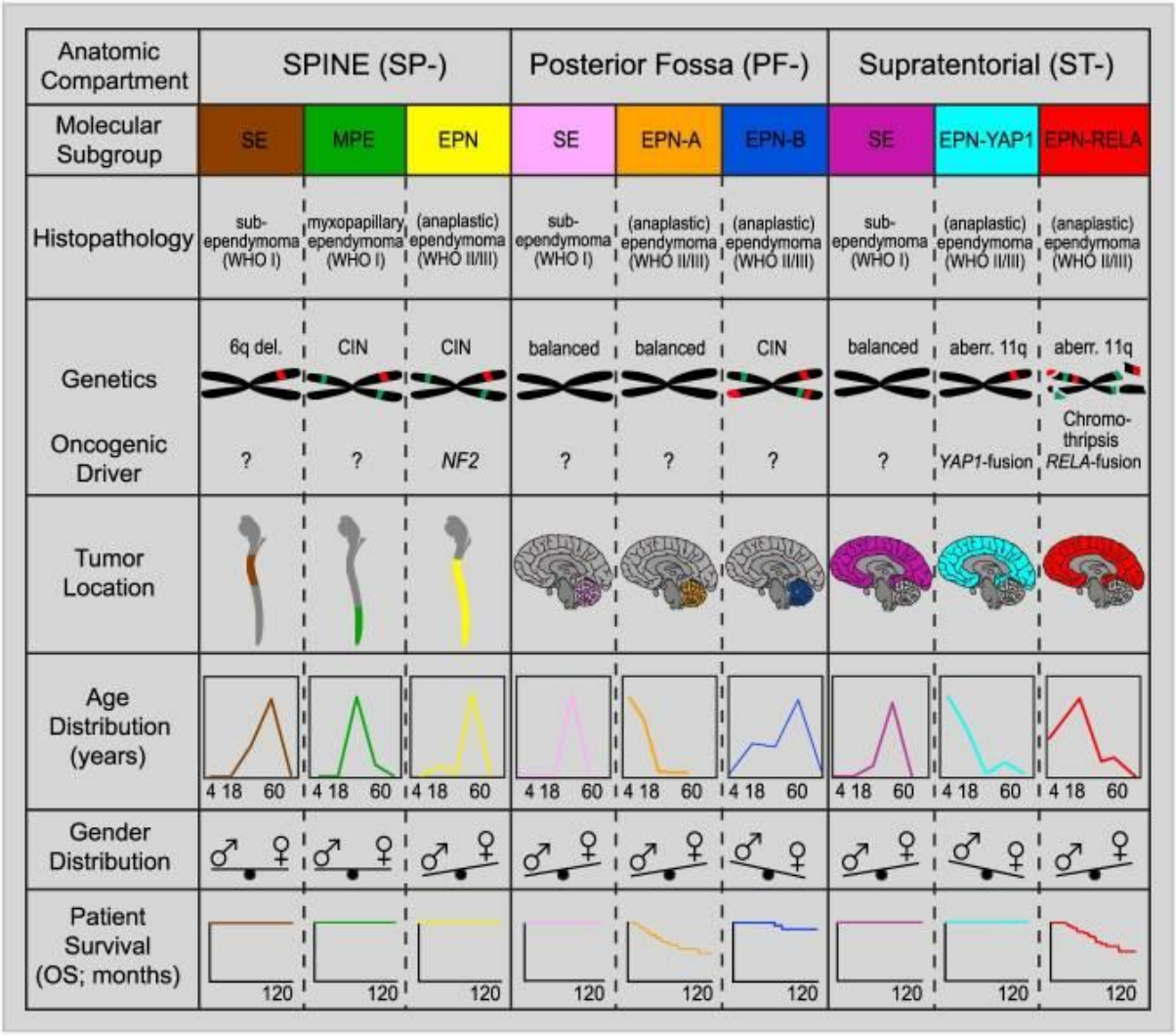


Figure 8: Subgroups of ependymomas. Based on DNA methylation data, nine molecular subgroups of ependymal tumors are defined that vary on histopathological, genetic, epigenetic, transcriptional, demographic and clinical levels. SE – subependymoma; MPE – mxopapillary ependymoma; EPN – ependymoma; YAP1 – yes-associated protein 1; RELA – v rel avian reticuloendotheliosis viral oncogene homolog A; NF2 – Neurofibromin-2; WHO – World Health Organization; del – deletion; aberr. – aberrant; CIN – chromosomal instability; OS – overall survival. [23]

Current therapy for ependymomas

The clinical presentation of ependymomas varies widely with patient age, tumor location, tumor size and tumor subtype. To diagnose ependymal tumors, magnetic resonance imaging (MRI) scans with contrast enhancement is the method of choice in most cases [24]. Subsequent classification of the tumor is achieved by histopathological and molecular diagnostic assessment of tissue specimen [4]. For newly diagnosed non-metastatic ependymomas, the current standard of care entails gross total resection (GTR) of the tumor through maximal safe surgery as first crucial step whenever feasible. In case of a WHO grade II- or III-classified ependymoma, local fractionated radiation therapy is usually applied postoperatively. If the tumor is classified as WHO grade I myxopapillary ependymoma, a

radiation therapy only follows a preceded subtotal tumor resection (STR). In any case, a long-term monitoring with serial enhanced MRI scans is crucial due to the risk of an asymptomatic or late recurrence. Upon diagnosis of a recurrent ependymoma, re-excision and re-irradiation are required. If the patient is not eligible for a local tumor therapy, chemotherapy is the treatment of choice [25, 26]. However, the role of chemotherapy as treatment option for ependymoma patients is controversial and has not yet been completely established neither for initial nor for recurrent tumors. Armstrong *et al.* (2018) published the results of the first prospective clinical trial addressing the application of chemotherapy in adult ependymoma patients [27]. In their single-armed phase II study, they demonstrated a beneficial effect of a combinatorial treatment with the chemotherapeutic agent temozolomide and the tyrosine kinase inhibitor lapatinib. These findings led to the inclusion of this therapy option into the National Comprehensive Cancer Network (NCCN) guidelines.

Besides relapse, metastasizing is another issue to face during treatment of ependymoma patients. A high risk for metastatic dissemination within the CNS commonly occurs in patients suffering from myxopapillary ependymomas. A study performed by Kraetzig *et al.* (2018) included spinal myxopapillary ependymoma cases and observed distant metastases in 57.9% of the patients. 36.4% of these metastases were found at initial diagnosis [28]. Metastases can develop along the entire neuroaxis. Usually the spine is affected by metastases, but sometimes they also occur in the brain. In case of a metastasizing disease progression, a GTR of the primary tumor is still the first goal of treatment. Despite metastases often being clinically stable, frequent follow-up MRI scans are essential for the detection and monitoring of distant metastasizing to initiate therapy in case of symptomatic signs of progression [28].

The described current standard of care that is recommended for the treatment of ependymoma patients leads to a five-year overall survival (OS) rate of 85.7% and a ten-year OS rate of 80.6%. These values are quite high compared to those observed for the entirety of all malignant brain and other CNS tumors that exhibit a five-year OS rate of 35.8% and a ten-year OS rate of 30.8% [3]. However, relapse remains a substantial challenge in the course of ependymoma treatment. Despite surgical re-excision and re-irradiation, a long-term control of the disease cannot be achieved in most cases [29, 30]. Therefore, patients with ependymomas are recommended to be included in clinical trials as often as possible [26].

Immunotherapy for ependymomas

The assurance of optimal clinical care and the performance of clinical trials to develop new therapeutic approaches are challenging for rare CNS tumors. This issue also concerns the rare ependymomas. Detailed insights into molecular and cellular mechanisms underlying ependymoma development will facilitate and improve prognosis and treatment of patients suffering from ependymal malignancies and are expected to lead to the establishment of novel therapy options.

The ependymoma HLA ligandome

The brain is often referred to as an immunologically privileged organ. However, it has long been shown that an induction of immune responses in the brain can be achieved [31]. Activated cells of the adaptive immune system are capable of accessing parts of the CNS to search for their respective antigens [32]. In case of a disrupted blood-brain barrier, neuroinflammation is possible. Brain tumors are often accompanied by a dysfunction of the blood-brain barrier, which enables immune infiltration into the CNS [33]. Thus, several studies addressing the development of new immunotherapeutic strategies for the treatment of CNS tumor patients are currently performed.

The application of immune checkpoint inhibitors (ICIs) has proven successful in clinical trials for various cancer entities [34]. Several recent studies are expanding the investigation of ICI applications to the research field of rare CNS malignancies. The effect of the programmed cell death protein 1 (PD-1) inhibitor nivolumab on patients with recurrent or refractory pediatric brain tumors was observed by Gorsi *et al.* (2019). This retrospective overview included one ependymoma patient. A beneficial effect of nivolumab was demonstrated in pediatric brain tumor patients with high tumor mutational burden and elevated programmed cell death protein ligand 1 (PD-L1) expression [35]. However, the application of PD-1/PD-L1 as therapeutic targets for ependymoma treatment is a controversial issue. Ependymal tumors included in a study performed by Dumont *et al.* (2017) displayed no expression of PD-L1 despite the presence of PD-1 expressing lymphocytes [36]. Still, some results speak for the use of PD-1 inhibitors in at least some ependymoma subgroups. Witt *et al.* (2018) suggested a benefit for patients diagnosed with supratentorial RELA fusion-positive ependymoma through tumor evasion and immunosuppression effects of PD-L1/PD-1-mediated T-cell exhaustion [37]. Nivolumab as treatment option is further observed in an ongoing phase II trial recruiting patients of ten different rare CNS tumor entities including ependymoma (NCT03173950) [38].

Besides checkpoint inhibition, another focus is on the discovery of tumor-associated antigens (TAAs) and neoantigens as targets in the field of cancer immunotherapy development. The expression of TAAs as potential targets for cancer vaccines is also of interest for ependymomas [39]. Yeung *et al.* (2013) searched for TAAs in ependymomas by immunohistochemical analysis. They showed an overexpression of ephrin type-A receptor 2 (EphA2), interleukin-13 receptor subunit alpha-2 (IL-13Rα2) and survivin [40] that have been demonstrated to be immunogenic in glioma patients of previous studies [39]. Another approach is the search for neoantigens presenting genomic mutations. A prediction of neoantigens by genomic analysis was performed in one ependymoma patient before and revealed fourteen expressed mutations that might produce potential HLA class I neoantigens [41]. Previous studies addressing the search for tumor-specific antigens in ependymomas were exclusively performed on the expression and prediction level. However, mass spectrometry (MS) is still the only unbiased method to reliably examine the repertoire of naturally HLA-presented peptides *in vivo* [42].

2.2 Contributions

The ependymoma study was performed in collaboration with different associates who contributed to this project.

The project was established in cooperation with Prof. Dr. Dr. med. Ghazaleh Tabatabai (Department of Neurology and Interdisciplinary Neuro-Oncology & Center for Neuro-Oncology, Comprehensive Cancer Center Tübingen-Stuttgart, University Hospital Tübingen & Hertie Institute for Clinical Brain Research, Eberhard Karls University Tübingen, Germany) and Dr. med. Irina Gepfner-Tuma (Department of Neurology and Interdisciplinary Neuro-Oncology & Center for Neuro-Oncology, Comprehensive Cancer Center Tübingen-Stuttgart, University Hospital Tübingen & Hertie Institute for Clinical Brain Research, Eberhard Karls University Tübingen, Germany; University Hospital Jena, Hans Berger Department of Neurology, Friedrich Schiller University Jena, Germany). They provided the tumor samples, collected the clinical data of the patients as well as the histological and molecular data of the tumors and supported the data interpretation.

Leon Bichmann (Institute for Cell Biology, Department of Immunology, Eberhard Karls University Tübingen, Germany; Applied Bioinformatics, Center for Bioinformatics and Department of Computer Science, Eberhard Karls University Tübingen, Germany) and Dr. Linus Backert (Immatics Biotechnologies GmbH, Tübingen, Germany) assisted the bioinformatic data analysis.

2.3 Materials and methods

In the following, an overview of the ependymoma cohort and the methods applied during the study is given. Several methods overlap with those from the OPSCC study (chapter 3). Detailed information and descriptions regarding deployed materials and methods can be assessed in the appendix (see 8.1).

Study cohort

The study cohort consisted of 22 ependymoma tissue samples from 21 different donor patients. Ependymoma tumors were surgically resected, immediately snap-frozen in liquid nitrogen (N_2) and subsequently stored at -80°C. The 10 female and 11 male patients were between 17 and 77 years old at the date of surgery. Written informed consent was received according to the Declaration of Helsinki protocol from all patients. The study was performed in accordance with the local ethical requirements (332/2017BO2). The cohort included twelve spinal, eight infratentorial and two supratentorial tumors classified as WHO grades I, II and III. Three tumors were recurrent diagnoses, of which one was infratentorially and two were supratentorially localized. The latter two were RELA fusion-positive

tumors. Both originated from one patient and were surgically resected within an interval of 13 months. HLA class I and class II typing was performed for all patients and HLA class I binders were determined. The following immunopeptidomic analyses were based on HLA class I binders and HLA class II-presented peptides. Detailed information about the study cohort are shown by **Supplementary Table 1**.

Analysis of HLA class I allelic distribution

HLA class I and class II typing was carried out by the HLA laboratory of the university medical center Tübingen (Universitätsklinikum Tübingen, UKT). To analyze the HLA class I allelic distribution, frequencies of alleles occurring within the ependymoma cohort were determined. A general German population was used for comparison to ascertain potential anomalies within the HLA allelic distribution of the ependymoma cohort. The cohort "Germany pop 8" (n = 39,689) as comprised in the Allele Frequency Net Database (www.allelefrequencies.net) served as reference dataset [43]. For statistical data analysis, the GraphPad Prism 6.1 software (GraphPad Software Inc) was used. 13 different HLA-A alleles, 24 different HLA-B alleles and 16 different HLA-C alleles occurred within the ependymoma cohort. Therefore, potentially significant differences between frequencies of individual HLA alleles within the patient cohort and the reference were observed by 13, 24 and 16 tables in the format 2x2 for HLA-A, -B and -C, respectively. Chi-square test was applied and, after Bonferroni correction, the levels of statistical significance were defined as p value ≤ 0.004 (≙ 0.05/13) for HLA-A, p value ≤ 0.002 (≙ 0.05/24) for HLA-B and p value ≤ 0.003 (≙ 0.05/16) for HLA-C. Relative risks were determined by logistic regression analysis and expressed by the odds ratio with a 95% confidence interval (CI).

HLA ligand isolation

Isolation of HLA-presented peptides was performed as described before by Nelde *et al.* (2019) [44]. Using antibodies specific for HLA molecules **(Appendix, Table 11)**, immunoaffinity chromatography was performed for HLA precipitation. Subsequently, acidic elution of HLA ligands followed by ultrafiltration was applied to obtain the peptide pool.

LC-MS/MS analysis

The pool of HLA-presented peptides was separated by reversed-phase nanoflow ultra-high-performance liquid chromatography (nanoUHPLC, UltiMate 3000 RSLCnano, Dionex) and analyzed by tandem mass spectrometry (MS/MS) in an on-line coupled LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific).

Data processing

Resulting MS raw data were processed using the Proteome Discoverer 1.4 software (Thermo Fisher Scientific). Database search for spectral alignment was performed based on the human reference proteome as comprised in the Swiss-Prot database (20,279 reviewed protein sequences; September 27th, 2013).

Analysis of HLA ligandomic yields

To observe the assumed amount of overall detectable HLA ligands (saturation value) and the proportion of the presently analyzed HLA ligands, saturation analysis was performed. This analysis was based on the rate of newly identified unique ligands per sample.

Furthermore, a potential correlation between sample mass and peptide yields was examined to address the variation of peptide yields between the individual tumor samples. The GraphPad Prism 6.1 software (GraphPad Software Inc) was used and Pearson´s correlation was applied for statistical analysis. The number of isolated HLA class I binders and of HLA class II-presented peptides were correlated against the respective sample masses of tumor tissues. A p value ≤ 0.05 was defined as level of statistical significance.

Tumor subgroup analysis

According to either WHO classification criteria or based on localization and molecular characteristics, the ependymomas were divided into different tumor subgroups. By comparison of immunopeptidomic subgroup datasets in Venn diagrams and unsupervised principle component analyses (PCA), potential differences between tumors of distinct subgroups were observed on the HLA ligandome level.

Comparative profiling against in-house immunopeptidomic databases

Comparative profiling of the entirety of isolated ependymoma HLA ligands was performed against an in-house benign database to search for frequent tumor-exclusive antigens within the ependymoma immunopeptidome. The benign database consists of 419 HLA class I ligandomes and 364 HLA class II ligandomes of benign samples from different tissue types **(Appendix, Table 15)**.
Similarly, comparative profiling of determined tumor-exclusive antigens against an in-house malignant database was performed to investigate their HLA presentation in the immunopeptidomes of other malignancies. After datasets from ependymomas of the present study were excluded, the malignant database comprised 773 HLA class I and 560 HLA class II ligandomes of malignant samples derived from different tumor entities **(Appendix, Table 16)**.

Search for potential TAAs

The ependymoma HLA ligandome dataset was investigated to identify potential TAAs. The focus of this study was on frequent antigens occurring in at least 3 ependymoma samples (corresponding to a frequency ≥ 10%) to detect TAAs that are universal across different patients. The screening for TAAs included the search for already described cancer-testis antigens (CTAs) as comprised in the CTdatabase (www.cta.lncc.br; state August 2019) [45]. Additionally, the tumor-exclusive antigens were screened for further TAAs. For this purpose, the HLA-presented peptides and corresponding source proteins of frequent tumor-exclusive antigens (frequency ≥ 10%) were evaluated. On the peptide level, this entailed the examination of several quality criteria including the spectrum, peptide spectrum matches (PSMs), cross correlation (Xcorr) and delta correlation (ΔCn). On the source protein level, the expression profiles in different healthy tissues and tumor entities were investigated using the databases The Genotype-Tissue Expression Project (GTEx; www.gtexportal.org) [46] and The Human Protein Atlas (www.proteinatlas.org) [47].

Functional annotation clustering

GO term enrichment for functional annotation clustering was performed with tumor-exclusive source proteins of HLA class I binders and HLA class II-presented peptides to reveal their involvement in cellular processes. The functional annotation tool of the database DAVID Bioinformatics Resources 6.8 [48, 49] was used.

2.4 Results

2.4.1 HLA class I allelic distribution

HLA class I and class II typing was performed for all 21 ependymoma patients included in the present study **(Supplementary Table 1)**. Based on the typing results, frequencies of the HLA-A, -B and -C alleles occurring in the patient cohort were determined **(Supplementary Tables 2, 3, 4)**. To reveal whether the cohort is representative with respect to the HLA class I allelic distribution, the HLA allele frequencies of the ependymoma patients were compared to the allele frequencies expected in a general German population as comprised in the Allele Frequency Net Database [43].

The majority of HLA class I allele frequencies was comparable between patients and the reference **(Figure 9)**. However, for some alleles, the patient frequencies differed from the expectation. This applied for example to the HLA alleles A*24:02, A*26:01, B*18:01, B*38:01 or C*12:03. Interestingly, the frequencies of two HLA alleles were significantly elevated in the study cohort compared to the reference database. The HLA allele A*26.01 had a frequency of F = 15.4% within the patients and of F = 3.6% within the reference population. Thereby, the allele A*26:01 had a significantly higher prevalence within the patient cohort (p value < 0.0001; OR = 4.9; 95% CI = 2.0 to 11.6). Additionally, the HLA allele C*12:03 displayed a significantly higher frequency within the study cohort with F = 17.9% in contrast to the reference with F = 6.3% (p value = 0.0028; OR = 3.3; 95% CI = 1.4 to 7.4). None of the investigated alleles had a significantly decreased frequency in the patient cohort compared to the reference dataset. The statistical results of all analyzed alleles are shown by the **Supplementary Tables 2, 3 and 4**.

According to these results, the HLA class I allelic distribution within the present study cohort was representative for most occurring alleles in comparison with a general German population with only two exceptions. The allele frequencies of HLA-A*26:01 and HLA-C*12:03 were significantly increased in the ependymoma patient cohort.

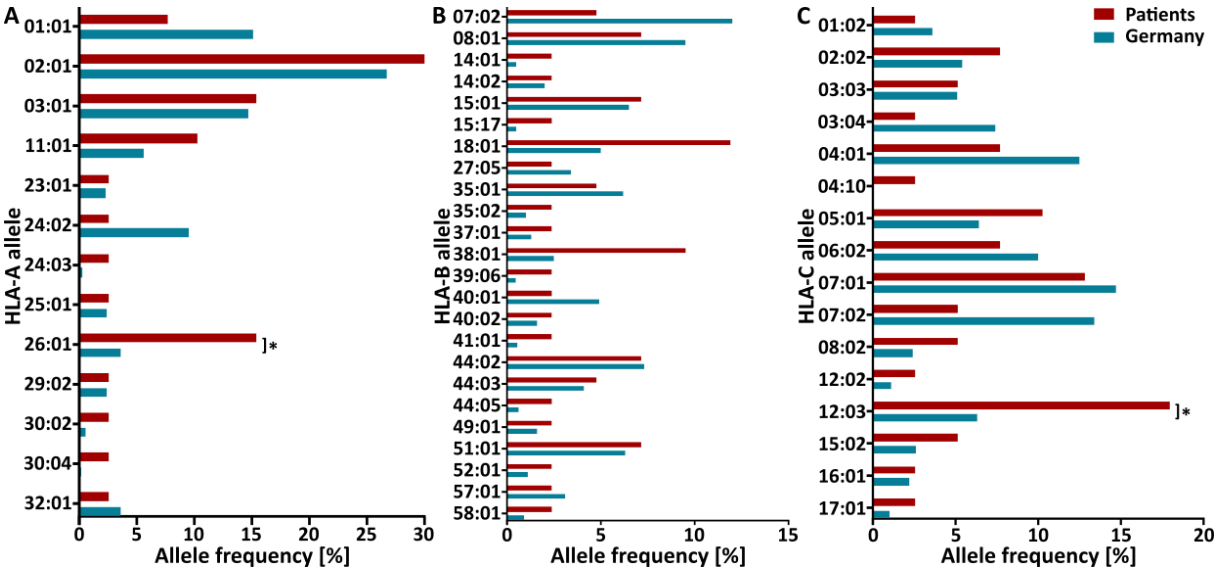


Figure 9: HLA class I allelic distribution in the ependymoma patient cohort and a German reference population. Allele frequencies of (A) HLA-A, (B) HLA-B and (C) HLA-C alleles occurring within the study cohort were determined. The Allele Frequency Net Database comprises the allele frequency data of the cohort "Germany pop 8" (n = 39,689), which served as reference dataset [43]. Most investigated alleles displayed allele frequencies comparable with those expected for a German population indicating a representative HLA class I allelic distribution in the patient cohort. Two deviations were revealed: HLA-A*26:01 (p value < 0.0001) and HLA-C*12:03 (p value = 0.0028) had a significantly higher prevalence within the patient cohort in comparison with the reference database. $p \leq 0.004$ for HLA-A, $p \leq 0.002$ for HLA-B and $p \leq 0.003$ for HLA-C alleles were applied as levels of statistical significance after Bonferroni correction. n(patients) = 21; n(Germany) = 39,689; * – significant

2.4.2 General characteristics of the ependymoma HLA ligandome

Peptide yields

HLA-presented peptides were isolated from ependymoma tissue samples by immunoaffinity chromatography using antibodies specific for HLA class I and class II molecules. Subsequently, isolated peptides were analyzed *via* LC-MS/MS. This revealed a dataset comprising peptides presented by HLA on the surface of ependymal tumor cells. **Figure 10** provides an overview of the yields for HLA class I- and class II-presented peptides in the 22 individual ependymoma samples. Altogether, 6,801 unique HLA class I-presented peptides derived from 4,671 source proteins and 3,855 unique HLA class II-presented peptides derived from 1,751 source proteins were detected.

HLA class I binder determination

Among the HLA class I ligandome, peptides which displayed a binding motif from one of the HLA allotypes of the respective donors were selected. These peptides were defined as HLA class I binders. In total, 6,185 unique binders were identified in the ependymomas. On this basis, the purity of each individual sample was calculated as percentage of binders among overall isolated HLA class I-presented peptides. Thereby, purities between 79% and 97% were revealed for the ependymoma samples of the

patient cohort **(Figure 10)**. Binder determination was not performed for the HLA class II ligandome due to the strong promiscuity of HLA class II molecules often resulting in an inaccurate assignment of peptides to specific HLA allotypes. All following analyses were performed on the basis of HLA class I binders and HLA class II-presented peptides.

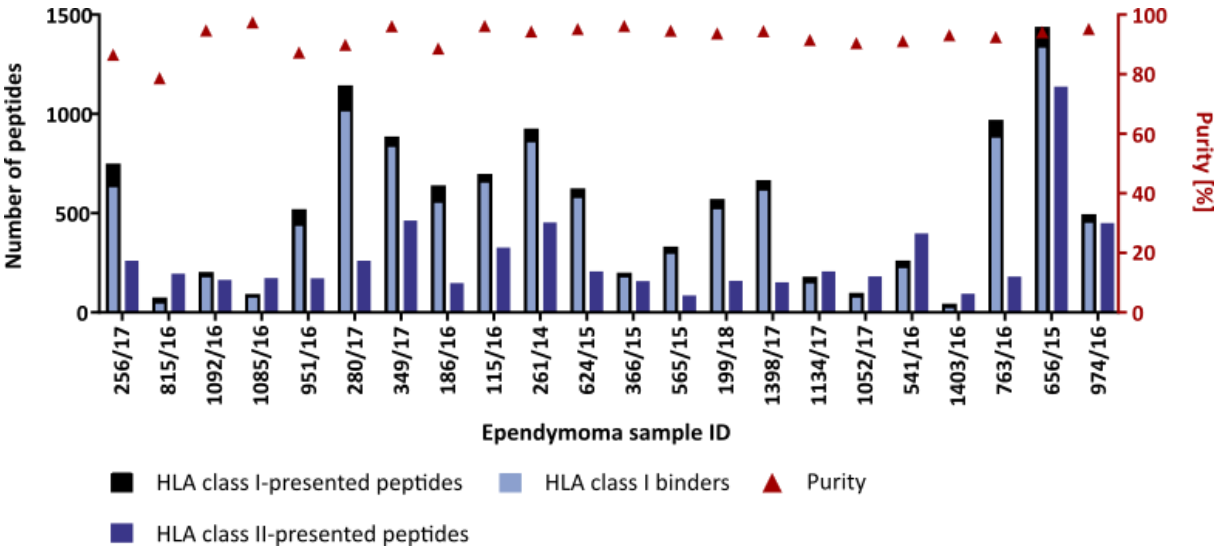


Figure 10: Yields of isolated HLA class I- and class II-presented peptides for individual ependymoma samples achieved by immunopeptidomic LC-MS/MS analysis. The peptide yields varied between 29 and 1,424 HLA class I- and between 83 and 1,127 HLA class II-presented peptides per individual sample. HLA class I binders were defined as HLA class I-presented peptides carrying a binding motif of an HLA allotype of the respective patient. The purity, which is defined as the proportion of binders among all isolated peptides, was determined for HLA class I and lay between 79% and 97% for the individual samples.

Correlation between sample masses and peptide yields

Peptide yields varied between individual samples from 27 to 1340 unique HLA class I binders and from 83 to 1127 unique HLA class II-presented peptides. The masses of the samples ranged from 30.2 mg to 459.6 mg **(Supplementary Table 5)**. Statistical analysis applying Pearson´s correlation was performed to observe a potential relation between available tissue masses and resulting HLA ligand yields. A positive correlation between sample masses and peptide yields was revealed for both HLA class I binders (p value = 0.0191; Pearson´s correlation coefficient r = 0.4951; 95% CI = 0.1 to 0.8) and HLA class II-presented peptides (p value = 0.0002; Pearson´s correlation coefficient r = 0.7112; 95% CI = 0.4 to 0.9). Linear regression analysis was applied due to the circumstance that the increase in the number of peptides for the present cohort size was still in the linear and not yet in the saturation phase **(Supplementary Figure 1)**. A p value ≤ 0.05 was used as level of statistical significance.

In conclusion, these results demonstrate that larger sample masses of ependymoma tissue enabled the identification of more unique HLA ligands.

Saturation analysis

In a saturation analysis, the size of the entire ependymoma HLA ligandome was estimated based on the identification rate of new unique antigens per additional sample. The present coverage rate was determined as percentage of the ligand count that was achieved with the investigated ependymoma samples among the maximal number of ligands (saturation value). A saturation value of 9,932 HLA class I binders was calculated. With the 6,185 presently analyzed unique binders, a proportion of 60% was covered **(Supplementary Figure 2A)**. For HLA class II-presented peptides, a saturation value of 8,475 peptides was determined. With 3,855 isolated HLA class II-presented peptides a proportion of 46% was covered **(Supplementary Figure 2C)**. The coverage rate was higher on the source protein level. A saturation value of 4,925 source proteins was calculated for HLA class I, of which 86% were covered by 4,228 identified source proteins **(Supplementary Figure 2B)**. For HLA class II source proteins, with a saturation value of 2,685 proteins and 1,751 identified proteins, a coverage rate of 65% was determined **(Supplementary Figure 2D)**.

Peptide length distribution

The peptide length distribution of isolated HLA ligands was observed within the length filter range applied during data processing. The majority of detected HLA class I binders corresponding to 71% had a peptide length of 9 amino acids (AAs) within the tolerated length range of 8-12 AAs **(Supplementary Figure 3A)**. The HLA class II-presented peptides were spread more widely within the allowed length range of 8-25 AAs and peaked at a length of 15 AAs corresponding to 18% **(Supplementary Figure 3B)**. These results coincide with the commonly expected length distribution for HLA class I and class II ligands [50, 51].

2.4.3 Ependymoma subgroup analysis

Overlaps of immunopeptidomes

As previously described, ependymomas are a group of heterogeneous tumors. According to the WHO classification scheme [4], the study cohort of the present project was categorized into four WHO grade I, sixteen WHO grade II and two WHO grade III tumors. Besides the histopathological grading scheme, a classification system based on tumor localization and moleculargenetic analyses has been established [23]. Therefore, the ependymal tumors of the present patient cohort were additionally subdivided according to their location into twelve spinal, eight infratentorial and two supratentorial tumors. Both supratentorial tumors were further characterized as RELA fusion-positive ependymoma on the genetic level.

To observe whether there are subgroup-associated differences in the immunopeptidomes of ependymomas, the immunopeptidomic dataset was divided according to both WHO grading and tumor localization and used for comparative analysis.

Comparison of the immunopeptidomes from ependymomas of different WHO grades revealed shared and exclusive antigens for all three subgroups on the levels of HLA class I binders, HLA class II-presented peptides and corresponding source proteins **(Figure 11)**. The overlap is smaller on the peptide than on the protein level, which is probably related to the HLA allotype variability between the different patients of the cohort. Additionally, the circumstance that both supratentorial tumors derived from the same patient might have a strong impact on the results of the comparative analysis and must be considered.

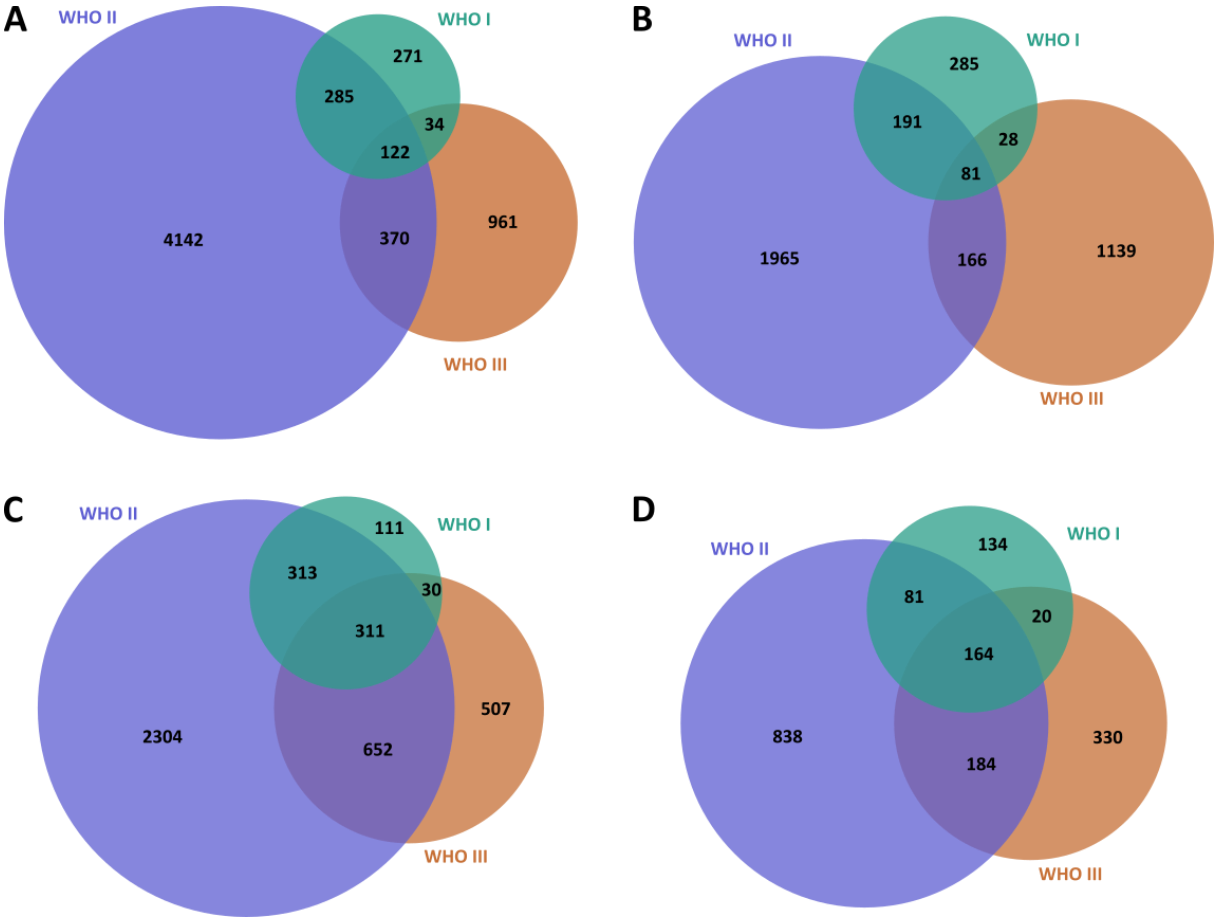


Figure 11: Comparison of HLA immunopeptidomes from ependymomas classified according to the WHO grading scheme. Data of subgroups were compared on the level of (A) HLA class I binders, (B) HLA class II-presented peptides, (C) source proteins of HLA class I binders and (D) source proteins of HLA class II-presented peptides. The overlap between the three groups was smaller on the peptide than on the protein level. n(WHO I) = 4; n(WHO II) = 16; n(WHO III) = 2; WHO – World Health Organization.

An equivalent comparison was performed based on ependymoma categorization according to the three different tumor locations: spinal, infratentorial and supratentorial **(Figure 12)**. The resulting pattern was similar to the comparison based on WHO grading. For HLA class I binders, HLA class II-presented peptides and corresponding source proteins, both shared and exclusive antigens were revealed in each subgroup. The overlap was smaller on the peptide level, which might again be related to the occurrence of various HLA allotypes within the patient cohort. Furthermore, as mentioned above, the influence of the two supratentorial tumors originating from the same donor patient might be an important factor for the results of the comparative analysis.

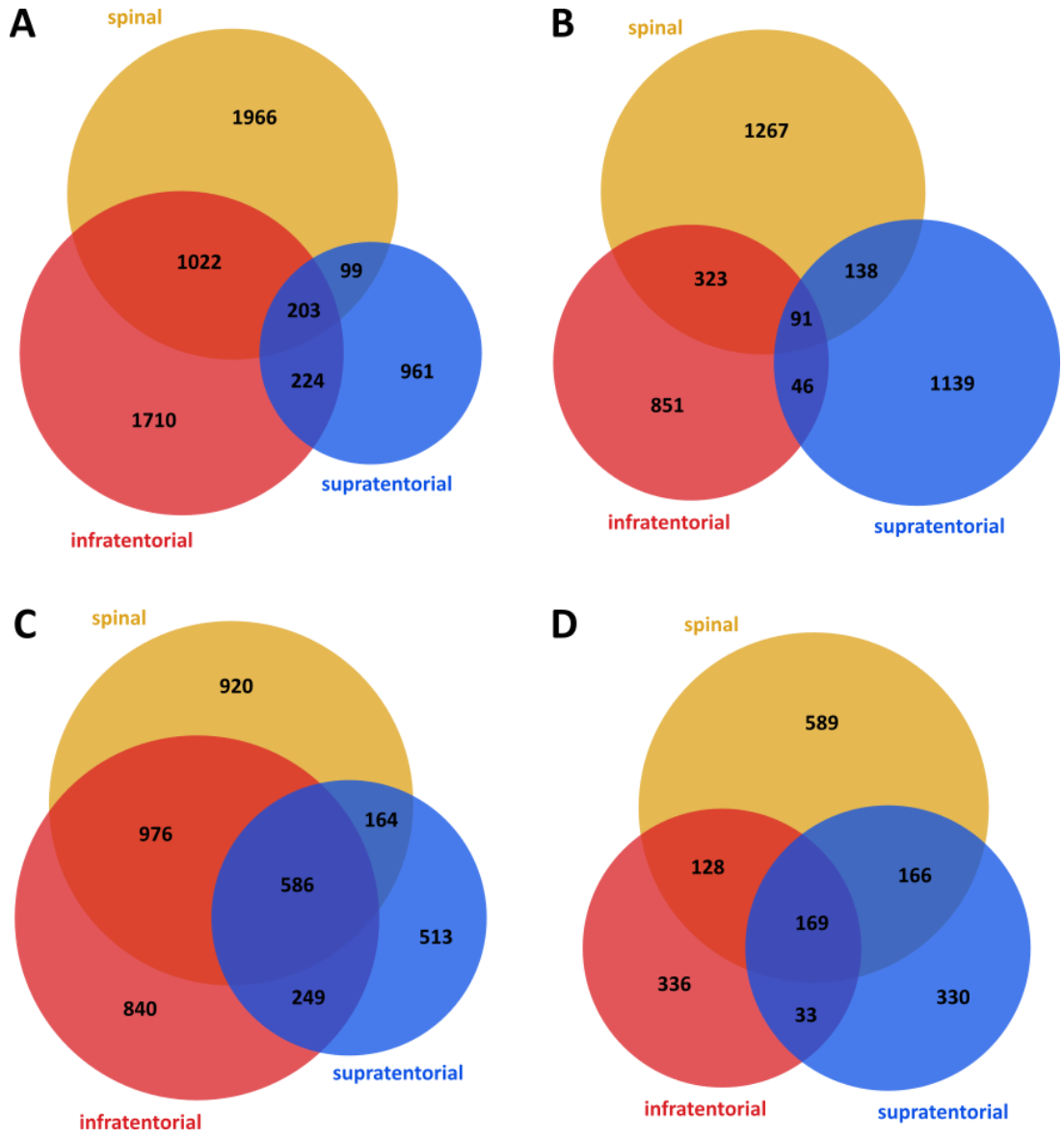


Figure 12: Comparison of HLA immunopeptidomes from ependymomas classified according to tumor localization. Data of subgroups were compared on the level of (A) HLA class I binders, (B) HLA class II-presented peptides, (C) HLA class I binder source proteins and (D) source proteins of HLA class II-presented peptides. The overlap between the three groups was smaller on the peptide than on the protein level. n(spinal) = 12; n(infratentorial) = 8; n(supratentorial) = 2.

Principle component analysis

To further analyze potential differences between the ependymoma subgroups, the immunopeptidomic data of the tumors were used for principle component analyses (PCAs), which were carried out by Leon Bichmann. **Figure 13** shows the scatter plot of the unsupervised PCA performed on the basis of the merged source proteins of HLA class I binders and HLA class II-presented peptides. One main cluster with only a few outliers was revealed. The samples did not clearly divide into distinct groups neither according to WHO grade nor according to tumor location These results are shown in more detail in the **Supplementary Figures 4 and 5** by scatter plots of unsupervised PCAs separately performed for HLA class I and class II. Only a slight trend of the WHO grade III, supratentorial tumor samples (sample IDs: 656-15 and 974-16) separating from the other samples can be assumed. The circumstance that both samples originated from the same patient must be considered in this context.

Since the PCAs did not reveal a clear clustering of ependymoma samples according to their assigned subgroups, all 22 tumor samples were treated as one cohort of ependymomas during further data analyses.

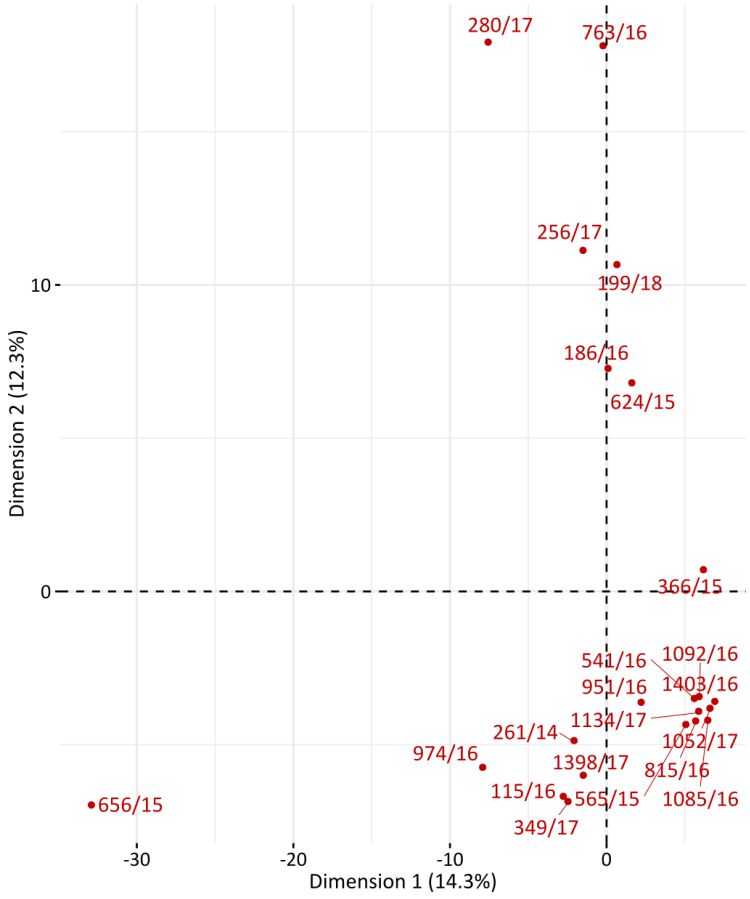


Figure 13: PCA based on immuno-peptidomic data of ependymomas. The merged source proteins of HLA class I binders and HLA class II-presented peptides served as basis for this unsupervised PCA. The main cluster in the lower right corner of the scatter plot as well as the outliers consisted of tumors from different subgroups, independent of WHO grade classification or tumor localization. The PCA was performed by Leon Bichmann. PCA – principle component analysis.

2.4.4 Identification of CTAs

Cancer-testis antigens (CTAs) are TAAs that are of interest for tumor immunotherapy development due to their characteristic expression and their immunogenicity [52]. A collection of described CTAs is comprised and accessible within the CTdatabase [45]. This database entails 277 CTAs (state August 2019) and served as basis to ascertain whether CTAs are presented by HLA molecules on the surfaces of ependymal tumor cells.

The ependymoma dataset was screened for HLA ligands derived from published CTAs as source proteins. This search revealed 19 CTA source proteins within the HLA class I dataset and 8 CTA source proteins within the HLA class II dataset. After quality control by antigen evaluation, 19 HLA class I binders derived from ten different CTAs and two HLA class II-presented peptides derived from two different CTAs remained **(Table 2, Supplementary Table 6)**. To select antigens that are universal for various patients, frequent antigens detected in at least three different samples (frequency ≥ 10%) were considered as antigens of interest. Four CTAs represented by HLA class I binders were identified in more than 10% of the tumor samples. This included the source proteins U3 small nucleolar ribonucleoprotein IMP3 (IMP3) represented by three different peptides in eight tumor samples, the ankyrin repeat domain-containing protein 45 (ANKRD45), of which the same peptide was found in six different samples, and the armadillo repeat-containing protein 3 (ARMC3) presented *via* two different peptides in three samples. Another frequently analyzed CTA was the lysine-specific demethylase 5B (KDM5B), which was found as source protein of four HLA class I binders and one HLA class II-presented peptide.

These results demonstrate that ependymomas present peptide sequences derived from previously described CTAs by both HLA class I and class II molecules on their cell surfaces. These antigens might bear the potential to serve as suitable targets for the development of novel immunotherapeutic approaches.

Table 2: CTAs detected within the ependymoma immunopeptidomic dataset. The search for CTAs was performed on the basis of the CTdatabase, which comprises 277 previously described CTAs (state August 2019) [45]. 19 HLA class I binders derived from CTAs as source proteins were revealed in the ependymoma dataset. Additionally, two identified HLA class II-presented peptides derived from CTAs. CTA – cancer-testis antigen; AC – accession; GN – gene name.

	Peptide			CTA source protein		
HLA class	**Sequence**	**HLA restriction**	**Represen-tation frequency**	**UniProt AC**	**GN**	**Represen-tation frequency**
I	ALLDKLYAL RSMEDFVTW EDYTRYNQL	A*02:01 B*58:01 B*08:02;C*14:02	32% 5% 5%	Q9NV31	IMP3	36%
	DVIRALAKY	A*25:01;A*26:01	27%	Q5TZF3	ANKRD45	27%
	SVAQQLLNGK ILNPYNLFL DPFAFIHKI DDWDNRTSY	A*03:01 A*02:01 B*51:01 B*18:01	5% 5% 14% 5%	Q9UGL1	KDM5B	23%
	IINDGFYDY KLPDFSWEL	A*01:01 A*02:01	5% 9%	Q5W041	ARMC3	14%
	SEFKAMDSF NEQDLANRF	B*40:02 B*44:05	5% 5%	P26232	CTNNA2	9%
	VEFPYQYDF DYPRYLFEI	B*18:01 A*24:02	5% 5%	Q14667	KIAA0100	9%
	RLFVTSGGLKK YPEEIVRYY	A*03:01 B*35:01	5% 5%	O75602	SPAG6	9%
	SLIQKVETY	A*02:01	5%	Q8TBZ0	CCDC110	5%
	DAVKFFVAV	B*51:01	5%	O60271	SPAG9	5%
	SFYEHIITV	B*52:01	5%	Q5T6F0	DCAF12	5%
II	DPFAFIHKI	-	9%	Q9UGL1	KDM5B	9%
	PFHIFKVKVTTERERMENIDSTIL	-	5%	Q86X24	HORMAD1	5%

2.4.5 Identification of ependymoma TAAs

To find novel TAAs of ependymomas, in addition to already published CTAs, a healthy reference dataset is necessary. In case of most brain tumors, it is impossible to resect a healthy tissue sample while removing the tumor without causing irreversible tissue damage. Thus, no corresponding benign samples of the ependymoma patients were available for the present study and an in-house benign database **(Appendix, Table 15)** was consulted as reference dataset.

Peptide level

By comparative profiling of HLA-presented peptides against the benign database, those peptides remained that have not been found on benign tissue so far. These peptides are further referred to as tumor-exclusive peptides. With the comparison on the peptide level, 889 tumor-exclusive HLA class I binders and 1,283 tumor-exclusive HLA class II-presented peptides were found within the ependymoma cohort **(Figure 14)**. Additionally, the frequencies of the detected peptides within the study cohort and the benign database are compared in the waterfall plots. To assure a universality of potential TAAs across different patients, peptides of interest were those that were tumor-exclusive and occurred in the patient cohort with a frequency of at least 10% (≥ 3 samples). 37 tumor-exclusive HLA class I binders and 11 tumor-exclusive HLA class II-presented peptides met these requirements **(Supplementary Tables 7 and 8)**. These frequently detected, tumor-exclusive HLA ligands and their corresponding source proteins were subsequently evaluated in more detail.

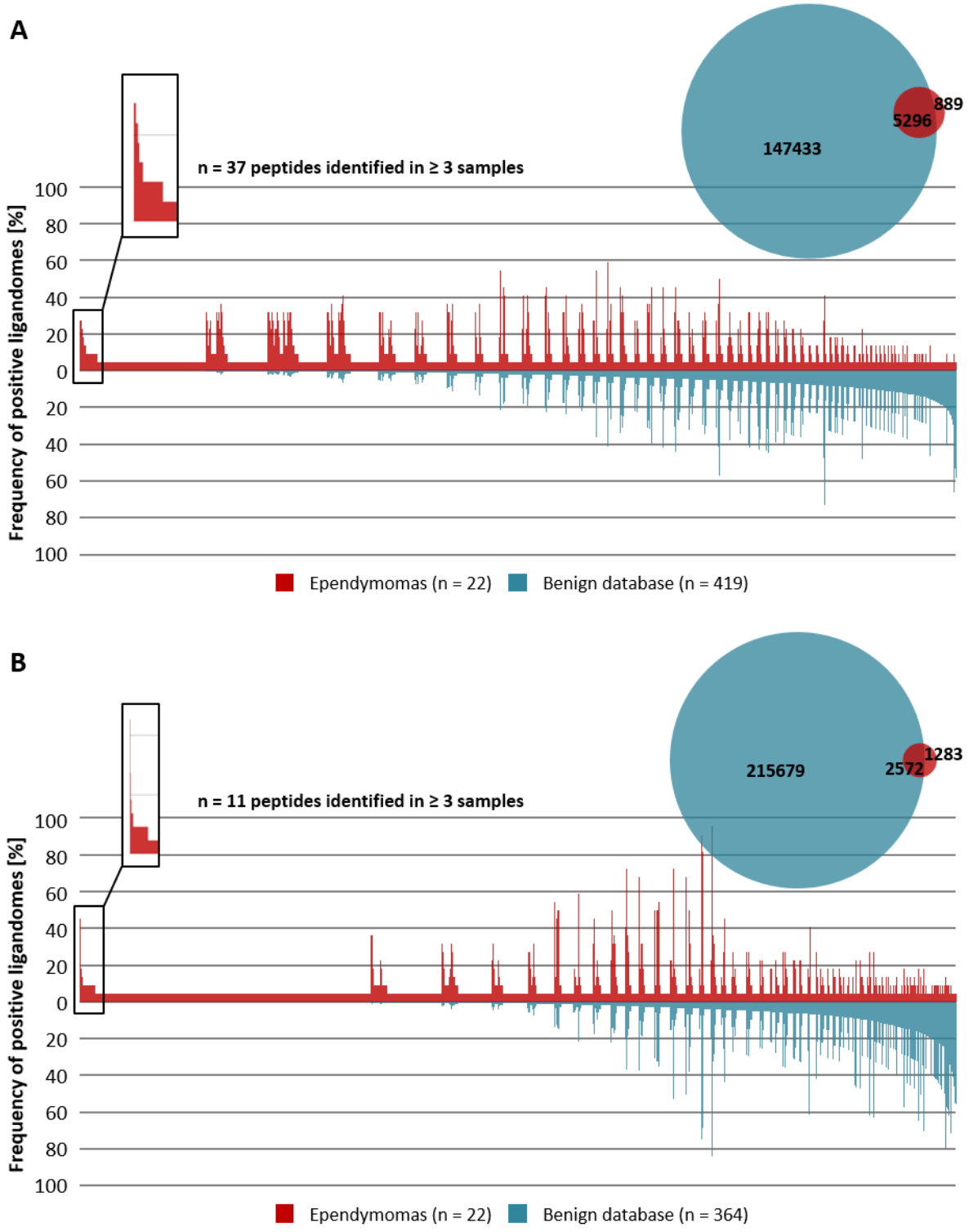


Figure 14: Comparative profiling of the HLA ligandomes of ependymomas against the in-house benign database on the peptide level. The analysis was performed for both (A) HLA class I binders and (B) HLA class II-presented peptides. This revealed tumor-exclusive peptides as shown by the Venn diagrams. The waterfall plots demonstrate the frequencies of detected peptides within the study cohort and the benign database in comparison. Unique ligands are plotted on the x-axis and the corresponding percentage of HLA ligandomes that contain the respective ligands on the y-axis. Tumor-exclusive peptides that occurred among the ependymomas with a frequency of at least 10% (≥ 3 samples) were of further interest and are highlighted on the far left of the waterfall plots.

Antigen evaluation was performed to select potential TAAs. First, the source proteins of highly frequent, tumor-exclusive peptides were observed by analyzing their RNA expression profiles. For this purpose, the databases Genotype-Tissue Expression Project (GTEx) [46] and The Human Protein Atlas [47] were consulted to receive the RNA expression profiles in different healthy tissues and tumor entities, respectively. Peptides derived from source proteins with strong RNA expression in several healthy tissues were excluded as TAAs. Some of the source proteins showed brain- and, thereby, glioma-restricted RNA expression profiles, which can be ascribed to a lack of sufficient and suitable control tissue samples in the benign database comprising only twelve brain samples. This led to a probably incorrect labeling of frequent HLA ligands as tumor-exclusive that derived from the following brain-specific proteins: glial fibrillary acidic protein (GFAP), zinc finger protein ZIC 2 (ZIC2), hepatocyte nuclear factor 3-alpha (FOXA1), paired box protein Pax-6 (PAX6), BDNF/NT-3 growth factors receptor (NTRK2), protein FAM135B (FAM135B), cLIP-associating protein 2 (CLASP2), BAI1-associated protein 3 (BAIAP3), zinc transporter ZIP12 (SLC39A12), glutamate receptor ionotropic, delta-2 (GRID2) and endothelin B receptor-like protein 2 (GPR37). For further inclusion of an antigen, the RNA expression of a source protein was required to be rather low in all healthy tissues (< 10 TPM) or exclusively elevated in testis as shown for previously described CTAs. Therefore, peptides from source proteins with CTA-like RNA expression profiles were of further interest during the search for novel TAAs.

The ten remaining antigens were further evaluated on the peptide level. For this purpose, several technical quality criteria were checked to confirm the peptide sequences that were annotated to the experimental MS spectra during data processing. Seven HLA class I binders and one HLA class II-presented peptide with a frequency above 10% met all quality requirements **(Table 3)**. They are suggested as potential TAAs. Interestingly, five of the six HLA class I-presented TAAs had an HLA-A*25:01 or HLA-A*26:01-restricted binding motif. They were found in 50% to 86% of samples with the respective allotypes. This indicates a relation to the high prevalence of the HLA allele A*26:01 within the patient cohort (see 2.4.1). Still, the circumstance that the benign database does only include ten HLA-A*26-positive samples overall and no HLA-A*26-positive brain samples, must be considered in this context.

Furthermore, the set of tumor-exclusive peptides included five of the priorly detected HLA class I binders and one HLA class II-presented peptide derived from previously published CTAs (see 2.4.4). This concerned only antigens with low frequencies appearing in just one ependymoma sample **(Supplementary Table 6)**.

Table 3: TAAs identified in ependymomas. After comparative profiling against an in-house benign database and detailed evaluation of the resulting tumor-exclusive peptides and corresponding source proteins, six HLA class I binders and one HLA class II-presented peptide remained as novel ependymoma TAAs. TAA – tumor-associated antigen; AC – accession; GN – gene name.

HLA class	Peptide				Source protein		
	Sequence	HLA restriction	Representation frequency in		UniProtKB AC	GN	Protein name
			cohort	allotype-positive samples			
I	**EVIEKTSYL**	A*25:01 A*26:01	27%	86%	Q9P1Z9	CCDC180	Coiled-coil domain-containing protein 180
	EVLNGQVSKY	A*25:01 A*26:01	27%	86%	Q9C0G6	DNAH6	Dynein heavy chain 6, axonemal
	EVTERLGEF	A*25:01 A*26:01	27%	86%	Q9C0G6	DNAH6	Dynein heavy chain 6, axonemal
	FLDSQITTV	A*02:01	27%	50%	Q6V702	CFAP299	Cilia- and flagella-associated protein 299
	ETVDENGRLY	A*25:01 A*26:01	23%	71%	Q8IYE1	CCDC13	Coiled-coil domain-containing protein 13
	EVFDGTVIREL	A*25:01 A*26:01	23%	71%	Q8N1V2	CFAP52	Cilia- and flagella-associated protein 52
	EVYDVSISEF	A*25:01 A*26:01	14%	43%	Q8TDW7	FAT3	Protocadherin Fat 3
II	**EIVLTQSPATLSLSPGER**	-	14%	-	P04433	IGKV3-11	Ig kappa chain V-III region VG (Fragment)

Source protein level

Comparative profiling of the ependymoma immunopeptidomic dataset against the in-house benign database was additionally performed on the source protein level. This analysis is supposed to reveal source proteins, of which no peptides were detected on benign tissue before, hereafter termed tumor-exclusive source proteins.

55 tumor-exclusive source proteins represented by HLA class I binders and 68 tumor-exclusive source proteins represented by HLA class II-presented peptides were identified **(Figure 15)**. The frequencies of source proteins within the ependymoma cohort and the benign database are illustrated by waterfall plots. Three of the tumor-exclusive HLA class I and none of the HLA class II source proteins occurred with a frequency of more than 10% among the patients. After detailed evaluation on the source protein and peptide level, only the cilia- and flagella-associated protein 299 (CFAP299) remained as potential TAA. CFAP299 was found represented by the HLA class I binder FLDSQITTV on six different tumor samples. This peptide has already been indicated as potential TAA during comparative profiling on the peptide level **(Table 3)**.

In conclusion, comparative profiling against the in-house benign database on the peptide level enabled the identification of several ependymoma TAAs. On the source protein level, no additional TAAs were identified but the results partly confirmed the previous analyses on the peptide level.

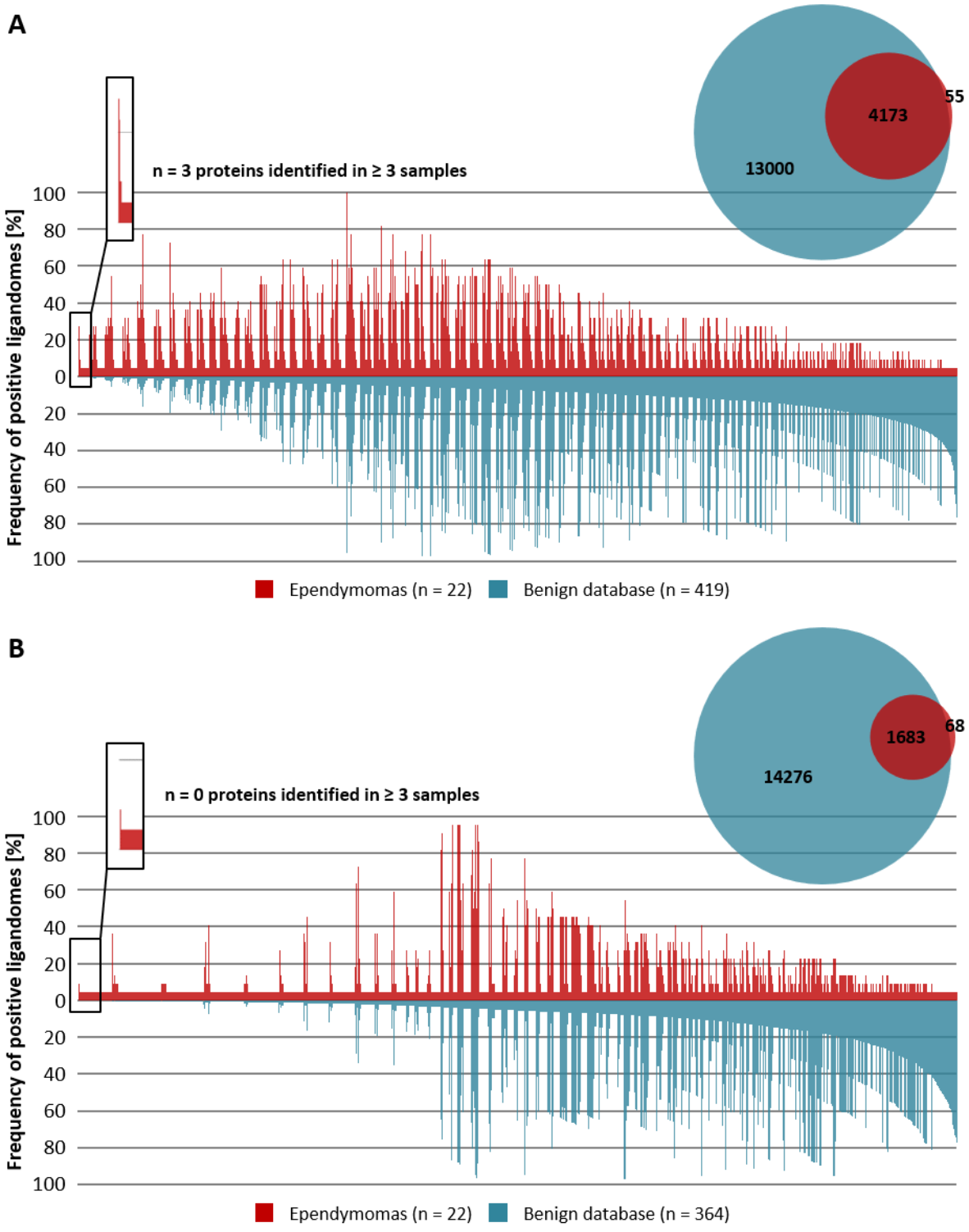


Figure 15: Comparative profiling of the HLA ligandomes of ependymomas against the in-house benign database on the source protein level. The analysis was performed for source proteins of both (A) HLA class I binders and (B) HLA class II-presented peptides. This analysis revealed several tumor-exclusive source proteins as shown by the Venn diagrams. The waterfall plots demonstrate the frequencies of detected source proteins within the study cohort and the benign database in comparison. Unique ligands are plotted on the x-axis and the corresponding percentage of HLA ligandomes that contain the respective ligands on the y-axis. Tumor-exclusive source proteins that occurred among the ependymomas with a frequency of at least 10% (≥ 3 samples) were of further interest and are highlighted on the far left of the waterfall plots.

2.4.6 Ependymoma TAAs in other malignancies

Comparative profiling of the ependymoma HLA ligandome dataset against the in-house benign database revealed tumor-exclusive antigens, which served as basis for the selection of potential TAAs (see 2.4.5). Subsequently, an in-house malignant database **(Appendix, Table 16)** was consulted, against which comparative profiling of the tumor-exclusive antigens was performed. Comparison with this malignant database aimed to observe the prevalence of the previously defined tumor-exclusive antigens and TAAs of ependymomas in the immunopeptidomes of other malignancies.

Thereby, 608 tumor-exclusive HLA class I binders and 580 tumor-exclusive HLA class I-presented peptides were revealed that have already been found within the HLA ligandomes of other tumor samples **(Figure 16A and C)**. This included four peptides earlier selected as potential ependymoma TAAs: EVIEKTSYL (CCDC180) has been found on one atypical teratoid rhabdoid tumor (ATRT) and two ovarian carcinomas, EVLNGQVSKY (DNAH6) on one ovarian carcinoma, FLDSQITTV (CFAP299) on one subependymoma and one ovarian carcinoma, ETVDENGRLY (CCDC13) on one acute myeloid leukemia (AML) sample and EVYDVSISEF (FAT3) on two glioblastomas (GBM). 281 tumor-exclusive HLA class I binders and 703 tumor-exclusive HLA class II-presented peptides were not represented in the malignant database including the potential ependymoma TAAs EVTERLGEF (DNAH6), EVFDGTVIREL (CFAP52) and EIVLTQSPATLSLSPGER (IGKV3-11) **(Figure 16B and D)**. Most HLA class I TAAs were shown to be restricted to the HLA allotype A*26:01. The malignant database comprised 40 samples derived from HLA-A*26:01-positive donors.

On the protein level, 52 tumor-exclusive source proteins of HLA class I binders have been found to be HLA-presented in other tumors before. This included the potential TAA CFAP299, which has been found on one ovarian carcinoma and one subependymoma represented by the same peptide as on the ependymomas and on one glioblastoma represented by another peptide. 3 tumor-exclusive antigens were not identified as source proteins in other malignancies before. 48 tumor-exclusive source proteins of HLA class II-presented peptides were already represented within the malignant database whereas 20 were not found on other tumor entities before.

Despite the low abundance of HLA-A*26:01-positive samples within the in-house malignant database, some antigens that were selected as potential TAAs of ependymomas (see 2.4.5) have been already detected in other cancer entities including some CNS tumors. This further supports their importance and consideration in the course of developing novel immunotherapeutic approaches for the treatment of ependymoma patients.

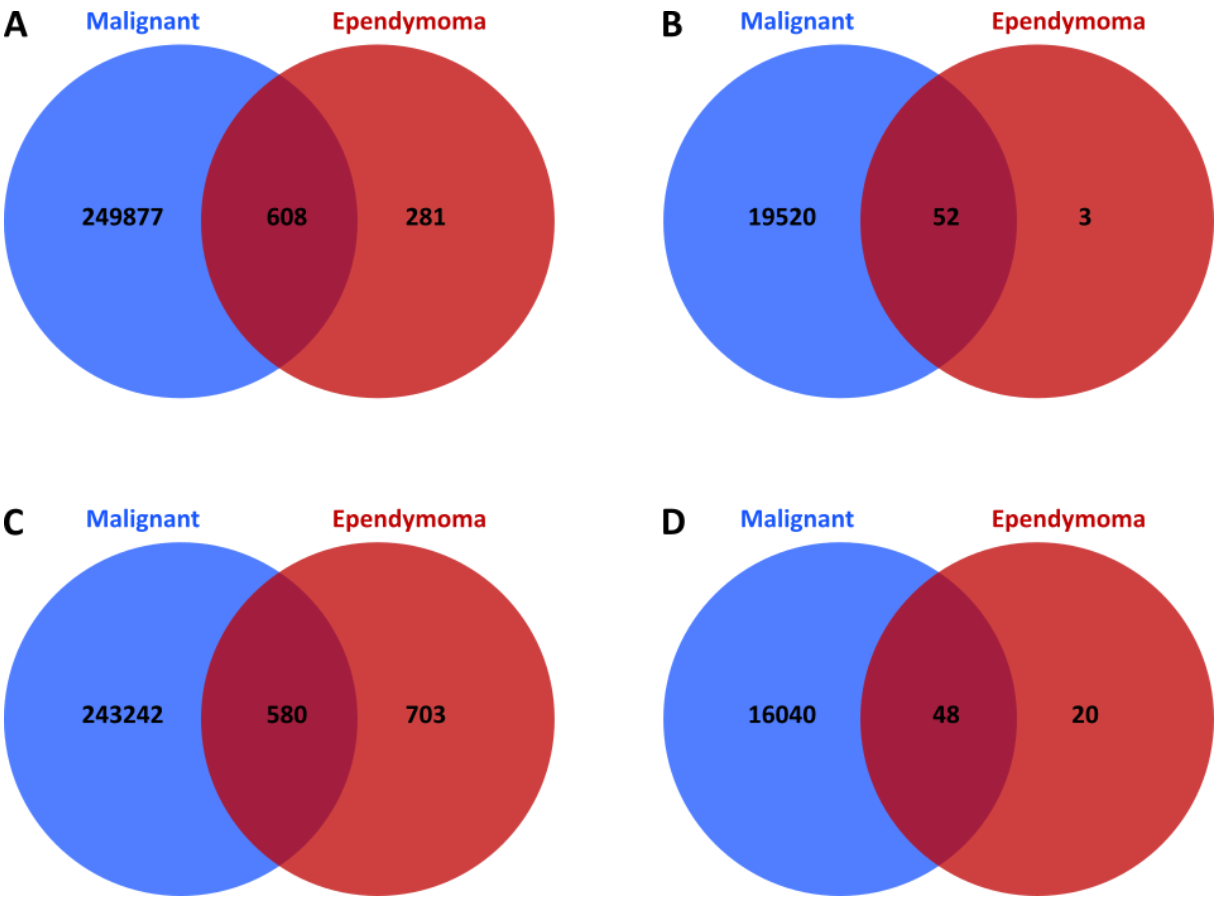


Figure 16: Comparison of tumor-exclusive antigens with an in-house malignant database. The tumor-exclusive antigens identified in the immunopeptidomic dataset of ependymomas were compared with an immunopeptidomic in-house malignant database. The Venn diagrams show the overlap between the malignant database and tumor-exclusive (A) HLA class I binders, (B) source proteins of HLA class I binders, (C) HLA class II-presented peptides and (D) source proteins of HLA class II-presented peptides. n(malignant HLA class I samples) = 773; n(malignant HLA class II samples) = 560; n(ependymomas) = 22.

2.4.7 Functional annotation clustering of tumor-exclusive antigens

Comparative profiling against the in-house benign database on the protein level revealed several tumor-exclusive source proteins (see 2.4.5) which might be involved in tumor-associated cellular processes. To investigate functional relations of these proteins, functional annotation clustering was performed with 55 and 68 tumor-exclusive source proteins for HLA class I and class II, respectively. The functional annotation tool of the database DAVID Bioinformatics Resources 6.8 [48, 49] was used for gene ontology (GO) term enrichment considering GO terms of biological processes (BP), molecular functions (MF) and cellular compartments (CC). Functional annotation clustering with tumor-exclusive source proteins of HLA class I binders revealed one annotation cluster with an enrichment score above 1 (enrichment score = 18.89) generated on the basis of 31 source proteins **(Table 4)**. This annotation cluster included BP, MF and CC GO terms associated with deubiquitination of proteins in the endoplasmic reticulum and the nucleus. The number and compilation of tumor-exclusive source

proteins of HLA class II-presented peptides were not sufficient to deliver results during functional annotation clustering.

Table 4: Result of GO term enrichment based on tumor-exclusive source proteins of HLA class I binders as identified in ependymomas. The functional annotation tool of the database DAVID Bioinformatics Resources 6.8 was applied [48, 49]. A minimal enrichment score of 1 was used as inclusion threshold. One annotation cluster with an enrichment score above 1 was annotated to 31 of the 55 tumor-exclusive source proteins. GO = gene ontology; AC – accession.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB ACs of source proteins		Proportion of tumor-exclusive antigens
Protein deubiquitination	18.89	GO:0036459 GO:0016579 GO:0006511 GO:0005783 GO:0005634	C9JPN9 C9J2P7 D6RCP7 P0C7I0 C9JJH3 D6R901 D6RJB6 Q0WX57 A6NCW7 Q6QN14 D6RBQ6 C9JLJ4 Q7RTZ2 C9JVI0 A6NCW0 D6R9N7	A8MUK1 D6RA61 Q8N5I3 Q49MI3 A6NJV1 Q8NCR6 Q9HCC6 Q8TBZ0 P26367 O60481 Q13536 Q6PJQ5 Q96PF2 Q8WXU2 C9JE40	56%

2.4.8 Single patient analysis of two recurrent ependymomas

The ependymoma cohort included the two RELA fusion-positive ependymomas 656/15 and 974/16 originating from the same patient. Ependymoma 656/15 was the first recurrent tumor of this patient whereas ependymoma 974/16 was surgically resected with an interval of 13 months. These two recurrences enabled the investigation of potential differences on the immunopeptidomic level that might emerge in tumors of one patient with temporal distance.

To observe this, the HLA-presented antigens identified in the two tumors were compared. The comparative analysis of the two datasets revealed shared but also unique HLA-presented antigens in both tumors on the peptide and the source protein level **(Figure 17)**. The overlap between the HLA

ligandomes was higher for HLA class I ligands than for HLA class II ligands. Despite sample 974/16 having a higher mass than sample 656/15 with 460 mg and 251 mg, respectively, the number of identified HLA ligands was lower in sample 974/16 **(Table 5)**. In 974/16, the number of HLA class I binders was decreased for all HLA allotypes of the patient. The loss of binders was slightly higher for HLA-B*07:02, -B*18:01, -C*07:02 and -C*12:03 than for HLA-A*03:01 **(Supplementary Figure 6)**. Furthermore, tumor 974/16 presented less tumor-exclusive antigens that were previously defined by comparison of the ependymoma dataset with the in-house benign database (see 2.4.5) than tumor 656/15. None of the peptides selected as potential TAAs were found in either of the two samples. This might be due to the circumstance that the patient was not positive for the respective HLA alleles A*02:01, A*25:01 or A*26:01, to which the HLA class I-presented TAAs were restricted. The following three HLA class I binders derived from CTAs (see 2.4.4) were found in sample 656/15: SVAQQLLNGK (KDM5B), DDWDNRTSY (KDM5B), VEFPYQYDF (KIAA0100). In contrast, immunopeptidomic analysis of sample 974/16 did not reveal any HLA-presented CTA **(Table 2)**.

This single patient analysis of two recurrent ependymomas indicates differences between the immunopeptidomes of temporally distant tumors occurring in one patient. Comparative analyses of further paired samples are necessary to confirm these observations.

Table 5: Comparison of HLA ligand yields achieved in two recurrent RELA fusion-positive ependymomas originating from one patient. The tumors were surgically resected with a temporal distance of 13 months. The sample masses were 460 mg and 251 mg for the first recurrent tumor 656/15 and the second recurrent tumor 974/16, respectively. However, the yields of HLA class I binders, HLA class II-presented peptides and corresponding source proteins were decreased in 974/16 compared to 656/15.

HLA class	Number of	656/15	974/16
I	Binders	1340	458
	Source proteins	1384	562
	Tumor-exclusive binders	129	35
	Tumor-exclusive source proteins	8	2
II	Peptides	1127	439
	Source proteins	579	288
	Tumor-exclusive peptides	423	90
	Tumor-exclusive source proteins	9	3

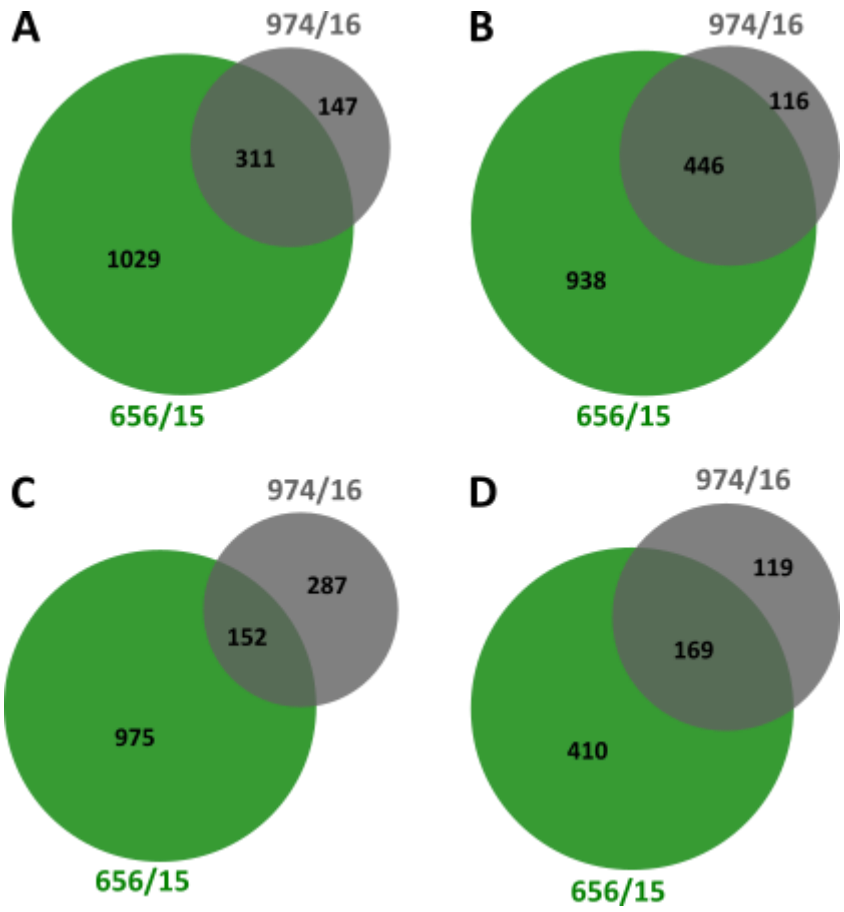


Figure 17: Comparative analysis of the HLA ligandomes of two recurrent ependymomas originating from one patient. The two RELA fusion-positive ependymoma samples 656/15 and 974/16 were recurrences and surgically resected from the same patient within an interval of 13 months. More shared (A) HLA class I binders and (B) source proteins of HLA class I binders than (C) HLA class II-presented peptides and (D) source proteins of HLA class II-presented peptides were identified among the two tumors.

2.5 Discussion

Ependymomas are CNS tumors occurring throughout all age groups. Due to their rarity, little is known about natural immune responses against ependymomas and there are only few attempts regarding immunotherapeutic approaches. In the present study, a cohort of ependymoma tumor samples was analyzed by LC-MS/MS on the immunopeptidomic level with the aim to identify potential target antigens for the development of novel immunotherapeutic strategies.

One interesting observation was already made by examining the HLA class I allelic distribution across the included patients of the study cohort. By statistical analysis, a significantly higher prevalence within the patient cohort compared to the German reference population was shown for the two alleles HLA-A*26:01 and HLA-C*12:03. This indicates an association between carrying one of these alleles and the probability of developing an ependymal malignancy. The term HLA-disease association describes the phenomenon that a disease has higher incidence in combination with the occurrence of a particular HLA allele. Such correlations have been shown for several combinations of diseases and HLA alleles [53]. Examples are correlations between ankylosing spondylitis and HLA-B*27:02 or HLA-B*27:05 [54], between narcolepsy and HLA-DQB1*06:02 [55] or between rheumatoid arthritis and HLA-DRB1*04:01 [56]. So far, only few associations have been made between cancerous diseases and specific HLA alleles. The nonclassical HLA-E molecule and its two alleles HLA-E*01:01 and HLA-E*01:03 were associated with several kinds of diseases including viral and bacterial infections, autoimmune disorders and also cancer [57]. Expression of HLA-E*01:03 is assumed to support the escape of cancer cells from the immune system. Thus, HLA-E*01:03-positve individuals are more susceptible of developing cervical cancer and acute leukemia than HLA-E*01:01-positve individuals [58, 59]. The data of the HLA class I allelic distribution analysis from the present ependymoma cohort suggest a predisposition to the emergence of an ependymoma for persons carrying the alleles HLA-A*26:01 or HLA-C*12:03. Certainly, these results need further confirmation by studies based on larger patient cohorts. But still, they indicate a novel potential HLA-disease association between the HLA alleles A*26:01 or HLA-C*12:03 and the development of ependymal malignancies.

Due to their heterogeneity, ependymomas are categorized into several tumor subgroups based on clinical, histopathological or molecular characteristics according to different classification systems. Classical grading uses histopathological criteria to assign ependymoma tumors to WHO grades I, II and III [4]. Presently, the tumor location and molecular features of ependymomas gain importance for tumor subgrouping in clinical application and research studies [23]. Merchant et al. (2019) and Upadhyaya et al. (2019) integrated the analysis of histopathologic grade, focal copy number gain on chromosome 1q and DNA methylation profiles into classification of children´s ependymomas. Both studies demonstrated significant variations in patient outcome in correlation with the assignment of

the tumors to different ependymoma subgroups [60, 61]. Such differences between distinct ependymoma subgroups on histopathological, molecular and location level also indicate an impact on the antigenic repertoire and the immunopeptidome of the tumors. The ependymoma cohort of the present study included samples of WHO grades I, II and III that derived from three different anatomic sites of the CNS. Even though some antigens were exclusively identified in ependymomas of a particular subgroup, the PCA did not reveal an obvious clustering of the samples according to their WHO classification or localization. However, the small cohort size does not allow definite conclusions about subgroup-specific characteristics on the HLA ligandome level. Each subgroup contained only two to sixteen samples from patients with different HLA allotype combinations. Another considerable factor is the circumstance that the two supratentorial tumors derived from the same patient. Thus, potential similarities or differences between the HLA ligandomes can rather be ascribed to patient or HLA allotype specificities than to the tumor classification. Since no clear clustering into the ependymoma subgroups was observed based on immunopeptidomic data, all 22 tumor samples were treated as one cohort in the present study. An enlargement of the cohort size and an inclusion of HLA allotype-matching samples from different patients are necessary to reveal potentially relevant subgroup-associated characteristics within the HLA ligandomes of ependymomas.

During the present study, the main objective was the search for TAAs within the ependymoma immunopeptidome that may serve as suitable target antigens for the development of novel immunotherapeutic strategies.

Three potential TAAs of ependymomas have already been described and suggested as targets for cancer vaccination therapy by Yeung *et al.* (2013) after demonstrating their enhanced expression in tumor tissue: ephrin type-A receptor 2 (EphA2), interleukin-13 receptor subunit alpha-2 (IL-13Rα2) and survivin [40]. None of them were represented by any HLA-presented peptides within the ependymoma immunopeptidomes of the present cohort and were thus not confirmed as potential TAAs. Protein expression data were not available for the present study so that alterations in the expression levels of these three proteins could not be analyzed. It must be considered that the correlation between protein expression and HLA-presentation is commonly moderate and high expression levels of proteins are no guarantee for their profound HLA-presentation [62]. Therefore, MS analysis is considered as method of choice to reliably confirm the HLA-presentation of a specific antigen [42]. Nevertheless, an HLA-presentation of the three mentioned antigens cannot be completely excluded. Antigens with a low abundance among the entire HLA ligandome might be below the detection level of the applied method. Still, a possibly low presentation rate of the three antigens might impede their recognition by T cells, which is no indication for their selection as potential immunotherapeutic targets.

Previously described cancer-testis antigens (CTAs) [45] were expected to display a more promising foundation during the search for potential immunotherapeutic targets in ependymomas. CTAs are a

group of proteins with a characteristic expression pattern, which is restricted to the human germ line under normal conditions and to several malignant tumors under the conditions of cancerous disease. They are considered as TAAs. Due to their immunogenicity, they play a crucial role for immunotherapy development [52].

Several HLA-presented peptides derived from CTAs were revealed in the ependymoma immunopeptidomic dataset. Many studies approved the suitability of CTAs as targets for cancer vaccination, which is why immunotherapeutic application of CTAs is subject of many different clinical trials [63]. Among the CTAs frequently identified in the ependymomas, especially the U3 small nucleolar ribonucleoprotein IMP3 has been addressed in several publications. A promising HLA-A*24-restricted epitope (KTVNELQNL) derived from IMP3 has been predicted and has shown to induce immune responses in lung and esophageal cancer [64]. Despite the ependymoma cohort containing two HLA-A*24-positive patients, none of the IMP3-derived HLA-presented peptides that were found in the ependymoma samples contained the epitope described by Suda *et al.* (2007) [64]. Furthermore, some studies addressed the search for CTAs in tumors of the CNS. An upregulated CTA expression has been shown to be accompanied by enhanced gliomagenesis in children with high-grade gliomas. This applied to the CTAs variable charge X-linked protein 3 (VCX3A) and IL-13Rα2 [65]. Moreover, Shraibman *et al.* (2019) identified HLA-presented peptides derived from well-known CTAs such as NY-ESO-1 and MAGE-A1 *via* LC-MS/MS in glioblastoma [66]. None of these CTAs that were associated with CNS malignancies were found within the ependymoma cohort. However, the studies show that CTAs play an important role in tumors of the CNS. Hence, CTAs identified in the ependymoma immunopeptidome might as well be of importance.

To identify novel TAAs additionally to previously described CTAs, comparative profiling of the ependymoma immunopeptidomic dataset against a benign database was performed and revealed a set of tumor-exclusive antigens. Only few of the HLA-presented CTAs found in the ependymomas were among the tumor-exclusive antigens. This observation indicates that many of the identified CTAs are also expressed and HLA-presented in benign tissue. Thus, their status as ependymoma TAAs and their suitability as targets for immunotherapeutic approaches require further examination to avoid potential on-target/off-tumor side effects.

Many tumor-exclusive peptides of the ependymoma immunopeptidome presented by HLA class I were restricted to the HLA allotype A*26:01. On the one hand, this supports the hypothesis of an HLA-disease association between HLA-A*26:01 and ependymoma. On the other hand, it must be considered that the in-house benign and malignant databases used for comparative profiling contained very few samples of HLA-A*26:01-positive donors. Only 2% (≙ 10 samples) of the samples included in the benign database and 7% (≙ 40 samples) of the samples included in the malignant database originated from HLA-A*26:01-positive donors. This can be ascribed to the low HLA-A*26:01

allele frequency of 3.6% within a normal German population (Allele Frequency Net database) [43]. None of the brain samples included in the benign database derived from an HLA-A*26-positive donor. In conclusion, the observed enrichment of HLA-A*26:01-restricted tumor-exclusive peptides and TAAs in the present ependymoma cohort might partly be related to a potential HLA-disease association. However, the lack of suitable HLA allotype-matching reference data might have been an essential impact factor leading to these results.

Nevertheless, the set of tumor-exclusive antigens beared the potential to contain promising TAAs and served as basis for further analysis. Functional annotation clustering was performed to reveal possible involvements of the tumor-exclusive source proteins in mechanisms of cancer development. GO term enrichment showed an involvement of source proteins from HLA class I binders in deubiquitination mechanisms. Proper functioning of (de)ubiquitination is essential for various cellular processes such as cell survival and differentiation as well as innate and adaptive immunity. In case of dysregulated ubiquitination systems, E3 ligases and deubiquitinases can operate as oncogenes and, thereby, contribute to tumor development [67]. This might explain an enhanced HLA presentation of peptides derived from deubiquitination-involved proteins in ependymal tumor tissue compared to benign tissue.

Tumor-exclusive antigens underwent a detailed evaluation process regarding HLA-presented peptides and corresponding source proteins, which resulted in the selection of eight novel TAAs as potential immunotherapeutic targets among the HLA ligandomes of ependymomas.

Five of the respective TAA source proteins have already been associated with tumorigenesis or tumor-associated processes in literature before. This applies to the source proteins coiled-coil domain-containing protein 13 (CCDC13) and dynein axonemal heavy chain 6 (DNAH6) that have been demonstrated to be involved in assembly and disassembly of cilia, termed ciliogenesis. CCDC13 is suggested to act as centriolar satellite protein participating in cilia formation and DNA stabilization [68]. Additionally, mutations of the source protein DNAH6 are suggested to cause ciliary dysfunction implicating a functional necessity of DNAH6 for ciliogenesis [69]. An intrusion of the normal cell cycle process within the affected tissue is an essential step during tumor development. Cilia are involved in cell signaling processes and, thereby, are able to alter cell cycle progression. Hence, a correctly functioning ciliogenesis is essential for normal implementations of cell cycle processes whereas interruptions of ciliogenesis were shown to contribute to the initiation of tumorigenesis [70]. Various transcription factors are involved in the transcriptional control of cilia formation. Key transcription factors are RFX1 and RFX2, which are both members of the regulatory factor X (RFX) gene family, and FOXJ1, which is a member of the forkhead/winged helix (FOX) family. All three are suggested to be involved in tumor formation, which indicates their usability as tumor biomarkers for cancer diagnosis, grading or outcome prediction in clinical applications [70, 71]. There is evidence that an altered

ciliogenesis is involved in ependymoma development and that different ependymoma subgroups display distinct ciliogenic mechanisms. Abedalthagafi *et al.* (2016) showed subgroup-depending expression patterns for FOXJ1 and RFX2. Both were highly expressed across myxopapillary and cellular WHO grade II ependymomas. In contrast, FOXJ1 expression was significantly decreased in the more aggressive forms of ependymomas, namely anaplastic and RELA fusion-positive ependymomas. A decreased FOXJ1 expression was additionally associated with worse patient outcome [72]. Furthermore, differences within the ciliary structures of anaplastic and WHO grade II ependymomas were observed by electron microscopy, though in a study based on only four tumor samples [73]. These changes in ciliogenic mechanisms of ependymomas could be mirrored within the HLA ligandome and might have led to the identification of TAAs derived from ciliogenesis-associated proteins in the present ependymoma study. Potential subgroup-specific differences in the HLA-presentation of ciliogenesis-associated antigens need to be addressed on the basis of larger ependymoma cohorts including a sufficient number of samples per tumor subgroup.

Three further source proteins of the selected ependymoma TAAs have been previously associated, not directly with ependymoma development, but with tumorigenesis in general.

Aoyama *et al.* (2019) ascribed promising characteristics of a potential immunohistochemical marker for malignancy to the coiled-coil domain-containing protein 180 (CCDC180) after comparable proteome analysis. They claimed that CCDC180 and the leucine-rich repeat-containing protein 4 (LRRC4) can be used to differentiate atypical lipomatous tumor/well-differentiated liposarcoma (ALT/WDL) from benign lipoma. Further, they showed that CCDC180 and LRRC4 expression might enhance the malignant potential of WDL cells and suggested the molecules as potential therapeutic targets of liposarcoma [74].

Another source protein that has been described as potential diagnostic marker is the cilia- and flagella-associated protein 52 (CFAP52), also commonly named WD repeat domain-containing protein 16 (WDR16) or WDRPUH (WD40-repeat protein upregulated in HCC). CFAP52 has functions in several physiological mechanisms such as signal transduction, RNA processing, cytoskeleton remodeling or cell proliferation [75, 76]. Silva *et al.* (2005) showed an increased CFAP52 expression in hepatocellular carcinoma (HCC) tissue, a primary epithelial malignancy of the liver, and indicated an involvement of CFAP52 in hepatocarcinogenesis [75]. Besides HCC, CFAP52 was also associated with microsatellite stable hereditary non-polyposis colorectal cancer (MSS HNPCC), where a relation with tumorigenesis was assumed after protein expression studies [76].

Protocadherin fat 3 (FAT3) is a transmembrane protein participating in tissue growth, tissue patterning, regulation of cell adhesion and planar cell polarity. Fat molecules perform tissue growth control at least partly *via* the Hippo pathway, a highly conserved kinase cascade. Anomalies in the

Hippo pathway are associated with malignancies such as ovarian cancer, breast cancer, sarcoma and HCC [77].

The circumstance that several antigens defined as TAAs in the ependymomas have been previously associated with cancer in literature supports their further consideration as potential immunotherapeutic targets for the development of novel treatment options. Another argument for their selection is their occurrence in other malignancies as shown by comparative analysis with a malignant immunopeptidomic database. Some of the ependymoma TAAs even occurred in the HLA ligandomes of other CNS cancer entities such as atypical teratoid rhabdoid tumor (ATRT), subependymoma and glioblastoma.

The ependymoma cohort contained two recurrent tumors surgically resected from the same patient within an interval of one year. Comparative analysis was performed to reveal differences on the immunopeptidomic level between the first and the second recurrence of the ependymoma. Across the complete ependymoma dataset, sample masses and yields of HLA ligands positively correlated. Contrarily, in the present case, the second recurrent tumor of the patient possessed the larger sample mass but resulted in a lower number of detected HLA ligands. This was accompanied by lower numbers of HLA-presented tumor-exclusive antigens and CTAs. As immune evasion strategy, some tumors are able to avoid T cell responses through a change or loss of HLA expression. Differences in HLA expression profiles have been demonstrated between primary tumors and corresponding metastases or recurrences. Metastatic tumors are capable of avoiding immune recognition and responses by losing HLA class I molecule expression independently from the primary tumor [78]. In relapsing acute myeloid leukemia (AML), a down-regulation of HLA class II gene expression and further changes in immune mechanism pathways have been revealed compared to the initial diagnosis [79]. Acquired immune evasion strategies are a possible explanation for the lower peptide yields of the second recurrence of the investigated ependymoma pair. In case of HLA class I, the loss of binders was not specific to one of the allotypes of the patient. Instead, a generally lower number of binders was identified across all allotypes. Additionally, the comparison between the immunopeptidomes of the tumors revealed shared as well as many individual HLA antigens. This might be due to changes of cellular or molecular mechanisms in the tumor cells of the different recurrences. However, differences between the HLA ligandome of the two samples caused by performance fluctuations during experimental procedures or MS analysis are also possible and cannot be ignored. Therefore, a final statement about immunopeptidomic changes in tumors recurring over time requires a larger data basis. Since the patient was not positive for the HLA alleles A*25:01, A*26:01 or A*02:01, none of the selected ependymoma TAAs were identified in the immunopeptidomes of the two tumors. Thus an analysis of potential changes in their HLA presentation was not possible. A cohort consisting of several paired ependymoma samples and the inclusion of the corresponding initial tumors is necessary to make

reliable conclusions about changes in the HLA ligandomes over the time course from initial to recurrent ependymomas. Relapse of ependymomas is still a major challenge during patient treatment and a long-term disease control is often not achievable [29, 30]. Unraveling potential differences in the HLA ligandome of initial and recurrent tumors might lead to an enhanced attention for the diagnosis state. This could impact the selection strategies of immunotherapeutic targets and the development of novel immunotherapeutic approaches possibly resulting in better adapted treatment options for relapsing ependymoma patients.

The strict assumption of the CNS being an immune privileged organ has long been revised. Cells of the immune system are able to access different parts of the brain under healthy and disease conditions [32]. Therefore, the progression in the development of novel immunotherapeutic approaches also concerns tumors of the CNS. Clinical studies have already proven the effectiveness of immunotherapeutic strategies in CNS tumors. During the phase I trial GAPVAC-101 of the Glioma Actively Personalized Vaccine Consortium (GAPVAC) an immune reaction was induced against glioblastomas upon highly individualized peptide vaccination [80]. In the present study, ependymal tumors have demonstrated to provide a wide range of HLA-presented antigens on their cell surfaces. This enabled the selection of ependymoma TAAs that might bear potential as target antigens for the development of immunotherapeutic strategies. To confirm the suitability of the selected antigens as immunotherapeutic targets, their capability to induce specific immune responses needs to be observed by immunogenicity testing. The present cohort consisted of ependymoma samples from several patients entailing a variety of HLA allotype combinations. A larger patient cohort would improve the present data density and enable the selection of additional TAAs covering more HLA allotypes. This could allow the establishment of a warehouse comprising target antigens for semi-personalized treatment of patients with frequent HLA allotypes. Enlargement of the ependymoma cohort would also lead to a better coverage of the distinct ependymoma subgroups. This would enable detailed analyses of subgroup-associated differences in the ependymoma immunopeptidomes to reveal a potential necessity of the development of subgroup-specific treatment options.

With the analysis of the immunopeptidome of ependymomas, an essential step in the direction of developing immunotherapeutic strategies for ependymoma patients has been made. Promising TAAs were identified, which should be further considered and investigated in respect of their suitability as immunotherapeutic target antigens.

2.6 References

[1] Wu J, Armstrong TS, and Gilbert MR. Biology and management of ependymomas. *Neuro Oncol* 18(7): 902–913 (2016).

[2] Taylor MD, Poppleton H, Fuller C, Su X, Liu Y, Jensen P, Magdaleno S, Dalton J, Calabrese C, Board J, Macdonald T, Rutka J, Guha A, Gajjar A, Curran T, and Gilbertson RJ. Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell* 8(4): 323–335 (2005).

[3] Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C, and Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro Oncol* 21(Supplement_5): v1-v100 (2019).

[4] Louis DN, Perry A, Reifenberger G, Deimling A von, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, and Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol* 131(6): 803–820 (2016).

[5] Armstrong TS, Vera-Bolanos E, and Gilbert MR. Clinical course of adult patients with ependymoma: results of the Adult Ependymoma Outcomes Project. *Cancer* 117(22): 5133–5141 (2011).

[6] Allen NJ and Lyons DA. Glia as Architects of Central Nervous System Formation and Function. *Science* 362(6411): 181–185 (2018).

[7] Clarke MF and Fuller M. Stem cells and cancer: two faces of eve. *Cell* 124(6): 1111–1115 (2006).

[8] Poppleton H and Gilbertson RJ. Stem cells of ependymoma. *Br J Cancer* 96(1): 6–10 (2007).

[9] Cavallaro U and Christofori G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nat Rev Cancer* 4(2): 118–132 (2004).

[10] Hitoshi S, Alexson T, Tropepe V, Donoviel D, Elia AJ, Nye JS, Conlon RA, Mak TW, Bernstein A, and van der Kooy D. Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells. *Genes Dev* 16(7): 846–858 (2002).

[11] Nyfeler Y, Kirch RD, Mantei N, Leone DP, Radtke F, Suter U, and Taylor V. Jagged1 signals in the postnatal subventricular zone are required for neural stem cell self-renewal. *EMBO J* 24(19): 3504–3515 (2005).

[12] Villano JL, Parker CK, and Dolecek TA. Descriptive epidemiology of ependymal tumours in the United States. *Br J Cancer* 108(11): 2367–2371 (2013).

[13] Bates JE, Choi G, and Milano MT. Myxopapillary ependymoma: a SEER analysis of epidemiology and outcomes. *J Neurooncol* 129(2): 251–258 (2016).

[14] Sonneland PR, Scheithauer BW, and Onofrio BM. Myxopapillary ependymoma. A clinicopathologic and immunocytochemical study of 77 cases. *Cancer* 56(4): 883–893 (1985).

[15] Scheithauer BW. Symptomatic subependymoma. Report of 21 cases with review of the literature. *J Neurosurg* 49(5): 689–696 (1978).

[16] Marks JE and Adler SJ. A comparative study of ependymomas by site of origin. *Int J Radiat Oncol Biol Phys* 8(1): 37–43 (1982).

[17] Guyotat J, Signorelli F, Desme S, Frappaz D, Madarassy G, Montange MF, Jouvet A, and Bret P. Intracranial ependymomas in adult patients: analyses of prognostic factors. *J Neurooncol* 60(3): 255–268 (2002).

[18] Raghunathan A, Wani K, Armstrong TS, Vera-Bolanos E, Fouladi M, Gilbertson R, Gajjar A, Goldman S, Lehman NL, Metellus P, Mikkelsen T, Necesito-Reyes MJT, Omuro A, Packer RJ, Partap S, Pollack IF, Prados MD, Robins HI, Soffietti R, Wu J, Miller CR, Gilbert MR, and Aldape KD. Histological predictors of outcome in ependymoma are dependent on anatomic site within the central nervous system. *Brain Pathol* 23(5): 584–594 (2013).

[19] Parker M, Mohankumar KM, Punchihewa C, Weinlich R, Dalton JD, Li Y, Lee R, Tatevossian RG, Phoenix TN, Thiruvenkatam R, White E, Tang B, Orisme W, Gupta K, Rusch M, Chen X, Li Y, Nagahawhatte P, Hedlund E, Finkelstein D, Wu G, Shurtleff S, Easton J, Boggs K, Yergeau D, Vadodaria B, Mulder HL, Becksfort J, Becksford J, Gupta P, Huether R, Ma J, Song G, Gajjar A, Merchant T, Boop F, Smith AA, Ding L, Lu C, Ochoa K, Zhao D, Fulton RS, Fulton LL, Mardis ER, Wilson RK, Downing JR, Green DR, Zhang J, Ellison DW, and Gilbertson RJ. C11orf95-RELA fusions drive oncogenic NF-κB signalling in ependymoma. *Nature* 506(7489): 451–455 (2014).

[20] Pietsch T, Wohlers I, Goschzik T, Dreschmann V, Denkhaus D, Dörner E, Rahmann S, and Klein-Hitpass L. Supratentorial ependymomas of childhood carry C11orf95-RELA fusions leading to pathological activation of the NF-κB signaling pathway. *Acta Neuropathol* 127(4): 609–611 (2014).

[21] DiDonato JA, Mercurio F, and Karin M. NF-κB and the link between inflammation and cancer. *Immunol Rev* 246(1): 379–400 (2012).

[22] Johnson RA, Wright KD, Poppleton H, Mohankumar KM, Finkelstein D, Pounds SB, Rand V, Leary SES, White E, Eden C, Hogg T, Northcott P, Mack S, Neale G, Wang Y-D, Coyle B, Atkinson J, DeWire M, Kranenburg TA, Gillespie Y, Allen JC, Merchant T, Boop FA, Sanford RA, Gajjar A, Ellison DW, Taylor MD, Grundy RG, and Gilbertson RJ. Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature* 466(7306): 632–636 (2010).

[23] Pajtler KW, Witt H, Sill M, Jones DTW, Hovestadt V, Kratochwil F, Wani K, Tatevossian R, Punchihewa C, Johann P, Reimand J, Warnatz H-J, Ryzhova M, Mack S, Ramaswamy V, Capper D, Schweizer L, Sieber L, Wittmann A, Huang Z, van Sluis P, Volckmann R, Koster J, Versteeg R, Fults D, Toledano H, Avigad S, Hoffman LM, Donson AM, Foreman N, Hewer E, Zitterbart K, Gilbert M, Armstrong TS, Gupta N, Allen JC, Karajannis MA, Zagzag D, Hasselblatt M, Kulozik AE, Witt O, Collins VP, Hoff K von, Rutkowski S, Pietsch T, Bader G, Yaspo M-L, Deimling A von, Lichter P, Taylor MD, Gilbertson R, Ellison DW, Aldape K, Korshunov A, Kool M, and Pfister SM. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell* 27(5): 728–743 (2015).

[24] Yuh EL, Barkovich AJ, and Gupta N. Imaging of ependymomas: MRI and CT. *Childs Nerv Syst* 25(10): 1203–1213 (2009).

[25] Rudà R, Reifenberger G, Frappaz D, Pfister SM, Laprie A, Santarius T, Roth P, Tonn JC, Soffietti R, Weller M, and Moyal EC-J. EANO guidelines for the diagnosis and treatment of ependymal tumors. *Neuro Oncol* 20(4): 445–456 (2018).

[26] Thorp N and Gandola L. Management of Ependymoma in Children, Adolescents and Young Adults. *Clin Oncol (R Coll Radiol)* 31(3): 162–170 (2019).

[27] Armstrong T, Yuan Y, Wu J, Mendoza T, Vera E, Omuro A, Lieberman F, Robins H, Gerstner E, Wu J, Wen P, Mikkelsen T, Aldape K, and Gilbert M. RARE-24. Objective response and clinical benefit in recurrent

ependymoma in adults: Final report of CERN 08-02: A phase II study of dose-dense temozolomide and lapatinib. *Neuro Oncol* 20(suppl_6): vi241-vi241 (2018).

[28] Kraetzig T, McLaughlin L, Bilsky MH, and Laufer I. Metastases of spinal myxopapillary ependymoma: unique characteristics and clinical management. *J Neurosurg Spine* 28(2): 201–208 (2018).

[29] Messahel B, Ashley S, Saran F, Ellison D, Ironside J, Phipps K, Cox T, Chong WK, Robinson K, Picton S, Pinkerton CR, Mallucci C, Macarthur D, Jaspan T, Michalski A, and Grundy RG. Relapsed intracranial ependymoma in children in the UK: Patterns of relapse, survival and therapeutic outcome. *Eur J Cancer* 45(10): 1815–1823 (2009).

[30] Massimino M, Miceli R, Giangaspero F, Boschetti L, Modena P, Antonelli M, Ferroli P, Bertin D, Pecori E, Valentini L, Biassoni V, Garrè ML, Schiavello E, Sardi I, Cama A, Viscardi E, Scarzello G, Scoccianti S, Mascarin M, Quaglietta L, Cinalli G, Diletto B, Genitori L, Peretta P, Mussano A, Buccoliero A, Calareso G, Barra S, Mastronuzzi A, Giussani C, Marras CE, Balter R, Bertolini P, Giombelli E, La Spina M, Buttarelli FR, Pollo B, and Gandola L. Final results of the second prospective AIEOP protocol for pediatric intracranial ependymoma. *Neuro Oncol* 18(10): 1451–1460 (2016).

[31] Walker PR, Calzascia T, Tribolet N de, and Dietrich PY. T-cell immune responses in the brain and their relevance for cerebral malignancies. *Brain Res Brain Res Rev* 42(2): 97–122 (2003).

[32] Ousman SS and Kubes P. Immune surveillance in the central nervous system. *Nat Neurosci* 15(8): 1096–1101 (2012).

[33] Weiss N, Miller F, Cazaubon S, and Couraud P-O. The blood-brain barrier in brain homeostasis and neurological diseases. *Biochim Biophys Acta* 1788(4): 842–857 (2009).

[34] Darvin P, Toor SM, Sasidharan Nair V, and Elkord E. Immune checkpoint inhibitors: Recent progress and potential biomarkers. *Exp Mol Med* 50(12): 1–11 (2018).

[35] Gorsi HS, Malicki DM, Barsan V, Tumblin M, Yeh-Nayre L, Milburn M, Elster JD, and Crawford JR. Nivolumab in the Treatment of Recurrent or Refractory Pediatric Brain Tumors: A Single Institutional Experience. *J Pediatr Hematol Oncol* 41(4): e235-e241 (2019).

[36] Dumont B, Forest F, Dal Col P, Karpathiou G, Stephan J-L, Vassal F, and Péoc'h M. PD1 and PD-L1 in ependymoma might not be therapeutic targets. *Clin Neuropathol* 36 (2017)(2): 90–92 (2017).

[37] Witt DA, Donson AM, Amani V, Moreira DC, Sanford B, Hoffman LM, Handler MH, Levy JMM, Jones KL, Nellan A, Foreman NK, and Griesinger AM. Specific expression of PD-L1 in RELA-fusion supratentorial ependymoma: Implications for PD-1-targeted therapy. *Pediatr Blood Cancer* 65(5): e26960 (2018).

[38] Penas-Prado M, Armstrong TS, and Gilbert MR. Progress in rare central nervous system tumors. *Curr Opin Neurol* 32(6): 895–906 (2019).

[39] Pollack IF, Jakacki RI, Butterfield LH, and Okada H. Ependymomas: Development of immunotherapeutic strategies. *Expert Rev Neurother* 13(10): 1089–1098 (2013).

[40] Yeung JT, Hamilton RL, Okada H, Jakacki RI, and Pollack IF. Increased expression of tumor-associated antigens in pediatric and adult ependymomas: Implication for vaccine therapy. *J Neurooncol* 111(2): 103–111 (2013).

[41] Miller CA, Dahiya S, Li T, Fulton RS, Smyth MD, Dunn GP, Rubin JB, and Mardis ER. Resistance-promoting effects of ependymoma treatment revealed through genomic analysis of multiple recurrences in a single patient. *Cold Spring Harb Mol Case Stud* 4(2) (2018).

[42] Caron E, Kowalewski DJ, Chiek Koh C, Sturm T, Schuster H, and Aebersold R. Analysis of Major Histocompatibility Complex (MHC) Immunopeptidomes Using Mass Spectrometry. *Mol Cell Proteomics* 14(12): 3105–3117 (2015).

[43] Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, Cid-Pavon GMD, Ramsbottom K, Ghattaoraya G, Alfirevic A, Middleton D, and Jones AR. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 48(D1): D783-D788 (2020).

[44] Nelde A, Kowalewski DJ, and Stevanović S. Purification and Identification of Naturally Presented MHC Class I and II Ligands. *Methods Mol Biol* 1988: 123–136 (2019).

[45] Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-T, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJG, Old LJ, Camargo AA, and Vasconcelos ATR. CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37(Database issue): D816-9 (2009).

[46] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45(6): 580–585 (2013).

[47] Thul PJ and Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci* 27(1): 233–244 (2018).

[48] Huang DW, Sherman BT, and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1): 1–13 (2009).

[49] Huang DW, Sherman BT, and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1): 44–57 (2009).

[50] van Endert, PM, Riganelli D, Greco G, Fleischhauer K, Sidney J, Sette A, and Bach JF. The peptide-binding motif for the human transporter associated with antigen processing. *J Exp Med* 182(6): 1883–1895 (1995).

[51] Neefjes J, Jongsma MLM, Paul P, and Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11(12): 823–836 (2011).

[52] Fratta E, Coral S, Covre A, Parisi G, Colizzi F, Danielli R, Nicolay HJM, Sigalotti L, and Maio M. The biology of cancer testis antigens: Putative function, regulation and therapeutic potential. *Mol Oncol* 5(2): 164–182 (2011).

[53] Holoshitz J. The quest for better understanding of HLA-disease association: scenes from a road less travelled by. *Discov Med* 16(87): 93–101 (2013).

[54] Reveille JD. Major histocompatibility genes and ankylosing spondylitis. *Best Pract Res Clin Rheumatol* 20(3): 601–609 (2006).

[55] Mignot E, Lin X, Arrigoni J, Macaubas C, Olive F, Hallmayer J, Underhill P, Guilleminault C, Dement WC, and Grumet FC. DQB1*0602 and DQA1*0102 (DQ1) are better markers than DR2 for narcolepsy in Caucasian and black Americans. *Sleep* 17(8 Suppl): S60-7 (1994).

[56] Holoshitz J. The rheumatoid arthritis HLA-DRB1 shared epitope. *Curr Opin Rheumatol* 22(3): 293–298 (2010).

[57] Kanevskiy L, Erokhina S, Kobyzeva P, Streltsova M, Sapozhnikov A, and Kovalenko E. Dimorphism of HLA-E and its Disease Association. *Int J Mol Sci* 20(21) (2019).

[58] Zheng H, Lu R, Xie S, Wen X, Wang H, Gao X, and Guo L. Human leukocyte antigen-E alleles and expression in patients with serous ovarian cancer. *Cancer Sci* 106(5): 522–528 (2015).

[59] Xu Y-P, Wieten L, Wang S-X, Cai Y, Olieslagers T, Zhang L, He L-M, Tilanus MGJ, and Hong W-X. Clinical significance of HLA-E genotype and surface/soluble expression levels between healthy individuals and patients with acute leukemia. *Leuk Lymphoma* 60(1): 208–215 (2019).

[60] Merchant TE, Bendel AE, Sabin ND, Burger PC, Shaw DW, Chang E, Wu S, Zhou T, Eisenstat DD, Foreman NK, Fuller CE, Anderson ET, Hukin J, Lau CC, Pollack IF, Laningham FH, Lustig RH, Armstrong FD, Handler MH, Williams-Hughes C, Kessel S, Kocak M, Ellison DW, and Ramaswamy V. Conformal Radiation Therapy for Pediatric Ependymoma, Chemotherapy for Incompletely Resected Ependymoma, and Observation for Completely Resected, Supratentorial Ependymoma. *J Clin Oncol* 37(12): 974–983 (2019).

[61] Upadhyaya SA, Robinson GW, Onar-Thomas A, Orr BA, Billups CA, Bowers DC, Bendel AE, Hassall T, Crawford JR, Partap S, Fisher PG, Tatevossian RG, Seah T, Qaddoumi IA, Vinitsky A, Armstrong GT, Sabin ND, Tinkle CL, Klimo P, Indelicato DJ, Boop FA, Merchant TE, Ellison DW, and Gajjar A. Molecular grouping and outcomes of young children with newly diagnosed ependymoma treated on the multi-institutional SJYC07 trial. *Neuro Oncol* 21(10): 1319–1330 (2019).

[62] Weinzierl AO, Lemmel C, Schoor O, Müller M, Krüger T, Wernet D, Hennenlotter J, Stenzl A, Klingel K, Rammensee H-G, and Stevanovic S. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol Cell Proteomics* 6(1): 102–113 (2007).

[63] Wei X, Chen F, Xin K, Wang Q, Yu L, Liu B, and Liu Q. Cancer-Testis Antigen Peptide Vaccine for Cancer Immunotherapy: Progress and Prospects. *Transl Oncol* 12(5): 733–738 (2019).

[64] Suda T, Tsunoda T, Daigo Y, Nakamura Y, and Tahara H. Identification of human leukocyte antigen-A24-restricted epitope peptides derived from gene products upregulated in lung and esophageal cancers as novel targets for immunotherapy. *Cancer Sci* 98(11): 1803–1808 (2007).

[65] Deng H, Zeng J, Zhang T, Gong L, Zhang H, Cheung E, Jones C, and Li G. Histone H3.3K27M Mobilizes Multiple Cancer/Testis (CT) Antigens in Pediatric Glioma. *Mol Cancer Res* 16(4): 623–633 (2018).

[66] Shraibman B, Barnea E, Kadosh DM, Haimovich Y, Slobodin G, Rosner I, López-Larrea C, Hilf N, Kuttruff S, Song C, Britten C, Castle J, Kreiter S, Frenzel K, Tatagiba M, Tabatabai G, Dietrich P-Y, Dutoit V, Wick W, Platten M, Winkler F, Deimling A von, Kroep J, Sahuquillo J, Martinez-Ricarte F, Rodon J, Lassen U, Ottensmeier C, van der Burg SH, Thor Straten P, Poulsen HS, Ponsati B, Okada H, Rammensee H-G, Sahin U, Singh H, and Admon A. Identification of Tumor Antigens Among the HLA Peptidomes of Glioblastoma Tumors and Plasma. *Mol Cell Proteomics* 18(6): 1255–1268 (2019).

[67] Mansour MA. Ubiquitination: Friend and foe in cancer. *Int J Biochem Cell Biol* 101: 80–93 (2018).

[68] Staples CJ, Myers KN, Beveridge RDD, Patil AA, Howard AE, Barone G, Lee AJX, Swanton C, Howell M, Maslen S, Skehel JM, Boulton SJ, and Collis SJ. Ccdc13 is a novel human centriolar satellite protein required for ciliogenesis and genome stability. *J Cell Sci* 127(Pt 13): 2910–2919 (2014).

[69] Li Y, Yagi H, Onuoha EO, Damerla RR, Francis R, Furutani Y, Tariq M, King SM, Hendricks G, Cui C, Saydmohammed M, Lee DM, Zahid M, Sami I, Leatherbury L, Pazour GJ, Ware SM, Nakanishi T, Goldmuntz E, Tsang M, and Lo CW. DNAH6 and Its Interactions with PCD Genes in Heterotaxy and Primary Ciliary Dyskinesia. *PLoS Genet* 12(2): e1005821 (2016).

[70] Walentek P. Ciliary transcription factors in cancer--how understanding ciliogenesis can promote the detection and prognosis of cancer types. *J Pathol* 239(1): 6–9 (2016).

[71] Choksi SP, Lauter G, Swoboda P, and Roy S. Switching on cilia: transcriptional networks regulating ciliogenesis. *Development* 141(7): 1427–1441 (2014).

[72] Abedalthagafi MS, Wu MP, Merrill PH, Du Z, Woo T, Sheu S-H, Hurwitz S, Ligon KL, and Santagata S. Decrease in FOXJ1 expression and its ciliogenesis program in aggressive ependymoma and choroid plexus tumours. *J Pathol* 238(4): 584–597 (2016).

[73] Alfaro-Cervelló C, Soriano-Navarro M, Ramírez M, Bernet L, Martinez Banaclocha M, Cano R, Reyes Santías RM, Forteza-Vila J, and García-Verdugo JM. Ultrastructural pathology of anaplastic and grade II ependymomas reveals distinctive ciliary structures--electron microscopy redux. *Ultrastruct Pathol* 39(1): 23–29 (2015).

[74] Aoyama T, Takasawa A, Takasawa K, Ono Y, Emori M, Murata M, Hayasaka T, Fujitani N, Osanai M, Yamashita T, Hasegawa T, and Sawada N. Identification of Coiled-Coil Domain-Containing Protein 180 and Leucine-Rich Repeat-Containing Protein 4 as Potential Immunohistochemical Markers for Liposarcoma Based on Proteomic Analysis Using Formalin-Fixed, Paraffin-Embedded Tissue. *Am J Pathol* 189(5): 1015–1028 (2019).

[75] Silva FP, Hamamoto R, Nakamura Y, and Furukawa Y. WDRPUH, a novel WD-repeat-containing protein, is highly expressed in human hepatocellular carcinoma and involved in cell proliferation. *Neoplasia* 7(4): 348–355 (2005).

[76] Chen W, Yuan L, Cai Y, Chen X, Chi Y, Wei P, Zhou X, and Shi D. Identification of chromosomal copy number variations and novel candidate loci in hereditary nonpolyposis colorectal cancer with mismatch repair proficiency. *Genomics* 102(1): 27–34 (2013).

[77] Sopko R and McNeill H. The skinny on Fat: an enormous cadherin that regulates cell adhesion, tissue growth, and planar cell polarity. *Curr Opin Cell Biol* 21(5): 717–723 (2009).

[78] Garrido F and Aptsiauri N. Cancer immune escape: MHC expression in primary tumours versus metastases. *Immunology* 158(4): 255–266 (2019).

[79] Christopher MJ, Petti AA, Rettig MP, Miller CA, Chendamarai E, Duncavage EJ, Klco JM, Helton NM, O'Laughlin M, Fronick CC, Fulton RS, Wilson RK, Wartman LD, Welch JS, Heath SE, Baty JD, Payton JE, Graubert TA, Link DC, Walter MJ, Westervelt P, Ley TJ, and DiPersio JF. Immune Escape of Relapsed AML Cells after Allogeneic Transplantation. *N Engl J Med* 379(24): 2330–2341 (2018).

[80] Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, Platten M, Tabatabai G, Dutoit V, van der Burg SH, Thor Straten P, Martínez-Ricarte F, Ponsati B, Okada H, Lassen U, Admon A, Ottensmeier CH, Ulges A, Kreiter S, Deimling A von, Skardelly M, Migliorini D, Kroep JR, Idorn M, Rodon J, Piró J, Poulsen HS, Shraibman B, McCann K, Mendrzyk R, Löwer M, Stieglbauer M, Britten CM, Capper D, Welters MJP, Sahuquillo J, Kiesel K, Derhovanessian E, Rusch E, Bunse L, Song C, Heesch S, Wagner C, Kemmer-Brück A,

Ludwig J, Castle JC, Schoor O, Tadmor AD, Green E, Fritsche J, Meyer M, Pawlowski N, Dorner S, Hoffgaard F, Rössler B, Maurer D, Weinschenk T, Reinhardt C, Huber C, Rammensee H-G, Singh-Jasuja H, Sahin U, Dietrich P-Y, and Wick W. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* 565(7738): 240–245 (2019).

3 The OPSCC HLA ligandome

3.1 Introduction

Head and neck squamous cell carcinomas (HNSCCs) are a group of heterogeneous tumors arising in the mucosal linings of the upper aerodigestive tract [1]. They comprise tumors of different anatomic compartments such as the oropharyngeal squamous cell carcinoma (OPSCC) occurring in the oropharynx. Localized posterior to the oral cavity, the oropharynx consists of the posterior one-third of the tongue (base of tongue) and the associated lingual tonsils, epiglottic vallecula, palatine tonsils, soft palate and superior constrictor muscle. According to clinical oncology, the oropharynx is subdivided into the four possible tumor locations base of tongue, soft palate, palatine tonsillar fossa and pharyngeal wall, of which the tonsils and the base of tongue are the most common anatomic subsites of OPSCCs [2, 3]. Risk factors for the development of an OPSCC can be the excessive consumption of alcohol, tobacco smoking or the infection with an oncogenic high-risk human papilloma virus (HPV) type [2]. Even though numbers of tumors caused by the classical risk factors alcohol and tobacco are declining, the incidence rates (IRs) of HNSCC increased in the past decades. This trend is induced by a rise of HPV infections, which took over the role as most common cause of OPSCCs and are presently responsible for 70% of newly diagnosed cases [4]. Between 1999 and 2017, the IR of oropharyngeal tumors significantly increased with an average annual rate of 2.8%. The current age-adjusted IR for oropharyngeal cancer is 2.5 patients per 100,000 individuals. With an IR of 4.3, men are more likely to develop cancer of the oropharynx than women with an IR of 0.9. The risk for developing oropharynx cancer increases with age. Among persons below an age of 65 years the IR is 1.8 whereas among persons of 65 years or older the IR is 7.2. The 5-year overall survival (OR) rate is indicated with 49% for patients with tumors of the oropharynx [5, 6]. In most cases, initial diagnosis of OPSCC follows clinical presentation of patients with a small primary lesion in combination with a large cystic neck mass. Final cancer diagnosis is confirmed by subsequent imaging *via* computer tomography (CT) [7].

Human papilloma virus

The small DNA virus HPV belongs to the group of papilloma viruses and has a specific tropism for the infection of squamous epithelial cells. Classification of HPVs is performed on the basis of differences in the nucleotide sequence of the gene L1 which codes for a viral capsid protein [8, 9]. According to the International HPV Reference Center (www.hpvcenter.se; state August 2020), 227 different HPV types are presently identified. They can be categorized into five genera with the Alpha, Beta and Gamma papillomaviruses building the three largest and the Mu- and Nu-papillomaviruses building the two smallest genera [10].

HPVs are further graded into the low-risk and the high-risk subgroup depending on the malignant potential of the skin or mucosal lesions which they cause upon infection. Infections with HPVs that are classified as low-risk types (e.g. HPV6 or HPV11) are commonly inconspicuous or cause benign epithelial tumors termed papillomas. They are usually handled by the natural immune system of the host but can also cause persisting warts whose treatment is challenging. Low-risk HPVs are rarely associated with cancer development [11]. In contrast, high-risk HPV types (e.g. HPV16 or HPV18) can initiate the development of premalignant squamous intraepithelial neoplasias that can progress towards malignant tumors and cancer [12]. The connection between HPV and cancer was already revealed in the 1970s [13]. Different tumor entities have been associated with HPV since. The majority of cervical cancers are caused by an infection with high-risk HPV types [14, 15] and the most common risk factor of OPSCCs are no longer alcohol and tobacco, but the infection with an oncogenic HPV type. This also led to a shift from predominantly older to an increasing number of younger, non-smoking patients [16]. The most common HPV subtype causing OPSCCs is HPV16 with a prevalence of 80% to 95% among HPV^+ OPSCCs [4].

Papillomaviruses are non-enveloped viruses and have an icosahedral structure with a diameter of 50 nm to 60 nm. It consists of two parts, the genome and the capsid. The viral capsid is built of the major and minor capsid proteins L1 and L2. Besides environing the virus and protecting the genome, the capsid is involved in the spread, transmission and survival of the virus. The viral genome is a double-stranded, circular DNA molecule of approximately 8 kbp (kilo base pairs). The genomic organization is similar in different HPV serotypes and comprises eight or nine open reading frames (ORFs) that can vary in size and position. The HPV genome is separated into the early and the late genomic region according to their respective expression periods during the viral life cycle. The early genomic region encodes regulatory proteins whereas the late genomic region encodes the capsid proteins L1 and L2 [17]. The early proteins E1 and E2 are expressed by conserved gene regions and are involved in virus replication whereas E4, E5, E6 and E7 are more diverse and play a role in activation of cell cycle, immune evasion and release of viruses [18, 19]. Due to complex processes including multiple promoters and splicing, the diversity of HPV-expressed proteins and molecules is extensive despite the small genome size [19]. The HPV genome is exemplarily depicted for the high-risk type HPV16 in **Figure 18**. In high-risk HPVs, the early proteins E5, E6 and E7 have oncogenic functions [20].

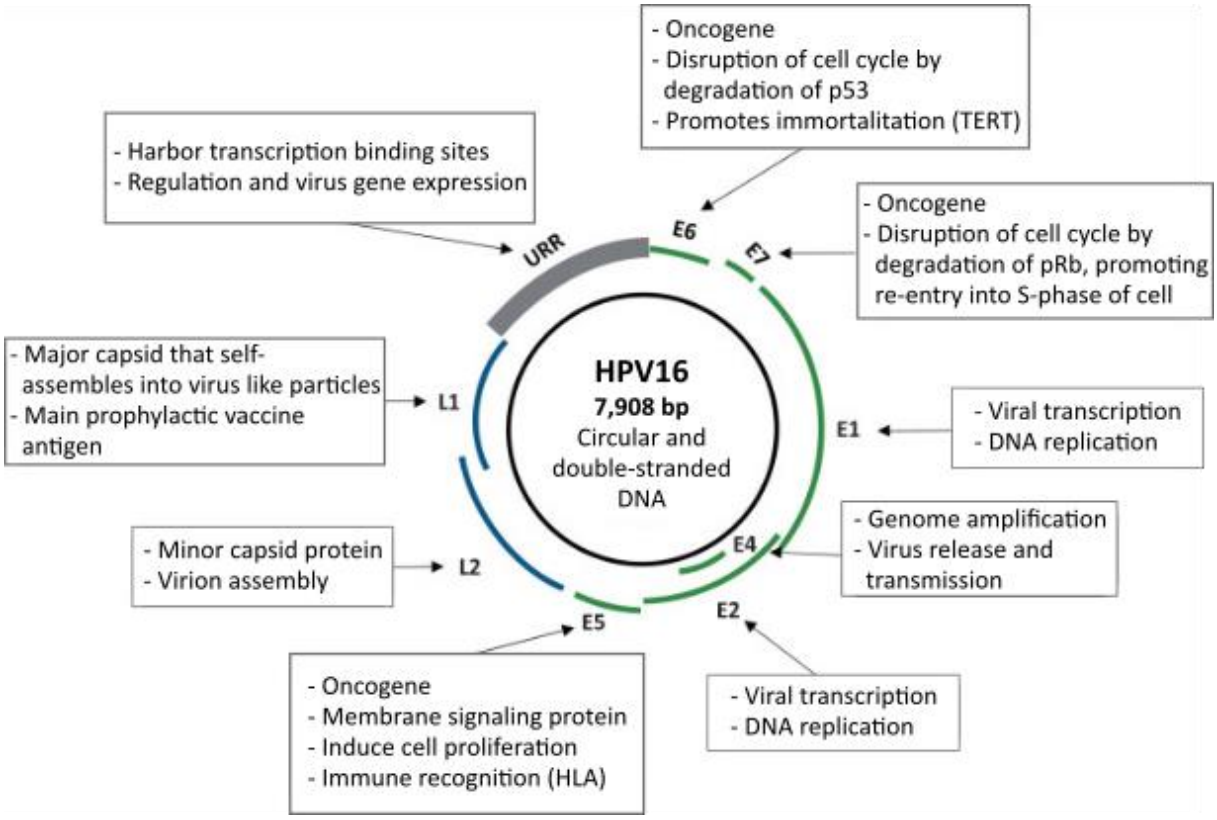


Figure 18: Organization of the HPV16 genome. The genome of HPV consists of the early and the late genomic regions. Proteins encoded by the early region (E1, E2, E4, E5, E6 and E7) are involved in various processes such as viral replication, cell cycle activation, immune evasion and virus release. Late proteins (L1 and L2) are responsible for building the viral capsid. In high-risk HPV types such as HPV16, the early proteins have additional oncogenic potential due to their ability to interfere with the host´s cell cycle processes. bp – base pair; p53 – cellular tumor antigen p53; TERT – telomerase reverse transcriptase; pRb - retinoblastoma protein; S-phase – synthesis phase. Modified from [20].

The HPV life cycle starts with the viral infection of epithelial basal stem cells, which is enabled by a wound or an epithelial trauma of the host [21]. Subsequently to the infection, the viral proteins E1 and E2 are important to trigger an initial phase of genome amplification. Once the copy number of the viral genome has reached a stable plateau, the early proteins E1 and E2 are no longer essential for replication processes and viral maintenance in the epithelium [22]. Especially in case of a low-risk HPV infection, wound healing processes accompanied by cell division drive amplification of cells carrying the HPV genome [19]. In case of high-risk HPV infections, the early proteins E5, E6 and E7 are capable to initiate the cell cycle and cell proliferation in the upper, basal and parabasal epithelium of the host. Thereby, amplification of the viral genome and packaging into infectious particles is enabled [18, 19]. Expression of the minor capsid protein L2 and exit from the cell cycle are essential steps for the completion of the viral life cycle. Final packaging of the HPV genome requires the additional expression of the major capsid protein L1 [23]. Within the superficial keratinocytes the viruses eventually maturate. These dying keratinocytes commute the environment from an oxidizing to a reducing milieu, which is beneficial for the viruses. The viral protein E4 accumulates to amyloid fibrils that are able to

disturb keratin structures. Further, E4 is suggested to be involved in the release of viral particles into the upper layer of the epithelium [24–26]. The newly generated HPV particles are now able to infect host cells and to initiate another viral life cycle.

HPV and HLA alleles

After an HPV infection, the immune response of the host plays an important role in the course of the disease. An infection with low-risk HPV types can cause lesions of skin or mucosa. In contrast, high-risk HPV types are able to induce the development of cancer through interference with regular processes of the host cell cycle in case the patient´s immune system cannot resolve the virus infection [17]. HLA class I and class II molecules are important compartments of the adaptive immune system and are responsible for the initiation of specific T cell responses. Therefore, the efficient presentation of viral antigens by HLA molecules of the host is an essential process enabling recognition and clearance of an HPV infection [27]. During this process, the polymorphism of the HLA gene region is of importance. HLA molecules with a high binding affinity for HPV-derived peptides are assumed to be associated with a lower probability of developing cancerous lesions whereas a lower binding affinity of HLA molecules for viral antigens might enhance the risk for cancer caused by an HPV infection [28]. The risks for HPV-triggered cervical cancer vary in dependence of different HLA alleles. HLA-B*07 was one of the first HLA alleles linked to an increased risk for cervical cancer induced by HPV [29]. As reviewed by Paaso *et al.* (2019), several HLA class I and class II alleles are presently described to be associated with an enhanced or decreased probability for a patient to develop a persistent HPV-infection resulting in the occurrence of cervical cancer **(Figure 19)** [27, 30–33]. Thus, the polymorphism of the HLA gene locus is suggested to serve as biomarker for the probability of patients with high-risk HPV infections to develop HPV-induced malignancies [27].

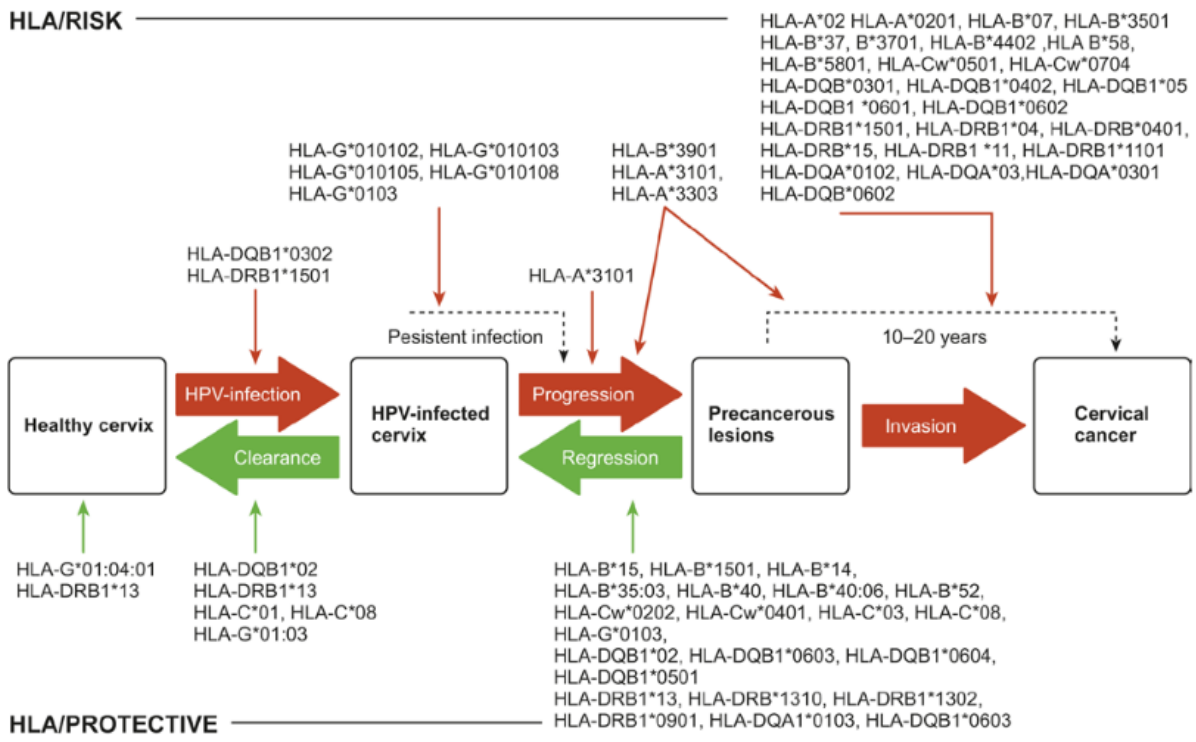


Figure 19: Connection between different HLA alleles and HPV infections resulting in the progression from persisting lesions towards cervical cancer. The occurrence of several HLA class I and class II alleles have been shown to impact the progression of HPV-induced precancerous lesions. Some HLA alleles were demonstrated to enhance the risk for cervical cancer after an HPV infection (HLA/RISK) whereas others were described with protective characteristics (HLA/PROTECTIVE). HLA – human leukocyte antigen; HPV – human papilloma virus. [27]

HPV$^+$ *versus* HPV$^-$ OPSCC

OPSCCs can be subdivided into two tumor subgroups according to their HPV status. Depending on whether the development of OPSCC is caused by an infection with a high-risk HPV type or by the abuse of tobacco and alcohol, distinct clinical and biological features are exhibited **(Table 6)**. On the demographic level, HPV$^+$ OPSCCs occur in a younger age group than HPV$^-$ OPSCCs [34]. Additionally, the response to treatment and long-term prognosis is better in HPV$^+$ than in HPV$^-$ OPSCC patients. Therefore, the patient´s HPV status has been included into the tumor grading scheme "TNM Classification of Malignant Tumors" [7, 35]. The identification of an HPV$^+$ tumor is often performed *via* immuno-histochemical staining of the tumor suppressor protein p16 in the tumor tissue. The two oncoproteins E6 and E7 of high-risk HPV types are able to degrade the tumor suppressor protein p53 and to disrupt DNA repair as well as apoptotic mechanisms of the host. As compensation strategy, the tumor cells overexpress the tumor suppressor protein p16, a cyclin-dependent kinase inhibitor, which is an attempt to inhibit cell proliferation. Thus, p16 can be used as histopathological marker to classify OPSCCs into p16$^+$ HPV$^+$ and p16$^-$ HPV$^-$ tumors [36]. Due to HPV$^+$ OPSCC being associated with better

treatment response and patient outcome compared to HPV⁻ OPSCC, p16 overexpression is used as prognostic marker in clinical applications [37].

Table 6: Comparison between HPV⁺ and HPV⁻ OPSCCs. Depending on the individual HPV status, the presentation of OPSCCs differs concerning clinical as well as the molecular characteristics. HPV – human papilloma virus; OPSCC – oropharyngeal squamous cell carcinoma; OS – overall survival; TP35 – tumor suppressor gene 53; p16 - cyclin-dependent kinase inhibitor 2A. [7]

Parameter	HPV⁺ OPSCC	HPV⁻ OPSCC
Patient age	Middle-aged	> 60 years
Risk factors	HPV infection	Smoking, alcohol
5-year OS rate	90%	40%
Tumor size	Smaller	Larger
Nodal burden	Multiple nodes	Fewer nodes
TP53 mutations	Infrequent	Frequent
p16 overexpression	Yes	No

Current therapy for OPSCC

The current standard of care for OPSCC patients depends on their annotated tumor stages. For a long period, OPSCC staging was mainly based on anatomic characteristics and the condition of the respective patient. Recently, the HPV status was additionally included into the staging system of OPSCCs [38, 39]. To decide on the appropriate form of therapy, OPSCC tumors are classified into the cancer stages I, II, III and IV.

Unresectable tumors are usually treated with radiation therapy. Whether a combination with systemic therapy is recommended, depends on the respective disease stage. For patients with stage IV disease, combinatory treatment with the cytostatic agent cisplatin and the immunotherapeutic antibody cetuximab, an epidermal growth factor receptor (EGFR) inhibitor, is recommended. The combinatory treatment with radiation therapy and cetuximab was shown to significantly increase the OS rate in patients with locoregionally HNSCC [40]. Systemic therapy concurrently applied to radiation therapy is also recommended for patients with stage III disease. In contrast, patients with stage I and II disease are not recommended for the reception of systemic therapy [41]. In case a tumor is surgically resectable, primary surgery is always the first choice of treatment. The immense progress of minimally invasive surgical techniques including the development of transoral robotic surgery (TORS) and transoral laser microsurgery (TLM) makes surgery a valuable tool for OPSCC therapy. Additionally, precise tumor resection facilitates pathological staging [42]. Subsequent to surgery, post-operative

radiation therapy is commonly applied. A potential combination with systemic therapy depends on the individual condition of a patient. This requires the individual evaluation of parameters such as the condition of tumor margins, the existence of extracapsular nodal extensions as well as potential risk factors like lymphovascular or perineural invasion. An initial chemotherapy is not recommended for OPSCC patients and should only be applied in combination with radiation therapy or subsequent to surgery [41].

Due to an increasing number of OPSCC cases caused by an HPV-infection instead of smoking or alcohol consumption, novel developments and adjustments of therapy recommendations are forming the current standard of care for HPV^+ patients. A positive HPV status was shown to be a strong prognostic factor for better survival, which is partly due to HPV^+ patients commonly being non-smokers and of younger age compared to HPV^- patients [43]. In contrast, patients with HPV^+ OPSCC have a higher probability to sustain treatment-related morbidity over long intervals of therapy. These circumstances led to the development of de-escalation strategies for the treatment of HPV^+ OPSCC patients. Thereby, the doses of radiation or systemic therapy are reduced to decrease the risks and side effects of therapy while high cure rates are tried to maintain. The positive results of several clinical studies support the application of de-escalation protocols [44–46].

Besides treatment of HPV-induced diseases, there is the possibility of prophylactic vaccination to prevent infections with HPV. Vaccination was shown to reduce the numbers of HPV-infected patients as well as the number of patients developing diseases caused by HPV infections. Especially among women, the vaccination rate is rising, which leads to a decrease in the IR of cervical cancer [4, 47, 48]. Furthermore, a reduced HPV incidence was observed in the oral cavity for HPV types covered by the vaccination. This indicates a potential benefit of prophylactic HPV vaccination also in regard to OPSCC [49]. In a long-term view, prophylactic vaccination against HPV is expected to markedly reduce the incidence of several HPV-induced diseases.

Immunotherapy for OPSCCs

In HNSCC patients, modifications of the natural immune system can be observed. Besides evading the immune reactions of the patients, the tumors are able to actively disrupt the antitumor responses [50]. HNSCC tumors are capable of activating several distinct processes that trigger a suppression of the immune reactions initiated against the tumor [51]. Decreased amounts of white blood cells are commonly observed in the peripheral bloodstream of HNSCC patients. The remaining white blood cells comprise an uncommonly high fraction of regulatory T (T_{reg}) cells with immunosuppressive function. Furthermore, an even higher proportion of immunosuppressive T_{reg} cells can be observed among the cell population of tumor infiltrating lymphocytes (TILs) identified within HNSCC tumors [52, 53]. Compared to HPV^- HNSCC tumors, HPV^+ tumors have higher levels of tumor infiltrating T_{reg} cells. Divergent opinions exist about the consequences for patient prognosis resulting from an enhanced

level of T_{reg} cells infiltrated into the tumor. On the one hand, higher numbers of tumor infiltrating T_{reg} cells are assumed as prognostic factor for an improved outcome of HNSCC patients [54]. On the other hand, this benefit is suggested to depend on an elevated rate of tumor infiltrating $CD8^+$ cytotoxic T cells as predominantly observed in HPV^+ OPSCC. The probability of a long-term, relapse-free survival has been shown to be increased for HPV^+ patients with an OPSCC tumor that is highly infiltrated by $CD8^+$ T cells [55].

$CD8^+$ TILs express immune checkpoint inhibitors that can serve as potential targets for an inhibitory checkpoint receptor blockade therapy. In the past years, various clinical trials addressed the benefits of immune checkpoint blocking agents like the programmed cell death protein 1 (PD1)-blocking agents pembrolizumab and nivolumab as novel therapy options against solid tumors including HNSCC [56].

The application of pembrolizumab resulted in significantly increased overall response rates in HNSCC patients [57]. Treatment of HNSCC patients with nivolumab was demonstrated to result in a significantly improved OS rate compared to treatment with standard therapy [58, 59]. These promising results for the application of immunotherapy led to the approval of pembrolizumab and nivolumab for the inclusion into standard of care protocols recommended for HNSCC patients with initial, recurrent or metastatic diseases [56, 60].

Several ongoing clinical trials address the investigation of applying different therapy options in combinatory treatments. The phase III trial NCT02741570 is currently in the recruiting phase and includes HPV^+ and HPV^- OPSCC patients among other HNSCC patients. Aim of this study is to observe the effects resulting from the combination of the checkpoint inhibitory agents nivolumab and ipilimumab in the treatment course of HNSCC. The effects of combining the checkpoint inhibitor pembrolizumab with chemoradiation therapy will be observed in the phase III trial NCT03040999 recruiting HPV^+ and HPV^- HNSCC patients.

Immunotherapeutic strategies provide promising tools for the development of novel treatment options for OPSCC patients. This especially applies to highly T cell infiltrated HPV^+ OPSCCs, which are rising in number among overall OPSCCs.

3.2 Contributions

The OPSCC study was performed in collaboration with different associates who contributed to this project.

The project was established in cooperation with PD Dr. med. Simon Laban and Jasmin Ezić, both from the Department of Oto-Rhino-Laryngology and Head and Neck Surgery (University Medical Center, Ulm, Germany). They provided the OPSCC and tonsil tissue samples, collected the clinical data of the patients and supported data interpretation.

PD Joannis Mytilineos (Institute for Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Transfusion Service, Baden Württemberg/Hessen, University Hospital Ulm, Germany) performed HLA typing.

Whole exome sequencing (WES) and RNA sequencing was performed by Dr. Martin Bens (Friedrich Schiller University Jena (FSU): Faculty of Biological Sciences; Leibniz Institute on Aging - Fritz Lipmann Institute (FLI): Platzer Research Group). Alignment of DNA and RNA reads as well as variant calling were carried out by Dr. Axel Fürstberger (Institute of Medical Systems Biology, Ulm University, Germany). Dr. Jaya Thomas (Cancer Sciences Unit, Faculty of Medicine, University of Southampton, UK) classified the OPSCC tumors according to the HPV status on the basis of RNA sequencing data.

Leon Bichmann (Institute for Cell Biology, Department of Immunology, Eberhard Karls University Tübingen, Germany; Applied Bioinformatics, Center for Bioinformatics and Department of Computer Science, Eberhard Karls University Tübingen, Germany) and Dr. Linus Backert (Immatics Biotechnologies GmbH, Tübingen, Germany) assisted the bioinformatic data analysis.

3.3 Materials and methods

The following section entails an overview of the OPSCC cohort and the methods applied during the study. Several methods overlap with those from the ependymoma study (chapter 2). Detailed information and descriptions regarding deployed materials and methods can be found in the appendix (see 8.1).

Study cohort

The study cohort comprised 40 OPSCC tissue samples from 40 different patients **(Supplementary Table 9)**. Additionally, five benign tonsil tissue samples were available as healthy reference **(Supplementary Table 10)**. The OPSCC tumors were surgically resected, subsequently snap-frozen in liquid nitrogen (N_2) and stored at -80°C. The age of the 7 female and 33 male patients ranged from 38

to 79 years at the date of diagnosis. In accordance with the Declaration of Helsinki, written informed consent was obtained from all patients. The study was performed according to the local ethical requirements (222/13, 90/15). The cohort included 22 HPV$^+$ samples, of which 19 were infected with HPV16 and three with HPV35, HPV58 and HPV59, respectively. 18 tumor samples were tested negative for an HPV infection. The majority of OPSCCs were located in the tonsils (34 samples) whereas four tumors were located at the base of the tongue, one at the lateral tongue and one at the soft palate. HLA class I and class II typing was carried out for 37 OPSCC patients and three tonsil donors. The HLA typing results were not yet available for three OPSCC patients and two tonsil donors, which is why HLA class I allotype assignment was performed for these cases on the basis of the identified HLA-presented peptides. For HLA class I, the binders among the HLA class I-presented peptides were determined. The following analyses of the OPSCC and tonsil immunopeptidomes were based on HLA class I binders and HLA class II-presented peptides.

Analysis of HLA class I allelic distribution

The allele frequencies of HLA class I alleles occurring within the OPSCC cohort were determined for an analysis of the HLA class I allelic distribution. The cohort “Germany pop 8” (n = 39,689) as comprised in The Allele Frequency Net Database (www.allelefrequencies.net) served as reference dataset [61]. For statistical data analysis, the GraphPad Prism 6.1 software (GraphPad Software Inc) was used. Alleles assigned by HLA ligands were excluded from the analysis and 13 different HLA-A alleles, 21 different HLA-B alleles and 16 different HLA-C alleles remained for the OPSCC cohort. Thus, 13, 21 and 16 tables in the format 2x2 for HLA-A, -B and -C, respectively, were generated. Chi-square test was applied and, after Bonferroni correction, p value ≤ 0.004 (≙ 0.05/13) for HLA-A, p value ≤ 0.002 (≙ 0.05/21) for HLA-B and p value ≤ 0.003 (≙ 0.05/16) for HLA-C were used as levels of statistical significance. Logistic regression analysis determined relative risks that were expressed by the odds ratio with a 95% confidence interval (CI). Additionally, the HLA class I allelic distribution was separately analyzed in the tumor subgroup of HPV$^+$ OPSCCs applying the same significance levels as for the complete cohort.

HLA ligand isolation

Isolation of HLA-presented peptides was carried out as described before by Nelde *et al.* (2019) [62]. HLA precipitation was performed by immunoaffinity chromatography using antibodies specific for HLA molecules **(Appendix, Table 11)**. This was followed by acidic elution and purification of HLA-presented peptides.

LC-MS/MS analysis

Reversed-phase nanoflow ultra-high-performance liquid chromatography (nanoUHPLC, UltiMate 3000 RSLCnano, Dionex) was applied for the separation of HLA ligand pools. HLA ligands were subsequently analyzed *via* tandem mass spectrometry (MS/MS) in an on-line coupled LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific).

Data processing

The Proteome Discoverer 1.4 software (Thermo Fisher Scientific) was used for processing of experimentally produced MS raw data. The human proteome as comprised within the Swiss-Prot database (20,279 reviewed protein sequences; September 27th, 2013) served as basis for database search and spectral alignment.

Analysis of HLA ligandomic yields

To observe the proportion of the identified HLA ligands within the assumed amount of overall detectable HLA ligands (saturation value), saturation analysis was performed. This analysis was based on the rate of newly identified unique ligands per sample.

Sample masses as well as yields of isolated HLA-presented peptides varied between individual tumors. Therefore, a potential correlation between sample mass and peptide yields was examined. For statistical analysis, the GraphPad Prism 6.1 software (GraphPad Software Inc) was used and Pearson´s correlation was applied. A p value ≤ 0.05 was defined as level of statistical significance.

Comparative profiling against in-house immunopeptidomic databases

Comparative profiling was performed against an in-house benign database to search for tumor-exclusive antigens in the immunopeptidomes of OPSCCs. This database consists of 419 HLA class I ligandomes and 364 HLA class II ligandomes from benign samples **(Appendix, Table 15)**.
Another comparative profiling was performed with the identified tumor-exclusive antigens against an in-house malignant database to reveal their appearance in the immunopeptidomes of other tumor entities. After the datasets from OPSCCs of the present study were excluded, the malignant database consisted of 768 HLA class I ligandomes and 555 HLA class II ligandomes from malignant samples **(Appendix, Table 16)**.

Search for potential TAAs

The OPSCC immunopeptidome was screened for potential tumor-associated antigens (TAAs). To identify TAAs that are universal across different patients, frequent antigens occurring in at least 3 patients (corresponding to a frequency ≥ 8%) were of major interest for this study. The screening for

TAAs included the search for previously described cancer-testis antigens (CTAs) presented by HLA molecules as comprised in the CTdatabase (www.cta.lncc.br; state August 2019) [63]. To identify additional TAAs, frequent tumor-exclusive antigens were analyzed in detail (frequency ≥ 8%). Thereby, HLA-presented peptides and corresponding source proteins of these tumor-exclusive antigens underwent evaluation to investigate their potential as TAAs. This entailed the examination of specific technical quality criteria from peptides including the spectrum quality, peptide spectrum matches (PSMs), cross correlation (Xcorr) and delta correlation (ΔCn). Additionally, the expression profiles from corresponding source proteins within different healthy tissues and tumor entities as comprised in the databases The Genotype-Tissue Expression Project (GTEx; www.gtexportal.org) [64] and The Human Protein Atlas (www.proteinatlas.org) [65] were investigated.

Whole exome sequencing

Whole exome sequencing (WES) was carried out on twelve pairs of tumors and corresponding peripheral blood mononuclear cells (PBMCs) by paired-end DNA sequencing technology. After alignment against the human reference genome hg19, variant calling was performed and variants were annotated to unravel somatic mutations of the tumors.

Search for neoantigens

The detected somatic mutations were used as basis to search for neoantigens among the HLA ligandomic data of the twelve DNA-sequenced tumor samples. The AA sequences of proteins encoded by mutated genes in combination with the human proteome of the Swiss-Prot database served as basis for the neoantigen search.

Search for HPV-derived antigens

Additionally, the immunopeptidomic dataset of the entire OPSCC cohort was screened for HLA-presented peptides derived from HPV proteins. The HPV proteome as comprised in the Swiss-Prot database (470 reviewed protein sequences, January 5th, 2018) was combined with the human proteome of the Swiss-Prot database and served as basis for this search.

RNA sequencing

RNA sequencing was carried out by single-end RNA sequencing technology on 15 OPSCC samples of the present study cohort. The alignment of reads was performed against the human reference genome hg19. Using normalized RPKM-counts, the correlation between RPKM-counts and counts of HLA-presented sequences per protein was examined. For this purpose, Pearson´s correlation analysis with a p value ≤ 0.05 as level of statistical significance and linear regression analysis were applied using the GraphPad Prism 6.1 software (GraphPad Software Inc).

Additionally, the determination of the HPV status of the tumors was based on the RNA read counts of the viral proteins L1, L2, E1, E2, E4, E5, E6 and E7.

Comparative analysis of HPV$^+$ and HPV$^-$ OPSCC

In an unsupervised and a supervised principle component analysis, a potential clustering into HPV$^+$ and HPV$^-$ OPSCC samples based on source proteins of HLA class I binders and HLA class II-presented peptides was investigated.

Additionally, comparative profiling of HPV$^+$ against HPV$^-$ tumors was performed to reveal HPV-associated differences within the HLA ligandomes.

Furthermore, differential expression analysis was carried out on the basis of the RNA sequencing data of 15 OPSCC samples using the DESeq2 algorithm [66]. Thereby, potential differences between HPV$^+$ and HPV$^-$ tumors were examined on the transcriptomic level. For a comparability of transcriptomic and immunopeptidomic data, differentially expressed genes that encoded proteins of the Swiss-Prot database were selected.

Immunopeptidomic and transcriptomic data were integrated by comparing HPV$^+$- and HPV$^-$-exclusive antigens revealed by comparative analysis with dysregulated genes revealed by differential expression analysis.

Functional annotation clustering

The functional annotation tool of the database DAVID Bioinformatics Resources 6.8 [67, 68] was applied for GO (gene ontology) term enrichment. Functional annotation clustering was performed for tumor-exclusive and HPV$^+$- and HPV$^-$-exclusive source proteins of both HLA class I binders and HLA class II-presented peptides. Additionally, the proteins encoded by differentially expressed genes were used as basis for functional annotation clustering.

3.4 Results

3.4.1 HLA class I allelic distribution

HLA typing results were available for 37 patients of the OPSCC cohort. These were used to analyze the HLA class I allelic distribution across the patient cohort. HLA allele frequencies were determined for the alleles represented within the study cohort and compared to expected allele frequencies of a normal German population [61]. Statistical analysis was performed applying chi-square tests to reveal potential deviations between patients and reference.

Most of the analyzed alleles had a similar frequency in the patient as in the reference dataset indicating a representative cohort in respect of the HLA class I allelic distribution **(Figure 20)**. However, the patient frequencies of a few alleles deviated from the expectation. This included for example the alleles HLA-B*51:01 or C*12:03. Of these, one significant deviation was revealed by statistical analysis. With an allele frequency of F = 16.2% compared to a frequency of F = 6.3% within the reference population, HLA-B*51:01 had a significantly higher prevalence within the patient cohort (p value = 0.0005; OR = 2.9; 95% CI = 1.6 to 5.3). None of the analyzed HLA class I alleles had a significantly lower allele frequency within OPSCC patients than in the reference dataset.

Associations between several HLA alleles and an increased or decreased risk for the development of a persisting HPV infection resulting in cancer were previously published [27]. Thus, the HLA class I allelic distribution was analyzed separately in the tumor subgroup of 22 HPV^+ OPSCCs. Statistical analysis showed no significantly increased or decreased frequencies for the observed alleles compared to the German reference population **(Supplementary Tables 11, 12 and 13)**. HLA-B*51:01 had a high prevalence in both HPV^+ and HPV^- patients with frequencies of F = 16.7% and F = 15.6%, respectively. However, the differences compared to the German reference (F = 6.3%) were not significant, provided that the significance level of the total cohort (p value ≤ 0.003) is applied (p value(HPV^+) = 0.0057; p value(HPV^-) = 0.0300). Larger numbers of samples per tumor subgroup might be necessary to reveal statistical significance in association with the HPV status.

In conclusion, analysis of the HLA class I allelic distribution revealed a mainly representative OPSCC patient cohort with one exception. The allele HLA-B*51:01 had a significantly higher prevalence within the total study cohort independent of the HPV status.

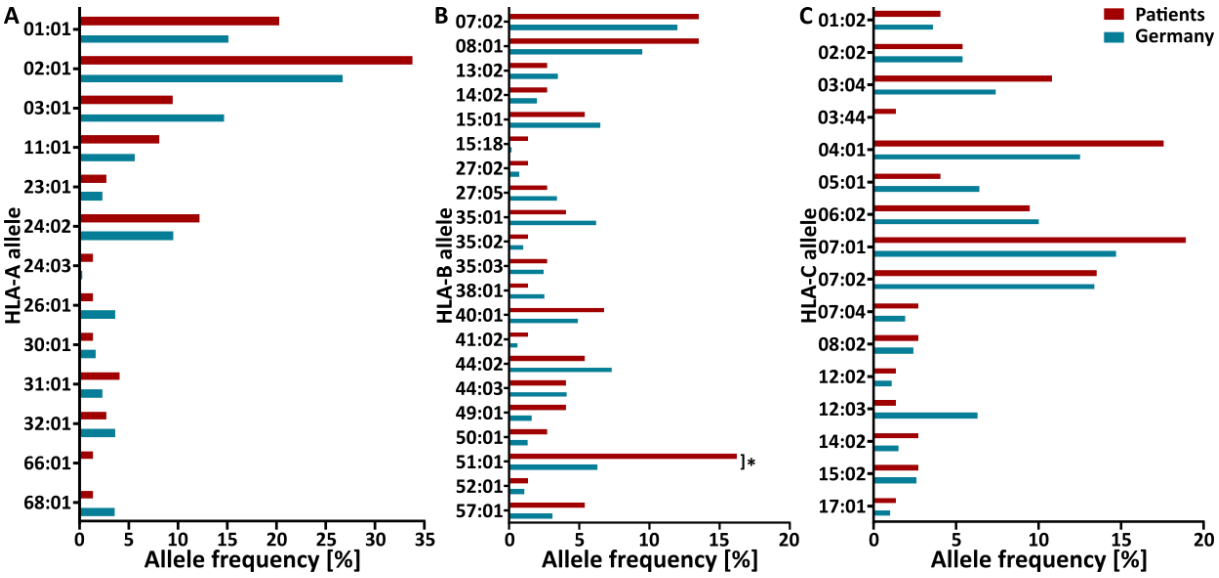


Figure 20: HLA class I allelic distribution in the OPSCC patient cohort and a German reference population. Allele frequencies of (A) HLA-A, (B) HLA-B and (C) HLA-C alleles occurring within the study cohort were determined. Most investigated alleles displayed allele frequencies comparable to those expected in a German population indicating a representative HLA class I allelic distribution of the patient cohort. One deviation was revealed: HLA-B*51:01 (p value = 0.0005) had a significantly higher prevalence within the patient cohort compared to the reference database. $p \leq 0.004$ for HLA-A, $p \leq 0.002$ for HLA-B and $p \leq 0.003$ for HLA-C alleles were applied as levels of statistical significance after Bonferroni correction. The German allele frequencies were used as comprised in the Allele Frequency Net Database (cohort "Germany pop 8") [61]. n(patients) = 37; n(Germany) = 39,689; * – significant.

3.4.2 General characteristics of the OPSCC HLA ligandome

Peptide yields

The HLA ligands of 40 OPSCC tissue samples were extracted *via* immunoaffinity purification. Analysis of the isolated peptides was performed by LC-MS/MS. This revealed a dataset of HLA ligands presented on the cell surfaces of OPSCC tumor tissue. Overall, 25,228 different HLA class I-presented peptides derived from 9,485 source proteins and 15,203 different HLA class II-presented peptides derived from 4,634 source proteins were detected. In average, 1,424 HLA class I-presented peptides per sample (1,1453 per HPV^+ and 1,1388 per HPV^- sample) and 702 HLA class II-presented peptides per sample (714 per HPV^+ and 687 per HPV^- sample) were identified **(Figure 21, Supplementary Table 14)**. Additionally, the immunopeptidomic data of benign tonsil samples were collected to serve as benign reference dataset. 8,201 different HLA class I ligands from 5,093 source proteins and 5,380 different HLA class II ligands from 1,832 source proteins were identified in these five tissue samples **(Figure 22, Supplementary Table 15)**.

Determination of HLA class I binders

In case of HLA class I, the binders among the isolated ligands were determined based on HLA typing or HLA allotype assignment and peptide binding motif restrictions. In OPSCCs, this revealed 22,769 different HLA class I binders derived from 8,565 source protein. The purity was calculated as percentage of binders among all identified ligands and was above 90% for the majority of the OPSCC samples **(Figure 21, Supplementary Table 14)**. In the five tonsil samples, 7,707 different HLA class I binders derived from 4,807 source proteins were identified with a purity between 94% and 96% in individual samples **(Figure 22, Supplementary Table 15)**. All following HLA class I ligandomic analyses were performed based on the dataset of HLA class I binders.

Correlation between sample masses and peptide yields

The peptide yields of individual samples ranged from 265 to 2,854 HLA class I binders and from 168 to 2,086 HLA class II-presented peptides. Moreover, the sample masses varied between 17.2 mg and 567.4 mg. Pearson´s correlation and non-linear regression analysis **(Supplementary Figure 7)** were performed and revealed significant (p value ≤ 0.05) positive correlations between sample masses and yields of isolated HLA class I binders (Pearson´s correlation coefficient r = 0.4645; p value = 0.0019; 95% CI = 0.2 to 0.7) as well as between sample masses and yields of HLA class II-presented peptides (Pearson´s correlation coefficient r = 0.6388; p value < 0.0001 ; 95% CI = 0.4 to 0.8). The achieved peptide yields in the five individual tonsil samples were between 1,602 and 2,586 HLA class I binders and between 1,026 and 2,012 HLA class II-presented peptides while the sample masses ranged from 162.1 mg to 487.2 mg. No significant correlation between sample masses and peptide yields was revealed in the tonsil cohort, which might be related to the small number of samples.
Taken together, these results showed that larger OPSCC sample masses allowed the identification of more unique HLA ligands.

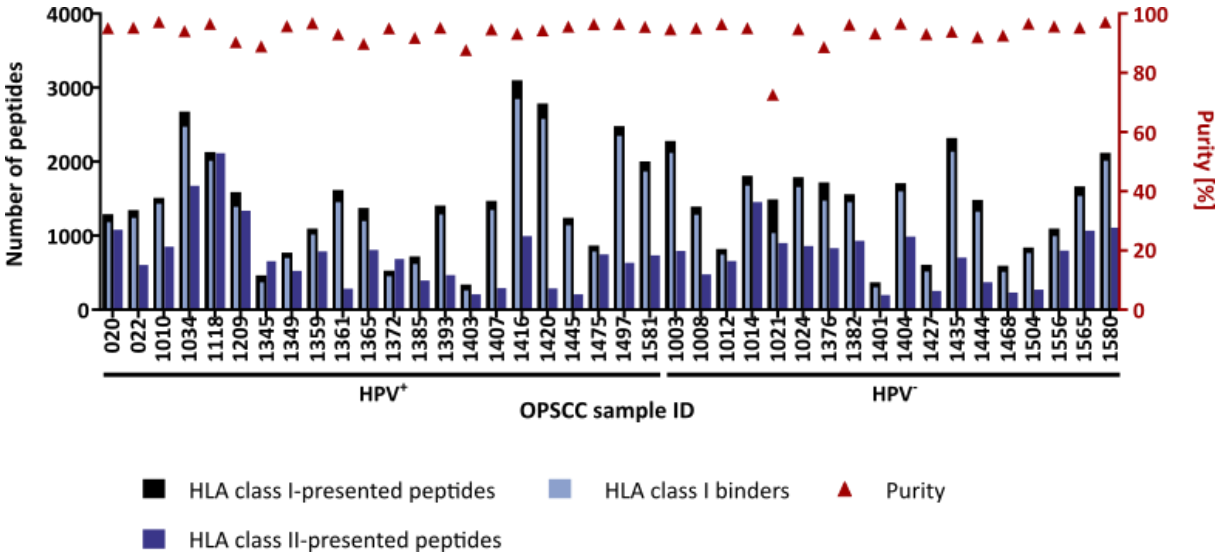


Figure 21: Yields of isolated HLA class I- and class II-presented peptides for individual OPSCC samples achieved by LC-MS/MS analysis. The peptide yields varied between 265 and 2,854 HLA class I- and between 168 and 2,086 HLA class II-presented peptides per individual sample. HLA class I binders were defined as HLA class I-presented peptides carrying a binding motif of an HLA allotype of the respective patient. The purity, which is defined as the proportion of binders among all isolated peptides, was determined for HLA class I and lay between 73% and 97% for the individual samples.

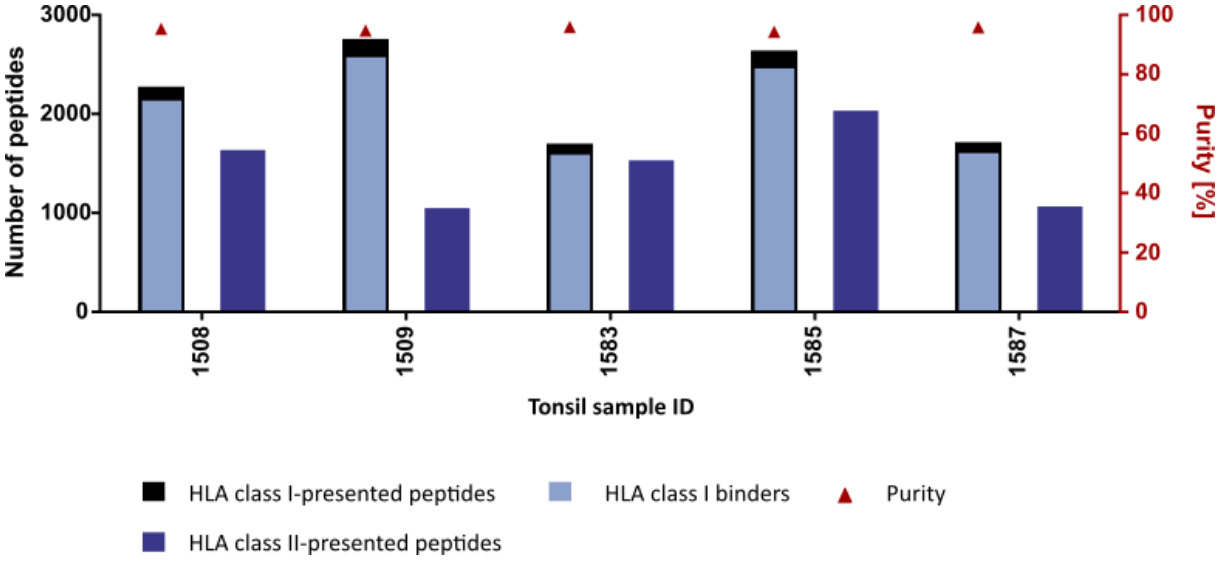


Figure 22: Yields of isolated HLA class I- and class II-presented peptides for individual tonsil samples achieved by LC-MS/MS analysis. The peptide yields varied between 1,602 and 2,586 HLA class I- and between 1,026 and 2,012 HLA class II-presented peptides per individual sample. HLA class I binders were defined as HLA class I-presented peptides carrying a binding motif of an HLA allotype of the respective patient. The purity, which is defined as the proportion of binders among all isolated peptides, was determined for HLA class I and lay between 94% and 96% for the individual samples.

Saturation analysis

Saturation analysis allowed an estimation on the sizes of both the OPSCC and the tonsil immunopeptidome. This calculation was based on the rate of newly identified unique antigens per additional sample.

The present coverage rate was determined as percentage of the ligand count that was achieved with the investigated OPSCC samples among the maximal number of ligands (saturation value). For HLA class I binders, a saturation value of 30,566 binders was determined. With 22,769 identified binders, a coverage rate of 75% was achieved **(Supplementary Figure 8A)**. A saturation value of 25,885 HLA class II-presented peptides was determined. With 15,203 identified HLA-presented peptides, a coverage rate of 59% was reached **(Supplementary Figure 8C)**. On the source protein level, the coverage rates were higher. The calculated saturation value was 8,712 source proteins of HLA class I binders. With 8,565 identified source proteins, 98% were covered **(Supplementary Figure 8B)**. A saturation value of 2,685 source proteins of HLA class II-presented peptides was determined and, with 1,751 source proteins, a coverage rate of 65% was reached **(Supplementary Figure 8D)**.

This analysis was equally performed for the immunopeptidomic dataset of the tonsils. Saturation values of 14,560 HLA class I binders and 10,141 HLA class II-presented peptides were calculated. Within five tonsil samples, 7,707 HLA class I binders and 5,380 HLA class II-presented peptides were identified corresponding to coverage rates of 53% each **(Supplementary Figure 9A and C)**. On the source protein level, the saturation values were 5,837 source proteins of HLA class I binders and 2,821 source proteins of HLA class II-presented peptides. With the identification of 4,807 source proteins of HLA class I binders and 1,832 source proteins of HLA class II-presented peptides, 82% and 65% were covered, respectively **(Supplementary Figure 9B and D)**.

Peptide length distribution

Peptide length filters were applied during data processing limiting the permitted number of amino acids (AAs) to 8-12 for HLA class I-presented peptides and to 8-25 for HLA class II-presented peptides. Within this range, the peptide length distribution of HLA ligands identified in OPSCCs was addressed **(Supplementary Figure 10)**. The majority of HLA class I binders (70%) had a length of 9 AAs. The HLA class II-presented peptides were spread more widely across the length range with a peak at a length of 15 AAs corresponding to 17%. A similar peptide length distribution was revealed for HLA-presented peptides isolated from tonsils: 71% of HLA class I binders consisted of 9 AAs whereas 19% of HLA class II ligands had a length of 15 AAs **(Supplementary Figure 11)**. These values correspond to the commonly expected length distribution for HLA class I and class II ligands [69, 70].

3.4.3 Identification of CTAs

In the course of the search for tumor-associated antigens (TAAs), the OPSCC ligandome dataset was screened for the presentation of cancer-testis antigens (CTAs). The collective of CTAs as comprised in the CTdatabase [63] was used as basis for this search.

Overall, 70 CTAs were revealed in the OPSCC immunopeptidomic dataset. Detailed antigen evaluation was applied and frequent antigens were of major interest to reveal universal CTAs across different OPSCC patients. 13 remaining CTAs had a frequency of at least 8% (≥ 3 samples) and served as source proteins of 38 different HLA class I binders meeting all peptide quality criteria **(Table 7)**. On the source protein level, the most frequent CTA was the U3 small nucleolar ribonucleoprotein IMP3. It was represented by four different peptides in 23 tumor samples corresponding to a representation frequency of 58%. On the peptide level, the IMP3-derived sequence ALLDKLYAL had the highest frequency (50%) among the CTA-derived HLA class I binders. This peptide was annotated to HLA-A*02:01 and was found in 95% of all HLA-A*02:01-positive patients of the study cohort. Within the HLA class II ligandome, no CTAs were identified that withstood the peptide quality checks in more than two samples. **Supplementary Table 16** shows further details about the frequent CTAs detected in the OPSCC HLA ligandome.

The results demonstrate that OPSCCs present previously described CTAs *via* CTA-derived peptides by HLA molecules on the tumor cell surfaces. These might be relevant for further consideration as potential targets in immunotherapeutic approaches for OPSCC patients.

Table 7: CTAs detected within the OPSCC immunopeptidomic dataset. The search for CTAs was performed on the basis of the CTdatabase, which comprises 277 previously described CTAs [63]. 38 HLA class I binders derived from CTAs with a frequency of at least 8% were revealed in the OPSCC immunopeptidome. No HLA class II-presented peptides that derived from CTAs met all quality criteria. CTA – cancer-testis antigen; AC – accession; GN – gene name.

Peptide			Source protein		
Sequence	**HLA restriction**	**Representation frequency**	**UniProtKB AC**	**GN**	**Representation frequency**
ALLDKLYAL	A*02:01	50%	Q9NV31	IMP3	58%
RSMEDFVTW	B*57:01	8%			
QHLQAAVAF	A*32:01; B*15:018	3%			
EDYTRYNQL	B*08:01	3%			
DPFAFIHKI	B*51:01	28%	Q9UGL1	KDM5B	43%
ILNPYNLFL	A*02:01;C*17:01	23%			
PFAFIHKI	B*51:01	3%			
RLIDLGVGL	A*02:01	18%			
QEAFGFEQA	B*50:01	3%			
SVAQQLLNGK	A*11:01	3%			
FIDASRLVY	A*01:01	30%	P26232	CTNNA2	33%
KVKAEVQNL	C*03:04	3%			
AYAIIKEEL	A*24:02;A*24:03;A*23:01	28%	Q6PL18	ATAD2	30%
LYPEVFEKF	A*24:02;A*23:01	10%			
KYLTVKDYL	A*23:01	3%			
GLSNHIAAL	A*02:01	3%			
SQNAIDHKI	B*13:02	3%			
SYQKVIELF	A*24:02;A*24:03;A*23:01	28%	Q96KB5	PBK	28%
KILDLETQL	A*02:01	18%	Q5BJF6	ODF2	20%
AEALSTLESW	B*44:02	3%			
AELESKTNTL	B*40:01	3%	Q53EZ4	CEP55	18%
QEEQTRVAL	B*40:01	3%			
SPKSPTAAL	B*07:02	10%			
EELLSQVQF	B*44:02	3%			
DEAVGVQKW	B*44:02;B*44:03	13%	Q14667	KIAA100	15%
DYPRYLFEI	A*24:02	5%			
DAVKFFVAV	B*51:01	13%	O60271	SPAG9	15%
NYADQISRL	C*07:02	3%			
LEEQASRQI	B*49:01	5%	Q99661	KIF2C	13%
IYNGKLFDLL	B*24:02	5%			
RLFPGLAI	A*03:01;C*06:02	5%			
SPIEKSGVL	B*07:02	8%	Q8NG31	CASC5	10%
IYVIPQPHF	A*23:01	3%			
ALLQEIVNI	A*02:01	5%	Q92600	RQCD1	8%
ILLDDTGLAYI	A*02:02	3%			
KIYQWINEL	A*32:01	3%			
KVLEYVIKV	A*02:01	8%	P43355	MAGEA1	8%
KEADPTGHSY	B*44:03	3%			

3.4.4 Identification of OPSCC TAAs

In addition to already published CTAs, the HLA ligandomic dataset of OPSCCs was screened for further TAAs. Comparative profiling of the OPSCC data against an in-house benign database **(Appendix, Table 15)** was performed to obtain those antigens that were not presented on benign tissue. These were further referred to as tumor-exclusive antigens. 83% of the OPSCCs included in the study originated from tonsillar location. Since the benign database did so far not include any HLA immunopeptidomic data of tonsillar tissue samples, five benign tonsils served as additional healthy reference to exclude tissue-specific antigens and improve the search for OPSCC TAAs.

Peptide level

By comparison of the OPSCC immunopeptidomic dataset with the benign database and the tonsil dataset, 3,746 tumor-exclusive HLA class I binders were revealed **(Figure 23A)** and 19,023 shared binders were excluded. The benign database was responsible for most of the exclusions, but 109 additional HLA class I binders were excluded by the tonsil dataset. To identify TAAs that are universal across several patients, the focus was on tumor-exclusive antigens occurring in at least three samples (frequency ≥ 8%). 101 tumor-exclusive HLA class I binders with a frequency above 8% were detected **(Supplementary Table 17)**. Two of the identified CTA-derived peptides (see 3.4.3) were among the tumor-exclusive HLA class I binders but with a frequency below 8% **(Supplementary Table 16)**. In case of HLA class II, 5,096 presented peptides remained after subtraction of the benign database and the tonsil dataset. The tonsil dataset was responsible for the exclusion of 250 HLA class II-presented peptides additional to those excluded by the benign database **(Figure 23B)**. 84 tumor-exclusive HLA class II-presented peptides occurred with a frequency above 8% among the OPSCC cohort **(Supplementary Table 18)**.

The frequent tumor-exclusive peptides resulting from comparative profiling against the benign datasets were further investigated to identify potential TAAs in the OPSCC immunopeptidome.

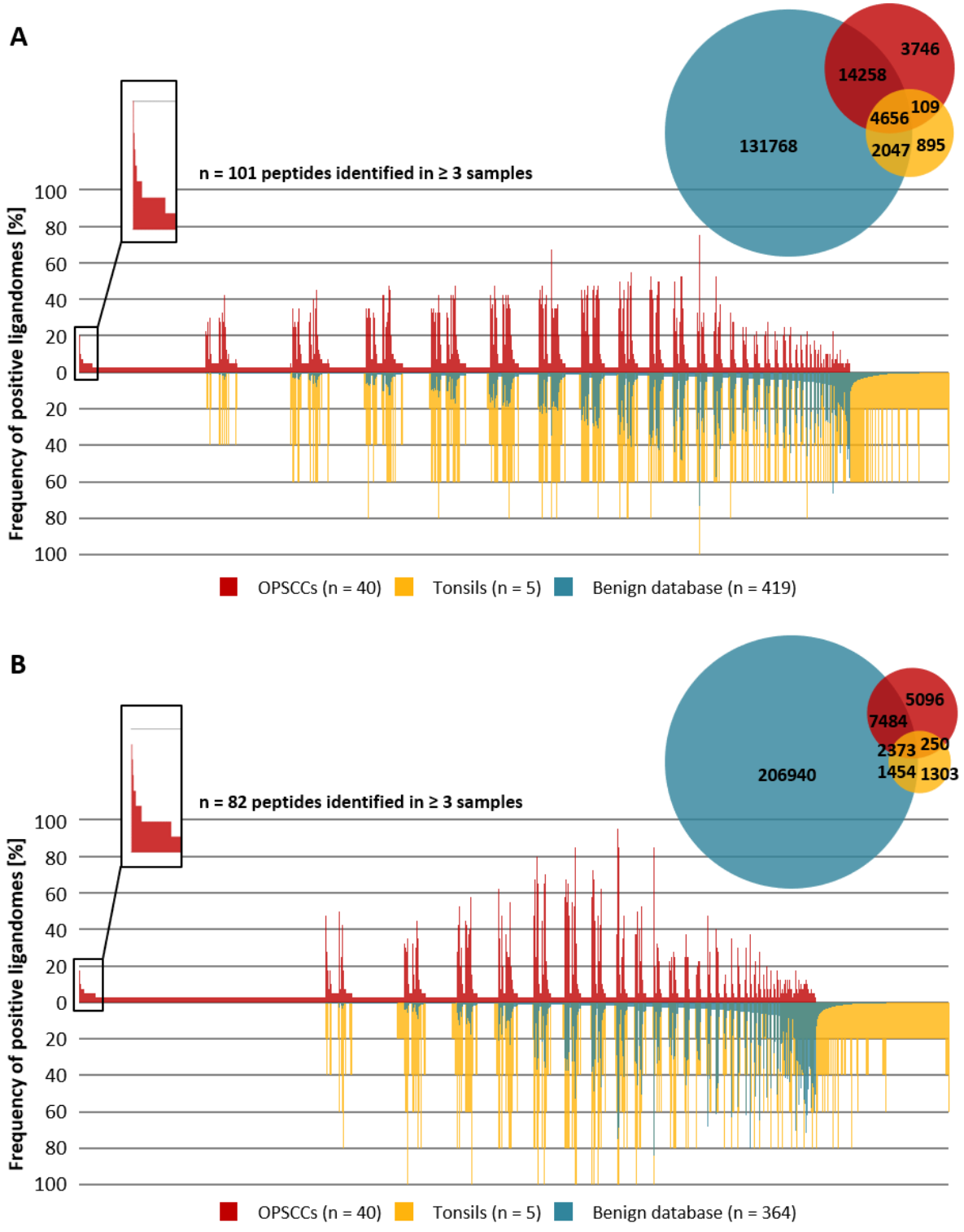


Figure 23: Comparative profiling of the HLA ligandomes of OPSCCs against an in-house benign database and the tonsil dataset on the peptide level. The analysis was performed for both (A) HLA class I binders and (B) HLA class II-presented peptides. This revealed tumor-exclusive peptides as shown by the Venn diagrams. The waterfall plots demonstrate the frequencies of detected peptides within the study cohort, the benign database and the tonsils dataset in comparison. Unique ligands are plotted on the x-axis and the corresponding percentage of HLA ligandomes that contain the respective ligands on the y-axis. Tumor-exclusive peptides that occurred among the OPSCCs with a frequency of at least 8% (≥ 3 samples) were of further interest and are highlighted on the far left of the waterfall plots.

To identify potential TAAs of OPSCCs, frequent tumor-exclusive HLA-presented antigens were at first evaluated on the source protein level. Their expression profiles in different healthy tissue types and tumor entities, which are accessible in the databases GTEx [64] and The Human Protein Atlas [65], respectively, were examined. Thereby, antigens were selected that derived from source proteins with low RNA expression levels in healthy tissues (≤ 10 TPM) or with elevated RNA expression levels solely in testis tissue resembling a CTA expression profile.

Eight HLA class I- and thirteen HLA class II-presented antigens with promising RNA expression profiles of source proteins remained and were further observed on the peptide level. To confirm the peptide sequences that were annotated to experimental MS spectra during data processing, several technical quality criteria were evaluated. Three HLA class I binders and five HLA class II-presented peptides withstood the quality checks **(Table 8)**. Two of the selected HLA class I-presented targets were restricted to the HLA allotype B*51:01. Both were identified in 33% of OPSCC samples with the respective allotype. This might be related to the significantly elevated prevalence of the HLA allele B*51:01 among the OPSCC patients of the study cohort (see 3.4.1). However, the circumstance that none of the five tonsil donors and only nine donors of the benign database were carriers of the allele HLA-B*51:01, must be considered in this context.

Only eight peptides remained that met all the criteria of the applied selection strategy. Therefore, a closer look was taken on the peptides derived from tissue-specific source proteins. 22 source proteins of frequent tumor-exclusive antigens have been shown to have an RNA expression restricted to the anatomic compartment of the tumor, in particular to esophageal tissue. This was often accompanied by an elevated expression in HNSCC tissue. Therefore, the focus was on source proteins with a tissue-specific expression profile in the tumor subsite that simultaneously displayed an altered HLA-presentation. In case these tissue-associated source proteins were represented by different peptides or occurred not at all in the benign reference datasets compared to the OPSCC immunopeptidome, the tumor-exclusive peptides were further investigated. Peptides with length variants in benign tissues were excluded. This screening strategy allowed the selection of six additional HLA class I binders from five different source proteins and one additional HLA class II-presented peptide as potential TAAs **(Table 9)**.

Altogether, fifteen TAAs were selected among the OPSCC tumor cohort that might have potential as target antigens for the development of novel immunotherapeutic approaches for the treatment of OPSCC patients.

Table 8: TAAs identified in OPSCCs. After comparative profiling against an in-house benign database and detailed evaluation of the resulting frequent tumor-exclusive peptides and corresponding source proteins, three HLA class I binders and five HLA class II-presented peptides remained as potential TAAs. TAA – tumor-associated antigen; AC – UniProtKB accession; GN – gene name.

HLA class	Peptide				Source protein		
	Sequence	HLA restriction	Representation frequency in		AC	GN	Protein name
			cohort	allotype-positive samples			
I	**DANPYDSVKKI**	B*51:01	10%	33%	O15205	UBD	Ubiquitin D
	VALPVYLLI	B*51:01	10%	33%	P57054	PIGP	Phosphatidyl-inositol N-acetylgluco-saminyltransferase subunit P
	HGTIKNQL	B*08:01	8%	30%	Q86XI2	NCAPG2	Condensin-2 complex subunit G2
II	**MLFFDKFANIVPF**	-	13%	-	O0220	TNFRSF10A	Tumor necrosis factor receptor superfamily member 10A
	LPFLDIAPLDIGG	-	10%	-	P08123	KIF4B	Chromosome-associated kinesin KIF4B
	TNKPSRLPFLDIAPLDIGGADQ	-	10%	-			
	TKPAIRRLARRGGVKR	-	10%	-	P62805	COL1A2	Collagen alpha-2(I) chain
	VDEEEVKFPGTNFDE	-	8%	-	Q9BVV6	KIAA0586	Protein TALPID3

Table 9: Tissue-related TAAs identified in OPSCCs. Tumor-exclusive antigens derived from source proteins with RNA expression profiles specific to the tumor subsite and altered HLA presentation in tumor and benign tissue were selected. Thereby, six additional HLA class I binders and one additional HLA class II-presented peptide were identified as potential TAAs. TAA – tumor-associated antigen; AC – UniProtKB accession; GN – gene name.

HLA class	Peptide				Source protein		
	Sequence	HLA restriction	Representation frequency in		AC	GN	Protein name
			cohort	allotype-positive samples			
I	**EETNPKGSGW**	B*44:02; B*44:03	13%	56%	Q13835	PKP1	Plakophilin-1
	KLAEISLGV	A*02:01	10%	19%	P32926	DSG3	Desmoglein-3
	YTDNWLAVY	A*01:01	8%	21%			
	YLDDPTNSWY	A*01:01	10%	29%	O60479	DLX3	Homeobox protein DLX-3
	GEAAKSVKL	B*40:01; B*41:02	8%	43%	Q9HCY8	S100A14	Protein S100-A14
	SVKITQVTW	B*57:01	8%	75%	Q15223	NECTIN1	Nectin-1
II	**TPAAQYVRIKENLAVG**	-	8%	-	Q14574	DSC3	Desmocollin-3

Protein level

Comparative profiling of the OPSCC immunopeptidome against the in-house benign database and the tonsil dataset was additionally performed on the source protein level. Antigens with source proteins shared between benign and tumor datasets were excluded. The benign database was responsible for the majority of exclusions. The tonsillar reference excluded four additional source proteins. Eventually, 99 tumor-exclusive source proteins of HLA class I binders were revealed **(Figure 24A)**. Three tumor-exclusive source proteins of HLA class I binders had a frequency of at least 8%: Homeobox protein DLX-3 (DLX3), putative uncharacterized protein UNQ6190/PRO20217 and podoplanin. The DLX3-derived peptide YLDDPTNSWY was already identified as potential TAA on the peptide level **(Table 9)**. The other two tumor-exclusive source proteins and their corresponding peptides did not withstand the quality checks. 218 tumor-exclusive source proteins of HLA class II-presented peptides were found in the OPSCC cohort **(Figure 24A)**. Two of them were identified with a frequency above 8%: phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2 (INPPL1) and neuronal acetylcholine receptor subunit alpha-10 (CHRNA10). However, both source proteins and corresponding peptides were excluded from further consideration as potential TAAs after the quality checks. None of the CTA source proteins previously detected in the OPSCC tumor cohort (see 3.4.3) were among the frequent tumor-exclusive antigens.

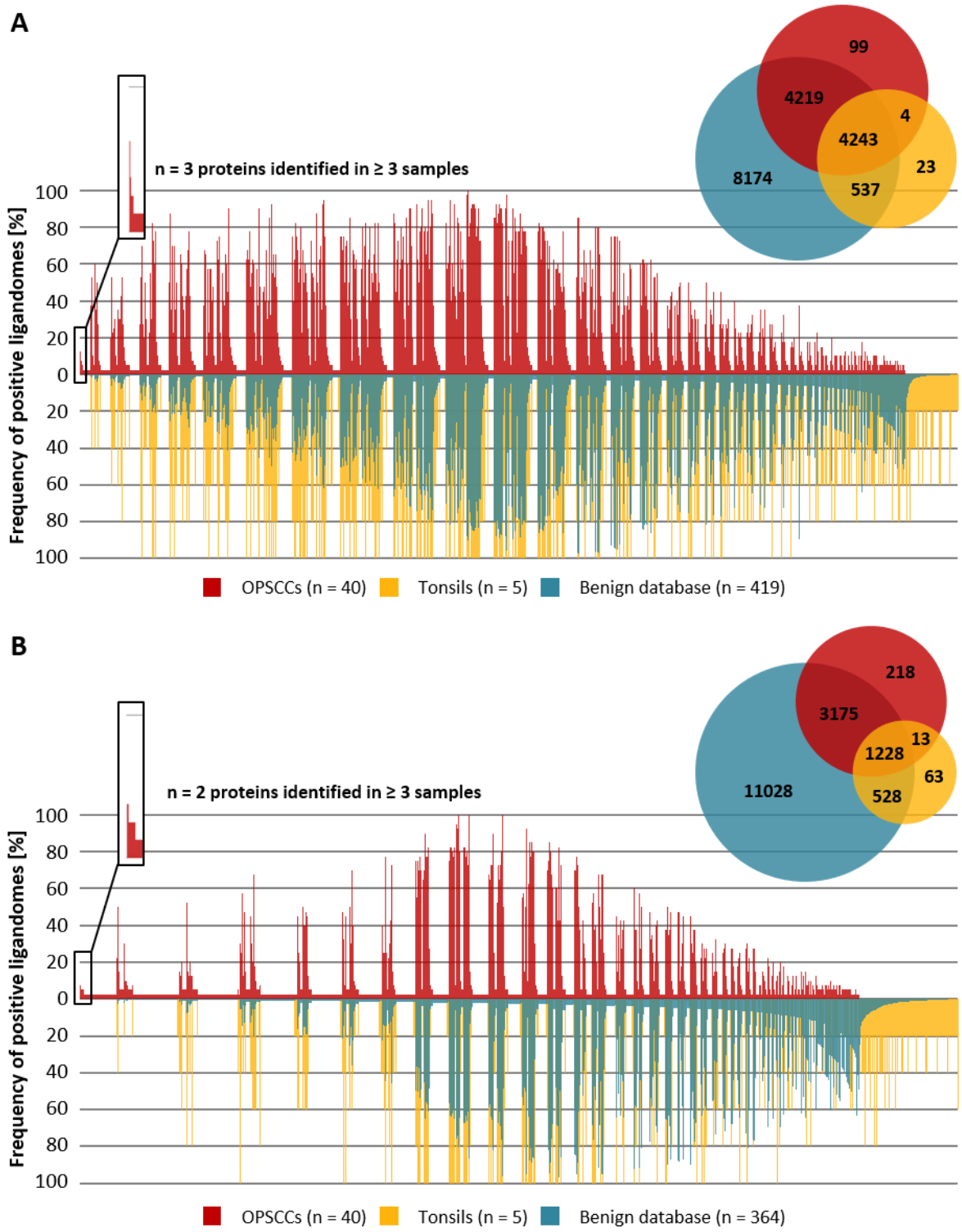


Figure 24: Comparative profiling of the HLA ligandomes of OPSCCs against the in-house benign database and the tonsil dataset on the source protein level. The analysis was performed for source proteins of both (A) HLA class I binders and (B) HLA class II-presented peptides. This revealed several tumor-exclusive source proteins as shown by the Venn diagrams. The waterfall plots demonstrate the frequencies of detected source proteins within the study cohort, the benign database and the tonsil dataset in comparison. Unique ligands are plotted on the x-axis and the corresponding percentage of HLA ligandomes that contain the respective ligands on the y-axis. Tumor-exclusive source proteins that occurred among the OPSCCs with a frequency of at least 8% (≥ 3 samples) were of further interest and are highlighted on the far left of the waterfall plots.

In summary, the comparative profiling of the OPSCC immunopeptidomic dataset against the benign references enabled the selection of several TAAs on the peptide level. Analyses on the source protein level revealed no additional TAAs but partly confirmed the selection.

3.4.5 OPSCC TAAs in other malignancies

Tumor-exclusive antigens of the OPSCC immunopeptidome were revealed by comparative profiling against an in-house benign database (see 3.4.4). These tumor-exclusive antigens were additionally used for comparative profiling against an in-house malignant database **(Appendix, Table 16)**. Thereby, the occurrence and frequencies of these antigens in other tumor entities were investigated.

On both the peptide and the source protein level, the comparison between the malignant database and the OPSCC dataset showed shared and OPSCC-exclusive antigens **(Figure 25)**. 2,493 tumor-exclusive HLA class I binders and 1,945 tumor-excusive HLA class II-presented peptides of the OPSCCs have been identified in the immunopeptidomes of other tumor entities before. In contrast, 1,253 tumor-exclusive HLA class I binders and 3,151 tumor-excusive HLA class II-presented peptides were solely identified in the OPSCC dataset. 90 and 156 tumor-exclusive source proteins of HLA class I binders and HLA class II-presented peptides, respectively, were detected on other tumor entities before whereas 9 and 62 tumor-exclusive source proteins of HLA class I binders and HLA class II-presented peptides, respectively, were OPSCC-exclusive.

The occurrence of the selected TAAs (see 3.4.4) within other malignancies was addressed in more detail. The majority of the 15 peptides of interest have previously been identified as HLA ligands in various tumor samples of different cancer entities **(Supplementary Table 19)**. Only one HLA class I binder (EETNPKGSGW) and one HLA class I-presented peptide (VDEEEVKFPGTNFDE), both identified as potential targets in several OPSCC samples, have not been found in any samples of the malignant database.

The circumstance that most tumor-exclusive antigens and TAAs of OPSCCs occurred in other tumor entities before, further confirms their selection and supports their consideration as potential immunotherapeutic targets.

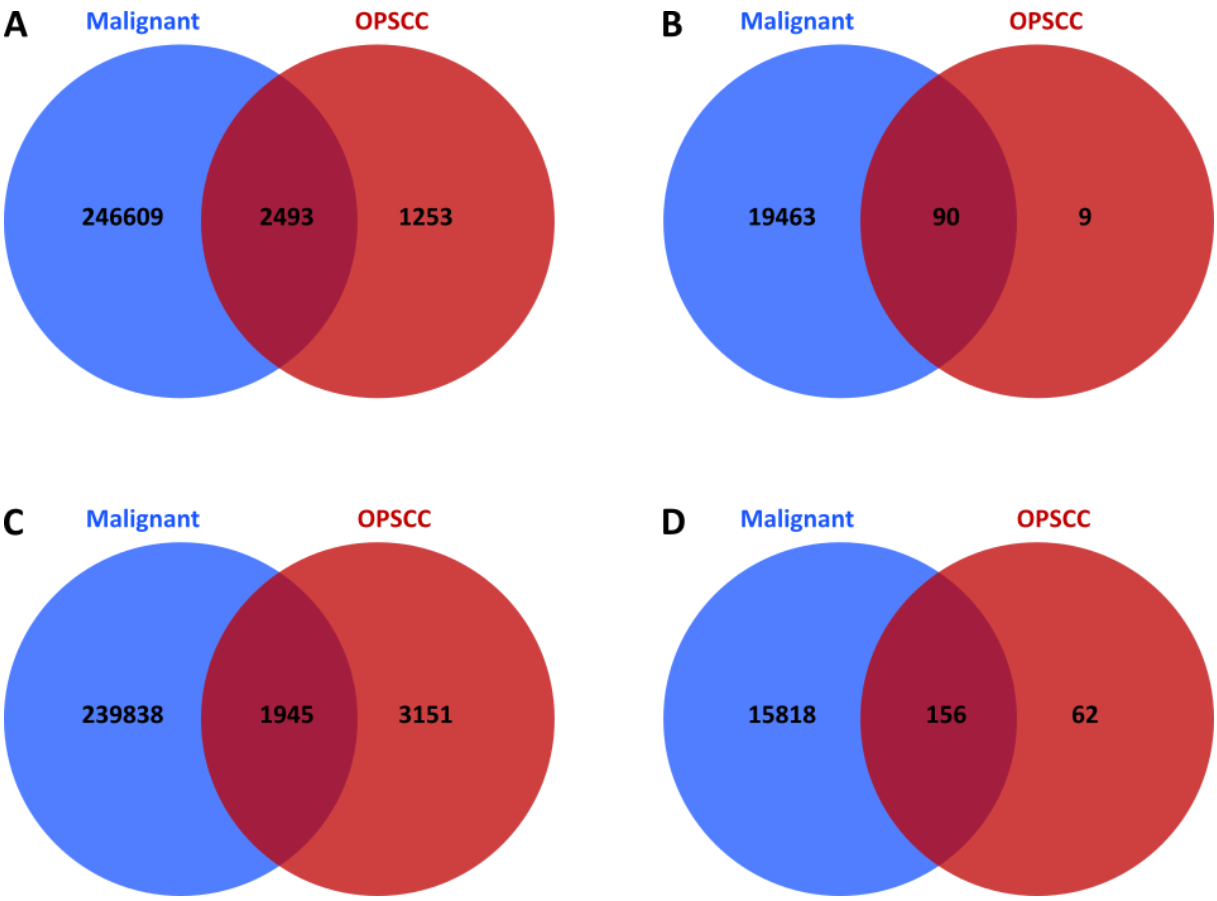


Figure 25: Comparison of tumor-exclusive antigens with in-house malignant database. The tumor exclusive antigens identified in the immunopeptidomic dataset of OPSSCCs were compared with an immunopeptidomic in-house malignant database. The Venn diagrams show the overlap between the malignant database and tumor-exclusive (A) HLA class I binders, (B) source proteins of HLA class I binders, (C) HLA class II-presented peptides and (D) source proteins of HLA class II-presented peptides. n(malignant HLA class I samples) = 768; n(malignant HLA class II samples) = 555; n(OPSCC) = 40.

3.4.6 Functional annotation clustering of tumor-exclusive antigens

To examine the functional relations of tumor-exclusive antigens, functional annotation clustering was performed using the functional annotation tool of the database DAVID Bioinformatics Resources 6.8 [67, 68]. Tumor-exclusive source proteins of HLA class I binders and HLA class II-presented peptides identified before (see 3.4.4) served as basis for GO term enrichment. The GO terms of molecular function (MF), biological process (BP) and cellular compartment (CC) were analyzed and a minimal enrichment score of 1 was used as threshold. Only one annotation cluster was revealed with an enrichment score of 1.33 on the basis of 99 tumor-exclusive source proteins from HLA class I binders. This indicated an involvement of 13 proteins in immune response mechanisms *via* activation of the complement system **(Table 10)**. GO term enrichment based on the 218 tumor-exclusive source proteins from HLA class II-presented peptides delivered no annotation cluster with an enrichment score ≥ 1.

Table 10: Result of GO term enrichment based on tumor-exclusive source proteins of HLA class I binders. One annotation cluster with an enrichment score ≥ 1 was annotated to 13 of the 99 tumor-exclusive source proteins. GO = gene ontology; AC - accession.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB ACs of source proteins	Proportion of tumor-exclusive antigens
Immune response *via* complement activation	1.33	GO:0006508 GO:0006955 GO:0050776 GO:0006956 GO:0006958 GO:0003823 GO:0038096 GO:0004252 GO:0038095 GO:0006898	P06318 P42830 Q16651 Q9Y256 P32971 P78310 P01710 P26842 Q9UQQ1 P01773 Q8N6M6 Q7Z4S9 Q6MZZ7	13%

3.4.7 Search for neoantigens

The results of WES were available for twelve tumor samples and corresponding PBMCs of the OPSCC patient cohort (sample IDs: 020, 022, 1003, 1008, 1014, 1021, 1118, 1209, 1345, 1349, 1359 and 1361). On the basis of these sequencing data, variant calling was performed to reveal somatic DNA mutations in the tumors. Between 11 and 96 mutations impacting the AA sequences of the encoded proteins were identified per sample. The mutated protein sequences were integrated into the human proteome of the Swiss-Prot database. By database search with the combinatory database, the HLA ligandomic datasets of the OPSCC samples were searched for peptides derived from these mutated proteins to identify HLA-presented neoantigens. The search for neoantigens resulted in zero matches. Neither HLA class I binders nor HLA class II-presented peptides were found within the twelve samples that comprised the altered sequences of source proteins encoded by mutated genes.

3.4.8 Search for HPV-derived antigens

Peptides derived from viral proteins can be presented by HLA molecules to induce an immune reaction against infected cells. Therefore, the HLA ligandomes of the OPSCC tumor samples were searched for antigens derived from HPV proteins. The human proteome and the HPV proteomes, both comprised in the Swiss-Prot database, were combined and served as basis for the screening of OPSCC immunopeptidomes for HPV-derived antigens. The proteomes of the HPV types HPV16, HPV35, HPV58 and HPV59, which occurred in the OPSCC cohort, were entailed in the viral proteome of the Swiss-Prot

database. The screening followed by antigen evaluation of potential matches was not successful. In the immunopeptidomes of the 22 HPV$^+$ and the 18 HPV$^-$ OPSCCs, no HLA class I binders or HLA class II-presented peptides were identified that derived from an HPV source protein.

3.4.9 The OPSCC transcriptome

For 15 OPSCC samples of the present cohort, RNA sequencing data were available (sample IDs: 020, 022, 1003, 1008, 1012, 1014, 1021, 1024, 1034, 1118, 1209, 1345, 1349, 1359, 1361). These data were used for integrative analysis of transcriptomic and immunopeptidomic data of OPSCCs. To allow a better comparability between both datasets, those expressed genes encoding proteins that are included in the human proteome of the Swiss-Prot database were selected. In total, they amounted to 18,123 different genes for the 15 samples. Comparison of these genes with the HLA-presented antigens showed the coverage rate of the HLA ligandome among the transcriptome **(Figure 26)**. The comparison was performed on the basis of the merged data of all 15 samples and revealed that the majority of antigenic source proteins was covered by the transcriptome. 95% of the source proteins from HLA class I binders were found within the expression data whereas 94% of the source proteins from HLA class II-presented peptides appeared within the transcriptome. For individual samples, the coverage rates of antigenic source proteins among expressed genes averaged 88% and 74% for HLA class I and HLA class II, respectively **(Supplementary Figures 12 and 13)**.

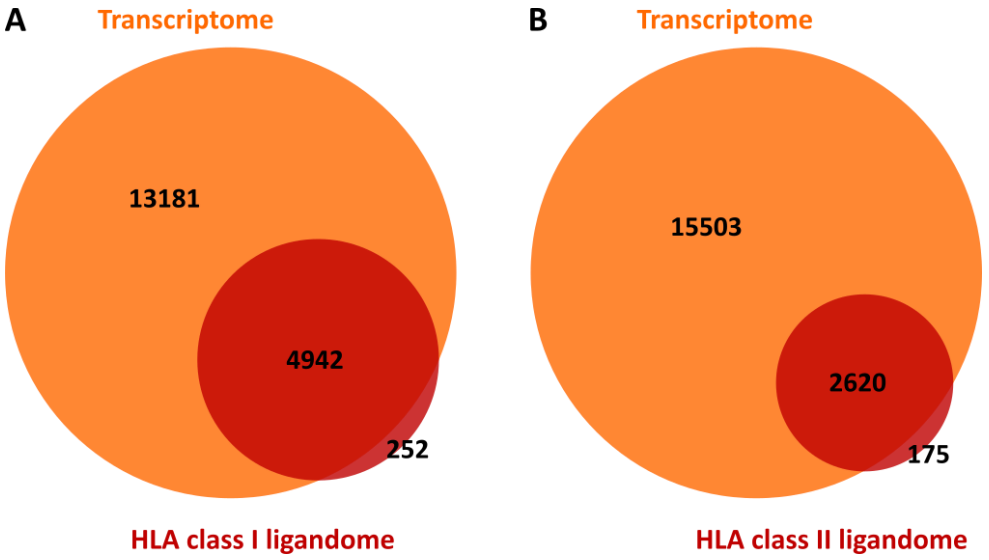


Figure 26: Comparison between the RNA transcriptome and HLA ligandome of OPSCCs. Expressed genes identified in the 15 OPSCC samples were merged. Those genes coding for proteins that are comprised within the Swiss-Prot database were used for comparative analysis with the merged antigenic source proteins of the OPSCC immunopeptidome. (A) 95% of the source proteins of HLA class I binders were covered by the expressed genes identified in the transcriptome. (B) 94% of the source proteins of HLA class II-presented peptides were found within the transcriptome. n(OPSCC) = 15.

To statistically analyze the relationship between the expression level of a gene and the number of different peptide sequences by which the encoded protein is represented within the immunopeptidome, Pearson´s correlation was applied. On the transcriptomic level, the data of expressed genes that encoded for proteins identified as antigenic sources in the immunopeptidomic dataset of OPSCCs were selected. On the immunopeptidomic level, the data of source proteins of both HLA class I binders and HLA class II-presented peptides were used. The analysis was performed on the basis of the mean reads per kilobase million (RPKM) of the genes and the mean numbers of unique HLA-presented peptide sequences per source protein across the 15 OPSCC samples. Significant positive correlations were revealed between source proteins of HLA class I binders and corresponding read counts (Pearson´s correlation coefficient r = 0.2955; p value < 0.0001; 95% CI = 0.27 to 0.32) as well as between source proteins of HLA class II-presented peptides and corresponding read counts (Pearson´s correlation coefficient r = 0.2918; p value < 0.0001; 95% CI = 0.26 to 0.33). Linear regression analysis further demonstrated the relationship of the immunopeptidomic and transcriptomic datasets **(Figure 27)**.

These results indicate that a strong gene expression might lead to an elevated HLA presentation of peptides derived from the respective encoded source protein.

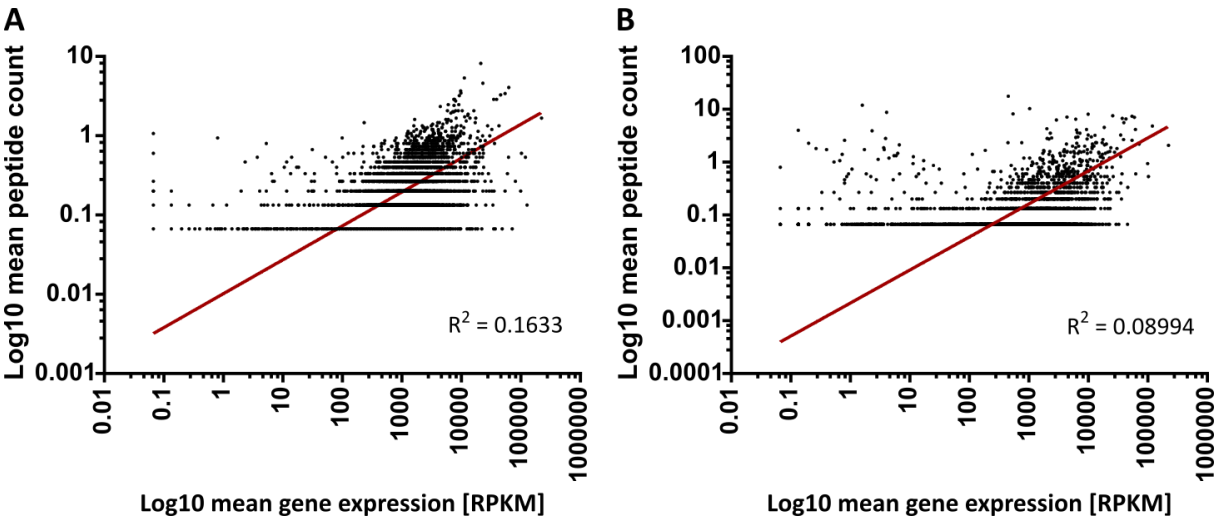


Figure 27: Linear regression analysis of gene expression levels and counts of unique HLA-presented sequences per source protein in OPSCCs. The mean numbers of unique (A) HLA class I binder sequences and (B) HLA class II-presented peptide sequences per source protein were plotted against the mean normalized read counts of corresponding genes. Higher gene expression levels were accompanied by elevated HLA presentation of the encoded antigenic source protein. n(OPSCC) = 15; RPKM – reads per kilobase million; R^2 = coefficient of determination (goodness of fit).

3.4.10 HPV$^+$ *versus* HPV$^-$ OPSCCs

One important risk factor for the development of an OPSCC is an infection with a high-risk HPV type. The present cohort comprised 40 tumor samples originating from 22 HPV$^+$ and 18 HPV$^-$ patients. No HLA-presented antigens derived from HPV proteins were detected in the immunopeptidomes of the OPSCCs (see 3.4.8). However, the infection status might have an impact on the general composition of the OPSCC HLA ligandome.

Therefore, principle component analyses (PCAs) were performed to unravel a potential clustering of the tumors according their HPV status on the basis of the immunopeptidomic dataset. The merged source proteins of HLA class I binders and of HLA class II-presented peptides were used as basis for these analyses. An unsupervised PCA did not result in a clear separation of the samples into HPV$^+$ and HPV$^-$ tumors **(Supplementary Figure 14)**. In contrast, by applying instructed data in regard of the HPV status, a definite clustering into HPV$^+$ or HPV$^-$ tumors occurred as demonstrated by the scatter plot of the supervised PCA **(Figure 28A)** and the corresponding heatmap **(Figure 28B)**.

These results indicate a potential influence of infection status on the HLA ligandomes of the OPSCC tumors.

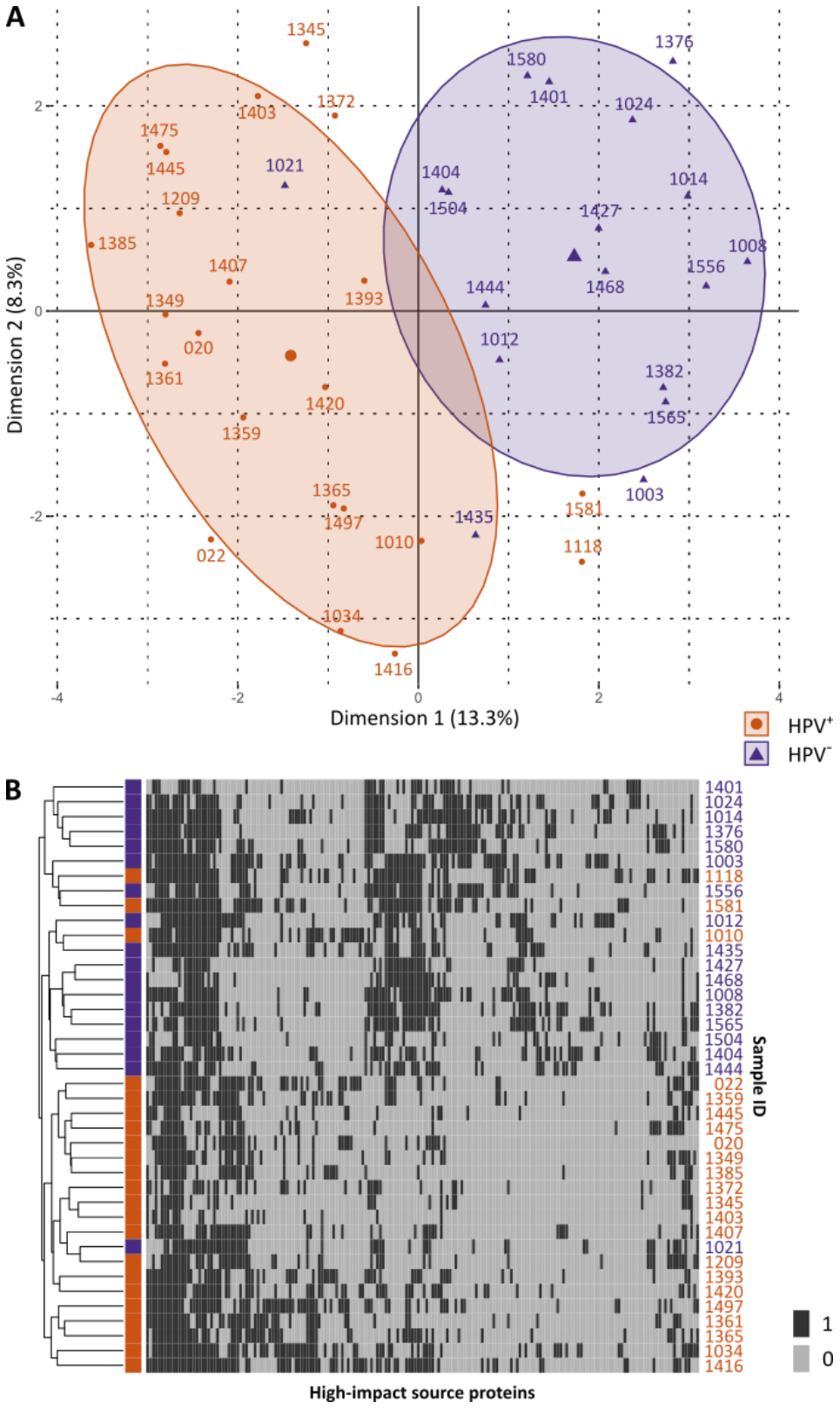


Figure 28: Supervised PCA on the basis of HLA ligands identified in HPV⁺ and HPV⁻ OPSCCs. The PCA was based on the merged source proteins of HLA class I binders and HLA class II-presented peptides from OPSCC samples. (A) The scatter plot of the PCA revealed a clear clustering into HPV⁺ and HPV⁻ tumors with only a few outliers such as OPSCC 1021 or 1435. (B) The heatmap, which was based on the source proteins with the highest impact on the separation into the two groups, further illustrates this clustering. n(HPV⁺ samples) = 22; n(HPV⁻ samples) = 18; n(high-impact proteins) = 190; PCA = principle component analysis.

The described trend, which suggested potential differences between the HLA ligandomes of HPV$^+$ and HPV$^-$ OPSCC tumors, was investigated in more detail by comparative profiling of the HPV$^+$ against the HPV$^-$ OPSCC immunopeptidomes.

On the peptide level, HLA class I binders and HLA class II-presented peptides were identified that were either shared or exclusively presented by HPV$^+$ or HPV$^-$ OPSCC samples **(Figure 29)**. Both HPV$^+$- and HPV$^-$-exclusive peptides comprised several frequent sequences occurring in at least three samples of the respective tumor subgroup. 653 and 190 frequent HLA class I binders were exclusively identified in HPV$^+$ or HPV$^-$ tumors, respectively. Additionally, 197 and 88 frequent HLA class II-presented peptides were exclusively detected in either the HPV$^+$ or the HPV$^-$ subgroup, respectively.

In addition to the analysis on the level of HLA-presented peptides, the corresponding source proteins were addressed to reveal differences between the immunopeptidomes of HPV$^+$ and HPV$^-$ OPSCCs. Thereby, the impact of HLA allotype-specific binding preferences is reduced. Comparative analysis on the source protein level demonstrated a higher overlap between the two subgroups than on the peptide level. Still, HPV$^+$- and HPV$^-$-exclusive source proteins of HLA class I binders and HLA class II-presented peptides were identified **(Figure 30)**. This included frequent source proteins occurring in at least three samples of the respective subgroup. They amounted to 305 HPV$^+$-exclusive and 67 HPV$^-$-exclusive source proteins of HLA class I binders. In addition, 79 HPV$^+$-exclusive and 29 HPV$^-$-exclusive source proteins of HLA class II-presented peptides were revealed.

These results further support the hypothesis that the immunopeptidomes of OPSCCs differ in their composition of antigens depending on the respective HPV status of the tumor.

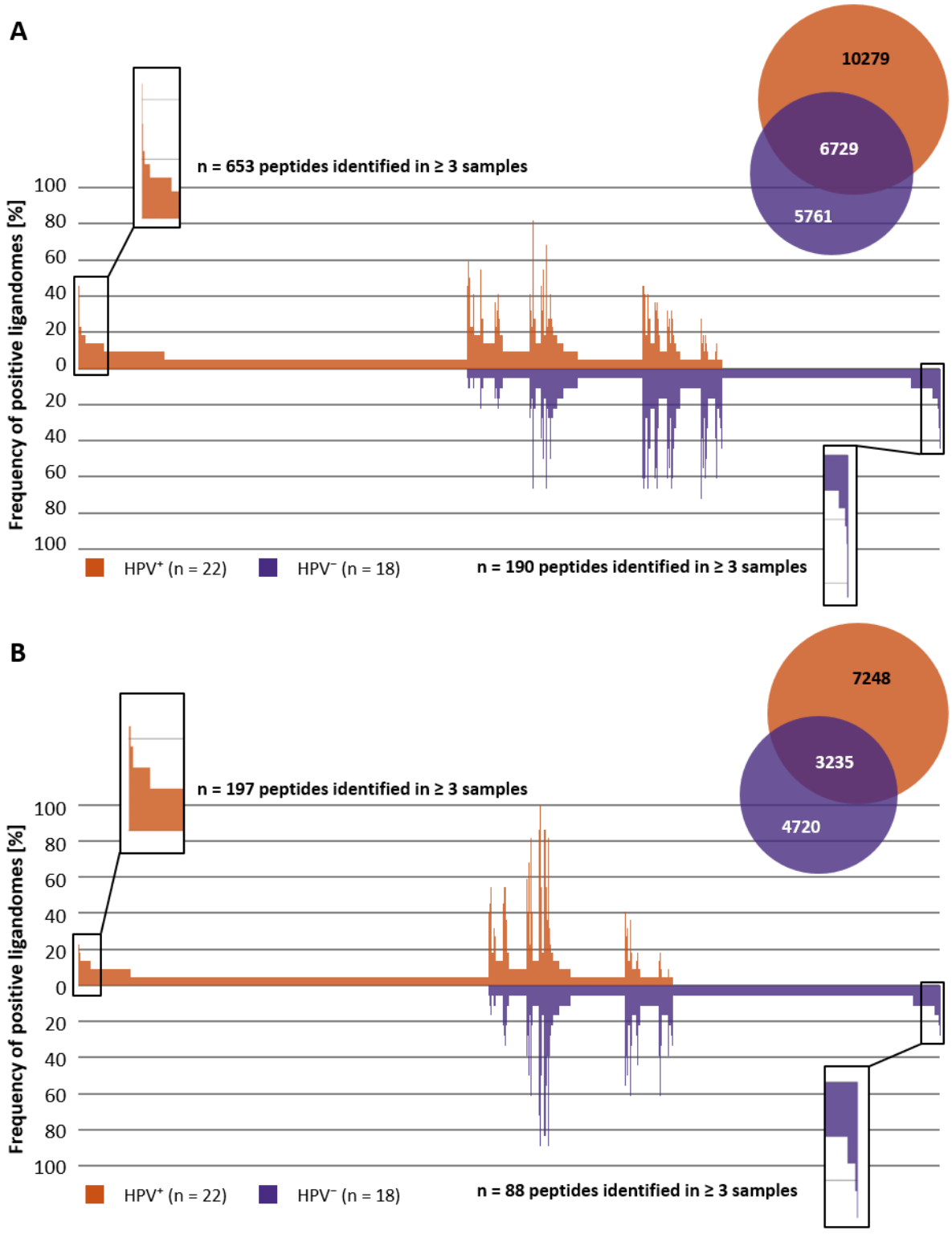


Figure 29: Comparative profiling of the HPV⁺ against the HPV⁻ OPSCC immunopeptidome on the peptide level. Several (A) HLA class I binders as well as (B) HLA class II-presented peptides were exclusively detected in either HPV⁺ or HPV⁻ OPSCCs. This included various frequent peptides found in at least three different samples of the respective subgroup. In the waterfall plots, unique ligands are plotted on the x-axis and the corresponding percentage of HLA ligandomes that contain the respective ligands on the y-axis. The frequent HPV⁺- and HPV⁻-exclusive peptides are highlighted on the far left and the far right, respectively.

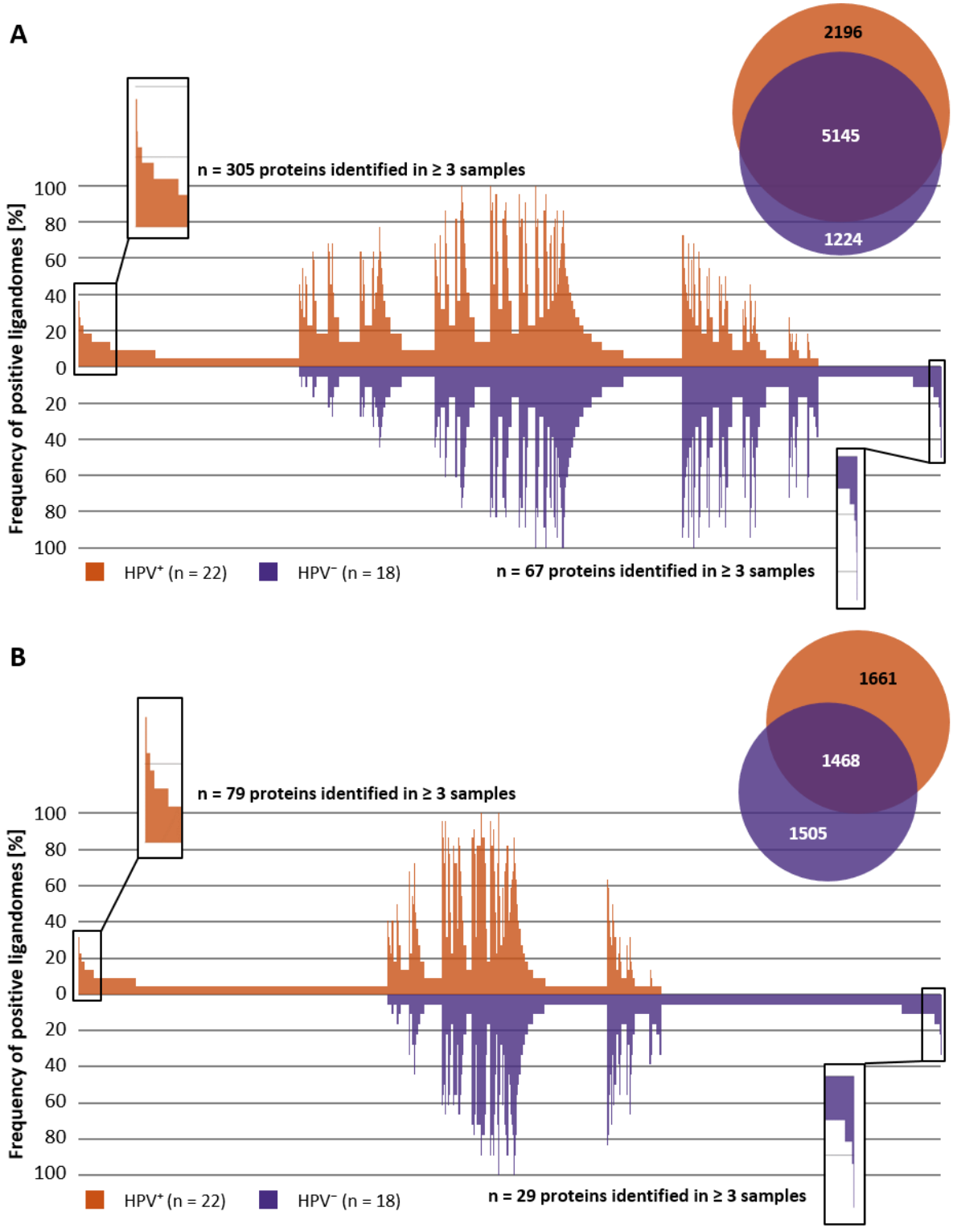


Figure 30: Comparative profiling of the HPV⁺ against the HPV⁻ OPSCC immunopeptidome on the source protein level. Several (A) source proteins of HLA class I binders as well as (B) source proteins of HLA class II-presented peptides were exclusively detected in either HPV⁺ or HPV⁻ OPSCCs. This included various frequent proteins found in at least three different samples of the respective subgroup. In the waterfall plots, unique ligands are plotted on the x-axis and the corresponding percentage of HLA ligandomes that contain the respective ligands on the y-axis. The frequent HPV⁺- and HPV⁻-exclusive proteins are highlighted on the far left and the far right, respectively.

An important issue was whether the described differences in the immunopeptidomes of HPV⁺ and HPV⁻ OPSCCs impact the HLA presentation of tumor-exclusive antigens and of selected TAAs (see 3.4.4). To address this, additional comparative analyses were performed **(Figure 31)**. Comparison of tumor-exclusive HLA ligands with HLA ligands from OPSCCs demonstrated that the majority of both tumor-exclusive HLA class I binders **(Figure 31A)** and HLA class II-presented peptides **(Figure 31B)** was solely detected in either HPV⁺ or HPV⁻ samples. The next step was to observe whether this also concerned the antigenic peptides selected as potential TAAs of OPSCCs **(Tables 8 and 9)**. Most of these peptides were found independently of the HPV status across both tumor subgroups. This applied for the TAAs DANPYDSVKKI, VALPVYLLI, HGTIKNQL, EETNPKGSGW, YTDNWLAVY, MLFFDKFANIVPF, LPFLDIAPLDIGG, TNKPSRLPFLDIAPLDIGGADQ, TKPAIRRLARRGGVKR and VDEEEVKFPGTNFDE. Still, five of the selected TAAs were exclusively identified among tumors of one subgroup. The peptides KLAEISLGV, GEAAKSVKL, SVKITQVTW and TPAAQYVRIKENLAVG were identified exclusively in the immunopeptidomes of HPV⁺ OPSCCs whereas the peptide YLDDPTNSWY was identified exclusively in HPV⁻ samples.

In summary, the entirety of comparative analyses between the immunopeptidomes of HPV⁺ and HPV⁻ OPSCC tumor samples unraveled similarities as well as clear differences in their composition of HLA-presented antigens. The differences partly concerned the selected TAAs and might be of importance for the selection of suitable immunotherapeutic target antigens.

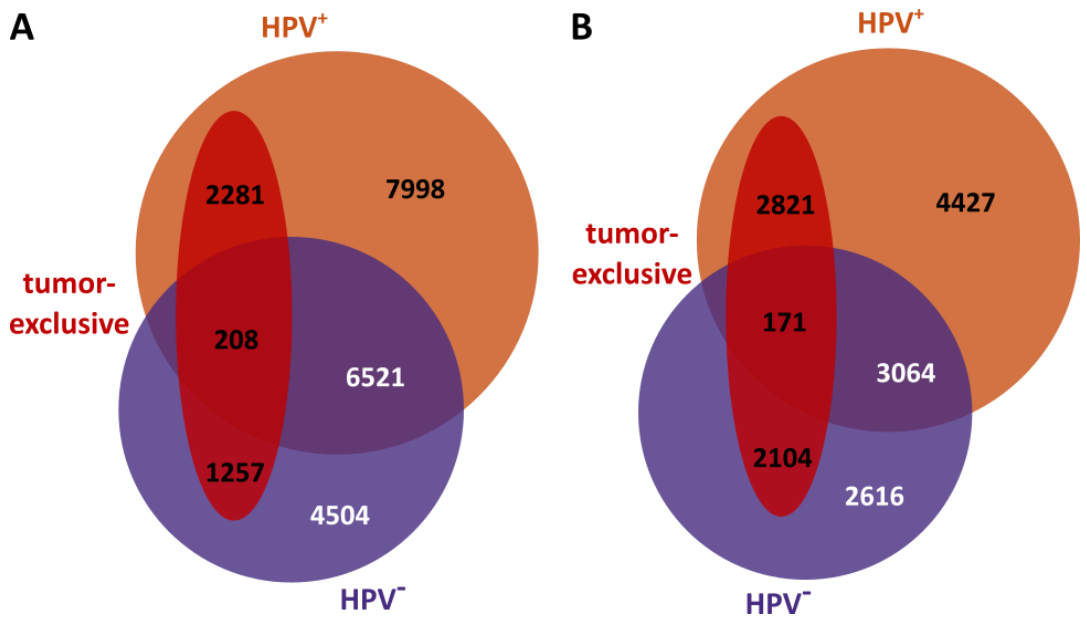


Figure 31: Proportion of tumor-exclusive HLA ligands among the immunopeptidomes of HPV⁺ and HPV⁻ OPSCCs. Tumor-exclusive peptides were compared to OPSCC HLA ligands (A) on the level of HLA class I binders and (B) on the level of HLA class II-presented peptides. In both cases, the majority of tumor-exclusive peptides was exclusively identified within the immunopeptidomes of either HPV⁺ or HPV⁻ OPSCC samples. The areas of tumor-exclusive peptides are unproportionally depicted in the Venn diagrams. n(HPV⁺ samples) = 22; n(HPV⁻ samples) = 18.

Functional annotation clustering

To reveal potential differences between biological processes in which HPV$^+$- and HPV$^-$-exclusive source proteins are involved, functional annotation clustering was performed using the functional annotation tool of the database DAVID Bioinformatics Resources 6.8 [67, 68]. The HPV$^+$- and HPV$^-$-exclusive source proteins of HLA class I binders and HLA class II-presented peptides served as basis for this analysis. An enrichment score of ≥ 1 was used as inclusion threshold for annotated clusters. In case of HLA class I, 19 functional clusters were revealed based on 2,196 HPV$^+$-exclusive source proteins **(Supplementary Table 20)** whereas 5 functional clusters were revealed based on 1,224 HPV$^-$-exclusive source proteins **(Supplementary Table 22)**. In case of HLA class II, functional annotation clustering with 1,661 HPV$^+$-exclusive source proteins revealed 11 different annotation clusters **(Supplementary Table 21)** and 1,505 HPV$^-$-exclusive source proteins revealed 22 different annotation clusters **(Supplementary Table 23)**. Interestingly, two functional clusters, which were revealed for the HPV$^+$-exclusive source proteins of HLA class II-presented peptides, indicated an involvement of the proteins in virus-related processes. This included DNA recombination and gene expression mechanisms in respect of genes coding for viral capsid proteins.

Integration of transcriptomic and immunopeptidomic data

The circumstance that the HLA ligandomes of HPV$^+$ and HPV$^-$ OPSCCs exhibited differences not only on the peptide but also on the protein level indicates additional differences among the transcriptomes of the tumor subgroups. Therefore, the RNA expression data were observed for potential variations associated with the HPV status. RNA sequencing data were available for 15 samples of the OPSCC cohort entailing nine HPV$^+$ and six HPV$^-$ samples. Based on the gene expression data of these tumor samples, differential expression analysis was performed using the DESeq2 algorithm [66]. A PCA based on the differential expression data of all detected genes already demonstrated a clustering into HPV$^+$ and HPV$^-$ samples **(Supplementary Figure 15)**. Further analysis revealed genes that were dysregulated depending on the HPV status of the patients. In particular, 680 genes were significantly up-regulated and 676 genes were significantly down-regulated among the HPV$^+$ samples in contrast to the HPV$^-$ tumor subgroup **(Figure 32)**.

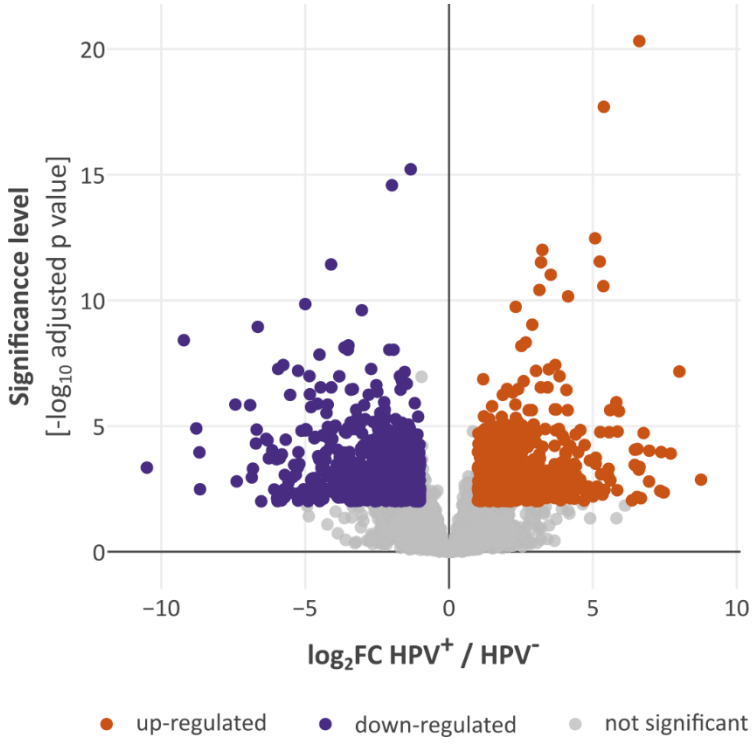


Figure 32: Volcano plot of the differential expression analysis based on RNA gene expression data of HPV⁺ and HPV⁻ OPSCCs. The analysis was performed using the DESeq2 algorithm [66]. 680 genes were significantly up-regulated whereas 676 genes were significantly down-regulated in the HPV⁺ tumor subgroup in comparison with the HPV⁻ tumor subgroup. n (HPV⁺ samples) = 9; n (HPV⁻ samples) = 6; FC – fold change.

Another PCA was performed based on the significantly dysregulated genes. This again revealed a clear separation of the two tumor subgroups, which becomes clear in the scatter plot of the PCA **(Figure 33A)** and the corresponding heatmap **(Figure 33B)**. The HPV⁻ tumors generated a more densely packed cluster than the HPV⁺ tumors, but there was no overlap between both groups.

To enable a better comparability between transcriptomic and immunopeptidomic data, those differentially expressed genes that encoded proteins of the Swiss-Prot database were selected. On the basis of the 592 and 574 resulting proteins encoded by up- and down-regulated genes, respectively, functional annotation clustering was performed. The functional annotation tool of the database DAVID Bioinformatics Resources 6.8 [67, 68] was used. For the proteins encoded by genes up-regulated in HPV⁺ samples, involvements in processes concerning DNA transcription, chemokine-mediated signaling or cell-cell adhesion were revealed **(Supplementary Table 24)**. Proteins encoded by genes down-regulated in HPV⁺ samples were shown to be involved in mechanisms like proteolysis, cytokine-mediated signaling or microfilament-based movement **(Supplementary Table 25)**.

In conclusion, the analysis of the transcriptomic data exhibited clear differences between HPV⁺ and HPV⁻ OPSCCs that might have an impact on the composition of the respective OPSCC immunopeptidomes. These results further confirm the relevance of the HPV status for the molecular and cellular mechanisms in OPSCC tumors.

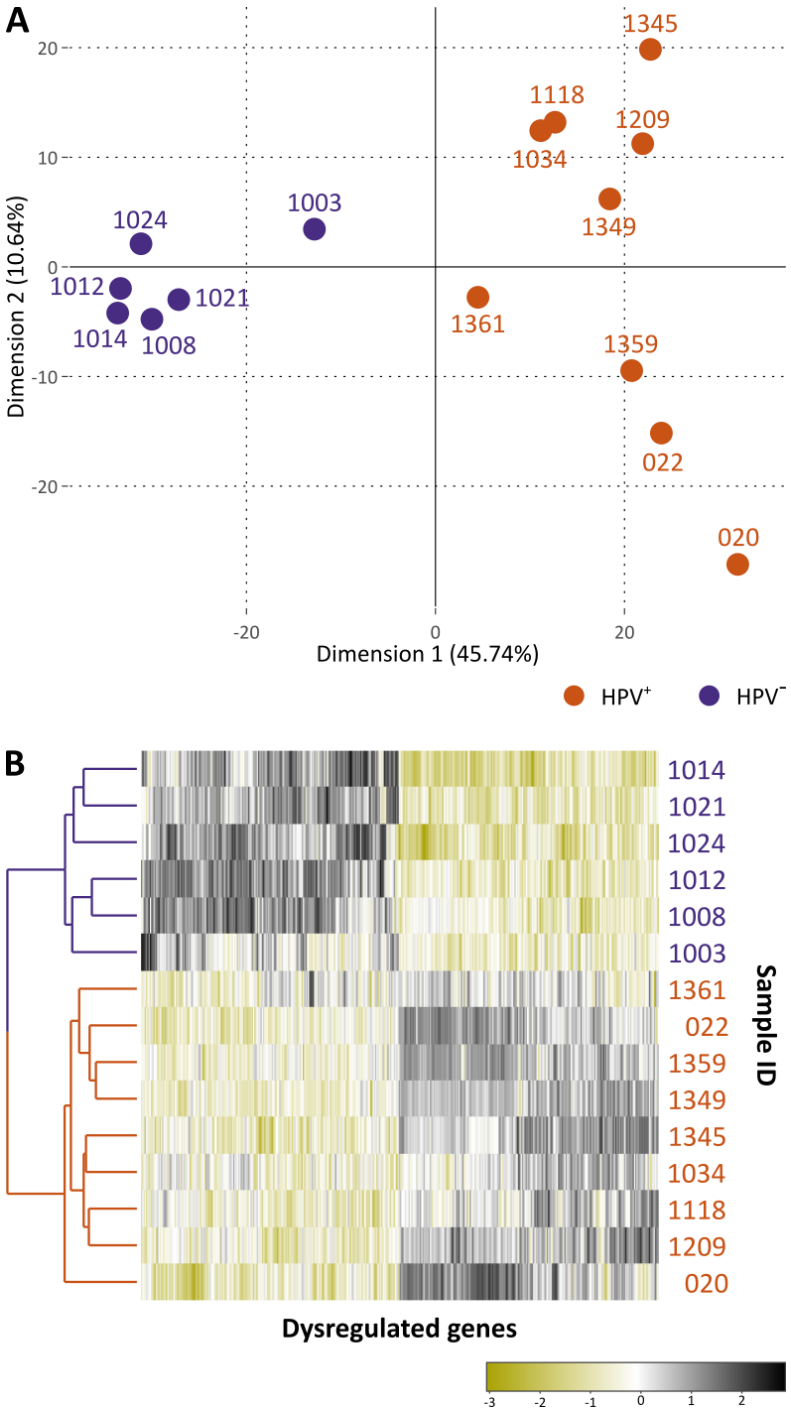


Figure 33: PCA and corresponding heatmap based on significantly dysregulated genes identified in HPV⁺ OPSCC samples compared to HPV⁻ samples by differential expression analysis. (A) The scatter plot of the PCA showed a clear clustering of the OPSCCs according to their HPV status. (B) This separation into HPV⁺ and HPV⁻ samples was further demonstrated by the corresponding heatmap. The DESeq2 algorithm was used for the differential expression analysis [66]. n (HPV⁺ samples) = 9; n (HPV⁻ samples) = 6; n (up-regulated genes) = 680; n (down-regulated genes) = 676; PCA – principle component analysis.

Subsequently, integrative analysis of the HLA ligandomic and transcriptomic data was performed for the OPSCC subcohort of 15 samples. The antigens of the HPV⁺ and HPV⁻ OPSCC immunopeptidomes were compared to the dysregulated genes identified by differential expression analysis of the HPV⁺

and HPV⁻ OPSCC transcriptomes. This was performed to reveal whether the HPV status-associated dysregulation of specific genes was mirrored within the HLA ligandomes of OPSCCs. Those dysregulated genes that encoded proteins from the Swiss-Prot database were selected. The merged source proteins of HLA class I binders and HLA class II-presented peptides were separated according to the HPV status of the tumors. On the basis of the dysregulated genes and both groups of source proteins, the comparative analysis was performed **(Figure 34)**. In general, the overlap between HLA-presented antigens and differentially expressed genes was quite small. Regarding HPV-associated antigen presentation, the overlap for genes up-regulated in HPV⁺ samples was slightly higher with HPV⁺-exclusive source proteins (n = 104) than with HPV⁻-exclusive source proteins (n = 25) **(Figure 34A)**. Additionally, the overlap for genes down-regulated in HPV⁺ samples was slightly higher with HPV⁻-exclusive antigens (n = 59) compared to the overlap with HPV⁺-exclusive antigens (n = 47) **(Figure 34B)**.

A correlation between HPV-associated differential gene expression and antigen presentation can be assumed. However, based on the small number of samples and the generally small overlap between both datasets, no final conclusion can be made. The analysis of the entire OPSCC cohort with corresponding transcriptomic and immunopeptidomic data of all 40 samples will be essential for a further confirmation of these results.

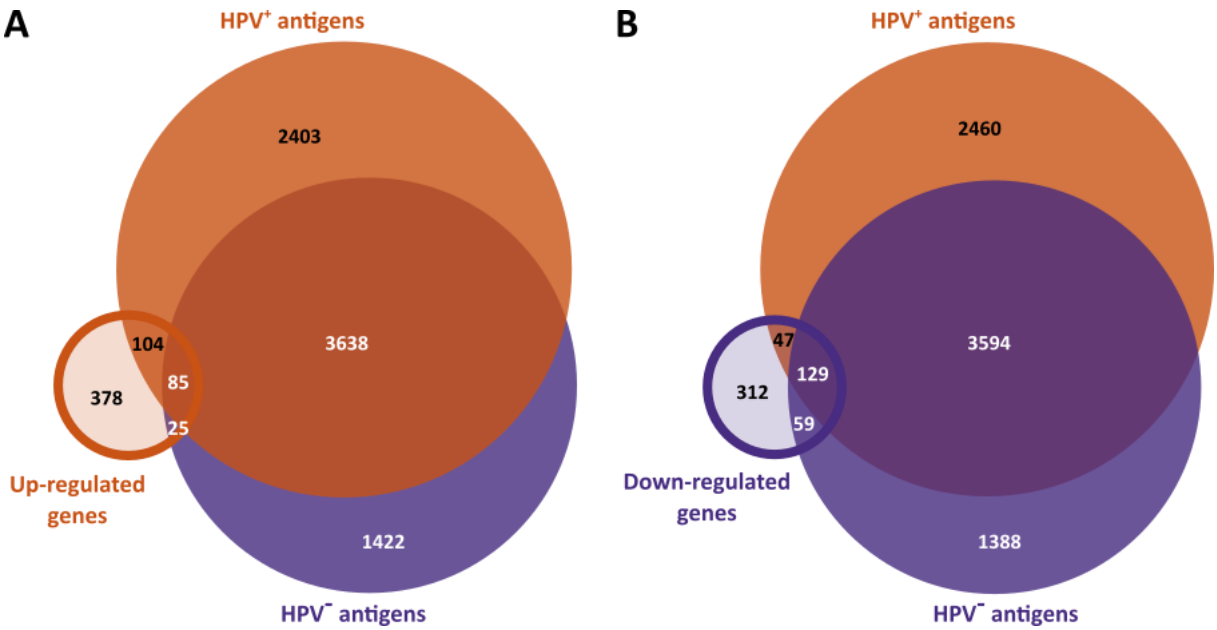


Figure 34: Comparison of antigens identified in the immunopeptidomes of HPV⁺ and HPV⁻ OPSCCs with genes differentially expressed depending on the HPV status. Source proteins of HLA class I binders and HLA class II-presented peptides were merged and compared with dysregulated genes coding for proteins that are comprised within the Swiss-Prot database. (A) Slightly more proteins that are encoded by genes up-regulated in HPV⁺ samples were shared with HPV⁺- than with HPV⁻-exclusive antigens. (B) Slightly more proteins that are encoded by genes down-regulated in HPV⁺ samples were shared with HPV⁻- than with HPV⁺-exclusive antigens. n(HPV⁺ samples) = 9; n(HPV⁻ samples) = 6.

3.5 Discussion

Oropharyngeal squamous cell carcinomas (OPSCCs) are a malignant tumor entity and are either caused by abusive cigarette or alcohol consumption or by an infection with a high-risk HPV type. A cohort of 22 HPV^+ and 18 HPV^- OPSCC patients served as basis for the present study which focused on the analysis of the tumor immunopeptidomes by MS-based methods and the selection of potential immunotherapeutic target antigens.

At first, the HLA class I allelic distribution was observed to reveal potential deviations in the OPSCC patient cohort from the allelic distribution that is expected under normal conditions. By comparison of the HLA allele frequencies with a reference database, statistical analysis revealed that the HLA allele B*51:01 showed significantly higher prevalence in the patient cohort than in the reference database. An elevated risk for developing a disease linked to specific HLA alleles, called HLA-disease association, was shown for several combinations of diseases and HLA alleles in the past [71]. In regard to OPSCC, no direct HLA-disease association has yet been published. However, an association between specific HLA alleles and the risk of developing cancer after a high-risk HPV infection has been described for cervix carcinoma [27]. These HLA-dependent risk variations are assumed to result from different binding affinities of HLA alleles for HPV-derived peptides. This influences the presentation of HPV-derived peptides and the resulting immune response, which decides about coping with the HPV infection or developing a persisting infection [28]. Similar HLA-disease associations might be possible for HPV-associated OPSCC as well. None of the previously described HLA alleles associated with a higher risk for HPV-induced cervical cancer had a significantly higher prevalence within the entire OPSCC cohort or the HPV^+ subgroup in comparison with the reference. The significantly increased frequency of HLA-B*51:01 was only shown across the entire cohort. In both, the HPV^+ and the HPV^- subgroup, the frequency of HLA-*51:01 was on a similar level but not significantly increased compared to the reference. In conclusion, a potential association between the HLA allele B*51:01 and an increased risk for the development of OPSCC can be assumed, which is possibly independent of an HPV infection. Larger tumor subgroups might be necessary to reveal potential HPV-related associations between particular HLA alleles and the occurrence of OPSCC.

The focus of the present study was on the search for TAAs that might have immunotherapeutic target potential among the immunopeptidomes of OPSCCs. Since CTAs are of interest for the selection of immunotherapeutic targets due to their restricted expression patterns and immunogenicity, the immunopeptidomic dataset was searched for previously published CTAs [63]. Thereby, several HLA-presented peptides were revealed that derived from CTA source proteins. This confirms the results of Atanackovic *et al.* (2006) who demonstrated a high CTA expression level in HNSCCs and suggested CTAs as potential targets for immunotherapeutic approaches [72]. A strong expression was shown especially

for the CTAs of the melanoma-associated antigen-A (MAGE-A) family in invasive HNSCC [73], which was also represented in the OPSCC HLA ligandomic dataset of the present study. Furthermore, the detected CTA U3 small nucleolar ribonucleoprotein IMP3 has been shown to serve as suitable target for HNSCC immunotherapy. An HLA-A*24-restricted epitope (KTVNELQNL) derived from IMP3 has been tested in combination with peptides from two other CTAs, the lymphocyte antigen 6K (LY6K) and the kinetochore protein Nuf2 (CDCA1). The clinical vaccination trial indicated an improved prognosis for HNSCC patients [74]. Despite the present cohort containing ten HLA-A*24-carrying OPSCC patients, the described peptide was not among the IMP3-derived peptides identified in the OPSCC immunopeptidome. However, the detected CTAs might bear potential as immunotherapeutic targets for OPSCCs which needs to be further investigated.

Besides published CTAs, the OPSCC immunopeptidome was screened for additional TAAs. Comparative profiling of the OPSCC dataset against in-house benign and malignant databases followed by detailed examination of tumor-exclusive antigens enabled the selection of fifteen HLA-presented peptides with characteristics of TAAs. In literature, nine corresponding source proteins of these potential targets have been associated with an involvement in cancer development before.

The TAA source protein Desmoglein-3 (DSG3) has been directly associated with HNSCC. DSG3 showed overexpression in HNSCC tumor tissue. Since inhibition of DSG3 was demonstrated to trigger suppression of the carcinogenic potential of tumor cells, the protein was suggested as therapeutic target [75]. DSG3 is further described as oncogene involved in cancer progression [76]. The correlation between highly expressed DSG3 and a shorter time period until the occurrence of a second primary malignancy suggests DSG3 as suitable prognostic biomarker [77].

Several further source proteins of selected OPSCC TAAs have been described as prognostic biomarkers for different tumor entities. They allow assumptions about disease progression and patient outcome and, thereby, enable precise adjustments of therapeutic strategies. The chromosome-associated kinesin KIF4B was identified as target source protein in the OPSCC cohort. An overexpression of kinesin family (KIF) members in general was shown to be involved in the progression of hepatocellular carcinoma (HCC). Thus, KIF members were suggested as indicators for prognosis and therapy. Especially KIF4B was linked to a decreased OS rate in HCC patients [78].

Additionally, the expression of the OPSCC TAA source protein Nectin-1 has been suggested as an indicative risk-factor for early recurrence due to its association with a significantly worse prognosis in respect of a progression-free survival for colorectal cancer patients [79].

Another TAA source protein, the condensin-2 complex subunit G2 (NCAPG2), was shown to be indicative for a poor prognosis in lung adenocarcinoma [80] and HCC [81] by promoting tumor proliferation.

The expression levels of the TAA source protein collagen alpha-2(I) chain (COL1A2) were demonstrated to be significantly higher in malignant tissue of gastric cancer than in premalignant and benign tissues. This was further linked to a lower OS rate. Therefore, COL1A is suggested as predictive marker for poor clinical outcome of gastric cancer patients [82, 83]. Inhibition of COL1A by gene silencing or inhibition of expression led to a decreased proliferation, migration and invasion of gastric cancer cells through the PI3K-Akt signaling pathway. These successful attempts for the application of COL1A2 as therapeutic target indicated a promising approach for the development of novel therapeutic options [84, 85].

The TAA source protein ubiquitin D (FAT10) was also shown to be upregulated in several tumor entities like glioma, HCC, breast, gastric, liver or colorectal cancer. It is suggested to be involved in carcinogenesis and metastasis by affecting signaling pathways and post-translational protein modification [86, 87].

Additionally, the TAA source protein plakophilin-1 (PKP1) was linked to cancer development. However, the opinions about its role during carcinogenesis are controversial. On the one hand, oncogenic functions are ascribed to PKP1 in squamous cell lung cancer [88]. It is suggested to post-transcriptionally regulate MYC, an oncogene and important target in antitumor therapy [89]. Thus, an inhibition of PKP1 might indirectly also affect MYC [88]. On the other hand, Haase *et al.* suggested PKP1 as novel tumor suppressor with characteristics of an indicative biomarker for advanced lung cancer [90].

Two other source proteins of selected TAAs were associated with tumor suppressive characteristics in different tumor entities before. This includes the homeobox protein DLX-3 and the desmosomal cadherin desmocollin-3 (DSC3). DLX-3 is known as an inhibitor of cell growth for basal skin cells [91]. Overexpression of DLX-3 was described as predisposing condition to initiate apoptotic cell death in different tumor cell lines including an esophageal squamous carcinoma cell line [92]. Since a loss of DLX-3 expression was shown to induce epidermal cell dysplasia resulting in cutaneous squamous cell carcinoma (cSCC), a cancer preventive role was ascribed to DLX-3 expression [93, 94]. DSC3 is essential for the maintenance of cell-cell adhesions between keratinocytes in the epidermis [95]. The protein was reported to have tumor suppressing functions in lung cancer cells [96] and colorectal cancer (CRC) cells [97] by inhibiting epidermal growth factors, which is why it is of interest for cancer therapy development.

The identification of S100A14 as source protein of a potential OPSCC TAA was contrary to expectations. Commonly, members of the S100 gene family are differentially expressed in various malignant tumor entities compared to benign tissue. Interestingly, the S100 gene family member S100A14 exhibits very diverged functions in different entities [98]. In tumor tissues of breast and ovarian carcinoma, S100A14 is usually overexpressed and indicates a role as potential prognostic marker for a poor patient outcome [99]. In contrast, in oral squamous cell carcinomas (OSCC), the chromosomal region encoding for

S100A14 is usually deleted and the protein significantly down-regulated [100, 101]. In patients with OSCC, the loss of S100A14 expression even serves as indicator for a reduced OS rate [102]. These findings are quite controversial to the observation of S10014A as frequently HLA-presented antigen in OPSCCs.

Many source proteins from selected TAAs of OPSCCs have been reported to be involved in processes of several cancer entities in the past. In addition, the integration of data from an immunopeptidomic malignant database revealed that the majority of selected OPSCC TAAs have been found as HLA-presented antigens in a variety of tumor entities before. These findings support their selection as OPSCC TAAs and suggests their further consideration and investigation for the application as targets in immunotherapeutic strategies against OPSCC.

For the tumors and corresponding PBMCs of twelve OPSCC patients of the present study cohort, WES data were available. These data were used to search for neoantigens carrying AA sequences with somatic mutations of the respective tumors. In the immunopeptidomic datasets of OPSCCs, no HLA-presented neoantigens were identified. The prevalence of neoantigens within the HLA ligandome of a tumor is highly dependent on the mutational burden of the respective cancer entity. The highest mutational burdens were revealed in cancer entities associated with chronic mutagenic exposures such as smoking or ultraviolet radiation. This applies for instance to lung cancer and melanoma [103]. The clinical application of neoantigens as immunotherapeutic targets showed success in several cancer vaccination studies with melanoma patients indicating their importance for certain cancer entities [104]. However, the search for HLA-presented neoantigens *via* MS analysis can be challenging and is often unsuccessful, especially in cancer entities with low mutational burden like hepatocellular carcinoma (HCC). In such cases, focusing on the selection of potential targets from the multitude of alternative antigens is suggested [105]. Mutagenic exposure often applies as cause for head and neck cancer and the resulting mutational burden is above overall average across the variety of cancer entities [103, 106]. However, no neoantigens were detected in the immunopeptidomes of the present OPSCC study. This might be either related to the absence of HLA-presented neoantigens or to technical limitations. The applied mass spectrometer and analysis methods might not be sufficiently sensitive for the detection of the probably low abundant neoantigens. An alternative method to identify mutated T cell epitopes is the *in silico* prediction of potentially presented neoantigens, which can be performed based on sequencing data but constitutes no final verification of their presentation [107].

Due to HPV infections displaying the cause for 22 OPSCCs of the present cohort, the HLA ligandomes of the tumors were searched for HPV-derived antigens. This revealed no viral peptides presented by HLA class I or class II molecules neither in HPV^+ nor in HPV^- samples. HPV is capable of interfering with the host´s immune response to facilitate an enhanced viral invasion into the epithelium and a persistent infection. Immune evasion strategies of HPV include the intrusion into antigen presentation

mechanisms. The viral oncogenic proteins E5, E6 and E7 are able to disrupt normal HLA class I processing, which leads to a reduced expression of HLA molecules on the surfaces of infected host cells. Thereby, the presentation of antigens to effector T cells, including the presentation of viral peptides, and the elimination of infected cells by immune responses is impeded [108–110]. A general loss of HLA surface expression seems not to be the reason for a lack of HLA-presented viral peptides among the HLA ligandomic dataset of the present study. The yields achieved by MS based immunopeptidomic analysis were quite high in HPV^+ OPSCC tissue samples for both HLA class I and class II-presented peptides. The average peptide yields were even slightly higher in the HPV^+ than in the HPV^- subgroup. Variations of peptide yields between distinct tissue samples are rather associated with differences in sample masses, which was shown to significantly correlate with peptide yields. Another important immune evasion strategy of HPV is to keep the production of viral proteins on a preferably low level. Thereby, potential HPV antigens stay low abundant among the repertoire of HLA-presented antigens and the probability of their recognition by T cells is markedly reduced [111]. In the present study, the quantity of HPV-derived antigens that are potentially presented by HLA molecules on the surfaces of OPSCC tumor cells might be below the detection level of the applied mass spectrometer and analysis methods. Prediction of potential T cell epitopes followed by targeted MS3 analysis is claimed as possible method for the detection of low-abundant HLA-presented peptides derived from viral proteins. With this method, the first directly detected HLA-presented HPV antigen was the HLA-A*02-restricted HPV16 peptide $E7_{11–19}$ identified in HPV16-transformed epithelial cell lines [112] and in primary cervix carcinomas [113]. Ten additional E7-derived peptides were identified by this detection method so far. Immunogenicity testing by ELISpot assays revealed their immunogenic potential [114].

In conclusion, naturally presented HPV antigens as well as neoantigens might occur in OPSCC tissues but were not identified by the applied methods of the present study. However, their presumably low abundance might impede their recognition by T cells. Therefore, the frequent TAAs selected from the OPSCC immunopeptidome probably bear more potential as immunotherapeutic targets.

Besides extending the antigenic repertoire of a tumor by viral proteins, viral infections have major additional impact on the host. Depending on the patient´s HPV status, cancerous diseases of the same entity display distinct characteristics concerning for instance patient age, nodal burden or disease staging [7]. In the present study, the observation of differences between HPV^+ and HPV^- OPSCC tumors was addressed on the HLA ligandomic level. This revealed antigens exclusively HLA-presented in either of the two tumor subgroups. A viral infection causes various changes in affected cells, the cellular environment and the immune system including cellular stress responses as well as alterations in cellular metabolism, protein expression, protein synthesis and signaling pathways. Due to the repertoire of HLA-presented antigens mirroring the functional cell state and indicating cellular stress,

the changes induced by a viral infection are reflected by the immunopeptidome on the surfaces of infected cells [115]. This was shown for infections with the human immunodeficiency virus (HIV) before. Ternette *et al.* demonstrated that HIV-infected cells harbor a collective of HLA class I-presented self-peptides that was generally modified compared to uninfected control cells [116]. This finding was confirmed by Ziegler *et al.* who revealed that the amount of virus-derived HLA-presented antigens was moderate but the antigen composition in general was altered upon an HIV infection entailing HLA-presented peptides exclusively identified in infected cells [117]. Not only the reactions of the host to viral infections but also viral processes lead to cellular changes. Immune evasion strategies of HPV are manifold and can include the suppression of innate immune sensing, pattern-recognition-receptor-(PRR-) related signal transduction, NFκB signaling or HLA-mediated antigen presentation [118]. The vast disruption of normal cellular processes caused by viral infections might also be reflected within the immunopeptidomes of HPV^+ OPSCC tumors impacting the antigenic repertoire that is presented by HLA molecules on the surface of the infected tumor cells. In the present study, differences in the antigenic composition of HPV^+ and HPV^- OPSCC samples were revealed. Additionally, the results of a functional annotation clustering annotated the involvement of several HPV^+-exclusive antigenic source proteins in virus-related DNA recombination and gene expression mechanisms. The absence of HPV-derived antigens in the OPSCC immunopeptidome confirms that differences on the immunopeptidomic level are only sparsely caused by the provision of viral source proteins. Cellular changes following a viral infection rather influence the general composition of host-self antigens. Therefore, the unraveled differences on the immunopeptidomic level of HPV^+ and HPV^- OPSCCs possibly result from alterations in cellular processes triggered by the HPV infection. This effect partly concerned the selected target antigens. Five of the HLA-presented peptides determined as OPSCC TAAs appeared to be exclusive for either HPV^+ or HPV^- tumors. This observation indicates a benefit for the selection of target antigens in consideration of the respective HPV status.

The massive cellular changes resulting from a viral infection are not only affecting the HLA ligandome but also the RNA gene expression profiles of infected cells. In case of a hepatitis B virus (HBV) infection, virus-associated changes in the expression levels of several genes in comparison with uninfected cells have been demonstrated [119]. Similar results were shown by Lamontagne *et al.* (2016) who applied differential expression analysis to HBV^+ and HBV^- hepatocytes. They observed an altered expression for various genes involved in cellular processes such as lipid biosynthesis, cellular metabolism or cell cycle regulation depending on the infection status of the cells [120]. In the present study, differential expression analysis was performed based on the transcriptomic data of the nine HPV^+ and six HPV^- OPSCC samples for which RNA sequencing results were available. This resulted in the detection of clear differences in the expression profiles of the two tumor subgroups. A set of genes that were significantly up- and down-regulated in dependence of the respective HPV status was unraveled. These

dysregulated genes encoded for proteins that showed distinct functional involvements, which might be a result of infection-associated alterations of cellular processes in the HPV$^+$ OPSCCs.

Differences between HPV$^+$ and HPV$^-$ OPSCC tumors were demonstrated on both the HLA ligandomic and the transcriptomic level. However, the overlap between dysregulated genes and differentially presented antigens was minor. The integration of RNA expression data into HLA ligandomic data is a challenging issue. Reconciling the two datasets revealed that most, but not all, source proteins of identified HLA-presented peptides were represented by the corresponding genes in the expression data. The circumstance that antigenic source proteins are not always reflected within the RNA expression data of the respective tissue is a frequent observation. The HLA ligandome certainly is affected but not completely mirrored by the transcriptome [121]. Weinzierl *et al.* (2007) detected several HLA-presented peptides in renal cell carcinoma (RCC) tissue samples for which no corresponding mRNA could be identified [122]. In accordance, Schuster *et al.* described only a slight correlation between HLA presentation and RNA expression [123]. Differences between the transcriptome and the HLA ligandome are very common. The sources for HLA-presented antigens are diverse and include short-lived proteins, long-lived proteins or DRiPs [124]. Long-lived proteins often remain within a cell for a certain time after their synthesis is completed and their encoding mRNA is degraded. Thereby, they display an important impact factor for deviations of the HLA ligandome from the transcriptome [125].

RNA expression data can assist the search for potential immunotherapeutic targets in tumors, especially if HLA presentation data are not available. For the OPSCC TAAs selected across the present cohort by MS-based immunopeptidomic analysis, the study of their corresponding RNA expression levels was not possible due to a presently incomplete RNA sequencing dataset and will be addressed at a later stage. However, the detection of a potential target antigen by mass spectrometry is still the most reliable method to confirm its presentation by HLA molecules. The correlation between peptide presentation and expression of the respective source proteins is limited [126], which is why the absence of corresponding RNA evidence should not serve as exclusion criterion during target selection.

In summary, the present study was performed to analyze the HLA ligandome of OPSCC with the aim of identifying potential immunotherapeutic target antigens. This resulted in the identification of several promising OPSCC TAAs. Comparative analyses of the immunopeptidomes of HPV$^+$ and HPV$^-$ OPSCCs revealed HPV status-associated differences that also concerned tumor-exclusive antigens and selected TAAs. This indicates that the consideration of the patient´s HPV status for the search for targets might be beneficial. Furthermore, an inclusion of more patients for both subgroups could enable a higher data density and the coverage of more frequent HLA allotypes in a population. Several further attempts are conceivable to search for additional target antigens. The search radius can be enlarged by considering antigenic source proteins that are similarly expressed but differentially presented in

malignant and benign tissue. Additionally, antigens with a low abundance in the benign references can be considered as target source.

Concerning sequencing analysis, the DNA exomic and RNA transcriptomic results of the 25 remaining OPSCC samples will be acquired and integrated into the current dataset. This will facilitate multi-dimensional data analysis, especially regarding the selected target antigens. Additionally, the OPSCC tumors will be analyzed in respect of the expression of the surface molecules CD8 and CD103. On this basis, tumors can be classified into immune hot and immune cold tumors providing information about the infiltration rate of $CD8^+$ $CD103^+$ T cells. These T cells were demonstrated to be essential for anti-cancer immune responses in OPSCCs. Their abundance, which is especially characteristic for HPV^+ OPSCCs, is described as prognostic factor for a better clinical outcome suggesting candidate patients for de-escalating therapy strategies [127, 128]. The correlation between HPV status and classification into immune hot or cold tumors will be observed in the current cohort. Potential differences in the immunopeptidomes and transcriptomes as well as the relevance for the selection of target antigens will be addressed.

In conclusion, the immunopeptidomic analysis of OPSCCs constitutes a solid foundation for the selection of potential immunotherapeutic targets. Before the performance of clinical studies regarding immunotherapy for OPSCC patients can be taken into consideration, the selected TAAs must be further evaluated. This demands MS validation of the corresponding synthetic peptides as well as immunogenicity testing. Peptides with positive results can be used for the compilation of an antigenic target warehouse, which in the best case covers several frequent HLA allotypes occurring in a population. Confirmed peptides can be considered for the application in clinical studies to observe their safety and benefit in immunotherapeutic approaches such as personalized or semi-personalized peptide vaccination in OPPSCC patients.

3.6 References

[1] Leemans CR, Snijders PJF, and Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer* 18(5): 269–282 (2018).

[2] Duvvuri U and Myers JN. Contemporary management of oropharyngeal cancer: anatomy and physiology of the oropharynx. *Curr Probl Surg* 46(2): 119–184 (2009).

[3] Fossum CC, Chintakuntlawar AV, Price DL, and Garcia JJ. Characterization of the oropharynx: anatomy, histology, immunology, squamous cell carcinoma and surgical resection. *Histopathology* 70(7): 1021–1029 (2017).

[4] Budu VA, Decuseară T, Balica NC, Mogoantă CA, Rădulescu LM, Chirilă M, Maniu AA, Mistra DM, Muşat GC, Oprişcan IC, and Georgescu MG. The role of HPV infection in oropharyngeal cancer. *Rom J Morphol Embryol* 60(3): 769–773 (2019).

[5] Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, and Cronin KA, $SEER Cancer Statistics Review, 1975-2017,. *SEER web site*, 2020, 2020.

[6] Siegel RL, Miller KD, and Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 70(1): 7–30 (2020).

[7] Parvathaneni U, Lavertu P, Gibson MK, and Glastonbury CM. Advances in Diagnosis and Multidisciplinary Management of Oropharyngeal Squamous Cell Carcinoma: State of the Art. *Radiographics* 39(7): 2055–2068 (2019).

[8] Villiers E-M de, Fauquet C, Broker TR, Bernard H-U, and zur Hausen H. Classification of papillomaviruses. *Virology* 324(1): 17–27 (2004).

[9] Bernard H-U, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, and Villiers E-M de. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 401(1): 70–79 (2010).

[10] Egawa N, Egawa K, Griffin H, and Doorbar J. Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses* 7(7): 3863–3890 (2015).

[11] Egawa N and Doorbar J. The low-risk papillomaviruses. *Virus Res* 231: 119–127 (2017).

[12] zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2(5): 342–350 (2002).

[13] zur Hausen H, Meinhof W, Scheiber W, and Bornkamm GW. Attempts to detect virus-secific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus. *Int J Cancer* 13(5): 650–656 (1974).

[14] Walboomers JMM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJF, Peto J, Meijer CJLM, and Muoz N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 189(1): 12–19 (1999).

[15] Kessler TA. Cervical Cancer: Prevention and Early Detection. *Semin Oncol Nurs* 33(2): 172–183 (2017).

[16] Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, Wentzensen N, Jordan RC, Altekruse S, Anderson WF, Rosenberg PS, and Gillison ML. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol* 29(32): 4294–4301 (2011).

[17] Vashisht S, Mishra H, Mishra PK, Ekielski A, and Talegaonkar S. Structure, Genome, Infection Cycle and Clinical Manifestations Associated with Human Papillomavirus. *Curr Pharm Biotechnol* 20(15): 1260–1280 (2019).

[18] Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, and Stanley MA. The biology and life-cycle of human papillomaviruses. *Vaccine* 30 Suppl 5: F55-70 (2012).

[19] Doorbar J, Egawa N, Griffin H, Kranjec C, and Murakami I. Human papillomavirus molecular biology and disease association. *Rev Med Virol* 25 Suppl 1: 2–23 (2015).

[20] Taberna M, Mena M, Pavón MA, Alemany L, Gillison ML, and Mesía R. Human papillomavirus-related oropharyngeal cancer. *Ann Oncol* 28(10): 2386–2398 (2017).

[21] Egawa K. Do human papillomaviruses target epidermal stem cells? *Dermatology* 207(3): 251–254 (2003).

[22] Egawa N, Nakahara T, Ohno S-I, Narisawa-Saito M, Yugawa T, Fujita M, Yamato K, Natori Y, and Kiyono T. The E1 protein of human papillomavirus type 16 is dispensable for maintenance replication of the viral genome. *J Virol* 86(6): 3276–3283 (2012).

[23] Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci* 110(5): 525–541 (2006).

[24] Brown DR, Kitchin D, Qadadri B, Neptune N, Batteiger T, and Ermel A. The human papillomavirus type 11 E1--E4 protein is a transglutaminase 3 substrate and induces abnormalities of the cornified cell envelope. *Virology* 345(1): 290–298 (2006).

[25] Wang Q, Griffin H, Southern S, Jackson D, Martin A, McIntosh P, Davy C, Masterson PJ, Walker PA, Laskey P, Omary MB, and Doorbar J. Functional analysis of the human papillomavirus type 16 E1=E4 protein provides a mechanism for in vivo and in vitro keratin filament reorganization. *J Virol* 78(2): 821–833 (2004).

[26] McIntosh PB, Martin SR, Jackson DJ, Khan J, Isaacson ER, Calder L, Raj K, Griffin HM, Wang Q, Laskey P, Eccleston JF, and Doorbar J. 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(16): 8196–8203 (2008).

[27] Paaso A, Jaakola A, Syrjänen S, and Louvanto K. From HPV Infection to Lesion Progression: The Role of HLA Alleles and Host Immunity. *Acta Cytol* 63(2): 148–158 (2019).

[28] Hildesheim A and Wang SS. Host and viral genetics and risk of cervical cancer: a review. *Virus Res* 89(2): 229–240 (2002).

[29] Ellis JR, Keating PJ, Baird J, Hounsell EF, Renouf DV, Rowe M, Hopkins D, Duggan-Keen MF, Bartholomew JS, and Young LS. The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. *Nat Med* 1(5): 464–470 (1995).

[30] Leo PJ, Madeleine MM, Wang S, Schwartz SM, Newell F, Pettersson-Kymmer U, Hemminki K, Hallmans G, Tiews S, Steinberg W, Rader JS, Castro F, Safaeian M, Franco EL, Coutlée F, Ohlsson C, Cortes A, Marshall M, Mukhopadhyay P, Cremin K, Johnson LG, Trimble CL, Garland S, Tabrizi SN, Wentzensen N, Sitas F, Little J, Cruickshank M, Frazer IH, Hildesheim A, and Brown MA. Defining the genetic susceptibility to cervical neoplasia-A genome-wide association study. *PLoS Genet* 13(8): e1006866 (2017).

[31] Moatter T, Aban M, Tabassum S, Shaikh U, and Pervez S. Molecular analysis of human leukocyte antigen class I and class II allele frequencies and haplotype distribution in Pakistani population. *Indian J Hum Genet* 16(3): 149–153 (2010).

[32] Ferguson R, Ramanakumar AV, Richardson H, Tellier P-P, Coutlée F, Franco EL, and Roger M. Human leukocyte antigen (HLA)-E and HLA-G polymorphisms in human papillomavirus infection susceptibility and persistence. *Hum Immunol* 72(4): 337–341 (2011).

[33] Madeleine MM, Brumback B, Cushing-Haugen KL, Schwartz SM, Daling JR, Smith AG, Nelson JL, Porter P, Shera KA, McDougall JK, and Galloway DA. Human leukocyte antigen class II and cervical cancer risk: a population-based study. *J Infect Dis* 186(11): 1565–1574 (2002).

[34] Bhatia A and Burtness B. Human Papillomavirus-Associated Oropharyngeal Cancer: Defining Risk Groups and Clinical Trials. *J Clin Oncol* 33(29): 3243–3250 (2015).

[35] Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, and Winchester DP. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 67(2): 93–99 (2017).

[36] Romagosa C, Simonetti S, López-Vicente L, Mazo A, Lleonart ME, Castellvi J, and Ramon y Cajal S. p16(Ink4a) overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene* 30(18): 2087–2097 (2011).

[37] Klussmann JP, Gültekin E, Weissenborn SJ, Wieland U, Dries V, Dienes HP, Eckel HE, Pfister HJ, and Fuchs PG. Expression of p16 Protein Identifies a Distinct Entity of Tonsillar Carcinomas Associated with Human Papillomavirus. *Am J Pathol* 162(3): 747–753 (2003).

[38] Yamashita Y, Ikegami T, Hirakawa H, Uehara T, Deng Z, Agena S, Uezato J, Kondo S, Kiyuna A, Maeda H, Suzuki M, and Ganaha A. Staging and prognosis of oropharyngeal carcinoma according to the 8th Edition of the American Joint Committee on Cancer Staging Manual in human papillomavirus infection. *Eur Arch Otorhinolaryngol* 276(3): 827–836 (2019).

[39] Kato MG, Baek C-H, Chaturvedi P, Gallagher R, Kowalski LP, Leemans CR, Warnakulasuriya S, Nguyen SA, and Day TA. Update on oral and oropharyngeal cancer staging – International perspectives. *World J Otorhinolaryngol Head Neck Surg* 6(1): 66–75 (2020).

[40] Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, Raben D, Baselga J, Spencer SA, Zhu J, Youssoufian H, Rowinsky EK, and Ang KK. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol* 11(1): 21–28 (2010).

[41] Sher DJ, Adelstein DJ, Bajaj GK, Brizel DM, Cohen EEW, Halthore A, Harrison LB, Lu C, Moeller BJ, Quon H, Rocco JW, Sturgis EM, Tishler RB, Trotti A, Waldron J, and Eisbruch A. Radiation therapy for oropharyngeal squamous cell carcinoma: Executive summary of an ASTRO Evidence-Based Clinical Practice Guideline. *Pract Radiat Oncol* 7(4): 246–253 (2017).

[42] Golusiński W and Golusińska-Kardach E. Current Role of Surgery in the Management of Oropharyngeal Cancer. *Front Oncol* 9: p. 388 (2019).

[43] Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, Westra WH, Chung CH, Jordan RC, Lu C, Kim H, Axelrod R, Silverman CC, Redmond KP, and Gillison ML. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 363(1): 24–35 (2010).

[44] Marur S, Li S, Cmelak AJ, Gillison ML, Zhao WJ, Ferris RL, Westra WH, Gilbert J, Bauman JE, Wagner LI, Trevarthen DR, Balkrishna J, Murphy BA, Agrawal N, Colevas AD, Chung CH, and Burtness B. E1308: Phase II Trial of Induction Chemotherapy Followed by Reduced-Dose Radiation and Weekly Cetuximab in Patients With HPV-Associated Resectable Squamous Cell Carcinoma of the Oropharynx- ECOG-ACRIN Cancer Research Group. *J Clin Oncol* 35(5): 490–497 (2017).

[45] Misiukiewicz K, Gupta V, Miles BA, Bakst R, Genden E, Selkridge I, Surgeon JT, Rainey H, Camille N, Roy E, Zhang D, Ye F, Jia R, Moshier E, Bonomi M, Hwang M, Som P, and Posner MR. Standard of care vs reduced-dose chemoradiation after induction chemotherapy in HPV+ oropharyngeal carcinoma patients: The Quarterback trial. *Oral Oncol* 95: 170–177 (2019).

[46] Hargreaves S, Beasley M, Hurt C, Jones TM, and Evans M. Deintensification of Adjuvant Treatment After Transoral Surgery in Patients With Human Papillomavirus-Positive Oropharyngeal Cancer: The Conception of the PATHOS Study and Its Development. *Front Oncol* 9: p. 936 (2019).

[47] Castle PE and Maza M. Prophylactic HPV vaccination: past, present, and future. *Epidemiol Infect* 144(3): 449–468 (2016).

[48] Markowitz LE, Liu G, Hariri S, Steinau M, Dunne EF, and Unger ER. Prevalence of HPV After Introduction of the Vaccination Program in the United States. *Pediatrics* 137(3): e20151968 (2016).

[49] Chaturvedi AK, Graubard BI, Broutian T, Pickard RKL, Tong Z-Y, Xiao W, Kahle L, and Gillison ML. Effect of Prophylactic Human Papillomavirus (HPV) Vaccination on Oral HPV Infections Among Young Adults in the United States. *J Clin Oncol* 36(3): 262–267 (2018).

[50] Whiteside TL. Immunobiology of head and neck cancer. *Cancer Metastasis Rev* 24(1): 95–105 (2005).

[51] Duray A, Demoulin S, Hubert P, Delvenne P, and Saussez S. Immune suppression in head and neck cancers: a review. *Clin Dev Immunol* 2010: p. 701657 (2010).

[52] Strauss L, Bergmann C, Gooding W, Johnson JT, and Whiteside TL. The frequency and suppressor function of CD4+CD25highFoxp3+ T cells in the circulation of patients with squamous cell carcinoma of the head and neck. *Clin Cancer Res* 13(21): 6301–6311 (2007).

[53] Jie H-B, Gildener-Leapman N, Li J, Srivastava RM, Gibson SP, Whiteside TL, and Ferris RL. Intratumoral regulatory T cells upregulate immunosuppressive molecules in head and neck cancer patients. *Br J Cancer* 109(10): 2629–2635 (2013).

[54] Mandal R, Şenbabaoğlu Y, Desrichard A, Havel JJ, Dalin MG, Riaz N, Lee K-W, Ganly I, Hakimi AA, Chan TA, and Morris LG. The head and neck cancer immune landscape and its immunotherapeutic implications. *JCI Insight* 1(17): e89829 (2016).

[55] Näsman A, Romanitan M, Nordfors C, Grün N, Johansson H, Hammarstedt L, Marklund L, Munck-Wikland E, Dalianis T, and Ramqvist T. Tumor infiltrating CD8+ and Foxp3+ lymphocytes correlate to clinical outcome and human papillomavirus (HPV) status in tonsillar cancer. *PLoS ONE* 7(6): e38711 (2012).

[56] Moskovitz J, Moy J, and Ferris RL. Immunotherapy for Head and Neck Squamous Cell Carcinoma. *Curr Oncol Rep* 20(2): p. 22 (2018).

[57] Chow LQM, Haddad R, Gupta S, Mahipal A, Mehra R, Tahara M, Berger R, Eder JP, Burtness B, Lee S-H, Keam B, Kang H, Muro K, Weiss J, Geva R, Lin C-C, Chung HC, Meister A, Dolled-Filhart M, Pathiraja K, Cheng JD, and Seiwert TY. Antitumor Activity of Pembrolizumab in Biomarker-Unselected Patients With Recurrent

and/or Metastatic Head and Neck Squamous Cell Carcinoma: Results From the Phase Ib KEYNOTE-012 Expansion Cohort. *J Clin Oncol* 34(32): 3838–3845 (2016).

[58] Ferris RL, Blumenschein G, Fayette J, Guigay J, Colevas AD, Licitra L, Harrington K, Kasper S, Vokes EE, Even C, Worden F, Saba NF, Iglesias Docampo LC, Haddad R, Rordorf T, Kiyota N, Tahara M, Monga M, Lynch M, Geese WJ, Kopit J, Shaw JW, and Gillison ML. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* 375(19): 1856–1867 (2016).

[59] Ferris RL, Blumenschein G, Fayette J, Guigay J, Colevas AD, Licitra L, Harrington KJ, Kasper S, Vokes EE, Even C, Worden F, Saba NF, Docampo LCI, Haddad R, Rordorf T, Kiyota N, Tahara M, Lynch M, Jayaprakash V, Li L, and Gillison ML. Nivolumab vs investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. *Oral Oncol* 81: 45–51 (2018).

[60] Ferris RL and Licitra L. PD-1 immunotherapy for recurrent or metastatic HNSCC. *The Lancet* 394(10212): 1882–1884 (2019).

[61] Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, Cid-Pavon GMD, Ramsbottom K, Ghattaoraya G, Alfirevic A, Middleton D, and Jones AR. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 48(D1): D783-D788 (2020).

[62] Nelde A, Kowalewski DJ, and Stevanović S. Purification and Identification of Naturally Presented MHC Class I and II Ligands. *Methods Mol Biol* 1988: 123–136 (2019).

[63] Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-T, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJG, Old LJ, Camargo AA, and Vasconcelos ATR. CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37(Database issue): D816-9 (2009).

[64] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45(6): 580–585 (2013).

[65] Thul PJ and Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci* 27(1): 233–244 (2018).

[66] Michael Love SA. DESeq2: Bioconductor, 2017.

[67] Huang DW, Sherman BT, and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1): 1–13 (2009).

[68] Huang DW, Sherman BT, and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1): 44–57 (2009).

[69] van Endert, PM, Riganelli D, Greco G, Fleischhauer K, Sidney J, Sette A, and Bach JF. The peptide-binding motif for the human transporter associated with antigen processing. *J Exp Med* 182(6): 1883–1895 (1995).

[70] Neefjes J, Jongsma MLM, Paul P, and Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11(12): 823–836 (2011).

[71] Holoshitz J. The quest for better understanding of HLA-disease association: scenes from a road less travelled by. *Discov Med* 16(87): 93–101 (2013).

[72] Atanackovic D, Blum I, Cao Y, Wenzel S, Bartels K, Faltz C, Hossfeld DK, Hegewisch-Becker S, Bokemeyer C, and Leuwer R. Expression of cancer-testis antigens as possible targets for antigen-specific immunotherapy in head and neck squamous cell carcinoma. *Cancer Biol Ther* 5(9): 1218–1225 (2006).

[73] Piotti KC, Scognamiglio T, Chiu R, and Chen Y-T. Expression of cancer/testis (CT) antigens in squamous cell carcinoma of the head and neck: evaluation as markers of squamous dysplasia. *Pathol Res Pract* 209(11): 721–726 (2013).

[74] Yoshitake Y, Fukuma D, Yuno A, Hirayama M, Nakayama H, Tanaka T, Nagata M, Takamune Y, Kawahara K, Nakagawa Y, Yoshida R, Hirosue A, Ogi H, Hiraki A, Jono H, Hamada A, Yoshida K, Nishimura Y, Nakamura Y, and Shinohara M. Phase II clinical trial of multiple peptide vaccination for advanced head and neck cancer patients revealed induction of immune responses and improved OS. *Clin Cancer Res* 21(2): 312–321 (2015).

[75] Chen Y-J, Chang JT, Lee L, Wang H-M, Liao C-T, Chiu C-C, Chen P-J, and Cheng A-J. DSG3 is overexpressed in head neck cancer and is a potential molecular target for inhibition of oncogenesis. *Oncogene* 26(3): 467–476 (2007).

[76] Chen Y-J, Lee L-Y, Chao Y-K, Chang JT, Lu Y-C, Li H-F, Chiu C-C, Li Y-C, Li Y-L, Chiou J-F, and Cheng A-J. DSG3 facilitates cancer cell growth and invasion through the DSG3-plakoglobin-TCF/LEF-Myc/cyclin D1/MMP signaling pathway. *PLoS ONE* 8(5): e64088 (2013).

[77] Bunbanjerdsuk S, Vorasan N, Saethang T, Pongrujikorn T, Pangpunyakulchai D, Mongkonsiri N, Arsa L, Thokanit N, Pongsapich W, Anekpuritanang T, Ngamphaiboon N, Jinawath A, Sunpaweravong S, Pisitkun T, Suktitipat B, and Jinawath N. Oncoproteomic and gene expression analyses identify prognostic biomarkers for second primary malignancy in patients with head and neck squamous cell carcinoma. *Mod Pathol* 32(7): 943–956 (2019).

[78] Chen J, Li S, Zhou S, Cao S, Lou Y, Shen H, Yin J, and Li G. Kinesin superfamily protein expression and its association with progression and prognosis in hepatocellular carcinoma. *J Cancer Res Ther* 13(4): 651–659 (2017).

[79] Tampakis A, Tampaki EC, Nonni A, Droeser R, Posabella A, Tsourouflis G, Kontzoglou K, Patsouris E, Flüe M von, and Kouraklis G. Nectin-1 Expression in Colorectal Cancer: Is There a Group of Patients with High Risk for Early Disease Recurrence? *Oncology* 96(6): 318–325 (2019).

[80] Zhan P, Xi G-M, Zhang B, Wu Y, Liu H-B, Liu Y-F, Xu W-J, Zhu Q, Cai F, Zhou Z-J, Miu Y-Y, Wang X-X, Jin J-J, Li Q, Lv T-F, and Song Y. NCAPG2 promotes tumour proliferation by regulating G2/M phase and associates with poor prognosis in lung adenocarcinoma. *J Cell Mol Med* 21(4): 665–676 (2017).

[81] Meng F, Zhang S, Song R, Liu Y, Wang J, Liang Y, Wang J, Han J, Song X, Lu Z, Yang G, Pan S, Li X, Liu Y, Zhou F, Wang Y, Cui Y, Zhang B, Ma K, Zhang C, Sun Y, Xin M, and Liu L. NCAPG2 overexpression promotes hepatocellular carcinoma proliferation and metastasis through activating the STAT3 and NF-κB/miR-188-3p pathways. *EBioMedicine* 44: 237–249 (2019).

[82] Li J, Ding Y, and Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J Surg Oncol* 14(1): p. 297 (2016).

[83] Rong L, Huang W, Tian S, Chi X, Zhao P, and Liu F. COL1A2 is a Novel Biomarker to Improve Clinical Prediction in Human Gastric Cancer: Integrating Bioinformatics and Meta-Analysis. *Pathol Oncol Res* 24(1): 129–134 (2018).

[84] Ao R, Guan L, Wang Y, and Wang J-N. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway. *J Cell Biochem* 119(6): 4420–4434 (2018).

[85] Wu J, Liu J, Wei X, Yu Q, Niu X, Tang S, and Song L. A feature-based analysis identifies COL1A2 as a regulator in pancreatic cancer. *J Enzyme Inhib Med Chem* 34(1): 420–428 (2019).

[86] Aichem A and Groettrup M. The ubiquitin-like modifier FAT10 in cancer development. *Int J Biochem Cell Biol* 79: 451–461 (2016).

[87] Xiang S, Shao X, Cao J, Yang B, He Q, and Ying M. FAT10: Function and Relationship with Cancer. *Curr Mol Pharmacol* 13(3): 182–191 (2020).

[88] Martin-Padron J, Boyero L, Rodriguez MI, Andrades A, Díaz-Cano I, Peinado P, Baliñas-Gavira C, Alvarez-Perez JC, Coira IF, Fárez-Vidal ME, and Medina PP. Plakophilin 1 enhances MYC translation, promoting squamous cell lung cancer. *Oncogene* 39(32): 5479–5493 (2020).

[89] Whitfield JR, Beaulieu M-E, and Soucek L. Strategies to Inhibit Myc and Their Clinical Applicability. *Front Cell Dev Biol* 5: p. 10 (2017).

[90] Haase D, Cui T, Yang L, Ma Y, Liu H, Theis B, Petersen I, and Chen Y. Plakophilin 1 is methylated and has a tumor suppressive activity in human lung cancer. *Exp Mol Pathol* 108: 73–79 (2019).

[91] Morasso MI, Markova NG, and Sargent TD. Regulation of epidermal differentiation by a Distal-less homeodomain gene. *J Cell Biol* 135(6 Pt 2): 1879–1887 (1996).

[92] Ferrari N, Paleari L, Palmisano GL, Tammaro P, Levi G, Albini A, and Brigati C. Induction of apoptosis by fenretinide in tumor cell lines correlates with DLX2, DLX3 and DLX4 gene expression. *Oncol Rep* 10(4): 973–977 (2003).

[93] Palazzo E, Kellett M, Cataisson C, Gormley A, Bible PW, Pietroni V, Radoja N, Hwang J, Blumenberg M, Yuspa SH, and Morasso MI. The homeoprotein DLX3 and tumor suppressor p53 co-regulate cell cycle progression and squamous tumor growth. *Oncogene* 35(24): 3114–3124 (2016).

[94] Palazzo E, Marconi A, Pincelli C, and Morasso MI. Do DLX3 and CD271 Protect Human Keratinocytes from Squamous Tumor Development? *Int J Mol Sci* 20(14) (2019).

[95] Spindler V, Heupel W-M, Efthymiadis A, Schmidt E, Eming R, Rankl C, Hinterdorfer P, Müller T, Drenckhahn D, and Waschke J. Desmocollin 3-mediated binding is crucial for keratinocyte cohesion and is impaired in pemphigus. *J Biol Chem* 284(44): 30556–30564 (2009).

[96] Cui T, Chen Y, Yang L, Knösel T, Huber O, Pacyna-Gengelbach M, and Petersen I. The p53 target gene desmocollin 3 acts as a novel tumor suppressor through inhibiting EGFR/ERK pathway in human lung cancer. *Carcinogenesis* 33(12): 2326–2333 (2012).

[97] Cui T, Yang L, Ma Y, Petersen I, and Chen Y. Desmocollin 3 has a tumor suppressive activity through inhibition of AKT pathway in colorectal cancer. *Exp Cell Res* 378(2): 124–130 (2019).

[98] Basnet S, Sharma S, Costea DE, and Sapkota D. Expression profile and functional role of S100A14 in human cancer. *Oncotarget* 10(31): 2996–3012 (2019).

[99] Hu L, Kong F, and Pan Y. Prognostic and clinicopathological significance of S100A14 expression in cancer patients: A meta-analysis. *Medicine (Baltimore)* 98(28): e16356 (2019).

[100] Lunde MLS, Roman E, Warnakulasuriya S, Mehrotra R, Laranne J, Vasstrand EN, and Ibrahim SO. Profiling of chromosomal changes in potentially malignant and malignant oral mucosal lesions from South and South-East Asia using array-comparative genomic hybridization. *Cancer Genomics Proteomics* 11(3): 127–140 (2014).

[101] Sapkota D, Bruland O, Bøe OE, Bakeer H, Elgindi OAA, Vasstrand EN, and Ibrahim SO. Expression profile of the S100 gene family members in oral squamous cell carcinomas. *J Oral Pathol Med* 37(10): 607–615 (2008).

[102] Pandey S, Osman TA, Sharma S, Vallenari EM, Shahdadfar A, Pun CB, Gautam DK, Uhlin-Hansen L, Rikardsen O, Johannessen AC, Costea DE, and Sapkota D. Loss of S100A14 expression at the tumor-invading front correlates with poor differentiation and worse prognosis in oral squamous cell carcinoma. *Head Neck* 42(8): 2088–2098 (2020).

[103] Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Imielinsk M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, and Stratton MR. Signatures of mutational processes in human cancer. *Nature* 500(7463): 415–421 (2013).

[104] Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L, Chen C, Olive O, Carter TA, Li S, Lieb DJ, Eisenhaure T, Gjini E, Stevens J, Lane WJ, Javeri I, Nellaiappan K, Salazar AM, Daley H, Seaman M, Buchbinder EI, Yoon CH, Harden M, Lennon N, Gabriel S, Rodig SJ, Barouch DH, Aster JC, Getz G, Wucherpfennig K, Neuberg D, Ritz J, Lander ES, Fritsch EF, Hacohen N, and Wu CJ. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 547(7662): 217–221 (2017).

[105] Löffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, Trautwein N, Hilke FJ, Zinser RS, Mühlenbruch L, Kowalewski DJ, Schuster H, Sturm M, Matthes J, Riess O, Czemmel S, Nahnsen S, Königsrainer I, Thiel K, Nadalin S, Beckert S, Bösmüller H, Fend F, Velic A, Maček B, Haen SP, Buonaguro L, Kohlbacher O, Stevanović S, Königsrainer A, and Rammensee H-G. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med* 11(1): p. 28 (2019).

[106] Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, Schrock A, Campbell B, Shlien A, Chmielecki J, Huang F, He Y, Sun J, Tabori U, Kennedy M, Lieber DS, Roels S, White J, Otto GA, Ross JS, Garraway L, Miller VA, Stephens PJ, and Frampton GM. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* 9(1): p. 34 (2017).

[107] Bassani-Sternberg M, Chong C, Guillaume P, Solleder M, Pak H, Gannon PO, Kandalaft LE, Coukos G, and Gfeller D. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *PLoS Comput Biol* 13(8): e1005725 (2017).

[108] Moody CA and Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer* 10(8): 550–560 (2010).

[109] Venuti A, Paolini F, Nasir L, Corteggio A, Roperto S, Campo MS, and Borzacchiello G. Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer* 10: p. 140 (2011).

[110] Ashrafi GH, Haghshenas M, Marchetti B, and Campo MS. E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain. *Int J Cancer* 119(9): 2105–2112 (2006).

[111] Steinbach A and Riemer AB. Immune evasion mechanisms of human papillomavirus: An update. *Int J Cancer* 142(2): 224–229 (2018).

[112] Riemer AB, Keskin DB, Zhang G, Handley M, Anderson KS, Brusic V, Reinhold B, and Reinherz EL. A conserved E7-derived cytotoxic T lymphocyte epitope expressed on human papillomavirus 16-transformed HLA-A2+ epithelial cancers. *J Biol Chem* 285(38): 29608–29622 (2010).

[113] Keskin DB, Reinhold B, Lee SY, Zhang G, Lank S, O'Connor DH, Berkowitz RS, Brusic V, Kim SJ, and Reinherz EL. Direct identification of an HPV-16 tumor antigen from cervical cancer biopsy specimens. *Front Immunol* 2: p. 75 (2011).

[114] 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. A Targeted LC-MS Strategy for Low-Abundant HLA Class-I-Presented Peptide Detection Identifies Novel Human Papillomavirus T-Cell Epitopes. *Proteomics* 18(11): e1700390 (2018).

[115] Gleimer M and Parham P. Stress Management. *Immunity* 19(4): 469–477 (2003).

[116] Ternette N, Block PD, Sánchez-Bernabéu Á, Borthwick N, Pappalardo E, Abdul-Jawad S, Ondondo B, Charles PD, Dorrell L, Kessler BM, and Hanke T. Early Kinetics of the HLA Class I-Associated Peptidome of MVA.HIVconsv-Infected Cells. *J Virol* 89(11): 5760–5771 (2015).

[117] Ziegler MC, Nelde A, Weber JK, Schreitmüller CM, Martrus G, Huynh T, Bunders MJ, Lunemann S, Stevanovic S, Zhou R, and Altfeld M. HIV-1 induced changes in HLA-C*03 : 04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS* 34(12): 1713–1723 (2020).

[118] Zhou C, Tuong ZK, and Frazer IH. Papillomavirus Immune Evasion Strategies Target the Infected Cell and the Local Immune System. *Front Oncol* 9: p. 682 (2019).

[119] Nakanishi F, Ohkawa K, Ishida H, Hosui A, Sato A, Hiramatsu N, Ueda K, Takehara T, Kasahara A, Sasaki Y, Hori M, and Hayashi N. Alteration in gene expression profile by full-length hepatitis B virus genome. *Intervirology* 48(2-3): 77–83 (2005).

[120] Lamontagne J, Mell JC, and Bouchard MJ. Transcriptome-Wide Analysis of Hepatitis B Virus-Mediated Changes to Normal Hepatocyte Gene Expression. *PLoS Pathog* 12(2): e1005438 (2016).

[121] Fortier M-H, Caron É, Hardy M-P, Voisin G, Lemieux S, Perreault C, and Thibault P. The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med* 205(3): 595–610 (2008).

[122] Weinzierl AO, Lemmel C, Schoor O, Müller M, Krüger T, Wernet D, Hennenlotter J, Stenzl A, Klingel K, Rammensee H-G, and Stevanovic S. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol Cell Proteomics* 6(1): 102–113 (2007).

[123] Schuster H, Peper JK, Bösmüller H-C, Röhle K, Backert L, Bilich T, Ney B, Löffler MW, Kowalewski DJ, Trautwein N, Rabsteyn A, Engler T, Braun S, Haen SP, Walz JS, Schmid-Horch B, Brucker SY, Wallwiener D, Kohlbacher O, Fend F, Rammensee H-G, Stevanović S, Staebler A, and Wagner P. The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl Acad Sci U S A* 114(46): E9942-E9951 (2017).

[124] Yewdell JW. The seven dirty little secrets of major histocompatibility complex class I antigen processing. *Immunol Rev* 207: 8–18 (2005).

[125] Wheatley DN, Giddings MR, and Inglis MS. Kinetics of degradation of "short-" and "long-lived" proteins in cultured mammalian cells. *Cell Biology International Reports* 4(12): 1081–1090 (1980).

[126] Milner E, Barnea E, Beer I, and Admon A. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Mol Cell Proteomics* 5(2): 357–365 (2006).

[127] Solomon B, Young RJ, Bressel M, Cernelc J, Savas P, Liu H, Urban D, Thai A, Cooper C, Fua T, Neeson P, Loi S, Porceddu SV, and Rischin D. Identification of an excellent prognosis subset of human papillomavirus-associated oropharyngeal cancer patients by quantification of intratumoral CD103+ immune cell abundance. *Ann Oncol* 30(10): 1638–1646 (2019).

[128] Hewavisenti R, Ferguson A, Wang K, Jones D, Gebhardt T, Edwards J, Zhang M, Britton W, Yang J, Hong A, and Palendira U. CD103+ tumor-resident CD8+ T cell numbers underlie improved patient survival in oropharyngeal squamous cell carcinoma. *J Immunother Cancer* 8(1) (2020).

4 General perspective

HLA ligandomic analyses are of high potential for both basic research and translational implementation. Large-scale immunopeptidomics are a research field underlying constant changes and novel developments. The possibility to analyze the HLA ligandomes of tumor tissues from a variety of cancer entities is a powerful tool in the identification of potential target antigens for the application in cancer immunotherapeutic strategies [1]. The selection of suitable target antigens remains challenging. The elaborate methods accompanying the isolation and MS analysis of HLA-presented antigens as well as the challenges for bioinformatic processing of large-scale datasets and integration of multi-dimensional datasets still bear limitations for clinical application. Increasing depth, coverage and confidence in peptide sequencing is necessary to cross present borders and enhance the possibilities of immunopeptidomics [2].

Current studies tend to extend their search radius for antigenic targets and focus on alternative source proteins besides the standard proteomic repertoire. Neoantigens, which present genomic mutations unique for a tumor, are of great interest due to their tumor specificity. There is evidence that neoantigens can be recognized by the immune system and neoantigen targeting can induce antitumor responses [3]. The detection of naturally HLA-presented neoantigens extracted from tumor tissue is challenging and rarely successful. Progression of in-depth and accurate MS methods will facilitate neoantigen identification [4]. Furthermore, post-translationally modified antigens are in the present focus of immunopeptidomic research. There is a wide range of post-translationally modified proteins like phosphorylated proteins that might play a substantial role for the repertoire of antigenic source proteins [5]. Aberrant cellular processes leading to altered post-translational protein modifications were shown to occur in cancer cells [6]. Improved methods will facilitate the analysis of interactions between HLA and phosphorylated peptides and provide further possibilities for the search of immunotherapeutic targets. Additional research focus is on the identification of HLA-presented cryptic peptides. These peptides originate from cryptic open reading frames encoded in untranslated regions, non-coding RNAs, intronic or intergenic regions or coding sequences with a shift in the canonical reading frame. Cryptic translation products are suggested to serve as source for HLA-presented antigens [7]. Novel enhancements in experimental protocols and data analysis tools enabled the identification of highly abundant HLA-presented cryptic antigens with potential relevance for the development of immunotherapeutic strategies [8].

The field of immunopeptidomic research is currently developing and far from exploiting the spectrum of possibilities. Improved HLA ligand isolation methods, LC-MS/MS devices with enhanced resolution and sensitivity, better acquisition methods, tailored bioinformatic data analysis tools and a rising amount of publicly accessible datasets will increase the knowledge about the naturally HLA-presented

antigenic repertoire. This will shape the opportunities arising for the development of novel cancer immunotherapeutic approaches.

References

[1] Freudenmann LK, Marcu A, and Stevanović S. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. *Immunology* 154(3): 331–345 (2018).

[2] Faridi P, Purcell AW, and Croft NP. In Immunopeptidomics We Need a Sniper Instead of a Shotgun. *Proteomics* 18(12): e1700464 (2018).

[3] Yarchoan M, Johnson BA, Lutz ER, Laheru DA, and Jaffee EM. Targeting neoantigens to augment antitumour immunity. *Nat Rev Cancer* 17(4): 209–222 (2017).

[4] Bassani-Sternberg M. Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens. *Methods Mol Biol* 1719: 209–221 (2018).

[5] Solleder M, Guillaume P, Racle J, Michaux J, Pak H-S, Müller M, Coukos G, Bassani-Sternberg M, and Gfeller D. Mass Spectrometry Based Immunopeptidomics Leads to Robust Predictions of Phosphorylated HLA Class I Ligands. *Mol Cell Proteomics* 19(2): 390–404 (2020).

[6] Cho Y, Kang HG, Kim S-J, Lee S, Jee S, Ahn SG, Kang MJ, Song JS, Chung J-Y, Yi EC, and Chun K-H. Post-translational modification of OCT4 in breast cancer tumorigenesis. *Cell Death Differ* 25(10): 1781–1795 (2018).

[7] Orr MW, Mao Y, Storz G, and Qian S-B. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res* 48(3): 1029–1042 (2020).

[8] Erhard F, Dölken L, Schilling B, and Schlosser A. Identification of the Cryptic HLA-I Immunopeptidome. *Cancer Immunol Res* 8(8): 1018–1026 (2020).

5 Publications

Löffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, Trautwein N, Hilke FJ, Zinser RS, Mühlenbruch L, Kowalewski DJ, Schuster H, Sturm M, Matthes J, Riess O, Czemmel S, Nahnsen S, Ingmar Königsrainer I, Thiel K, Nadalin S, Beckert S, Bösmüller H, Fend F, Velic A, Maček B, Haen SP, Buonaguro L, Kohlbacher O, Stevanović S, Königsrainer A, HEPAVAC Consortium, and Rammensee H-G. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Medicine* 11(1): p28 (2019).

Löffler MW, Nussbaum B, Jäger G, Jurmeister PS, Budczies J, Pereira PL, Clasen S, Kowalewski DJ, Mühlenbruch L, Königsrainer I, Beckert S, Ladurner R, Wagner S, Bullinger F, Gross TH, Schroeder C, Sipos B , Königsrainer A, Stevanović S, Denkert C, Rammensee H-G, Gouttefangeas C , Haen SP. A Non-interventional Clinical Trial Assessing Immune Responses After Radiofrequency Ablation of Liver Metastases From Colorectal Cancer. *Frontiers in Immunology* 10: 2526 (2019).

Wang J, Jelcic I, Mühlenbruch L, Haunerdinger V, Toussaint NC, Zhao Y, Cruciani C, Faigle W, Naghavian R, Foege M, Binder TMC, Eiermann T, Opitz L, Fuentes-Font L, Reynolds R, Kwok WW, Nguyen J, Lee J-H, Lutterotti A, Münz C, Rammensee H-G, Hauri-Hohl M, Sospedra M, Stevanović S, Martin R. Tightly Linked HLA-DR15 Molecules Jointly Shape an Autoreactive T Cell Repertoire in Multiple Sclerosis. *Cell* 183(5): 1264-1281 (2020).

Ghosh M, Hartmann H, Jakobi M, März L, Bichmann L, Freudenmann LK, Mühlenbruch L, Segan S, Rammensee H-G, Schneiderhan-Marra N, Shipp C, Stevanović S, Joos TO. The impact of biomaterial cell contact on the immunopeptidome. *Frontiers in Bioengineering and Biotechnology. Accepted.*

6 Danksagung

An dieser Stelle möchte ich mich herzlich bei all denjenigen bedanken, die mich während meiner Promotion begleitet und unterstützt haben.

Ein besonderer Dank geht an Prof. Dr. Stefan Stevanović und Prof. Dr. Hans-Georg Rammensee für die Möglichkeit, meine Promotion zu absolvieren, und ihre wissenschaftliche Unterstützung. Ich bedanke mich außerdem bei Prof. Dr. Oliver Planz für fachliche Unterstützung und die Bereitschaft als Zweitgutachter zu fungieren.

Meinen Kooperationspartnern Prof. Dr. Dr. Ghazaleh Tabatabai und Dr. Irina Gepfner-Tuma sowie PD Dr. Simon Laban und Jasmin Ezić danke ich für die Bereitstellung der Proben und die gemeinsame wissenschaftliche Arbeit an den Projekten.

Ein großer Dank gilt meinen Kollegen für die wissenschaftlichen Gespräche und Ratschläge, aber auch für nicht-fachliche Unterstützung: Lena, Marion, Michi, Ana, Annika, Jens, Leon, Maren, Tati, Moni, Miri, Camille. Vielen Dank an die gesamte Arbeitsgruppe Stevanović und alle Elche der Abteilung Immunologie für die freundliche Arbeitsatmosphäre.

Ein besonderes Dankeschön geht an meine Familie und meine Freunde, die mich in den letzten Jahren begleitet und immer unterstützt haben: Petra, Andreas, Lisa, Laurens, Lena, Catrin, Hannah, Meli und viele mehr.

7 Abbreviations

AA	amino acid
AcN	acetonitrile
APC	antigen presenting cell
BP	biological process
CAR	chimeric antigen receptor
CC	cellular compartment
CI	confidence interval
CID	collision-induced dissociation
CLIP	major histocompatibility complex class II-associated li peptide
CNBr	cyanogen bromide
CNS	central nervous system
CSC	cancer stem cell
CTA	cancer-testis antigen
CTL	cytotoxic T lymphocyte
Da	Dalton
DDA	data-dependent acquisition
DRiP	defective ribosomal protein
EGFR	epidermal growth factor receptor
ER	endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
ERAP	endoplasmic reticulum aminopeptidase
ESI	electrospray ionization
FDR	false discovery rate
FT	Fourier transformation
GO	gene ontology
GTEx	Genotype-Tissue Expression Project
GTR	gross total resection
HLA	human leukocyte antigen
HPV	human papilloma virus
ICI	immune checkpoint inhibitor
IFN	interferon
IL	interleukin
IR	incidence rate
IT	ion trap
kb	kilobase
LC-MS/MS	liquid chromatography

Abbreviations

Ii	invariant chain
m/z	mass-to-charge
mAb	monoclonal antibody
MF	molecular function
MHC	major histocompatibility complex
MIC	MHC class I polypeptide-related sequence
MRI	magnetic resonance imaging
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NK cell	natural killer cell
OPSCC	oropharyngeal squamous cell carcinoma
OR	odds ratio
OS	overall survival
PBMC	peripheral blood mononuclear cell
PBS	phosphate-buffered saline
PCA	principle component analysis
PD1	programmed cell death protein 1
PDL1	programmed cell death ligand 1
PLC	peptide-loading complex
ppm	parts per million
PSM	peptide spectrum match
RELA	v-rel avian reticuloendotheliosis viral oncogene homolog A
RG	radial glia cell
RPKM	reads per kilobase million
rpm	rounds per minute
STR	subtotal resection
TAA	tumor-associated antigens
TAP	transporter associated with antigen processing
TCR	T cell receptor
TFA	trifluoroacetic acid
T_H cell	T helper cell
TIL	tumor-infiltrating lymphocyte
TLR	Toll-like receptor
TNF	tumor necrosis factor
TPM	transcripts per kilobase million
T_{reg} cell	regulatory T cell
TSA	tumor-specific antigen
WES	whole exome sequencing
WHO	World Health Organization

8 Appendix

8.1 Materials and methods

HLA typing

HLA class I and class II typing was performed based on DNA sequencing on blood samples of the patients. For the ependymoma cohort, this was carried out by the HLA laboratory of the university medical center Tübingen (Universitätsklinikum Tübingen, UKT). For the OPSCC cohort, this was carried out by PD Joannis Mytilineos (Institute for Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Transfusion Service, Baden Württemberg/Hessen, University Hospital Ulm, Germany).

HLA class I allelic distribution

To analyze the HLA class I allelic distribution, frequencies of HLA alleles occurring within the respective patient cohorts were determined using following equation:

$$F = \frac{n(\text{allele})}{n(\text{total})},$$

where F is the allele frequency, n(allele) is the count of the allele of interest within the cohort and n(total) is the count of either all HLA-A, -B, or -C alleles, respectively, within the cohort. Statistical analysis of the HLA class I allelic distribution was performed in accordance to the analysis methods applied by Krini *et al.* (2012) [1]. Assigned HLA-alleles from the OPSCC study were excluded from this analysis. To ascertain potential anomalies within the HLA allelic distributions of the cohorts, a general German population was used for comparison. The cohort “Germany pop 8” (n(donors) = 39,689), which is comprised in The Allele Frequency Net Database (www.allelefrequencies.net) [2], served as reference dataset. A double chromosome set was assumed for the reference dataset leading to n(total) = 79,378 HLA-A, -B or -C alleles. For statistical data analysis, the GraphPad Prism 6.1 software (GraphPad Software Inc, San Diego, USA) was used. Whether the frequencies of individual HLA alleles of the patient cohort significantly differed compared to the reference, was examined by 2x2 tables. By this, the count of each individual allele was compared to the count of other alleles within the patient or the reference cohort. Chi-square test was applied due to the high number of alleles in the reference dataset. The levels of statistical significance were defined after Bonferroni correction for HLA-A, -B and -C, respectively. Afterwards, the relative risks of the individual HLA alleles were determined by logistic regression analysis and expressed by the odds ratios with a 95% confidence interval (CI). Results from HLA alleles that were only represented once within a patient cohort were assumed as unreliable and excluded from further consideration.

Isolation of HLA ligands

HLA class I- and class II-presented peptides were isolated *via* standard immunoaffinity chromatography, which was performed as described by Nelde *et al.* (2019) [3].

For the extraction of HLA molecules, the pan-HLA class I-specific mouse monoclonal antibody (mAb) W6/32, the pan-HLA class II-specific mouse mAb Tü-39 and the HLA-DR-specific mouse mAb L243 were used **(Table 11)**. All antibodies were produced in-house by Claudia Falkenburger and stored at 4°C with thiomersal as preservative until usage.

Table 11: Antibodies specific for HLA molecules were used for immunoaffinity chromatography.

Clone	Specificity	Isotype
W6/32	pan-HLA class I-specific (HLA-A, -B, -C)	$IgG_{2\alpha}$
Tü39	pan-HLA class II-specific (HLA-DR, -DP, -DQ)	$IgG_{2\alpha}$
L243	HLA-DR-specific	$IgG_{2\alpha}$

Prior to their application for HLA precipitation, the mAbs W6/32, Tü39 and L243 were covalently linked to Cyanogen bromide (CNBr)-activated Sepharose™ beads (GE Healthcare Life-Sciences, Uppsala, Sweden). This procedure was performed at room temperature and all centrifugation steps were conducted without brake for 4 min at 300 rpm. 40 mg Sepharose beads were used per 1 mg of antibody. Sepharose beads were covered with 45 ml of 1 mM HCl (Carl Roth, Karlsruhe, Germany), inverted and rotated for 30 min. Centrifugation was performed and the supernatant was discarded. Coupling buffer (0.5 M NaCl (Merck Millipore, Billerica, USA) / 0.1 M $NaHCO_3$ (Merck Millipore), pH 8.3) and the respective antibody were added. After 120 min of rotation followed by centrifugation, the sepharose beads with coupled antibody were covered with 45 ml 0.2 mM glycine (Merck Millipore) for blocking of unspecific binding sites. Another period of rotation for 60 min and centrifugation followed before the sepharose beads were washed twice with phosphate-buffered saline (PBS; in-house produced by Claudia Falkenburger). The supernatant was discarded after centrifugation and PBS was added to the sepharose beads to reach an antibody concentration of 1 mg/ml.

The following lysate preparation was performed at 4°C. After determination of sample masses, solubilization buffer, which is comprised of 10 mM CHAPS (PanReac AppliChem, Darmstadt, Germany) and protease inhibitor (cOmplete; Roche, Basel, Switzerland) in PBS, was added to the tissues for chemical lysis. Tissues were sliced into small pieces and samples ≥ 100 mg were further homogenized through pottering (homogenizers/bench drilling machines: RB 18, Rotwerk, München, Germany; PBD 40, Robert Bosch GmbH, Stuttgart, Germany). Between two shaking periods at 300 rpm for 60 min each, samples were sonicated (Branson Sonifier 250, Emerson Industrial Automation, Danbury, USA) to assist cell membrane destruction. After centrifugation of the cell lysate at 4000 rpm for 45 min,

samples with an initial mass ≥ 100 mg were filtered using 5 µm sterile filters (Millex®-SV low protein binding PVDF Durapore® syringe filter unit, Merck Millipore). Subsequently, the chromatography setup was established. The peristaltic pumps (Rotarus smart 30, Hirschmann, Eberstadt, Germany) were set to 10 rpm. Two serially connected columns per sample were prepared with the sepharose-coupled antibodies, one for HLA class I precipitation with the mAb W6/32 (1 ml) and the other one for HLA class II precipitation with the mAbs Tü39 (0.5 ml) and L243 (0.5 ml) in combination **(Table 11)**. The antibody-containing columns were equilibrated with solubilization buffer for 1 h. Overnight, immunoaffinity chromatography was performed by circulating the sample lysates through the columns of the installation.

After washing the columns with PBS and ddH_2O on the next day, trifluoroacetic acid (TFA; Sigma-Aldrich, St. Louis, USA) was used to elute HLA molecules and ligands. Four elution steps were performed. In the first one, the samples were shaken with 10 µl 10% TFA and 100 µl 0.2% TFA for at least 30 min. In the following three elution steps, 100 µl of 0.2% TFA were added and the samples were shaken again for at least 15 min. Peptides were separated from the larger HLA molecules by ultrafiltration using 3 kDa and 10 kDa centrifugal filter units for HLA class I and class II ligands, respectively. Afterwards, possible hydrophobic peptides were dissolved from the filter using 400 µl AB_E buffer, which was comprised of 32.5% MS grade acetonitrile (AcN; J.T. Baker, Center Valley, USA) and 0.2% MS grade TFA in MS grade H_2O (J.T. Baker). By lyophilization (dry freezers: KF-2-110 cold trap, H. Saur Laborbedarf, Reutlingen, Germany; Sublimator VaCo 2, Zirbus technology, Bad Grund, Germany), samples were narrowed down to a volume of 30-50 µl. Using ZipTip C_{18} pipette tips (Merck Millipore), peptides were extracted and eluted in 35 µl AB_E buffer. To evaporate AcN, samples were vacuum-centrifuged (SpeedVac vacuum concentrator, BACHOFER, Reutlingen, Germany) until a volume of 5 µl was reached. Subsequently, samples were brought to a volume of 25 µl with loading buffer (1% MS grade AcN, 0.05% MS grade TFA in MS grade H_2O). The HLA ligand pool were either directly mass spectrometrically analyzed or stored at -80°C and thawed in an ultrasonic bath (Ultrasonic cleaner: JSP, Los Angeles, USA) for 2.5 min before analysis.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of HLA ligands

Isolated HLA ligands were separated by reversed-phase nanoflow ultra-high-performance liquid chromatography (nanoUHPLC, UltiMate 3000 RSLCnano, Dionex, Sunnyvale, USA) and, subsequently, analyzed in an on-line coupled LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, USA) by tandem mass spectrometry (MS/MS) applying data-dependent acquisition (DDA). Per sample, three replicates comprising 5 µl volume each were injected. Peptide trapping was performed in a 75 µm × 2 cm trapping column (Acclaim PepMap RSLC, Dionex) at a flow rate of 4 µl/min, followed by peptide separation in a 50 µm × 25 cm separation column (Acclaim PepMap RSLC, Dionex). By a

gradient ranging from 2.4% to 32.0% of AcN over the course of 90 min at a flow rate of 0.175 µl/min, peptides were eluted from the separation column. The positively charged peptides were ionized *via* electrospray ionization (ESI) and released into the mass spectrometer. Parallel MS1 and MS2 analyses were performed. Mass-to-charge (m/z) ratios of precursor ions were determined by orbitrap Fourier transformation (FT). Fragment ions were produced by top five collision-induced dissociation fragmentation (CID) and analyzed in the ion trap (IT). In **Table 12**, the described parameter settings are summarized as they were applied for the MS/MS analysis of HLA class I and class II ligands.

Table 12: Parameter settings applied for MS/MS analysis of HLA class I and class II ligands. The analysis was performed in an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). An AcN gradient from 2.4% to 32% was applied over a time course of 90 min. MS1 analysis of precursor ions was performed in the orbitrap, MS2 analysis of product ions was performed in the IT. Differences between HLA class I and class II ligand analysis consisted in the analyzed mass windows and the permitted charge states of precursor ions selected for fragmentation. AcN – acetonitrile; FT – Fourier transformation; CID – collision induced dissociation; IT – ion trap.

Parameter		HLA class I ligands	HLA class II ligands
Acquisition time		130 min	
Gradient	**Time**	90 min	
	AcN concentration	2.4% – 32%	
Mass window (m/z ratios)		400 – 650 m/z	300 – 1500 m/z
MS1	**Resolution**	60,000	
	Charge state	2+ and 3+	2+ to 5+
	Analyzer	Orbitrap FT	
MS2	**Fragmentation**	CID	
	Analyzer	IT	

Data processing

LC-MS/MS analysis of HLA ligands resulted in the generation of raw MS spectra that were processed with the Proteome Discoverer 1.4 software (Thermo Fisher Scientific). This entails database search based on the human proteome as deposited in the Swiss-Prot database (20,279 reviewed protein sequences; September 27th, 2013) without enzymatic restriction. For alignment of theoretical and experimental MS spectra, the SEQUEST HT search engine was used. Precursor mass tolerance was set to 5 ppm and fragment ion mass tolerance was set to 0.5 Da. Oxidation of the AA methionine (M+15.995 Da) was permitted as the only dynamic modification since this modification can occur during sample preparation. Technical replicates were both co-processed and individually processed to ensure technical accuracy. An overview of settings applied for data processing is shown by **Table 13**.

Table 13: Settings applied for data processing in the Proteome Discoverer 1.4 software (Thermo Fisher Scientific). Database search for spectral annotation was performed based on the Swiss-Prot database (20,279 reviewed protein sequences; September 27th, 2013).

Parameter	Setting for HLA class I and class II ligands
Search engine	SEQUEST HT
Precursor mass tolerance	5 ppm
Fragment mass tolerance	0.5 Da
Dynamic modification	Oxidation / +15.995 Da (methionine)
Maximum equal modifications / peptide	3

After processing, specific filter settings were applied to identify HLA ligands **(Table 14)**. A false discovery rate (FDR) of 5% was used, which was estimated by the Percolator algorithm 2.04 [4]. Protein inference was disabled to allow multiple protein annotations of peptides. To select HLA class I and class II ligands, different peptide lengths were permitted.

Table 14: Filters applied after processing for identification of HLA ligands. FDR estimation was assisted by the Percolator algorithm 2.04. The permitted peptide lengths differed between HLA class I and class II ligands. FDR – false discovery rate; AA – amino acid.

Parameter	HLA class I ligands	HLA class II ligands
Minimum peptide confidence	Medium (5% FDR)	
Peptide length	8 – 12 AAs	8 – 25 AAs
Search engine rank	1	

HLA class I allotype assignment

The HLA typing results were not available for three samples of the OPSCC cohort at the time of data analysis. Therefore, the in-house software Ligandosphere was used for HLA class I allotype assignment based on the detected HLA class I-presented peptides. The 2-digit allotype assignment was performed using a collection of allotypical peptides, which has previously been generated in-house by Michael Ghosh (Department of Immunology, Institute for Cell Biology, Eberhard Karls University Tübingen, Germany; NMI Natural and Medical Sciences Institute at the Eberhard Karls University Tübingen, Germany) for all HLA class I allotypes with a frequency above 5% in the German population. HLA ligands have been selected that occurred in more than 95% of tissues carrying a particular allotype and, simultaneously, occurred in less than 5% of tissues being negative for this allotype. Among these, the ten most frequent peptides have been defined as allotypical peptides. By means of the software Ligandosphere, lists of HLA class I-presented peptides were searched for allotypical peptides for 2-digit HLA allotyping. Assuming genetic heterozygosity, two allotypes were eventually assigned for HLA-A, -B and -C, respectively.

Appendix

HLA class I binder determination

Among all identified HLA class I-presented peptides, HLA class I binders were determined. The term HLA class I binders was used for those HLA class I-presented peptides that displayed a peptide binding motif of one of the HLA allotypes from the respective patient. To identify these binders, the in-house software Ligandosphere was used. Detected HLA class I-presented peptides were scored against the HLA allotypes of the corresponding patients. To determine HLA class I binders, the algorithm NetMHCpan-4.0 [5] was applied with a maximal NetMHC binding rank of 2% and a minimal SYFPEITHI score of 50% as binding affinity thresholds. The purity of a sample was calculated as the proportion of binders among the total number of HLA class I-presented peptides. All following HLA class I ligandomic analyses were performed on the basis of HLA class I binders.

Analysis of HLA ligandomic yields

To analyze the coverage rates of the ependymoma and OPSCC immunopeptidomes that were reached with the present cohorts, saturation analyses were performed. For this purpose, following in-house R scripts written by Dr. Linus Backert (Immatics Biotechnologies GmbH, Tübingen, Germany) were used:

- "saturation_analysis_peptides_Binders_ClassI.R" – for HLA class I binders
- "saturation_analysis_peptides_ClassII.R" – for HLA class II-presented peptides
- "saturation_analysis_proteins_ofBinders_ClassI.R" – for source proteins of HLA class I binders
- "saturation_analysis_proteins_ClassII.R" – for source proteins of HLA class II-presented peptides

Thereby, mean numbers of newly detectable antigens per sample were determined based on 1,000 random samplings. By exponential regression, assumed amounts of maximally detectable HLA ligands (saturation value) were estimated for both ependymomas and OPSCCs. This enabled the determination of the present coverage rates of the HLA ligandomes and allowed an estimation about the proportion that has not yet been detected. The coverage rates were calculated as the percentages of the reached ligand count from the saturation value.

To investigate a potential correlation between yields of isolated HLA-presented peptides and sample masses of tumors, the GraphPad Prism 6.1 software (GraphPad Software Inc) was used for statistical data analysis. Pearson´s correlation (two-tailed test of significance) was applied to both the ependymoma and the OPSCC dataset. The number of isolated HLA class I binders and HLA class II-presented peptides were correlated against the respective sample masses. A $p\ value \leq 0.05$ was defined as level of statistical significance.

Comparative profiling against in-house immunopeptidomic databases

During the search for tumor-associated antigens (TAAs) within the immunopeptidomes of the study cohorts, comparative profiling against an in-house benign reference database **(Table 15)** was performed to identify tumor-exclusive HLA-presented antigens. In case of OPSCCs, the benign reference was extended by the immunopeptidomic dataset of five tonsil samples. For comparative profiling against the database, following in-house R scripts were used written by Leon Bichmann (Institute for Cell Biology, Department of Immunology, Eberhard Karls University Tübingen, Germany; Applied Bioinformatics, Center for Bioinformatics and Department of Computer Science, Eberhard Karls University Tübingen, Germany):

- "sql_Sequence_benign_counter_classI_groupedtissue_exactMatch_case-insensitive.R" – for HLA class I binders
- "sql_Sequence_benign_counter_classII_groupedtissue_ExactMatch_case-insensitive.R" – for HLA class II-presented peptides
- "sql_protein_benign_counter_classI_groupedtissue.R" – for source proteins of HLA class I binders
- "sql_protein_benign_counter_classII_groupedtissue.R" – for source proteins of HLA class II-presented peptides

Appendix

Table 15: Overview of the in-house benign database. The database comprises the HLA class I ligandomic data of 419 benign samples derived from 33 different tissue types (TueDB_Class1_190425) and the HLA class II ligandomic data of 364 benign samples from 32 different tissues types (TueDB_Class2_190425).

Tissue type	Number of samples	
	HLA class I	HLA class II
Adrenal Gland	10	10
Aorta	10	10
Blood	98	85
Bone Marrow	27	23
Brain	12	12
Cell Line	1	11
Cerebellum	11	8
Colon	8	10
Esophagus	10	4
Gallbladder	4	11
Heart	11	13
Kidney	11	13
Liver	14	14
Lung	14	17
Lymph Node	17	5
Mamma	5	10
Muscle	10	2
Ovary	31	9
Pancreas	9	6
Prostate	18	12
Skin	11	6
Small Intestine	6	6
Small Intestine (Jejunum)	6	1
Spinal Cord	1	12
Spleen	12	8
Stomach	8	1
Stomach (Fundus)	2	4
Thyroid	13	13
Tongue	9	9
Trachea	7	7
Transplant Kidney	1	10
Urinary Bladder	10	0
Uterus	2	2
Total	**419**	**364**

An in-house malignant database **(Table 16)** was used to compare the immunopeptidomes of the study cohorts with other malignancies. For comparative profiling against the database, following in-house R scripts were used written by Leon Bichmann:

- "sql_sequence_counter_malignant_classI_exactMatch_case-insensitive.R" – for HLA class I binders
- "sql_sequence_counter_malignant_classII_exactMatch_case-insensitive.R" – for HLA class II-presented peptides
- "sql_protein_counter_malignant_classI.R" – for source proteins of HLA class I binders
- "sql_protein_counter_malignant_classII.R" – for source proteins of HLA class II-presented peptides

Appendix

Table 16: Overview of the in-house malignant database. The database comprises the HLA class I ligandomic data of 796 malignant samples from 38 different cancer entities (TueDB_Class1_190425) and the HLA class II ligandomic data of 583 malignant samples from 30 different cancer entities (TueDB_Class2_190425).

Cancer entity	Number of samples	
	HLA class I	HLA class II
Acute lymphocytic leukemia (ALL)	4	2
Acute myeloid leukemia (AML)	97	84
Anaplastic astrocytoma	1	1
Atypical teratoid rhabdoid tumor (ATRT)	10	0
Bladder cancer	3	3
Breast cancer	57	57
Chronic lymphocytic leukemia (CLL)	98	69
Chronic myeloid leukemia (CML)	21	20
Colon carcinoma	1	1
Colorectal cancer	36	1
Ependymoma	23	23
Esophageal cancer	3	0
Fibrosis	4	4
Gastric cancer	18	0
Gastrointestinal stromal tumor	1	0
Germ cell tumor	2	5
Glioblastoma	50	41
Hepatocellular carcinoma (HCC)	26	4
Leiomyosarcoma	1	0
Melanoma	1	1
Meningioma	33	33
Merkel-cell carcinoma	1	0
Multiple myeloma	13	13
Neuroblastoma	16	16
Non-small-cell lung carcinoma (NSCLC)	18	18
Oligodendroglioma	1	1
Oropharyngeal squamous cell carcinoma (OPSCC)	28	28
Osteosarcoma	3	3
Ovarian carcinoma	89	78
Pancreatic cancer	1	1
Paraganglioma	1	1
Parotid carcinoma	1	1
Polycythemia vera	13	11
Primary myelofibrosis	8	8
Prostate carcinoma	28	0
Renal cell carcinoma (RCC)	79	54
Sarcoma	1	1
Subependymoma	5	5
Total	**796**	**583**

Antigen evaluation

To select potential TAAs from the immunopeptidomes of ependymomas end OPSCCs, several evaluation steps on the level of both antigenic source proteins and HLA-presented peptides were applied to frequent tumor-exclusive antigens that were found in at least three samples of the respective cohort.

The RNA expression profiles of the source proteins in healthy tissues and different tumor entities were investigated. The expression data for healthy tissues were taken from the database Genotype-Tissue Expression Project (GTEx; www.gtexportal.org), which comprises transcriptomic data of 54 different non-diseased tissue types using the results from nearly 1000 individual donors [6]. For further consideration of an antigen, the source protein was required to exhibit a median RNA expression level ≤ 10 TPM (transcripts per million) in all healthy tissues except for testis. This aimed to avoid the selection of antigens that were mistakenly defined as tumor-exclusive due to a lack of sufficient and suitable reference data. Additionally, the Human Protein Atlas (www.proteinatlas.org) was consulted to observe the RNA expression profiles of source proteins in malignant tissues. This database comprises the transcriptomic data of 17 cancer types collected from 8000 different patients [7].

During processing of MS data, peptide sequences were annotated to experimental MS spectra. To control and verify this annotation, peptides of interest were evaluated. At first, multimapping peptides, whose source protein cannot be clearly defined, were excluded. Another crucial step is the manual inspection of the experimental MS spectra. Additionally, several technical characteristics of HLA-presented peptides were investigated. The term peptide spectrum matches (PSMs) describes the number of identified peptide spectra for a sequence. At least two PSMs were required for the inclusion of a peptide. A cross correlation (Xcorr) value ≥ 1.5 was used as threshold to ensure a sufficient goodness of fit between experimental and theoretical spectra. The delta correlation (ΔCn) compares the fit of the primary and the secondary sequence candidate to the experimental data and, thereby, is a measure of the specificity of the fit. ΔCn values ≥ 0.2 were permitted during peptide evaluation. Additionally, HLA class II-presented peptides consisting of 8 AAs were excluded since HLA class II ligands with this length are very uncommon. In case the identified peptides of interest met all described criteria, they were selected as potential TAAs.

Whole exome sequencing

For the tumor samples of the OPSCC project, whole exome sequencing (WES) was performed by Dr. Martin Bens (Leibniz-Institute on Aging, Fritz-Lipmann-Institute, Jena, Germany). 12 pairs of OPSCC tumors and corresponding PBMCs were available for DNA sequencing (sample IDs: 020, 022, 1003, 1008, 1014, 1021, 1118, 1209, 1345, 1349, 1359, 1361). The AllPrep DNA/RNA Mini Kit (Qiagen, Venlo, Netherlands) was used for DNA extraction. Sequencing of exome samples was performed applying the

Illumina's next-generation sequencing methodology [8]. Total DNA was quality checked using the 4200 TapeStation System (Genomic DNA ScreenTape; Agilent, Santa Clara, USA) and quantified using Quant-iT™ PicoGreen™ (Thermo Fisher Scientific). Preparation of libraries was performed on the basis of 3 µg input material using the exome SureSelect Human All Exon V6 (manufacturer's instructions; Agilent). Quantification and quality check of the libraries was performed using the 4200 TapeStation System (D1000 ScreenTape; Agilent). Libraries were pooled and sequenced on the NextSeq 500 System (High Output Flow Cell; Illumina, San Diego, USA) running in the 150 cycle mode (2x 75 bp paired-end). For converting sequence information into FASTQ format the conversion software bcl2fastq (v2.20.0.422; Illumina) was used. Alignment of reads was performed against the human reference genome hg19 by Dr. Axel Fürstberger (Institute of Medical Systems Biology, Ulm University, Ulm, Germany) using the BWA-MEM algorithm of the Burrow-Wheeler Aligner (BWA) software package. Duplicates were removed from the exome data and scores were recalibrated.

Search for neoantigens

On the basis of the WES data, variant calling was performed by Dr. Axel Fürstberger to identify somatic DNA mutations in the 12 OPSCC samples compared to the corresponding PBMCs (germline). The tumors were searched for short mutations including single nucleotide, insertion and deletion alterations using the variant calling tool Mutect2 (reference hg19 plus germline *versus* tumor). Detected variants were annotated *via* the bioinformatic software tool ANNOVAR (ANNOtate VARiation) to identify the somatic mutations.

Using the Python script „MapMutationsOnGenes.py" written by Leon Bichmann, files in FASTA format were generated for each individual tumor sample comprising the human proteome (Swiss-Prot database; 20,279 reviewed protein sequences; September 27th, 2013) in combination with the respective variants. This expanded the database search radius by the AA sequences of the mutated proteins encoded by the mutated genes. The resulting FASTA files were used in another processing of the MS raw data to search for HLA-presented mutations, termed neoantigens, in the 12 tumor samples.

RNA sequencing

RNA sequencing was performed by Dr. Martin Bens on 40 OPSCC samples of the present study cohort. Using the AllPrep DNA/RNA Mini Kit (Qiagen), total RNA was extracted and RNA sequencing was performed by Illumina's next-generation sequencing methodology [8]. Quantification and quality checking of total RNA was performed using the 2100 Bioanalyzer Instrument (RNA 6000 Pico; Agilent)., Libraries were prepared on the basis of 500 ng input material using TruSeq Stranded mRNA (manufacturer's instructions; Illumina). The 2100 Bioanalyzer Instrument (RNA 7500 kit; Agilent) was used for quantification and quality checking of the libraries. Subsequently, libraries were pooled and

sequenced in one lane on the HiSeq 2500 System (Illumina) running in 51 cycle/single-end/high output mode. The conversion of sequencing information into FASTQ format was performed using the software bcl2fastq (v2.20.0.422; Illumina). Alignment of reads was performed by Dr. Axel Fürstberger for 15 OPSCC samples (sample IDs: 020, 022, 1003, 1008, 1012, 1014, 1021, 1024, 1034, 1118, 1209, 1345, 1349, 1359, 1361). The bioinformatic tool Bowtie 2 was used for sequence alignment against the human reference genome hg19. Duplicates were removed from RNA sequencing data and scores were recalibrated. For read counting carried out on BAM files, the bioinformatic tool htseq-count was applied. By assignment of UniProt accessions to the identified genes, those genes were selected that are comprised in the Swiss-Prot database as used for MS spectra annotation. To integrate immunopeptidomic and transcriptomic data, the coverage of antigenic source proteins by expressed genes was examined in comparative analyses. Using normalized RPKM-counts, a potential correlation between counts of HLA-presented sequences per source protein and RPKM-counts of corresponding genes was examined. For this purpose, Pearson's correlation, with $P \geq 0.05$ as level of statistical significance, and linear regression analysis were applied using the GraphPad Prism 6.1 software (GraphPad Software Inc).

HPV status determination

The classification of the OPSCC samples according to their HPV status was carried out by Dr. Jaya Thomas (Cancer Sciences Unit, Faculty of Medicine, University of Southampton, UK) on the basis of RNA sequencing data. RNA reads for all 40 OPSCC samples were aligned against a human reference genome. RNA reads that were not aligned in this step were used for an alignment against sequences from the HPV genome. The read counts of all HPV genes (L1, L2, E1, E2, E4, E5, E6, E7) were quantified applying BLAST search. A positive HPV status was declared in case the sum of all HPV read counts was ≥ 1000. Thereby, the OPSCC samples were separated into HPV^+ and HPV^- tumor samples and the respective HPV types were annotated.

Search for HLA-presented HPV-derived peptides

The HLA ligandomic dataset of the OPSCC cohort was screened for HLA-presented peptides derived from HPV source proteins. The standard FASTA file used for database search and spectral annotation of the isolated HLA-presented peptides comprised the human proteome (Swiss-Prot database; 20,279 reviewed protein sequences; September 27th, 2013). This FASTA file was extended by the protein sequences of the HPV proteome (Swiss-Prot database, 470 reviewed protein sequences, January 5th, 2018). The added database of viral protein sequences comprised the proteomes of 76 different HPV types including the proteomes of the four oncogenic HPV types that were relevant for the present OPSCC cohort (HPV16, HPV35, HPV58 and HPV59). The generated combinatory

database was used as basis for another processing of the MS raw data. Thereby, the immunopeptidomes of the OPSCC samples were searched for HLA-presented HPV-derived peptides.

Comparative analysis of HPV$^+$ and HPV$^-$ OPSCCs

To reveal a potential clustering of the OPSCC samples according to their HPV status, PCAs were performed, for which an in-house R script ("PCA.R") written by Leon Bichmann was used. An unsupervised and a supervised PCA were carried out on the basis of the merged source proteins of HLA class I binders and HLA class II-presented peptides. For the supervised PCA, a corresponding heatmap was generated based on the source proteins with the highest impact (importance threshold ≥ 60) on a clustering into HPV$^+$ and HPV$^-$ tumors.

In a next step overlapping analyses and comparative profiling of HPV$^+$ against HPV$^-$ OPSCC samples were performed. These analyses were based on HLA class I binders, HLA class II-presented peptides and corresponding source proteins. Thereby, HPV$^+$- and HPV$^-$-exclusive antigens and their frequency among the respective tumor subgroup were revealed.

The differences between OPSCC tumors associated with their HPV status were further analyzed on the RNA-expression level. Differential expression analysis was performed for the 15 samples of which annotated RNA sequencing data were available. The Differential Expression Browser DEBrowser (v1.16.2) was used. Unnormalized read counts of expressed genes and meta data that give information about the HPV status of the samples were uploaded as data basis. For differential expression analysis, the R package DESeq2 [9] was applied. Differentially expressed genes were revealed and PCAs were performed based on either all identified genes or on genes significantly dysregulated in HPV$^+$ compared to HPV$^-$ samples. The dysregulated genes were further used as basis for volcano plotting and heatmap generation. Additionally, UniProtKB accessions were assigned to the up- and down-regulated genes and those comprised within the Swiss-Prot database were used for a comparison with HPV$^+$- and HPV$^-$-exclusive HLA-presented antigens.

Functional annotation clustering

To investigate the functional relations of antigenic source proteins, functional annotation clustering was performed. The functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11] was used for gene ontology (GO) term enrichment. The reference gene list Homo sapiens as deposited within DAVID was selected as background. The GO terms of biological processes (GOTERM_BP_DIRECT), cellular compartments (GOTERM_CC_DIRECT) and molecular functions (GOTERM_MF_DIRECT) were permitted. Classification stringency was set to medium. As the enrichment score of a gene group indicates its importance, only gene groups with an enrichment score ≥ 1 were considered in accordance with recommendations from Huang *et al.* (2009) [11]. Functional annotation clustering with the described parameters was applied for tumor-exclusive source proteins

of HLA class I binders and HLA class II-presented peptides from both the ependymoma and the OPSCC immunopeptidomes. This analysis was additionally performed based on the HPV^+- and HPV^--exclusive source proteins of HLA class I binders and HLA class II-presented peptides from OPSCCs. Furthermore, functional annotation clustering was carried out on the basis of proteins encoded by dysregulated genes in dependence of the OPSCC HPV status.

Software and Databases

The analysis of the immunopeptidomic data from ependymomas as well as of the immunopeptidomic, exmomic and transcriptomic data from OPSCCs required the application of different software and several bioinformatic tools **(Table 17)**. Additionally, various databases were consulted **(Table 18)**.

Table 17: Overview of software and bioinformatic tools used for the present work.

Software/Tool	Reference/Company
BioVenn	Hulsen *et al.* (2008), [12]
DAVID Bioinformatics Resources 6.8	Huang *et al.* (2009), [10, 11]
GraphPad Prism 6.0	GraphPad Software Inc, San Diego, USA
jVenn	Bardou *et al.* (2014), [13]
Ligandosphere	in-house software (Dr. Linus Backert)
Percolator 2.04 (2012)	Käll *et al.* (2007), [4]
Proteome Discoverer 1.4.1.14	Thermo Fisher Scientific, Waltham, USA
RStudio	RStudio Inc, Boston, USA
Scintilla and SciTE	Open source software
Xcalibur 4.0.27.10 (2015	Thermo Fisher Scientific, Waltham (USA)
SQLite 3.27.2	Open source tool
DB Browser for SQLite (DB4S)	Open source tool
DEBrowser (v1.16.2)	Open source tool

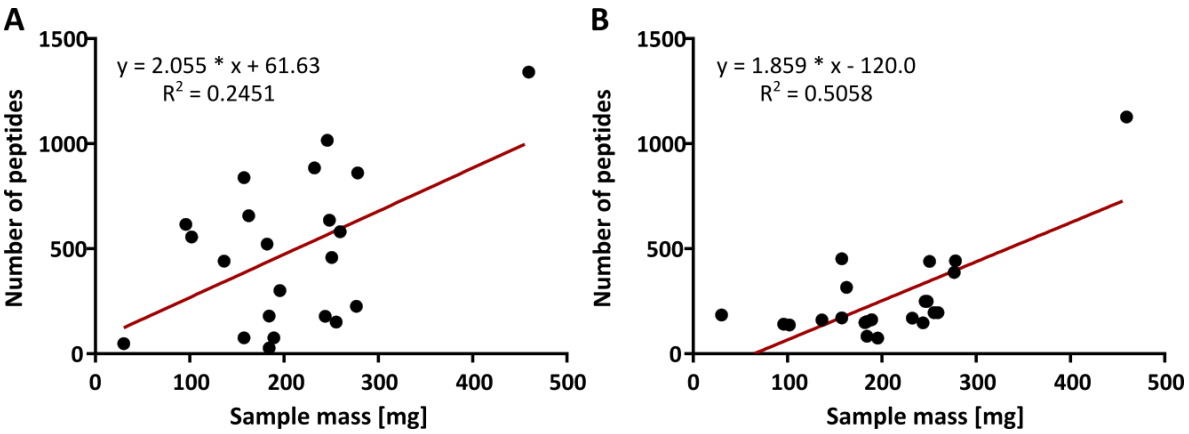
Appendix

Table 18: Overview of databases consulted for the present work.

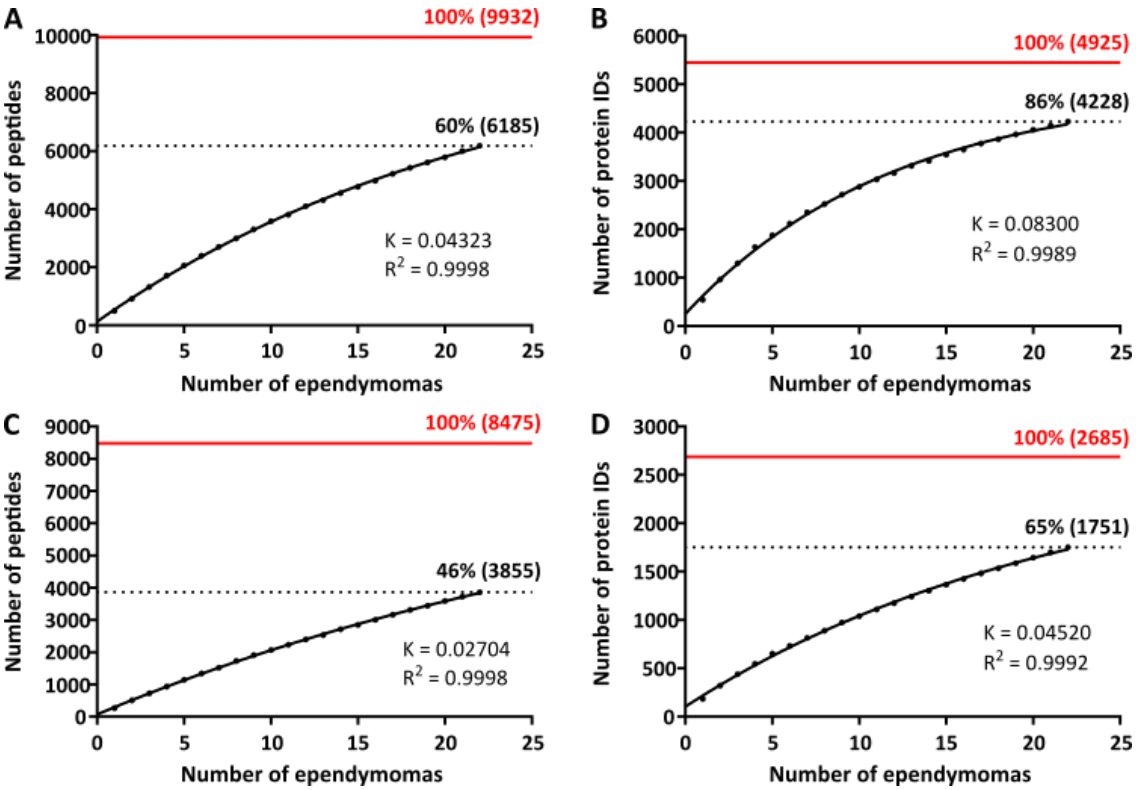
Database	Reference	Website
Allele Frequency Net Database	Gonzalez-Galarza *et al.* (2020), [2]	www.allelefrequencies.net
SYFPEITHI	Rammensee *et al.* (1999) [14]	www.syfpeithi.de
NetMHC 4.0	Andretta *et al.* (2016), [15]	www.cbs.dtu.dk/services/NetMHC/
NetMHC pan 4.0	Jurtz *et al.* (2017), [5]	www.cbs.dtu.dk/services/NetMHC/
UniProt (Swiss-Prot database)	UniProt Consortium (2007), [16]	www.uniprot.org
IPD-IMGT/HLA database	Robinson (2020), [17]	www.ebi.ac.uk/ipd/imgt/hla/
MHC Motif Viewer	Rapin *et al.* (2010), [18]	www.cbs.dtu.dk/biotools/MHCMotifViewer
TueDB_Class1_190425	in-house	-
TueDB_Class2_190425	in-house	-
CTdatabase	Almeida *et al.* (2009), [19]	www.cta.lncc.br
GTEx	GTEx Consortium (2013), [6]	www.gtexportal.org
The Human Protein Atlas	Thul and Lindskog (2018), [7]	www.proteinatlas.org

8.2 Supplementary figures

8.2.1 Supplementary figures of chapter 2 – Ependymomas

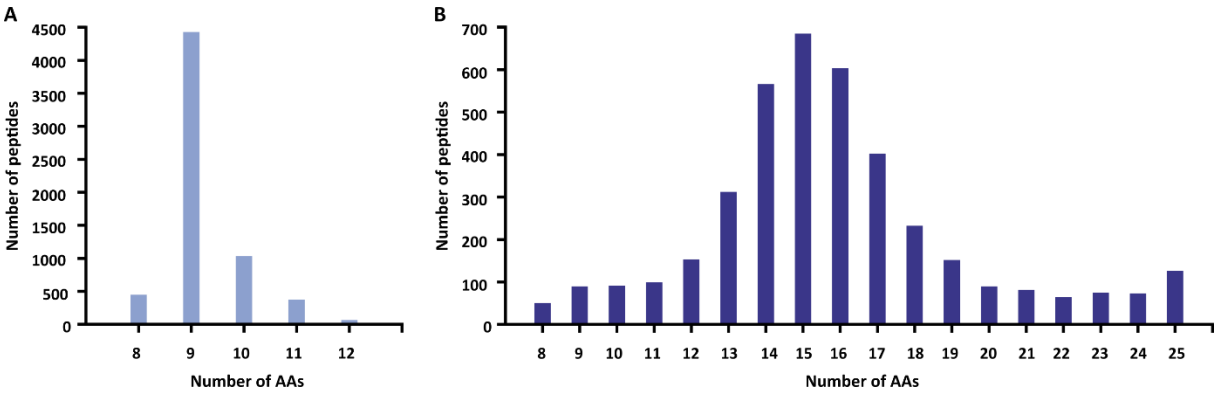


Supplementary Figure 1: Linear regression analysis of ependymoma sample masses and yields of HLA ligands. The graphs of the linear regression analysis display that larger tumor samples enabled the detection of more unique (A) HLA class I binders and (B) HLA class II-presented peptides. Linear regression was applied since the increase in the number of peptides was still in the linear phase with the present number of sample and not yet in the saturation phase. n(ependymomas) = 22; R^2 = coefficient of determination (goodness of fit).

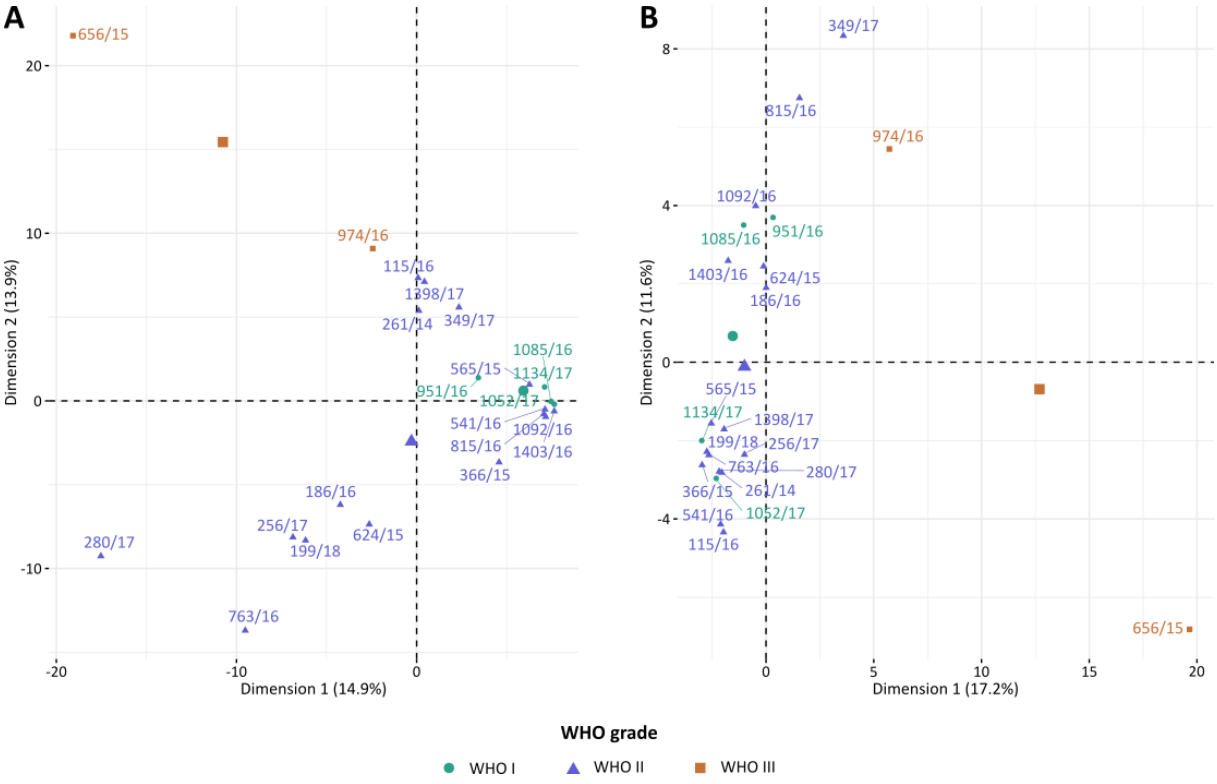


Supplementary Figure 2: Saturation analysis of HLA ligands from ependymomas. On the basis of additionally identified unique ligands per sample, the saturation values were estimated. The present coverage rates that were achieved with the immunopeptidomic analysis of ependymomas were determined for (A) HLA class I binders, (B) source proteins of HLA class I binders, (C) HLA class II-presented peptides and (D) source proteins of HLA class II-presented peptides. n(ependymomas) = 22; red – saturation value; K = dissociation constant; R^2 = coefficient of determination (goodness of fit).

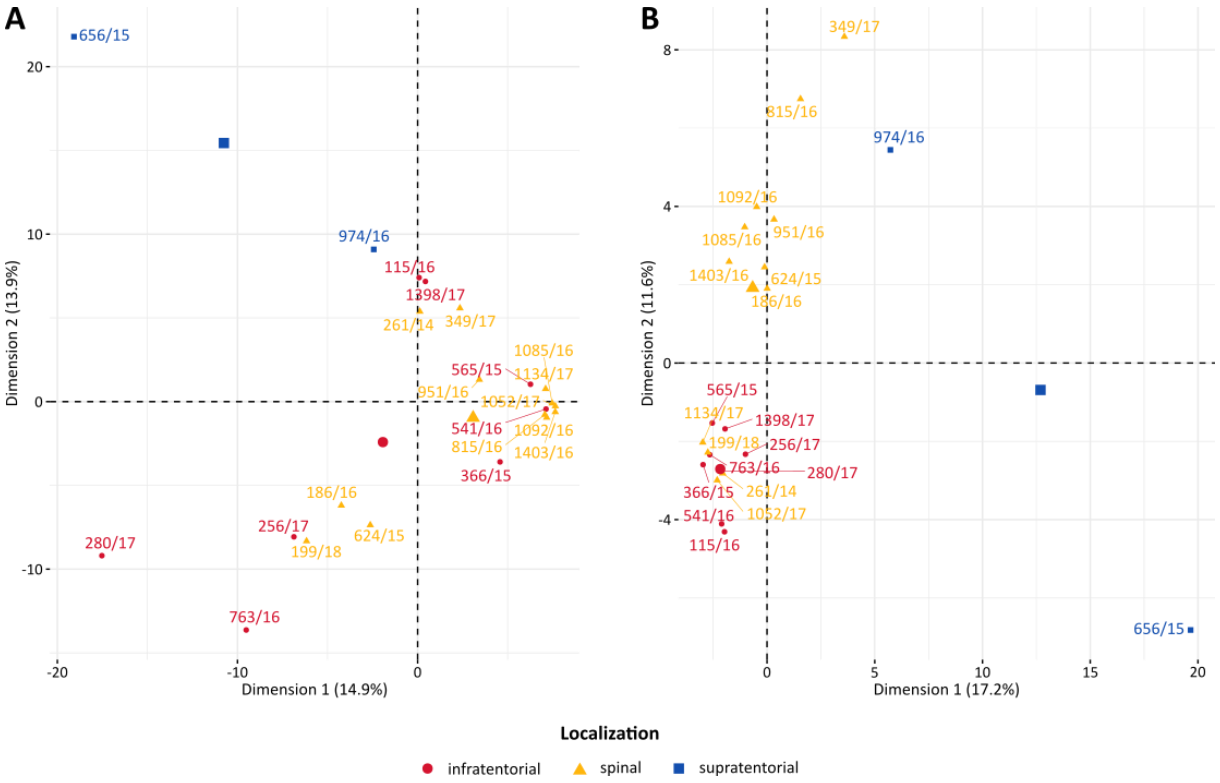
Appendix



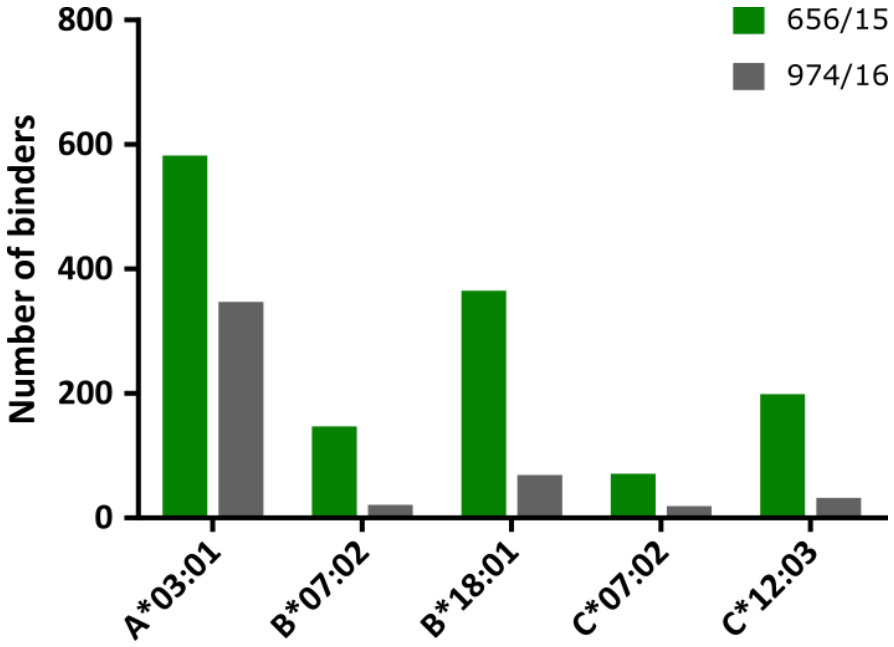
Supplementary Figure 3: Length distribution of HLA-presented peptides isolated from ependymomas. (A) The majority of HLA class I binders (71%) consisted of nine AAs. (B) HLA class II-presented peptides peaked with 18% at a length of 15 AAs. n(ependymomas) = 22; AA – amino acid.



Supplementary Figure 4: PCAs based on immunopeptidomic data of ependymomas categorized according to WHO classification. The source proteins of (A) HLA class I binders and (B) HLA class II-presented peptides served as basis for these unsupervised PCAs. No clear clustering according to the ependymoma subgroups was revealed. The PCAs were performed by Leon Bichmann. n(WHO I) = 4; n(WHO II) = 16; n(WHO III) = 2; PCA – principle component analysis.

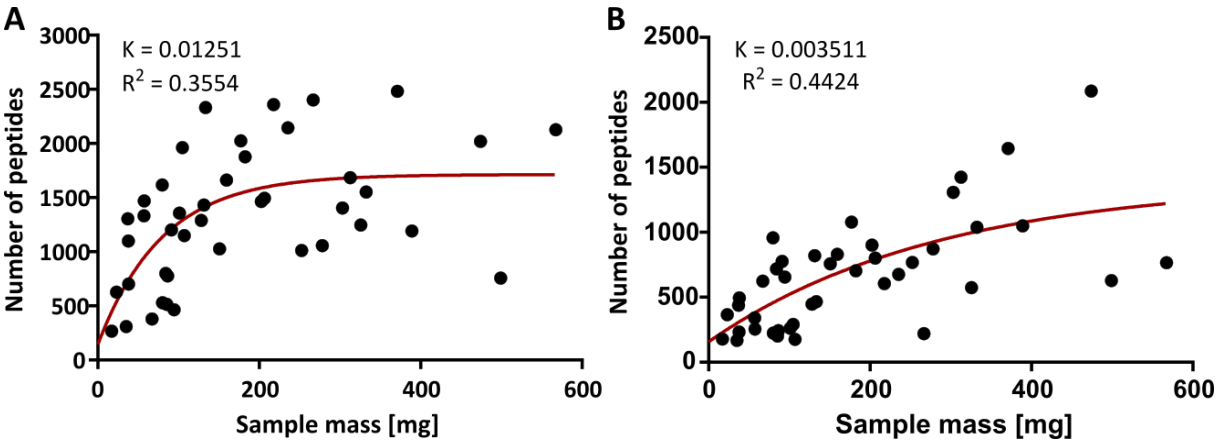


Supplementary Figure 5: PCAs based on immunopeptidomic data of ependymomas categorized according to tumor location. The source proteins of (A) HLA class I binders and (B) HLA class II-presented peptides served as basis for these unsupervised PCAs. No clear clustering according to the ependymoma subgroups was revealed. The PCAs were performed by Leon Bichmann. n(spinal) = 12; n(infratentorial) = 8; n(supratentorial) = 2; PCA – principle component analysis.

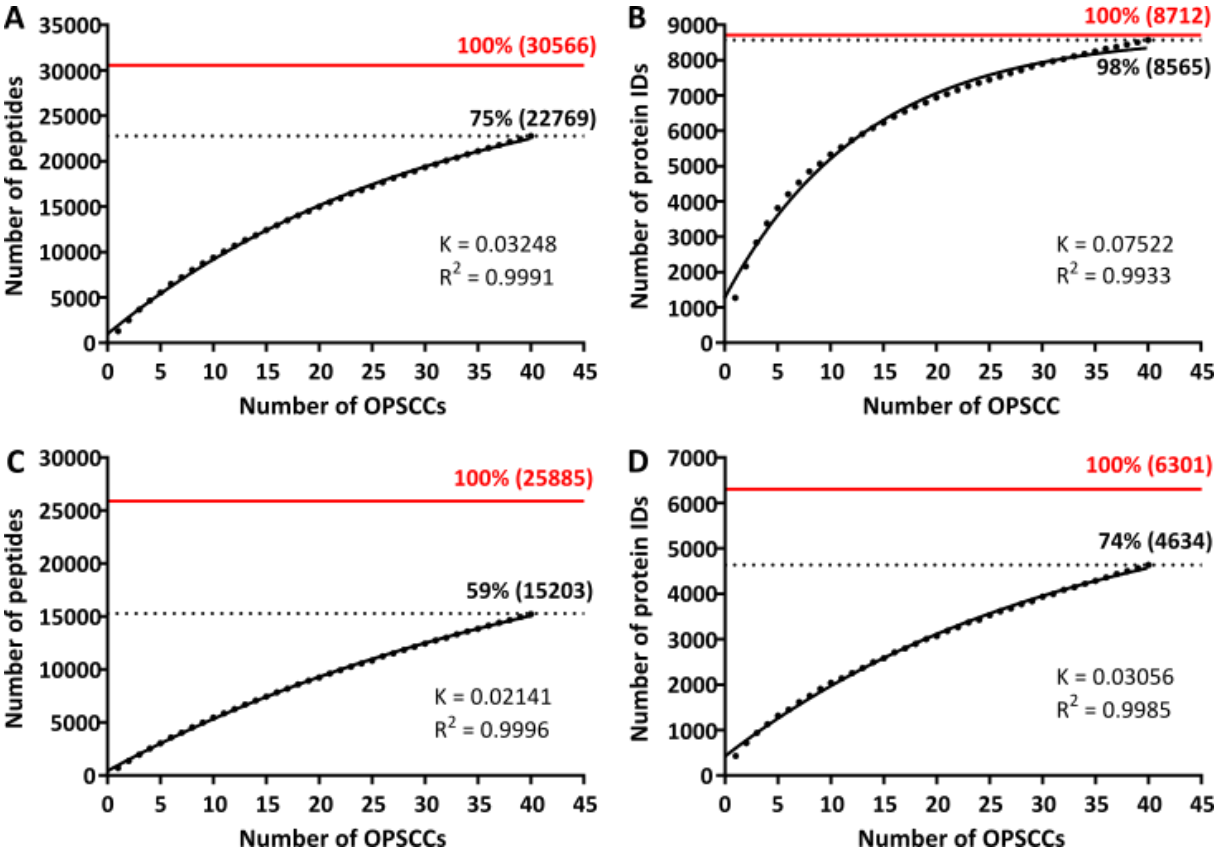


Supplementary Figure 6: Distribution of binders among HLA class I allotypes in two ependymomas from the same patient. Both 656/15 and 974/16 were recurrent tumors originating from one patient. The total number of identified HLA class I binders was lower in the second (974/16) than in the first (656/15) recurrence. This decrease occured across binders of all five HLA class I allotypes of the patient. The loss of binders was slightly higher for HLA-B*07:02, -B*18:01, -C*07:02 and -C*12:03 than for HLA-A*03:01.

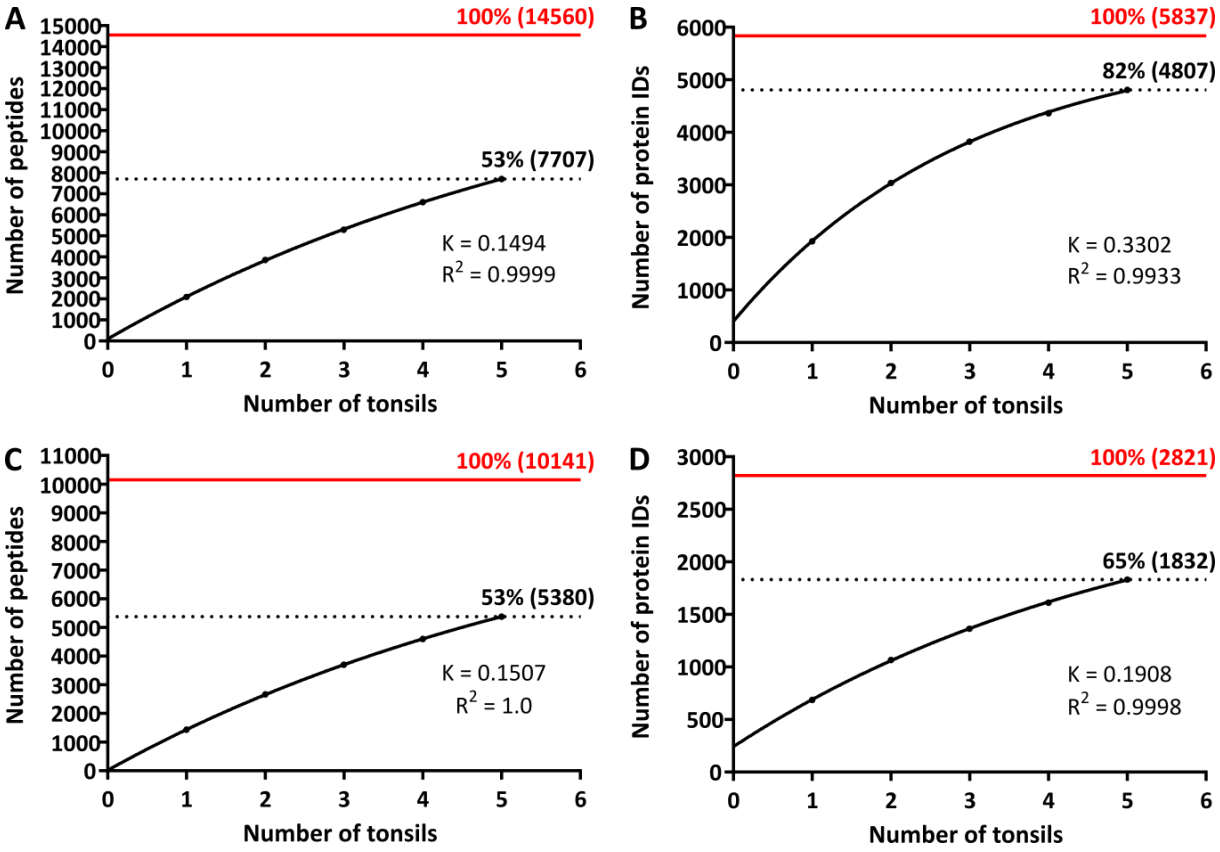
8.2.2 Supplementary figures of chapter 3 – OPSCCs



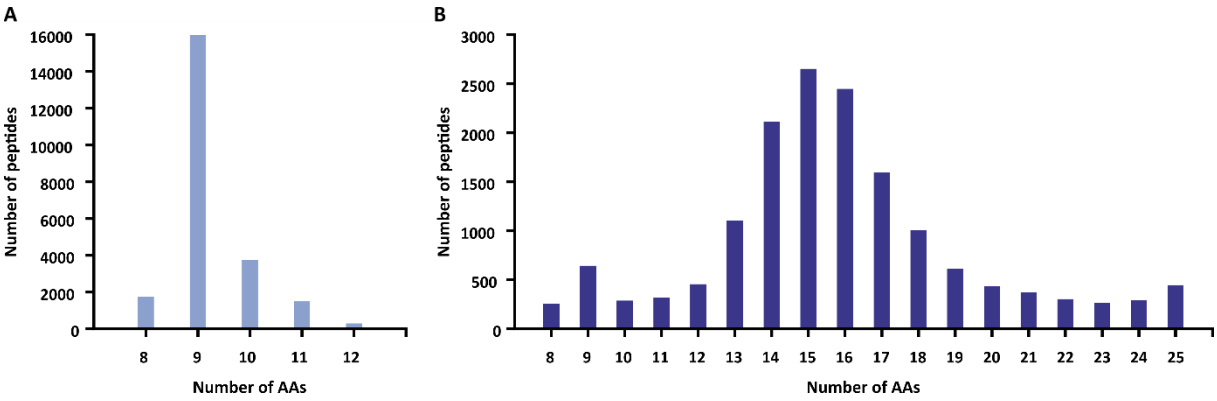
Supplementary Figure 7: Non-linear regression analysis of OPSCC sample masses and yields of HLA ligands. The graphs of the non-linear regression analysis display that larger tumor samples enabled the detection of more unique (A) HLA class I binders and (B) HLA class II-presented peptides. Non-linear regression analysis was applied since the increase in the number of peptides approaches the saturation phase. n(OPSCCs) = 40; R^2 = coefficient of determination (goodness of fit).



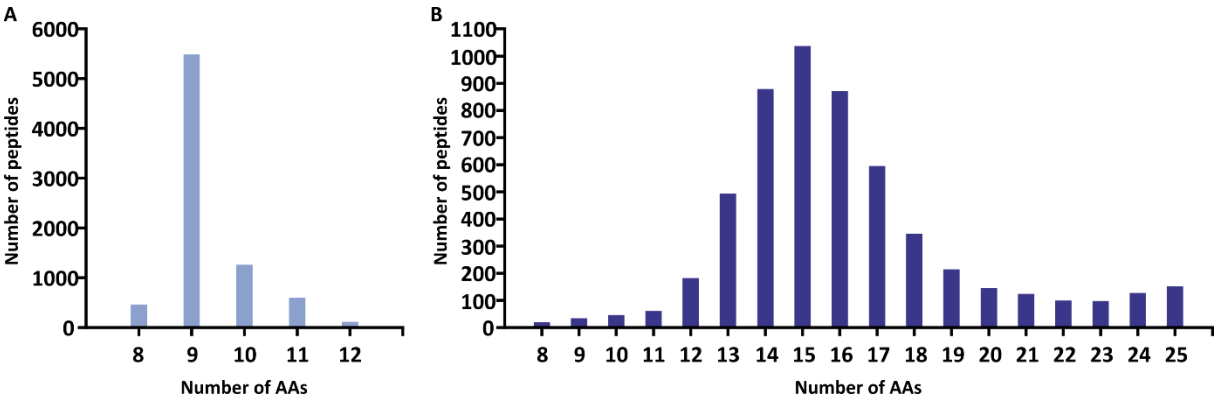
Supplementary Figure 8: Saturation analysis of HLA ligands from OPSCCs. On the basis of additionally identified unique ligands per sample, the saturation values were estimated. The present coverage rates that were achieved with the immunopeptidomic analysis of OPSCCs were determined for (A) HLA class I binders, (B) source proteins of HLA class I binders, (C) HLA class II-presented peptides and (D) source proteins of HLA class II-presented peptides. n(OPSCCs) = 40; red – saturation value; K = dissociation constant; R^2 = coefficient of determination (goodness of fit).



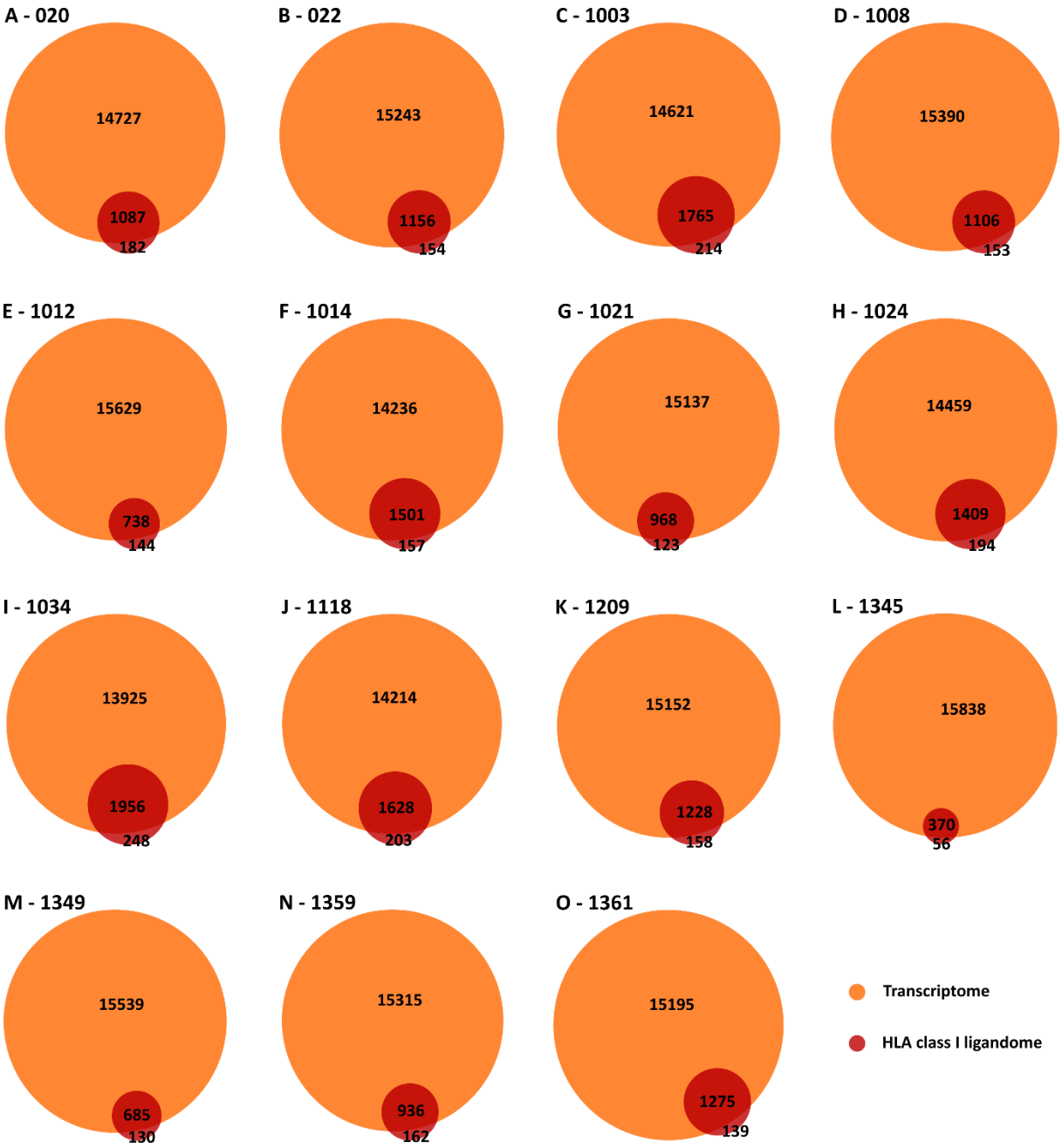
Supplementary Figure 9: Saturation analysis of HLA ligands from tonsils. On the basis of additionally identified unique ligands per sample, the saturation values were estimated. The present coverage rates that were achieved with the immunopeptidomic analysis of tonsils were determined for (A) HLA class I binders, (B) source proteins of HLA class I binders, (C) HLA class II-presented peptides and (D) source proteins of HLA class II-presented peptides. n(tonsils) = 5; red – saturation value; K = dissociation constant; R^2 = coefficient of determination (goodness of fit).



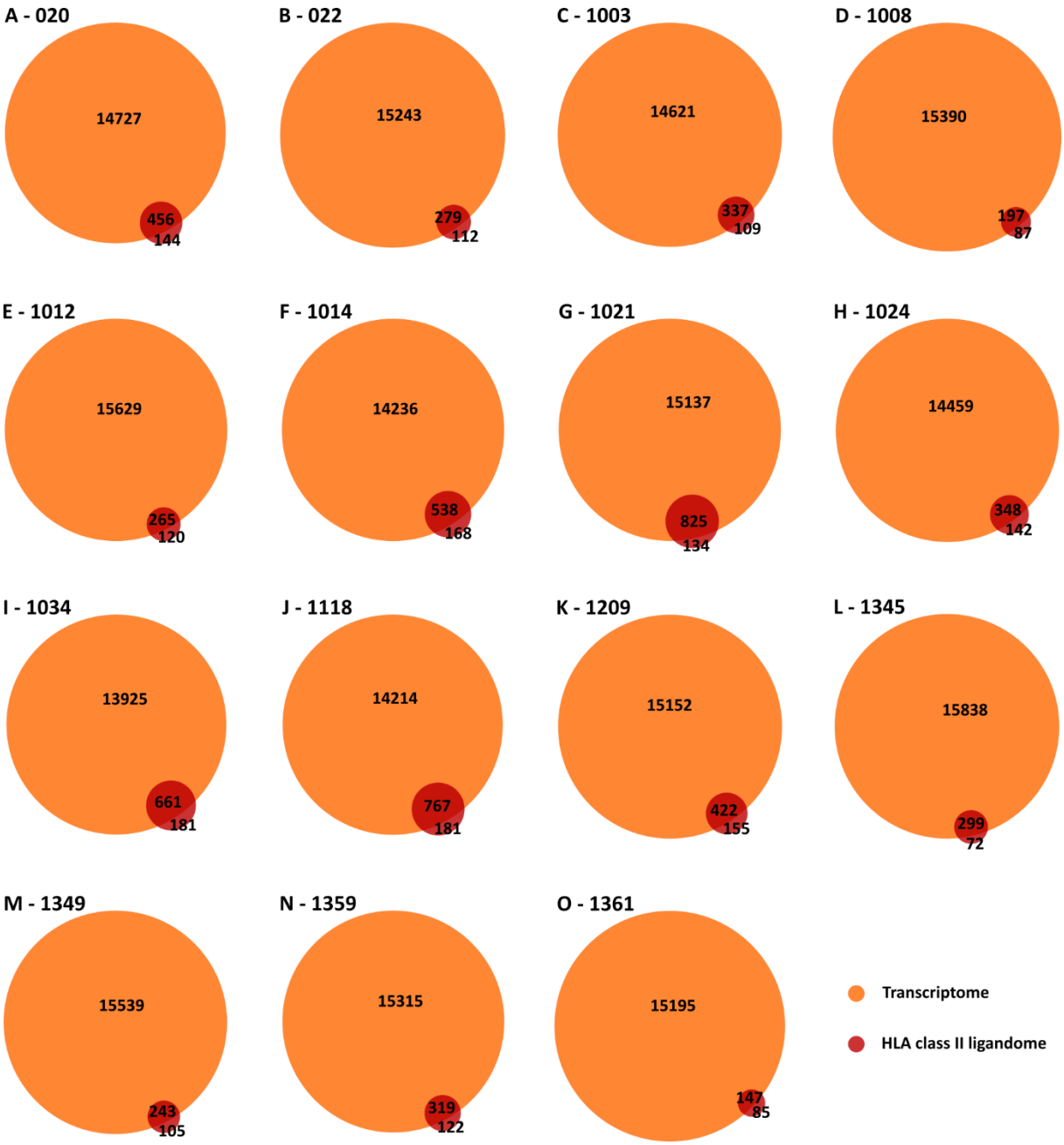
Supplementary Figure 10: Length distribution of HLA-presented peptides isolated from OPSCCs. (A) The majority of HLA class I binders (70%) consisted of nine AAs. (B) HLA class II-presented peptides peaked with 17% at a length of 15 AAs. n(OPSCCs) = 40; AA – amino acid.



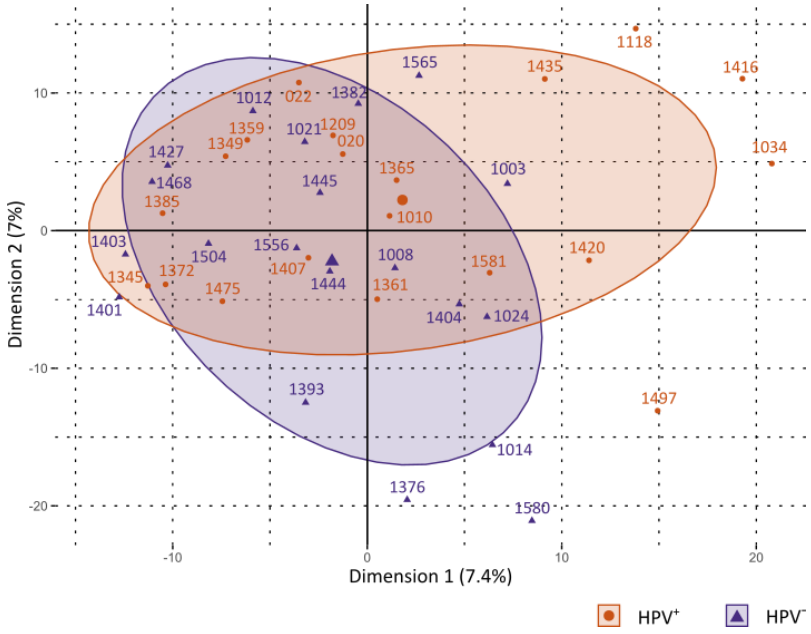
Supplementary Figure 11: Length distribution of HLA-presented peptides isolated from tonsils. (A) The majority of HLA class I binders (71%) consisted of nine AAs. (B) HLA class II-presented peptides peaked with 19% at a length of 15 AAs. n(tonsils) = 5; AA – amino acid.



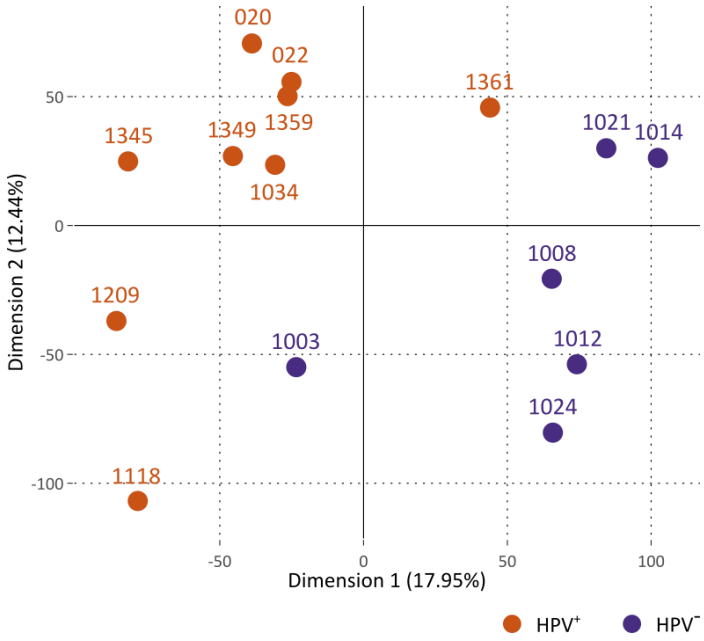
Supplementary Figure 12: Comparison between the RNA transcriptome and HLA class I ligandome of OPSCCs. Genes coding for proteins that are comprised within the Swiss-Prot database were used for comparative analysis with the antigenic source proteins of the immunopeptidomes. (A) – (O) show the coverage of the source proteins of HLA class I binders among the expressed genes for 15 different samples of the OPSCC cohort. The coverage rates ranged from 84% to 91% for the individual samples.



Supplementary Figure 13: Comparison between the RNA transcriptome and HLA class II ligandome of OPSCCs. Those genes coding for proteins that are comprised within the Swiss-Prot database were used for comparative analysis with the antigenic source proteins of the immunopeptidome. (A) – (O) show the coverage of the source proteins of HLA class II-presented peptides among the expressed genes for 15 different samples of the OPSCC cohort. The coverage rates ranged from 63% to 86% for the individual samples.



Supplementary Figure 14: Unsupervised PCA on the basis of the OPSCC immuno-peptidome. Merged source proteins of identified HLA class I binders and HLA class II-presented peptides were used as data basis. No clear clustering into HPV⁺ and HPV⁻ tumors was revealed. n(HPV⁺) = 20; n(HPV⁻) = 20.



Supplementary Figure 15: PCA generated after differential expression analysis of HPV⁺ and HPV⁻ OPSCCs. The transcriptomic data were available for 15 OPSCC samples. Differential expression analysis was performed by the DESeq2 algorithm on the basis of unnormalized read counts. The expression data of all mapped genes were used for this PCA. n(HPV⁺) = 9; n(HPV⁻) = 6.

8.3 Supplementary tables

8.3.1 Supplementary tables of chapter 2 – Ependymomas

Supplementary Table 1: Overview of the ependymoma cohort and patient characteristics. 22 tumor samples from 21 different patients were included in the study. The tumors 656/15 and 974/16 originated from the same patient.

Sample number	Sample ID	Sex	Age at surgery	WHO grade	Location	Diagnosis	HLA class I typing	HLA class II typing
1	**256/17**	f	52	II	infratentorial	initial	A*02:01;A*25:01 B*14:01;B*18:01 C*08:02;C*12:03	DRB1*07:01;DRB1*15:01 DQB1*02:02;DQB1*06:02
2	**815/16**	m	17	II	spinal	initial	A*02:01;A*03:01 B*38:01;B*49:01 C*12:03;C*06:02	DRB1*11:04;DRB1*16:01 DQB1*03:01;DQB1*05:02
3	**1092/16**	f	42	II	spinal	initial	A*02:01;A*03:01 B*44:02;B*51:01 C*05:01;C*15:02	DRB1*01:01;DRB*14:01 DQB1*05:01;DQB1*05:03
4	**1085/16**	m	57	I	spinal	initial	A*02:01;A*32:01 B*07:02;B*40:01 C*07:02;C*12:03	DRB1*01:01;DRB1*15:01 DQB1*05:01;DQB1*06:02
5	**951/16**	f	33	I	spinal	initial	A*02:01;A*11:01 B*15:01;B*51:01 C*03:03;C*15:02	DRB1*04:01;DRB*12:01 DQB1*03:01
6	**280/17**	m	77	II	infratentorial	initial	A*03:01;A*26:01 B*35:01;B*38:01 C*04:01;C*12:03	DRB1*13:01 DQB1*06:03
7	**349/17**	m	38	II	spinal	initial	A*01:01 B*08:01;B*35:02 C*07:01;C*04:01	DRB1*04:01;DRB1*11:04 DQB1*03:02;DQB1*03:01

Supplementary Table 1 continued.

Sample number	Sample ID	Sex	Age at surgery	WHO grade	Location	Diagnosis	HLA class I typing	HLA class II typing
8	**186/16**	f	54	II	spinal	initial	A*11:01;A*26:01 B*44:02;B*52:01 C*05:01;C*12:02	DRB1*13:01;DRB1*15:02 DQB1*06:03;DQB1*06:01
9	**115/16**	m	22	II	infratentorial	initial	A*01:01;A*03:01 B*15:01;B*08:01 C*03:03;C*07:01	DRB1*13:01;DRB1*15:02 DQB1*06:03;DQB1*06:01
10	**261/14**	m	30	II	spinal	initial	A*03:01;A*24:02 B*40:02;B*51:01 C*02:02;C*01:02	DRB1*11:01;DRB1*13:01 DQB1*03:01;DQB1*06:03
11	**624/15**	f	45	II	spinal	initial	A*02:01;A*26:01 B*15:17;B*44:05 C*07:01;C*02:02	DRB1*13:02;DRB1*16:01 DQB1*06:04;DQB1*05:01
12	**366/15**	f	67	II	infratentorial	initial	A*02:01;A*26:01 B*44:03;B*08:01 C*04:01;C*07:01	DRB1*03:01;DRB1*07:01 DQB1*02:01;DQB1*02:02
13	**565/15**	m	71	II	infratentorial	initial	A*02:01;A*23:01 B*41:01;B*57:01 C*17:01;C*06:02	DRB1*04:01;DRB1*07:01 DQB1*03:02;DQB1*03:03
14	**199/18**	f	26	II	spinal	initial	A*11:01;A*26:01 B*27:05;B*38:01 C*02:02;C*12:03	DRB1*04:04;DRB1*14:01 DQB1*03:02;DQB1*05:03
15	**1398/17**	m	22	II	infratentorial	initial	A*24:03;A*30:04 B*14:02;B*18:01 C*08:02;C*07:01	DRB1*11:04 DQB1*03:01
16	**1134/17**	f	30	I	spinal	initial	A*01:01;A*29:02 B*37:01;B*44:03 C*06:02;C*16:01	DRB1*07:01;DRB1*10:01 DQB1*02:02;DQB1*05:01

Supplementary Table 1 continued.

Sample number	Sample ID	Sex	Age at surgery	WHO grade	Location	Diagnosis	HLA class I typing	HLA class II typing
17	**1052/17**	m	52	I	spinal	initial	A*02:01;A*30:02 B*18:01;B*39:06 C*12:03;C*05:01	DRB1*03:01;DRB1*16:01 DQB1*02:01;DQB1*05:02
18	**541/16**	m	33	II	infratentorial	initial	A*02:01 B*18:01B*44:02 C*05:01	DRB1*03:01;DRB1*15:01 DQB1*02:01;DQB1*06:02
19	**1403/16**	f	47	II	spinal	initial	A*02:01;A*11:01 B*15:01;B*35:01 C*03:04;C*04:10	DRB1*04:01;DRB1*07:01 DQB1*03:02;DQB1*02:02
20	**763/16**	f	51	II	infratentorial	recurrence	A*02:01;A*26:01 B*38:01;B*58:01	DRB1*07:01;DRB1*13:01 DQB1*03:03;DQB1*06:03
21	**656/15**	m	41	III	supratentorial	recurrence	A*03:01 B*07:02;B*18:01 C*07:02;C*12:03	DRB1*04:03;DRB1*16:01 DQB1*03:05;DQB1*05:02
22	**974/16**		43	III	supratentorial	recurrence		

Supplementary Table 2: Allelic distribution of HLA-A alleles in ependymoma patients and the German population. Chi-square test was applied and a p value ≤ 0.004 was used as level of statistical significance after Bonferroni correction. The Allele Frequency Net Database was consulted for German allele frequencies [2]. Statistical results from alleles with a patient frequency of F = 0% (zero hits) or F = 2.564% (one hit) were assumed as unreliable and were not announced. CI – confidence interval; * – significant; ns – not significant.

HLA-A allele	Allele frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% CI for odds ratio
01:01	15.100%	7.692%	0.1795	ns	-	-
02:01	26.700%	30.769%	0.6372	ns	-	-
03:01	14.700%	15.385%	0.6172	ns	-	-
11:01	5.600%	10.256%	0.2263	ns	-	-
23:01	2.320%	2.564%	-	-	-	-
24:02	9.500%	2.564%	-	-	-	-
24:03	0.255%	2.564%	-	-	-	-
25:01	2.400%	2.564%	-	-	-	-
26:01	**3.600%**	**15.385%**	**> 0.0001**	*****	**4.868**	**2.038 to 11.630**
29:02	2.400%	2.564%	-	-	-	-
30:02	0.547%	2.564%	-	-	-	-
30:04	0.163%	2.564%	-	-	-	-
32:01	3.600%	2.564%	-	-	-	-

Supplementary Table 3: Allelic distribution of HLA-B alleles in ependymoma patients and the German population. Chi-square test was applied and a p value ≤ 0.002 was used as level of statistical significance after Bonferroni correction. The Allele Frequency Net Database was consulted for German allele frequencies [2]. Statistical results from alleles with a patient frequency of F = 0% (zero hits) or F = 2.381% (one hit) were assumed as unreliable and were not announced. CI – confidence interval; * – significant; ns – not significant.

HLA-B allele	Allele Frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% CI for odds ratio
07:02	12.000%	4.762%	0.2903	ns	-	-
08:01	9.500%	7.143%	0.5441	ns	-	-
14:01	0.485%	2.381%	-	-	-	-
14:02	2.000%	2.381%	-	-	-	-
15:01	6.500%	7.143%	0.9319	ns	-	-
15:17	0.487%	2.381%	0.0895	ns	-	-
18:01	5.000%	11.905%	0.0086	ns	-	-
27:05	3.400%	2.381%	-	-	-	-
35:01	6.200%	4.762%	0.6492	ns	-	-
35:02	1.000%	2.381%	-	-	-	-
37:01	1.300%	2.381%	-	-	-	-
38:01	2.500%	9.524%	0.0051	ns	-	-
39:06	0.472%	2.381%	-	-	-	-
40:01	4.900%	2.381%	-	-	-	-
40:02	1.600%	2.381%	-	-	-	-
41:01	0.542%	2.381%	-	-	-	-
44:02	7.300%	7.143%	0.9021	ns	-	-
44:03	4.100%	4.762%	0.8814	ns	-	-
44:05	0.623%	2.381%	-	-	-	-
49:01	1.600%	2.381%	-	-	-	-

Appendix

Supplementary Table 3 continued.

HLA-B allele	Allele Frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% Confidence Interval
51:01	6.300%	7.143%	0.8876	ns	-	-
52:01	1.100%	2.381%	-	-	-	-
57:01	3.100%	2.381%	-	-	-	-
58:01	0.904%	2.381%	-	-	-	-

Supplementary Table 4: Allelic distribution of HLA-C alleles in ependymoma patients and the German population. Chi-square test was applied and a p value ≤ 0.003 was used as level of statistical significance after Bonferroni correction. The Allele Frequency Net Database was consulted for German allele frequencies [2]. Statistical results from alleles with a patient frequency of F = 0% (zero hits) or F = 2.564% (one hit) were assumed as unreliable and were not announced. CI – confidence interval; * – significant; ns – not significant; n.a. – not available.

HLA-C allele	Allele frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% CI of odds ratio
01:02	3.600%	2.564%	-	-	-	-
02:02	5.400%	7.692%	0.5871	ns	-	-
03:03	5.100%	5.128%	0.9486	ns	-	-
03:04	7.400%	2.564%	-	-	-	-
04:01	12.500%	7.692%	0.3157	ns	-	-
04:10	n.a.	2.564%	-	-	-	-
05:01	6.400%	10.256%	0.3801	ns	-	-
06:02	10.000%	7.692%	0.5669	ns	-	-
07:01	14.700%	12.821%	0.6506	ns	-	-
07:02	13.400%	5.128%	0.2529	ns	-	-
08:02	2.400%	5.128%	0.3001	ns	-	-
12:02	1.100%	2.564%	-	-	-	-
12:03	**6.300%**	**17.949%**	**0.0028**	*****	**3.253**	**1.435 to 7.376**
15:02	2.600%	5.128%	0.3596	ns	-	-
16:01	2.200%	2.564%	-	-	-	-
17:01	1.000%	2.564%	-	-	-	-

Supplementary Table 5: Overview of HLA ligand yields from ependymomas. HLA ligands were analyzed by LC-MS/MS. The numbers of HLA ligands are depicted for 22 individual ependymoma tissue samples. The purity was calculated by the percentage of HLA class I binders among the total number of HLA class I-presented peptides.

Sample number	Sample ID	Tissue mass [mg]	HLA class I					HLA class II	
			Presented peptides	Source proteins	Binders	Source proteins of binders	Purity [%]	Presented peptides	Source proteins
1	**256/17**	248	735	793	636	700	87	249	169
2	**815/16**	30	61	114	48	66	79	184	274
3	**1092/16**	184	189	263	179	202	95	153	149
4	**1085/16**	189	78	259	76	97	97	162	145
5	**951/16**	137	505	684	441	508	87	160	177
6	**280/17**	246	1129	1113	1015	1033	90	249	116
7	**349/17**	158	872	1017	838	871	96	452	337
8	**186/16**	102	626	672	555	620	89	137	133
9	**115/16**	163	683	782	657	757	96	316	187
10	**261/14**	278	912	947	861	910	94	442	244
11	**624/15**	260	610	661	581	637	95	195	160
12	**366/15**	244	185	207	178	196	96	147	78
13	**565/15**	196	317	374	300	356	95	74	56
14	**199/18**	182	557	615	522	584	94	148	82
15	**1398/17**	96	652	803	616	767	95	140	94
16	**1134/17**	255	165	192	151	178	92	196	102
17	**1052/17**	158	84	92	76	84	91	171	115
18	**541/16**	277	247	276	225	255	91	387	242
19	**1403/16**	184	29	39	27	37	93	83	106
20	**763/16**	232	956	937	884	865	93	169	93
21	**656/15**	251	1424	1451	1340	1384	94	1127	579
22	**974/16**	460	481	699	458	562	95	439	288

Supplementary Table 6: CTAs identified in the ependymoma immunopeptidome. The CTdatabase [19] was used as basis for the identification of CTAs. CTA – cancer-testis antigen; red – tumor-exclusive.

HLA class	Peptide				Source protein			
	Sequence	HLA restriction	Sample	Representation frequency	UniProtKB AC	Gene name	Protein name	Representation frequency
I	ALLDKLYAL	A*02:01	366-15 541-16 565-15 624-15 763-16 951-16 1092-16	32%	Q9NV31	IMP3	U3 small nucleolar ribonucleoprotein IMP3	36%
	RSMEDFVTW	B*58:01	763-16	5%				
	EDYTRYNQL	B*08:02;C*14:02	1398-17	5%				
	DVIRALAKY	A*25:01;A*26:01	186-16 256-17 280-17 624-15 763-16 199-18	27%	Q5TZF3	ANKRD45	Ankyrin repeat domain-containing protein 45	27%
	DPFAFIHKI	B*51:01	951-16 1092-16 261-14	14%	Q9UGL1	KDM5B	Lysine-specific demethylase 5B	23%
	ILNPYNLFL	A*02:01	541-16	5%				
	SVAQQLLNGK	A*03:01	656-15	5%				
	DDWDNRTSY	B*18:01	656-15	5%				
	KLPDFSWEL	A*02:01	541-16 199-18	9%	Q5W041	ARMC3	Armadillo repeat-containing protein 3	14%
	IINDGFYDY	A*01:01	1134-17	5%				
	SEFKAMDSF	B*40:02	261-14	5%	P26232	CTNNA2	Catenin alpha-2	9%
	NEQDLANRF	B*44:05	624-15	5%				

Supplementary Table 6 continued.

HLA class	Peptide				Source protein			
	Sequence	HLA restriction	Sample	Representation frequency	UniProtKB AC	Gene name	Protein name	Representation frequency
I	VEFPYQYDF	B*18:01	656-15	5%	Q14667	KIAA0100	UPF0378 protein KIAA0100	9%
	DYPRYLFEI	A*24:02	261-14	5%				
	RLFVTSGGLKK	A*03:01	115-16	5%	O75602	SPAG6	Sperm-associated antigen 6	9%
	YPEEIVRYY	B*35:01	280-17	5%				
	SLIQKVETY	A*02:01	541-16	5%	**Q8TBZ0**	**CCDC110**	**Coiled-coil domain-containing protein 110**	5%
	DAVKFFVAV	B*51:01	261-14	5%	O60271	SPAG9	C-Jun-amino-terminal kinase-interacting protein 4	5%
	SFYEHIITV	B*52:01	186-16	5%	Q5T6F0	DCAF12	DDB1- and CUL4-associated factor 12	5%
II	DPFAFIHKI	-	1085-16 1092-16	9%	Q9UGL1	KDM5B	Lysine-specific demethylase 5B	9%
	PFHIFKVKVTTERERMENIDSTIL	-	366-15	5%	Q86X24	HORMAD1	HORMA domain-containing protein 1	5%

Appendix

Supplementary Table 7: Tumor-exclusive HLA class I binders of the ependymoma immunopeptidome. Peptides with a frequency ≥ 10% (≥ 3 samples) within the ependymoma cohort are depicted. AC – accession; green – selected TAAs.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
EIIDNSQGFY	A*25:01;A*26:01	186/16 256/17 280/17 624/15 763/16 199/18	27%	Q9Y617
EVAPPGLITNF	A*25:01;A*26:01	186/16 256/17 280/17 624/15 763/16 199/18	27%	Q13304
EVIEKTSYL	A*25:01;A*26:01	186/16 256/17 280/17 366/15 763/16 199/18	27%	Q9P1Z9
EVISYYSQY	A*25:01;A*26:01	186/16 256/17 280/17 624/15 763/16 199/18	27%	Q8N485
EVLNGQVSKY	A*25:01;A*26:01	186/16 256/17 280/17 624/15 763/16 199/18	27%	Q9C0G6
EVTERLGEF	A*25:01;A*26:01	186/16 256/17 280/17 624/15 763/16 199/18	27%	Q9C0G6
FLDSQITTV	A*02:01	256/17 541/16 624/15 763/16 1052/17 1092/16	27%	Q6V702
YQDLLNVKL	B*38:01;C*04:01;C*08:02	256/17 280/17 763/16 815/16 1398/17 199/18	27%	P14136

Supplementary Table 7 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
DTAGQELAWLY	A*25:01;A*26:01	256/17 280/17 624/15 763/16 199/18	23%	Q149M9
DVFGAQSPFGGY	A*25:01;A*26:01	186/16 280/17 624/15 763/16 199/18	23%	P05091
EIRTQYEAM	A*25:01;A*26:01;B*08:01	115/16 256/17 280/17 349/17 763/16	23%	P14136
ETVDENGRLY	A*25:01;A*26:01	256/17 280/17 624/15 763/16 199/18	23%	Q8IYE1
EVFDGTVIREL	A*25:01;A*26:01	186/16 280/17 366/15 763/16 199/18	23%	Q8N1V2
EVFGDIMIDY	A*25:01;A*26:01	256/17 280/17 624/15 763/16 199/18	23%	P48449
EVFGRSEQY	A*25:01;A*26:01	256/17 280/17 366/15 763/16 199/18	23%	O95409
DLYTGEFLSF	A*25:01;A*26:01	256/17 624/15 763/16 199/18	18%	O75129
ETSDWNSYY	A*25:01;A*26:01	186/16 624/15 763/16 199/18	18%	P55317
EVVSKIAQY	A*25:01;A*26:01	186/16 256/17 280/17 763/16	18%	P26367

Appendix

Supplementary Table 7 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
NEFIITQH	B*18:01	256/17 656/15 1052/17 1398/17	18%	P23470;P23471
TQDEFHPFI	B*38:01;C*04:01;C*08:02	256/17 280/17 763/16 199/18	18%	Q12857;O00712; P08651;Q14938
WLFDHPAQV	A*02:01	624/15 763/16 951/16 1092/16	18%	Q9NYC9
DHSQAFLQAI	B*14:01;B*38:01	256/17 280/17 763/16	14%	O95714;Q9BVR0
DLPDDLIGYF	A*25:01;A*26:01	256/17 624/15 763/16	14%	Q15036
DTSSVKSEY	A*25:01;A*26:01	256/17 280/17 763/16	14%	Q5SYE7
DVYNNSDLFY	A*26:01	624/15 763/16 199/18	14%	Q9BXC9
EIIQHIQLY	A*26:01	186/16 624/15 199/18	14%	Q49AJ0
EIMPGLIQGY	A*25:01;A*26:01	256/17 763/16 199/18	14%	O75122
ELMEFEIAL	A*02:01	541/16 624/15 763/16	14%	O75800
ETAMETLINVF	A*25:01;A*26:01	256/17 280/17 763/16	14%	P23297
EVIGLKGMGRY	A*26:01	186/16 280/17 763/16	14%	O94812
EVTDEDTVKRY	A*25:01;A*26:01	256/17 280/17 763/16	14%	Q9UBH6
EVVQRVSLL	A*25:01;A*26:01	256/17 280/17 763/16	14%	Q504Y0
EVYDVSISEF	A*25:01;A*26:01	256/17 280/17 763/16	14%	Q8TDW7

Supplementary Table 7 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
NVVDFTTRY	A*26:01	186/16 280/17 763/16	14%	O43424
SKDPGENYNL	C*08:02;B*38:01	256/17 280/17 763/16	14%	P15289
SMIDDLLGV	A*02:01	256/17 624/15 763/16	14%	Q96M63
SSYGGTLRY	B*15:17;C*12:03	624/15 656/15 974/16	14%	O15230

Supplementary Table 8: Tumor-exclusive HLA class II-presented peptides of the ependymoma immunopeptidome. Peptides with a frequency ≥ 10% (≥ 3 samples) within the ependymoma cohort are depicted. AC – accession; green – selected TAAs.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
GSFTLSLHEDELFTLTTLTTG	115/16 186/16 256/17 261/14 366/15 624/15 199/18 1398/17 1052/17 541/16	45%	P54803
RMPPPLPTRVDFSLAGALN	115/16 261/14 280/17 763/16 1134/17 1052/17	27%	P14136
GSFTLSLHEDELFTLTTLTT	115/16 261/14 199/18 1052/17	18%	P54803
LEIINEDDVEAYVGLR	115/16 366/15 1052/17 541/16	18%	Q16620
SIKTDSTFSGFLVYSDWHS	256/17 261/14 1092/16 199/18	18%	Q9BXJ0

Supplementary Table 8 continued.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
DKWIQKAPMTTVRGLHCM	115/16 261/14 541/16	14%	Q9P2N7
EIVLTQSPATLSLSPGER	1085/16 1092/16 1403/16	14%	P04433
GVQQYVPEEWAEYPRPI	656/15 951/16 1398/17	14%	O60883
LPTRVDFSLAGALN	115/16 280/17 763/16	14%	P14136
NEYYIADASEDQVFV	349/17 951/16 1403/16	14%	Q92673
SSVPGVRLLQDSVDFSLAD	261/14 1134/17 1052/17	14%	P08670

8.3.2 Supplementary tables of chapter 3 – OPSCCs

Supplementary Table 9: Overview of the OPSCC cohort and patient characteristics. 40 tumor samples from 22 HPV^{+} and 18 $^{-}$ patients were included in the study. red – assigned HLA class I allotypes.

Sample number	Sample ID	Sex	Age at diagnosis	HPV status	Tumor subsite	HLA class I typing	HLA class II typing
1	**020**	m	60	+ (HPV16)	tonsil	A*03:01;A*66:01 B*41:02;B*51:01 C*15:02;C*17:01	DRB1*04:02;DRB1*13:03 DQB1*03:01;DQB1*03:02
2	**022**	m	54	+ (HPV16)	tonsil	A*02:01;A*11:01 B*50:01;B*51:01 C*04:01;C*06:02	DRB1*04:07;DRB1*07:01 DQB1*02:02;DQB1*03:01
3	**1010**	f	59	+ (HPV16)	tonsil	A*02:01;A*24:02 B*50:01;B*51:01 C*01:02;C*06:02	DRB1*07:01;DRB1*11:01 DRB3*02:02 DRB4*01:03 DQB1*02:02;DQB1*03:01 DPB1*03:01;DPB1*04:01
4	**1034**	m	58	+ (HPV16)	tonsil	A*31:01;A*32:01 B*15:18;B*40:01 C*03:04;C*07:04	DRB1*07:01;DRB1*13:02 DQB1*02:02;DQB1*06:04
5	**1118**	m	63	+ (HPV16)	tonsil	A*02:01;A*02:01 B*07:02;B*40:01 C*03:04;C*07:02	DRB1*04:01;DRB1*11:03 DRB3*02:02 DRB4*01:03 DQB1*03:01;DQB1*03:01 DPB1*03:01;DPB1*04:02
6	**1209**	m	64	+ (HPV16)	tonsil	A*03:01;A*11:01 B*07:02;B*52:01 C*07:02;C*12:02	DRB1*15:01;DRB1*15:02 DQB1*06:01;DQB1*06:02 DPB1*04:01;DPB1*04:01
7	**1345**	m	61	+ (HPV16)	tonsil	A*01:01;A*01:01 B*07:02;B*07:02 C*07:02;C*07:02	DRB1*01:01;DRB1*15:01 DRB5*01:01:01G DQB1*05:01;DQB1*06:02 DPB1*02:01;DPB1*04:01
8	**1349**	m	63	+ (HPV58)	tonsil	A*02:01;A*11:01 B*44:02;B*51:01 C*04:01;C*05:01	DRB1*04:04;DRB1*08:01 DRB4*01:01;DRB4*01:03 DQB1*03:02;DQB1*04:02 DPB1*05:01;DPB1*20:01
9	**1359**	m	57	+ (HPV16)	tonsil	A*02:01;A*03:01 B*35:01;B*51:01 C*02:02;C*04:01	DRB1*01:01;DRB1*13:01 DRB3*01:01 DQB1*05:01;DQB1*06:03 DPB1*04:02;DPB1*05:01
10	**1361**	m	69	+ (HPV16)	tonsil	A*11:01;A*24:02 B*49:01;B*57:01 C*06:02;C*07:01	DRB1*08:01;DRB1*12:01 DRB3*02:11 DQB1*03:01;DQB1*04:02 DPB1*02:01;DPB1*13:01
11	**1365**	m	57	+ (HPV16)	tonsil	A*02:01;A*30:01 B*49:01;B*57:01 C*03:04;C*07:01	DRB1*04:023;DRB1*07:01 DRB4*01:01;DRB4*01:03 DQB1*03:02;DQB1*03:03 DPB1*02:01;DPB1*04:01

Appendix

Supplementary Table 9 continued.

Sample number	Sample ID	Sex	Age at diagnosis	HPV status	Tumor subsite	HLA class I typing	HLA class II typing
12	**1372**	m	53	+ (HPV16)	tonsil	A*01:01;A*03:01 B*07:02;B*14:02 C*07:02;C*08:02	DRB1*01:02;DRB1*04:01 DRB4*01:03 DQB1*03:02;DQB1*05:01 DPB1*03:01;DPB1*04:01
13	**1385**	m	41	+ (HPV16)	tonsil	A*03:01;A*31:01 B*35:01;B*51:01 C*04:01;C*15:02	DRB1*04:07;DRB1*10:01 DRB4*01:01 DQB1*03:01;DQB1*05:01 DPB1*03:01;DPB1*04:01
14	**1393**	m	68	+ (HPV16)	tonsil	A*01:01;A*24:03 B*08:01;B*35:01 C*04:01;C*07:01	DRB1*01:01;DRB1*12:01 DRB3*01:01 DQB1*03:01;DQB1*05:01 DPB1*02:01;DPB1*04:01
15	**1403**	m	70	- (HPV35)	tongue	A*01:01;A*32:01 B*35:03;B*44:03 C*04:01;C*04:01	DRB1*07:01;DRB1*11:01 DRB3*02:02 DRB4*01:01 DQB1*02:02;DQB1*03:01 DPB1*02:01;DPB1*03:01
16	**1407**	m	68	+ (HPV16)	tonsil	A*03:01;A*24:02 B*15:01;B*51:01 C*03:04;C*14:02	DRB1*04:01;DRB1*11:04 DRB3:02:02 DRB4*01:03 DQB1*03:01;DQB1*03:02 DPB1*02:01;DPB1*03:01
17	**1416**	m	76	+ (HPV16)	tonsil	A*02:01;A*03:01 B*27:05;B*44:02 C*02:02;C*07:04	DRB1*11:01;DRB1*13:01 DRB3*02:02 DQB1*03:01;DQB1*06:03 DPB1*04:01;DPB1*19:01
18	**1420**	m	53	+ (HPV16)	tonsil	A*23:01;A*26:01 B*35:03;B*44:03 C*04:01;C*04:01	DRB1*04:01;DRB1*12:01 DRB3*02:02 DRB4*01:03 DQB1*03:02;DQB1*03:12 DPB1*02:01;DPB1*04:01
19	**1445**	m	59	+ (HPV59)	tonsil	A*03;A*26 B*07;B*40 C*01;C*03	-
20	**1475**	m	72	+ (HPV16)	tonsil	A*11:01;A*24:02 B*07:02;B*27:05 C*02:02;C*07:02	DRB1*04:08;DRB1*09:01 DRB4*01:03 DQB1*03:01;DQB1*03:03 DPB1*03:01;DPB1*04:01
21	**1497**	m	54	+ (HPV16)	tonsil	A*01:01;A*23:01 B*40:01;B*49:01 C*03:04;C*07:01	DRB1*10:01;DRB1*15:01 DRB5*01:01 DQB1*05:01;DQB1*06:02 DPB1*02:01;DPB1*06:01
22	**1581**	m	38	+ (HPV16)	base of tongue	A*01:01;A*02:01 B*08:01;B*57:01 C*06:02;C*07:01	DRB1*03:01;DRB1*13:03 DRB3*01:01 DQB1*02:01;DQB1*03:01 DPB1*01:01;DPB1*04:01

Supplementary Table 9 continued.

Sample number	Sample ID	Sex	Age at diagnosis	HPV status	Tumor subsite	HLA class I typing	HLA class II typing
23	**1003**	f	54	-	tonsil	A*01:01;A*02:01 B*51:01;B*57:01 C*04:01;C*06:02	DRB1*08:01;DRB1*15:01 DRB5*01:01 DQB1*04:02;DQB1*06:02 DPB1*03:01;DPB1*04:01
24	**1008**	m	52	-	tonsil	A*01:01;A*02:01 B*08:01;B*40:01 C*03:04;C*07:01	DRB1*03:01;DRB1*13:02 DRB3*01:01;DRB3*03:01 DQB1*02:01;DQB1*06:04 DPB1*01:01;DPB1*04:01
25	**1012**	f	52	-	base of tongue	A*02;A*03 B*07;B*44 C*05;C*07	-
26	**1014**	m	56	-	tonsil	A*01:01;A*01:01 B*08:01;B*08:01 C*07:01;C*07:01	DRB1*01:01;DRB1*03:01 DRB3*01:01 DQB1*02:01;DQB1*05:01 DPB1*04:01;DPB1*04:01
27	**1021**	m	60	-	tonsil	A*02:01;A*11:01 B*44:02;B*51:01 C*04:01;C*05:01	DRB1*04:04;DRB1*08:01 DRB4*01:03;DRB4*01:01 DQB1*03:02;DQB1*04:02 DPB1*05:01;DPB1*20:01
28	**1024**	m	70	-	tonsil	A*01;A*29 B*08;B*44 C*07;C*16	-
29	**1376**	f	54	-	tonsil	A*01:01;A*24:02 B*08:01;B*14:02 C*07:01;C*08:02	DRB1*01:02;DRB1*03:01 DRB3*01:01 DQB1*02:01;DQB1*05:01 DPB1*01:01;DPB1*03:01
30	**1382**	m	49	-	lateral tongue	A*02:01;A*02:01 B*07:02;B*15:01 C*03:44;C*07:02	DRB1*01:01;DRB1*15:01 DQB1*05:01;DQB1*06:02
31	**1401**	m	56	-	soft palate	A*01:01;A*02:01 B*08:01;B*40:01 C*03:04;C*07:01	DRB1*03:01;DRB1*11:01 DRB3*02:02 DQB1*02:01;DQB1*03:01 DPB1*04:01;DPB1*04:01
32	**1404**	f	53	-	tonsil	A*24:02;A*31:01 B*07:02;B*44:02 C*05:01;C*07:02	DRB1*04:01;DRB1*04:04 DRB4*01:03 DQB1*03:01;DQB1*03:02 DPB1*04:01;DPB1*06:01
33	**1427**	m	54	-	tonsil	A*02:01;A*02:01 B*08:01;B*13:02 C*06:02;C*07:01	DRB1*03:01;DRB1*07:01 DRB3*01:01 DRB4*01:03 DQB1*02:01;DQB1*02:02 DPB1*04:01;DPB1*04:02
34	**1435**	f	58	-	tonsil	A*02:01;A*68:01 B*44:03;B*51:01 C*04:01;C*14:02	DRB1*07:01;DRB1*13:01 DRB3*01:01 DRB4*01:01 DQB1*02:02;DQB1*06:03

Appendix

Supplementary Table 9 continued.

Sample number	Sample ID	Sex	Age at diagnosis	HPV status	Tumor subsite	HLA class I typing	HLA class II typing
35	**1444**	m	54	-	tonsil	A*02:01;A*24:02 B*15:01;B*35:02 C*01:02;C*04:01	DRB1*07:01;DRB1*09:01 DRB4*01:01;DRB4*01:03 DQB1*02:02;DQB1*03:03 DPB1*02:01;DPB1*03:01
36	**1468**	m	77	-	tonsil	A*02:01;A*02:01 B*15:01;B*51:01 C*03:04;C*07:01	DRB1*01:01;DRB1*11:03 DRB3*02:02 DQB1*03:01;DQB1*05:01 DPB1*02:01;DPB1*10:01
37	**1504**	m	58	-	tonsil	A*02:01;A*24:02 B*07:02;B*27:02 C*02:02;C*07:02	DRB1*15:01;DRB1*16:01 DRB5*01:01;DRB5*02:02 DQB1*05:02;DQB1*06:02 DPB1*02:01;DPB1*04:02
38	**1556**	m	76	-	base of tongue	A*01:01;A*02:01 B*08:01;B*51:01 C*01:02;C*07:01	DRB1*04:01;DRB1*13:01 DRB3*02:02 DRB4*01:03 DQB1*03:01;DQB1*06:03 DPB1*04:01;DPB1*04:01
39	**1565**	f	79	-	base of tongue	A*02:01;A*02:01 B*07:02;B*13:02 C*06:02;C*07:02	DRB1*07:01;DRB1*13:01 DRB3*01:01 DRB4*01:01 DQB1*02:01;DQB1*06:03 DPB1*03:01;DPB1*11:01
40	**1580**	m	52	-	tonsil	A*01:01;A*24:02 B*08:01;B*38:01 C*07:01;C*12:03	DRB1*03:01;DRB1*09:01 DRB3*01:01 DRB4*01:01 DQB1*02:01;DQB1*03:03 DPB1*03:01;DPB1*04:01

Supplementary Table 10: Overview of the tonsil cohort and donor characteristics. Five healthy tonsil tissue samples were included in the study serving as benign reference. red – assigned HLA class I allotypes.

Sample number	Sample ID	Sex	Age (at surgery)	HLA class I typing	HLA class II typing
1	**1508**	f	19	A*01;A*32 B*07;B*08 C*07;C*07	-
2	**1509**	m	27	A*03;A*24 B*15;B*58 C*03;C*07	-
3	**1583**	f	32	A*02:01;A*02:01 B*27:05;B*44:02 C*02:02;C*05:01	DRB1*04:01;DRB1*04:04 DRB4*01:03 DQB1*03:02;DQB1*03:02 DPB1*03:01;DPB1*04:01
4	**1585**	f	28	A*01:01;A*32:01 B*07:02;B*57:01 C*06:02;C*07:02	DRB1*07:01;DRB1*15:01 DRB4*01:03 DRB5*01:01 DQB1*03:03;DQB1*06:02 DPB1*04:01;DPB1*13:01
5	**1587**	m	32	A*03:01;A*24:02 B*07:02;B*07:02 C*07:02;C*07:02	DRB1*15:01;DRB1*15:01 DRB5*01:01 DQB1*06:02;DQB1*06:01 DPB1*02:01;DPB1*04:01

Appendix

Supplementary Table 11: Allelic distribution of HLA-A alleles in OPSCC patients and the German population. Statistical analysis was performed on the basis of both the overall cohort and the HPV⁺ subgroup. Chi-square test was applied and a p value ≤ 0.004 was used as level of statistical significance after Bonferroni correction. The Allele Frequency Net Database was consulted for German allele frequencies [2]. Statistical results from alleles with a patient frequency of F = 0% (zero hits) and F = 1.351% (one hit in overall cohort) or F = 2.381% (one hit in HPV⁺ subgroup) were assumed as unreliable and were not announced. CI – confidence interval; * – significant; ns – not significant; gray-shaded – HPV⁺ patient subgroup; n.a. – not available.

HLA-A allele	Allele frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% CI of odds ratio
01:01	15.100%	20.270%	0.2144	ns	-	-
		16.667%	0.7768	ns	-	-
02:01	26.700%	33.784%	0.1686	ns	-	-
		21.429%	0.4401	ns	-	-
03:01	14.700%	9.459%	0.2031	ns	-	-
		16.667%	0.7191	ns	-	-
11:01	5.600%	8.108%	0.3483	ns	-	-
		11.905%	0.0757	ns	-	-
23:01	2.320%	2.703%	0.8272	ns	-	-
		4.762%	0.2936	ns	-	-
24:02	9.500%	12.162%	0.4351	ns	-	-
		9.524%	0.9958	ns	-	-
24:03	0.255%	1.351%	-	-	-	-
		2.381%	-	-	-	-
26:01	3.600%	1.351%	-	-	-	-
		2.381%	-	-	-	-
30:01	1.620%	1.351%	-	-	-	-
		2.381%	-	-	-	-
31:01	2.300%	4.054%	0.3147	ns	-	-
		4.762%	0.2875	ns	-	-
32:01	3.6%	2.703%	0.6786	ns	-	-
		4.762%	0.6863	ns	-	-
66:01	n.a.	1.351%	-	-	-	-
		2.381%	-	-	-	-
68:01	3.540%	1.351%	-	-	-	-
		0.000%	-	-	-	-

Supplementary Table 12: Allelic distribution of HLA-B alleles in OPSCC patients and the German population. Statistical analysis was performed on the basis of both the overall cohort and the HPV⁺ subgroup. Chi-square test was applied and a p value ≤ 0.002 was used as level of statistical significance after Bonferroni correction. The Allele Frequency Net Database was consulted for German allele frequencies [2]. Statistical results from alleles with a patient frequency of F = 0% (zero hits), F = 1.351% (one hit in overall cohort) or F = 2.381% (one hit in HPV⁺ subgroup) were assumed as unreliable and were not announced. CI – confidence interval; * – significant; ns – not significant; gray-shaded – HPV⁺ patient subgroup.

HLA-B allele	Allele Frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% CI of odds ratio
07:02	12.000%	13.514%	0.6887	ns	-	-
		14.286%	0.6485	ns	-	-
08:01	9.500%	13.514%	0.2393	ns	-	-
		4.762%	0.2951	ns	-	-

Supplementary Table 12 continued.

HLA-B allele	Allele Frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% Confidence Interval
13:02	3.470%	2.703%	0.7186	ns	-	-
		0.000%	-	-	-	-
14:02	2.000%	2.703%	0.6664	ns	-	-
		2.381%	-	-	-	-
15:01	6.500%	5.405%	0.7025	ns	-	-
		2.381%	-	-	-	-
15:18	0.190%	1.351%	-	-	-	-
		2.381%	-	-	-	-
27:02	0.730%	1.351%	-	-	-	-
		0.000%	-	-	-	-
27:05	3.400%	2.703%	0.7407	ns	-	-
		4.762%	0.6264	ns	-	-
35:01	6.200%	4.054%	0.4442	ns	-	-
		7.143%	0.7999	ns	-	-
35:02	1.000%	1.351%	-	-	-	-
		0.000%	-	-	-	-
35:03	2.460%	2.703%	0.8930	ns	-	-
		4.762%	0.3359	ns	-	-
38:01	2.500%	1.351%	-	-	-	-
		0.000%	-	-	-	-
40:01	4.900%	6.757%	0.4598	ns	-	-
		7.143%	0.5010	ns	-	-
41:02	0.620%	1.351%	-	-	-	-
		2.381%	-	-	-	-
44:02	7.300%	5.405%	0.5310	ns	-	-
		4.762%	0.5272	ns	-	-
44:03	4.100%	4.054%	0.9843	ns	-	-
		4.762%	0.8286	ns	-	-
49:01	1.600%	4.054%	0.0928	ns	-	-
		7.143%	0.0042	ns	-	-
50:01	1.320%	2.703%	0.2979	ns	-	-
		4.762%	0.0509	ns	-	-
51:01	**6.300%**	**16.216%**	**0.0005**	*****	**2.879**	**1.550 to 5.345**
		16.667%	0.0057	ns	-	-
52:01	1.100%	1.351%	-	-	-	-
		2.381%	-	-	-	-
57:01	3.100%	5.405%	0.2530	ns	-	-
		7.143%	0.1309	ns	-	-

Appendix

Supplementary Table 13: Allelic distribution of HLA-C alleles in OPSCC patients and German population. Statistical analysis was performed on the basis of both the overall cohort and the HPV⁺ subgroup. Chi-square test was applied and a p value ≤ 0.004 was used as level of statistical significance after Bonferroni correction. The Allele Frequency Net Database was consulted for German allele frequencies [2]. Statistical results from alleles with a patient frequency of F = 0% (zero hits), F = 1.351% (one hit in overall cohort) or F = 2.381% (one hit in HPV⁺ subgroup) were assumed as unreliable and were not announced. CI – confidence interval; * – significant; ns – not significant; gray-shaded – HPV⁺ patient subgroup.

HLA-C allele	Allele frequency		Chi-square test			
	Germany	Patients	p value	Significance	Odds ratio	95% CI of odds ratio
01:02	3.600%	4.054%	0.8342	ns	-	-
		2.381%	-	-	-	-
02:02	5.400%	5.405%	0.9982	ns	-	-
		7.143%	0.6173	ns	-	-
03:04	7.400%	10.811%	0.2627	ns	-	-
		11.905%	0.2649	ns	-	-
03:44	0.003%	1.351%	-	-	-	-
		0.000%	-	-	-	-
04:01	12.500%	17.568%	0.2536	ns	-	-
		21.429%	0.0803	ns	-	-
05:01	6.400	4.054%	0.4098	ns	-	-
		2.381%	-	-	-	-
06:02	10.000%	9.459%	0.8768	ns	-	-
		9.524%	0.9180	ns	-	-
07:01	14.700%	18.919%	0.3058	ns	-	-
		11.905%	0.6090	ns	-	-
07:02	13.400%	13.514%	0.9772	ns	-	-
		14.286%	0.8663	ns	-	-
07:04	1.920%	2.703%	0.6239	ns	-	-
		4.762%	0.1798	ns	-	-
08:02	2.400%	2.703%	0.8649	ns	-	-
		2.381%	-	-	-	-
12:02	1.100%	1.351%	-	-	-	-
		2.381%	-	-	-	-
12:03	6.300%	1.351%	-	-	-	-
		0.000%	-	-	-	-
14:02	1.520%	2.703%	0.4064	ns	-	-
		2.381%	-	-	-	-
15:02	2.600%	2.703%	0.9558	ns	-	-
		4.762%	0.3789	ns	-	-
17:01	1.000%	1.351%	-	-	-	-
		2.381%	-	-	-	-

Supplementary Table 14: Overview of HLA ligand yields from OPSCCs. HLA ligands were analyzed by LC-MS/MS. The numbers of HLA ligands are depicted for 40 individual OPSCC tissue samples. The purity was calculated by the percentage of HLA class I binders among the total number of HLA class I-presented peptides. For samples 1416 and 1420, two biological replicates were used. Their results were merged for data analysis.

Sample number	Sample ID	Tissue mass [mg]	HLA class I					HLA class II	
			Presented peptides	Source proteins	Binders	Source proteins of binders	Purity [%]	Presented peptides	Source proteins
1	**020**	389	1253	1312	1192	1269	95	1050	600
2	**022**	326	1307	1362	1246	1310	95	574	391
3	**1010**	132	1471	1481	1430	1446	97	820	473
4	**1034**	371	2637	2311	2481	2204	94	1644	842
5	**1118**	474	2092	1892	2019	1831	97	2086	920
6	**1209**	303	1553	1502	1403	1386	90	1308	577
7	**1345**	67	425	469	378	426	89	624	371
8	**1349**	38	732	841	701	815	96	494	348
9	**1359**	151	1059	1142	1025	1098	97	758	441
10	**1361**	57	1578	1507	1468	1414	93	255	232
11	**1365**	91	1336	1291	1200	1226	90	777	472
12	**1372**	94	488	493	464	473	95	656	408
13	**1385**	23	682	795	626	737	92	366	286
14	**1407**	101	1431	1426	1355	1362	95	262	244
15	**1416**	133	2452	2173	2330	2088	95	965	466
		105	2067	1920	1962	1831	95	537	290
16	**1420**	267	2494	2181	2400	2106	96	257	219
		38	1166	1132	1099	1076	94	277	233
17	**1435**	236	2282	2024	2143	1917	94	675	419
18	**1475**	84	827	892	798	865	96	717	383
19	**1497**	218	2442	2253	2358	2171	97	605	446
20	**1581**	182	1964	1808	1876	1736	96	704	447
21	**1003**	567	2243	2067	2127	1979	95	765	446
22	**1008**	128	1354	1269	1288	1259	95	447	284

Supplementary Table 14 continued.

Sample number	Sample ID	Tissue mass [mg]	HLA class I					HLA class II	
			Presented peptides	Source proteins	Binders	Source proteins of binders	Purity [%]	Presented peptides	Source proteins
23	**1012**	499	783	906	755	882	96	628	386
24	**1014**	313	1770	1730	1683	1658	95	1424	706
25	**1021**	278	1452	1431	1055	1091	73	871	959
26	**1024**	159	1753	1788	1662	1603	95	830	490
27	**1376**	206	1683	1638	1493	1529	89	802	434
28	**1382**	203	1520	1457	1463	1411	96	901	458
29	**1393**	37	1367	1360	1303	1297	95	438	279
30	**1401**	35	330	371	308	347	93	168	154
31	**1403**	17	302	380	265	338	88	178	149
32	**1404**	80	1673	1649	1616	1597	97	957	517
33	**1427**	80	567	656	528	636	93	224	148
34	**1444**	57	1445	1437	1331	1348	92	341	301
35	**1445**	107	1201	1256	1148	1199	96	177	193
36	**1468**	85	555	609	514	566	93	201	139
37	**1504**	87	803	866	776	845	97	242	163
38	**1556**	252	1057	1123	1012	1081	96	767	432
39	**1565**	333	1627	1466	1551	1390	95	1038	611
40	**1580**	177	2083	1808	2023	1875	97	1079	447

Supplementary Table 15: Overview of HLA ligand yields from tonsils. HLA ligands were analyzed by LC-MS/MS. The numbers of HLA ligands are depicted for five individual tonsil tissue samples. The purity was calculated by the percentage of HLA class I binders among the total number of HLA class I-presented peptides.

Sample number	Sample ID	Tissue mass [mg]	HLA class I					HLA class II	
			Presented peptides	Source proteins	Binders	Source proteins of binders	Purity [%]	Presented peptides	Source proteins
1	**1508**	378	2247	2097	2144	2006	95	1614	770
2	**1509**	302	2728	2366	2586	2272	95	1026	559
3	**1583**	210	1670	1635	1602	1579	96	1509	649
4	**1585**	487	2610	2350	2463	2229	94	2012	900
5	**1587**	162	1685	1599	1615	1542	96	1044	524

Supplementary Table 16: CTAs identified in the OPSCC immunopeptidome. The CTdatabase [19] was used as basis for the identification of CTAs. CTAs are depicted that were represented as source proteins in the HLA class I ligandome with a frequency of at least 8%. AC – accession; CTA – cancer-testis antigen; red – tumor-exclusive.

Peptide				Source protein			
Sequence	**HLA restriction**	**Sample**	**Representation frequency**	**UniProtKB AC**	**Gene name**	**Protein name**	**Representation frequency**
ALLDKLYAL	A*02:01	022 1003 1008 1010 1012 1118 1349 1359 1365 1382 1416 1427 1435 1444 1468 1475 1504 1556 1565 1581	50%	Q9NV31	IMP3	U3 small nucleolar ribonucleoprotein IMP3	58%
RSMEDFVTW	B*57:01	1003 1361 1581	8%				
QHLQAAVAF	A*32:01;B*15:018	1034	3%				
EDYTRYNQL	B*08:01	1376	3%				

Supplementary Table 16 continued.

Peptide				Source protein			
Sequence	**HLA restriction**	**Sample**	**Representation frequency**	**UniProtKB AC**	**Gene name**	**Protein name**	**Representation frequency**
DPFAFIHKI	B*51:01	020 022 1003 1010 1349 1359 1385 1407 1435 1468 1556	28%	Q9UGL1	KDM5B	Lysine-specific demethylase 5B	43%
ILNPYNLFL	A*02:01;C*17:01	020 022 1003 1012 1118 1359 1504 1556 1565	23%				
RLIDLGVGL	A*02:01	022 1003 1012 1118 1382 1504 1565	18%				
PFAFIHKI	B*51:01	020	3%				
QEAFGFEQA	B*50:01	1010	3%				
SVAQQLLNGK	A*11:01	1021	3%				

Supplementary Table 16 continued.

Peptide				Source protein			
Sequence	**HLA restriction**	**Sample**	**Representation frequency**	**UniProtKB AC**	**Gene name**	**Protein name**	**Representation frequency**
FIDASRLVY	A*01:01	1003 1008 1014 1024 1345 1359 1372 1376 1385 1393 1403 1580	30%	P26232	CTNNA2	Catenin alpha-2	30%
KVKAEVQNL	C*03:04	1365	3%				
AYAIIKEEL	A*24:02;A*24:03;A*23:01	1010 1361 1376 1393 1404 1407 1444 1475 1497 1504 1580	28%	Q6PL18	ATAD2	ATPase family AAA domain-containing protein 2	30%
LYPEVFEKF	A*24:02;A*23:01	1010 1404 1497 1504	10%				
KYLTVKDYL	A*23:01	1497	3%				
GLSNHIAAL	A*02:01	1565	3%				
SQNAIDHKI	B*13:02	1565	3%				

Supplementary Table 16 continued.

Peptide				Source protein			
Sequence	**HLA restriction**	**Sample**	**Representation frequency**	**UniProtKB AC**	**Gene name**	**Protein name**	**Representation frequency**
SYQKVIELF	A*24:02;A*24:03;A*23:01	1010 1361 1376 1393 1404 1407 1420 1444 1475 1497 1504	28%	Q96KB5	PBK	Lymphokine-activated killer T-cell-originated protein kinase	28%
KILDLETQL	A*02:01	1003 1012 1382 1427 1468 1504 1565	18%	Q5BJF6	ODF2	Outer dense fiber protein 2	20%
AEALSTLESW	B*44:02	1416	3%				
SPKSPTAAL	B*07:02	1345 1372 1382 1565	10%	Q53EZ4	CEP55	Centrosomal protein of 55 kDa	18%
AELESKTNTL	B*40:01	1008	3%				
QEEQTRVAL	B*40:01	1034	3%				
EELLSQVQF	B*44:02	1416	3%				

Supplementary Table 16 continued.

Peptide				Source protein			
Sequence	**HLA restriction**	**Sample**	**Representation frequency**	**UniProtKB AC**	**Gene name**	**Protein name**	**Representation frequency**
DEAVGVQKW	B*44:02;B*44:03	1024 1349 1404 1416 1420	13%	Q14667	KIAA100	Protein KIAA0100	15%
DYPRYLFEI	A*24:02	1376 1404	5%				
DAVKFFVAV	B*51:01	1003 1010 1349 1359 1556	13%	O60271	SPAG9	C-Jun-amino-terminal kinase-interacting protein 4	15%
NYADQISRL	C*07:02	1404	3%				
LEEQASRQI	B*49:01	1361 1365	5%	Q99661	KIF2C	Kinesin-like protein KIF2C	13%
IYNGKLFDLL	B*24:02	1361 1444	5%				
RLFPGLAI	A*03:01;C*06:02	1416 1565	5%				
SPIEKSGVL	B*07:02	1372 1382 1565	8%	Q8NG31	CASC5	Kinetochore scaffold 1	10%
IYVIPQPHF	A*23:01	1497	3%				
ALLQEIVNI	A*02:01	1003 1565	5%	Q92600	RQCD1	CCR4-NOT transcription complex subunit 9	8%
ILLDDTGLAYI	A*02:02	1003	3%				
KIYQWINEL	A*32:01	1034	3%				
KVLEYVIKV	A*02:01	1427 1435 1468	8%	P43355	MAGEA1	Melanoma-associated antigen 1	8%
KEADPTGHSY	B*44:03	1435	3%				

Supplementary Table 17: Tumor-exclusive HLA class I binders of the OPSCC immunopeptidome. Peptides with a frequency ≥ 8% (≥ 3 samples) within the OPSCC cohort are depicted. AC – accession; green – selected TAAs.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
RTEFNLNQY	A*01:01	1008 1014 1024 1376 1393 1556 1580 1581	20%	Q99715
YSELASHVVSY	A*01:01	1008 1014 1024 1345 1372 1401 1580 1581	20%	Q9UMD9
DQYKFLAV	B*51:01;B*52:01	020 022 1010 1021 1209 1349 1407	18%	Q9NRX3
EMEQQNQEY	A*01:01	1003 1008 1014 1024 1376 1401 1580	18%	P02533;Q04695
KYMYFTVVM	A*23:01;A*24:02	1010 1361 1376 1407 1420 1504 1580	18%	Q8WV24
RLLEGEDAHL	A*02:01	1003 1010 1012 1118 1382 1427	15%	P02533;P08779;Q04695
DGVLIWKI	B*51:01	020 022 1010 1359 1556	13%	Q9BUZ4
DIYYSLNIY	A*01:01	1014 1345 1376 1556 1581	13%	Q4G0N8

Appendix

Supplementary Table 17 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
EETNPKGSGW	B*44:02;B*44:03	1024 1404 1416 1420 1435	13%	Q13835
GVLENIFGV	A*02:01	1003 1012 1118 1382 1504	13%	Q9H6A9
KLAELEEAL	A*02:01	1118 1382 1504 1565 1581	13%	P13647
LSFVDTRTL	B*57:01;C*15:02	1003 1361 1365 1385 1581	13%	P08123
LTRAGILTF	B*57:01;C*12:02	1003 1209 1361 1365 1581	13%	P12111
RLDDLKMTV	A*02:01	1003 1008 1118 1382 1581	13%	Q13753
STAKSATWTY	A*01:01;B*57:01	1014 1361 1401 1580 1581	13%	Q9H3D4;O15350
VLFPNLKTV	A*02:01	1008 1010 1382 1468 1581	13%	Q01954
AEMQFGELL	B*40:01;B*44:02;B*49:01	1034 1118 1361 1416	10%	O95361
AEYIEKVVY	B*44:02;B*44:03	1024 1403 1420 1435	10%	P07942
ALTDIVSQV	A*02:01	1003 1118 1565 1581	10%	Q9NQU5

Supplementary Table 17 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
DANPYDSVKKI	B*51:01	020 022 1010 1435	10%	O15205
EMEAQNQEY	A*01:01	1345 1401 1497 1581	10%	P19012
IPYKGGNTM	B*35:01;B*35:03;B*51:01	1010 1393 1407 1420	10%	Q99715
KLAEISLGV	A*02:01	022 1010 1118 1365	10%	P32926
KVLETKWTL	A*24:02;A*32:01;B*57:01	1034 1361 1365 1581	10%	P02538
LPVDHISLI	B*51:01	020 1010 1349 1407	10%	Q15758
MSFVQKGSW	B*57:01	1003 1361 1365 1581	10%	P02461
NEVFGLFQKL	B*44:02;B*44:03	1024 1416 1420 1435	10%	P15924
NWPSRPYLF	A*24:02	1376 1404 1504 1580	10%	O43795
QLDEKSSQL	A*02:01	1010 1365 1382 1581	10%	Q9C099
RPAQGVVTTL	B*07:02;B*51:01	1021 1118 1209 1382	10%	P02751
RPLLLALL	B*07:02	1012 1118 1372 1382	10%	Q9P2B2
RTDIARTEY	A*01:01	1003 1014 1580 1581	10%	Q14204

Supplementary Table 17 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
SMLNNIINL	A*02:01	1008 1118 1468 1565	10%	Q13835
VALPVYLLI	B*51:01	020 022 1003 1359	10%	P57054
VLDSHIHAY	A*01:01	1003 1014 1393 1580	10%	P23471
VPFSHVNI	B*51:01	020 1003 1407 1435	10%	Q8N766
YLDDPTNSWY	A*01:01	1014 1376 1556 1580	10%	O60479
YLEEFITNI	A*02:01	1003 1382 1468 1565	10%	Q9Y2D4
AEALVSKGL	B*44:02;B*44:03	1024 1416 1435	8%	Q7KZF4
AEKELVQSL	B*40:01;B*41:02;B*44:02	020 1008 1416	8%	O15460
ALLGSAFQL	A*02:01	1118 1359 1435	8%	Q96MG2
ALTDIDLQL	A*02:01	1003 1012 1118	8%	Q6UVK1
ARLIPIIVL	C*06:02	022 1003 1581	8%	B0I1T2
AVANIVNSV	A*02:01	1382 1435 1565	8%	Q9BRQ8
DAHEFLNYL	B*51:01	020 022 1003	8%	O75317;P62068
DASTFTINI	B*51:01	022 1003 1010	8%	Q02809;O00469
DATETTITI	B*51:01	1010 1435 1556	8%	P02751
DAWIEHDVW	B*35:01	1359 1385 1393	8%	Q9UHE8;Q6NZ63

Supplementary Table 17 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
DSFLGQTSI	B*51:01	1003 1010 1407	8%	Q9Y6Y8
EAVGREALEL	C*03:04	1034 1118 1497	8%	Q5T4D3
FLQDLEQRL	A*02:01	1118 1382 1565	8%	Q8IW93
FTKRKFGLMKK	A*03:01;A*11:01	1021 1416 1445	8%	Q02078;Q02080; Q06413;Q14814
GEAAKSVKL	B*40:01;B*41:02	020 1034 1497	8%	Q9HCY8
GLIEDYEALL	A*02:01	022 1003 1118	8%	Q96SN8
GLWEDGRSTLL	A*02:01	1012 1468 1556	8%	Q96BA8
GTEDELDKY	A*01:01	1014 1024 1376	8%	P09493;P06753
GYIDNVTLI	A*24:02;A*24:03	1376 1393 1404	8%	Q13753
HEDPEVKFNW	B*44:02;B*44:03	1024 1420 1435	8%	P01857
HGTIKNQL	B*08:01	1014 1580 1581	8%	Q86XI2
IAFLMINAV	B*51:01	020 022 1010	8%	O95807
IAFSATRTI	B*51:01	020 1003 1435	8%	P02746
IEDDPKMMW	B*44:02;B*44:03	1012 1024 1435	8%	P20929
KEAVKNLEW	B*44:02;B*44:03	1024 1420 1435	8%	P12110
KTIQEVAGY	B*57:01	1003 1365 1581	8%	P00533
LEEAQRVIL	B*40:01	1008 1034 1445	8%	P12111
LLLTGMLVGGI	A*02:01	1118 1565 1581	8%	Q63ZE4

Supplementary Table 17 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
MWKVSALLF	A*23:01;A*24:02	1376 1404 1420	8%	Q86YL7
NEIGQVLHF	B*44:02;B*44:03	1024 1416 1435	8%	P36952
NTYKSIQMM	A*02:01	1349 1416 1435	8%	P29508
PLPLALLL	B*51:01	020 022 1003	8%	Q6UXQ8
PSKTAENATFY	A*01:01	1024 1497 1556	8%	Q9UI66
QEDRQLINAL	B*40:01	1008 1034 1445	8%	P12111
QENIVRILL	B*40:01;B*44:02;B*44:03	1034 1416 1435	8%	P57078
QLNEKVAQL	B*08:01	1014 1427 1581	8%	Q16850
QTDQSHIGQY	A*01:01	1008 1024 1376	8%	Q8WZ42
RESSSSAERQW	B*44:02;B*44:03	1416 1420 1435	8%	Q68CZ2
REVVDPEVFF	B*40:01	1034 1118 1497	8%	Q13835
REYENELAKV	B*49:01	1361 1365 1497	8%	P15924
RLLDSVSRL	A*02:01	1003 1382 1427	8%	Q13753
RQINVGNAL	B*15:01	1382 1407 1444	8%	P12111
RQLEVLQSI	A*02:01;A*03:01;A*11:01;B*13:01	1021 1209 1565	8%	O95067
RSFPGSKEY	B*15:01;B*57:01	1361 1382 1581	8%	Q14517
SALALIIPAL	A*02:01;C*17:01	020 022 1010	8%	Q495N2
SEAESRIFW	B*44:02;B*44:03	1349 1416 1420	8%	Q9NYZ1;Q96ET8

Supplementary Table 17 continued.

Sequence	HLA restriction	Sample ID	Representation frequency	UniProtKB AC of source protein
SEPNDVFFKL	B*40:01	1008 1034 1445	8%	P12111
SEVVDVAKL	B*40:01	1008 1034 1445	8%	Q9HD67
SIFEGLLSGV	A*02:01	022 1012 1581	8%	O95377
SLAALVVHV	A*02:01	022 1010 1581	8%	Q9NYQ8
SPEDGIHEL	B*35:02;B*35:03	1403 1420 1444	8%	P02751
SVKITQVTW	B*57:01	1361 1365 1581	8%	Q15223
SYLQAANAL	A*24:02;A*24:03	1376 1393 1407	8%	P12111
TENSAKLHW	B*44:02;B*44:03	1024 1349 1420	8%	P12111
VAAPRWVL	B*08:01	1014 1024 1393	8%	A6NJ16
VEDSALLMQTL	B*40:01	1008 1034 1445	8%	Q9Y5V3
VLVPYEPPQV	A*02:01	1008 1010 1565	8%	Q9H3D4
VMAPRTVL	B*08:01	1014 1376 1401	8%	P01889;P30460; P30462;Q95365; P30475;P30480; P30486;Q29836; Q31612;Q31610; Q31610
YEIDDVERL	B*40:01	1008 1034 1445	8%	Q9ULI4;Q2KJY2
YPFHKQPPTYV	B*51:01	020 1003 1359	8%	Q9BU23
YQDPHSTAV	A*02:01	1003 1365 1565	8%	P00533
YTDNWLAVY	A*01:01	1014 1372 1581	8%	P32926
YTFRYPLSL	C*07:01	1361 1365 1581	8%	P24347

Appendix

Supplementary Table 18: Tumor-exclusive HLA class II-presented peptides of the OPSCC immunopeptidome. Peptides with a frequency ≥ 8% (≥ 3 samples) within the OPSCC cohort are depicted. AC – accession; green – selected TAAs; grey – no reliable HLA class II ligand due to length.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
IRQFTSSSSIKGSSG	1012 1024 1118 1372 1385 1404 1565	18%	Q04695
GMQDLVEDFKNKYEDE	1008 1014 1376 1401 1580 1581	15%	P02538
MQDLVEDFKNKYEDE	1008 1014 1376 1401 1580 1581	15%	P13647;P02538
NMQDLVEDFKNKYEDEIN	1008 1014 1376 1401 1580 1581	15%	P13647
TPLNLQIDPTIQRVRAE	1008 1014 1376 1401 1580 1581	15%	P02538
DPASAKIEGNLIFDPNNYL	020 1359 1376 1435 1556	13%	P04114
GMQDLVEDFKNKYEDEIN	1008 1014 1401 1580 1581	13%	P02538
LTSTSGPGFHLMLPFI	020 1012 1359 1416 1565	13%	O94905
MLFFDKFANIVPF	1012 1404 1420 1565 1580	13%	O00220

Supplementary Table 18 continued.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
MQDLVEDFKNKYED	1014 1376 1401 1580 1581	13%	P13647;P02538
TPLNLQIDPSIQRVRT	1014 1376 1401 1580 1581	13%	P13647
TPLNLQIDPTIQRVRA	1014 1376 1401 1580 1581	13%	P02538
DLDSIIAEVKAQYEE	1008 1014 1376 1581	10%	P35908;P13647; P02538;P04259; P48668
ELKKKLKEKAKERREKEMLERLEK	1118 1416 1497 1556	10%	Q9UIG0
ENDVIISINGQSVVS	1014 1365 1372 1376	10%	Q92743
FWTQLMLLLWKNFM	022 1024 1404 1565	10%	Q8IZY2
IIDENTVHMSWAKPVDPI	1024 1034 1365 1580	10%	Q99715
IQDFIEAEDDLSSFR	1024 1349 1445 1565	10%	Q15063
LDSELRNMQDLVEDFK	020 1349 1404 1445	10%	P13647
LPFLDIAPLDIGG	022 1012 1565 1580	10%	P08123
LPIIDLAPVDVGGTD	1024 1365 1420 1580	10%	P05997

Supplementary Table 18 continued.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
NMQDLVEDFKNKYE	1014 1401 1580 1581	10%	P13647
NMQDLVEDFKNKYED	1014 1401 1580 1581	10%	P13647
NMQDLVEDLKNKYEDE	1008 1014 1580 1581	10%	P04259;P48668
NMQDLVEDLKNKYEDEI	1008 1014 1580 1581	10%	P04259
PVHWQFGQLDQHPIDG	1361 1393 1420 1580	10%	P09486
QAHLKYILSDSSPAPEF	1365 1404 1580 1581	10%	P23786
QDLVEDFKNKYEDE	1014 1401 1580 1581	10%	P13647;P02538; Q5XKE5
SPYFKTIEDLRNK	1024 1372 1385 1404	10%	P02533;P08779
SVNEHEDGDGDGDSDEGDD	1012 1382 1393 1404	10%	Q2VIQ3
TKPAIRRLARRGGVKR	1010 1118 1416 1565	10%	P62805
TNKPSRLPFLDIAPLDIGGADQ	022 1012 1024 1580	10%	P08123
TPLNLQIDPTIQRVR	1014 1376 1580 1581	10%	P02538
WDNEMRVTEYLVVYTPT	1024 1118 1382 1420	10%	P24821

Supplementary Table 18 continued.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
WIPDIQIDPNGLSFNP	1012 1118 1372 1565	10%	P01042
AENEFVTLKKDVDAAY	020 1361 1581	8%	P02538;P04259; P48668
ARLPIIDLAPVDVGGTD	1024 1365 1580	8%	P05997
DAEDWFFSKTEELNR	1010 1024 1034	8%	Q04695
DAEDWFFSKTEELNRE	1010 1024 1034	8%	Q04695
DGEDDCGDGLDESDSICG	1372 1382 1404	8%	Q9NZR2
DLVEDFKNKYEDE	1014 1376 1580	8%	P13647;P02538; Q5XKE5
DNANILLQIDNARLAADD	1003 1209 1382	8%	P13645;Q04695
DPDDTHAYNVADFESLSR	1209 1565 1580	8%	Q99715
DPELLKLF	1014 1209 1581	8%	Q7Z5Q5
DSAPVELILSDETLPAPEFSP	1012 1118 1372	8%	P04217
DTRIFFVNPAPPYLWPA	1010 1118 1416	8%	O00468
EKNKKISLLHSSKEKLRRERIKYCC	1003 1435 1475	8%	Q9NX45
FADFTFATGKIIG	1444 1475 1580	8%	Q13093
GDEPQYLDLPSTATSVNIPD	1014 1382 1468	8%	P02751
GMQDLVEDFKNKYED	1014 1401 1580	8%	P02538
GPTAPRDVQYFLYIRNS	022 1349 1359	8%	P15509

Appendix

Supplementary Table 18 continued.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
GRRDLRFQPVSIGRWG	022 1034 1349	8%	Q99102
HRGLRHFWGLR	1359 1416 1420	8%	P62269
IKNIISEYKSAIQSQKRRRPRYRKR	1014 1209 1475	8%	Q6RI45
ISKMLFVEPILEVSSLPT	020 1118 1359	8%	P05155
KAHLGTAL	020 022 1012	8%	P62266
KNKHKRKKVKLA	1435 1565 1580	8%	P62979
KPSRLPFLDIAPLDIGGADQ	022 1012 1580	8%	P08123
KQAMKAILTDQPMI	020 1118 1416	8%	Q8IZV5
LDLTYWIDGTRHVVS	1376 1580 1581	8%	P35442
LDSELRNMQDLVEDF	020 1349 1445	8%	P13647
LNVLRIINEPTAAAIAYGLD	1118 1372 1407	8%	P34931;P08107; P54652;P17066; P11142
LPFLDIAPLDIGGAD	022 1012 1580	8%	P08123
LQNKILTATVDNANIL	1024 1372 1580	8%	Q04695
LTDEINFLRALYD	1361 1393 1420	8%	P02538;P04259; P48668
NMIRITAVCKVPDE	1012 1416 1580	8%	Q08188
NMQDLVEDLKNKYEDEIN	1008 1580 1581	8%	P04259;P48668
QDLVEDFKNKYEDEI	1008 1014 1581	8%	P13647;P02538; Q5XKE5
QGKVEEDLELLDKSFET	020 1118 1359	8%	Q9P1Z9

Supplementary Table 18 continued.

Sequence	Sample ID	Representation frequency	UniProtKB AC of source protein
QQLTREATQAEIEADR	1010 1024 1580	8%	Q13753
QSPTWLLYSHPVGRR	022 1034 1556	8%	Q8WWR8
REKHLMYLLEHLHPFLQRQQLDYG	1118 1359 1468	8%	O60513
RRAKFKFPGRQK	1008 1420 1580	8%	Q96L21;P27635
SWKDDTERTNRLVLLGRNLDKDILK	1003 1468 1581	8%	Q9BRT8;Q8IUF1
TKDQDLIKQYNTLIEEMLQV	1118 1497 1556	8%	Q8IWV7
TKHEISEMNRMIQR	1010 1401 1416	8%	P13647
TLCEKLTVSLSDPDPVF	1012 1024 1565	8%	Q96NH3
TPAAQYVRIKENLAVG	1118 1372 1416	8%	Q14574
TQLMPFGCLLDYVREH	020 1014 1565	8%	P00533
VDEEEVKFPGTNFDE	1014 1580 1581	8%	Q9BVV6
VIAAVKIFPRFFMVAKQCSAG	1359 1365 1435	8%	Q17RY6
WEWVPWEELPP	1003 1012 1416	8%	Q9NV35
YNPSISIVGTLEAEKE	1349 1365 1404	8%	P23229
YTSALRPVADTDQTLNV	1024 1407 1565	8%	Q9GZZ6

Appendix

Supplementary Table 19: TAAs identified in the OPSCC cohort and their occurrence in the immunopeptidomes from other cancer entities of an in-house malignant database. BrCa – breast cancer; CRC – colorectal cancer; HCC – hepatocellular carcinoma; NSCLC – non-small-cell lung carcinoma; OvCa – ovarian carcinoma; RCC – renal cell carcinoma; AML – acute myeloid leukemia; CLL – chronic lymphocytic leukemia.

HLA class	Sequence	HLA restriction	Representation frequency in		Tumor entities
			Cohort	Malignant database	
I	**DANPYDSVKKI**	B*51:01	10%	3%	BrCa CRC Germ cell tumor Glioblastoma HCC Melanoma Meningioma Neuroblastoma NSCLC Osteosarcoma OvCa RCC
	VALPVYLLI	B*51:01	10%	4%	AML BrCa Glioblastoma Meningioma NSCLC Osteosarcoma OvCa RCC
	HGTIKNQL	B*08:01	8%	2%	AML CLL CRC Glioblastoma NSCLC OvCa
	EETNPKGSGW	B*44:02;B*44:03	13%	0%	-
	KLAEISLGV	A*02:01	10%	0.3%	Bladder cancer BrCa
	YTDNWLAVY	A*01:01	8%	0.4%	NSCLC
	YLDDPTNSWY	A*01:01	10%	0.5%	CRC NSCLC Osteosarcoma
	GEAAKSVKL	B*40:01;B*41:02	8%	1.4%	BrCa CRC Gastric cancer OvCa
	SVKITQVTW	B*57:01	8%	0.1%	AML
II	**MLFFDKFANIVPF**	-	13%	1%	Meningioma RCC
	LPFLDIAPLDIGG	-	10%	1%	BrCa OvCa RCC

Supplementary Table 19 continued.

HLA class	Sequence	HLA restriction	Representation frequency in		Tumor entities
			cohort	Malignant database	
	TNKPSRLPFLDIAPLDIGGADQ	-	10%	5%	Bladder cancer BrCa Glioblastoma Meningioma Neuroblastoma NSCLC Osteosarcoma OvCa RCC
	TKPAIRRLARRGGVKR	-	10%	2%	Meningioma NSCLC OvCa RCC
	VDEEEVKFPGTNFDE	-	8%	0%	-
	TPAAQYVRIKENLAVG	-	8%	0.4%	BrCa OvCa

Appendix

Supplementary Table 20: Functional annotation clustering for HPV$^+$-exclusive HLA class I ligands. GO term enrichment was performed with the functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11] on the basis of HPV$^+$-exclusive source proteins of HLA class I binders. An enrichment score ≥ 1 was used as threshold. HPV – human papilloma virus; GO – gene ontology.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV$^+$-exclusive antigens
Lipid antigen presentation by HLA	2.43	GO:0030883 GO:0030884 GO:0048007 GO:0071723 GO:0030881	P29017 P06126 P29016	P15813 P15812	0.228%
Intracellular protein transport	2.04	GO:0031338 GO:1902017 GO:0017137 GO:0090630 GO:0006886	Q92609 Q0IIM8 Q8IV04 Q9Y3P9 O60447 Q4KMP7 Q86TI0 Q96C34 B7ZAP0 Q9P2M4 Q9NVG8 Q8WUA7 P54922 Q9BU20 Q96BZ9 Q13099 Q9ULP9 Q92696 Q96CV9 Q9BWQ6 Q5T5C0 Q8TDW5 Q6P1M3 Q8TB24 Q9H2M9	Q8IYU2 Q9Y2T2 Q8TBA6 Q8N3F8 P29317 Q9ULV4 Q9H6U6 P32248 Q9UIA9 Q9BXS5 Q9H115 Q9P299 Q9BZE9 Q5T9L3 Q4G0F5 Q9NRS6 Q9Y6B7 Q07866 Q8IV36 Q9UJY4 Q8TEX9 P53677 Q9UBQ0 Q9H2T7 P83436	2.277%
DNA damage response	1.92	GO:0010606 GO:0090503 GO:0030014 GO:0004535 GO:0000289 GO:0006977	Q9ULM6 Q504Q3 Q9NZN8 Q96LI5 Q58A45 Q8TCS8 Q9UIV1 Q9Y3B8 Q8TF46	Q9NPD3 P26651 Q8NE35 Q9UKZ1 P06730 Q01094 Q07812 P30307 Q86X55	0.820%
ATP biosynthesis	1.76	GO:0042776 GO:0000276 GO:0046933 GO:0015986 GO:0006754 GO:0005753 GO:0015078 GO:0022857	P10606 P48201 O75964 P48047 P56134 Q9UJZ1 O75947 P36542 P05496	Q7Z4Y8 Q9Y487 P30613 P01137 Q8NHE4 Q495N2 Q96S37 Q9BYT1 Q6NT16	0.820%
Heparin biosynthesis	1.64	GO:0015016 GO:0015012 GO:0030210	O95803 P52848 P52849	Q9H3R1 Q9Y662 Q9ULK5	0.273%

Supplementary Table 20 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins	Proportion of HPV$^+$-exclusive antigens
Phototransduction	1.56	GO:0031683 GO:0007603 GO:0019001 GO:0001750	Q14344 P29992 P30542 P19087 P11488 P78363 P09471 P35913 A8MTJ3 O60840 P50148 Q15051 P38405 Q8IWZ6	0.638%
Protein folding	1.52	GO:0003755 GO:0000413 GO:0061077 GO:0005528 GO:0006457	A2BFH1 P10909 Q9Y536 P25686 Q8WUA2 O75718 P30414 O75347 Q92623 P11488 Q9NWM8 P49257 F5H284 Q8WW22 Q8N6N2 Q58FG0 Q96AY3 Q9NYA1 P26885 A8MTJ3 Q75LS8 Q13371 O43447 O75695 Q96BP3 P09471 Q8NHS0 P19087	1.275%
mRNA transcription termination and export from nucleus	1.41	GO:0031124 GO:0006369 GO:0006405 GO:0006406	Q09161 P82979 O94913 Q86U42 Q8IYB3 P62306 Q96FV9 P62304 Q01081 P06730 Q92989 P43487 Q16629 P52594 Q05519 Q8NFH3 P61326 Q86US8 Q9P2I0 Q53GS7 Q96J01 Q9BW27	1.002%
Histone demethylation	1.33	GO:0071558 GO:0071557 GO:0032452 GO:0051213	O15550 Q9BY66 O15054 Q14865 Q6ZMT4 Q92833 O14607	0.319%
mRNA decay process	1.26	GO:0030014 GO:0060213 GO:1900153	P26651 Q9NZN8 Q8NE35 Q96LI5 Q9UKZ1 Q9HCJ0 Q9UIV1	0.319%
Mitochondrial respiratory activity	1.24	GO:0008137 GO:0006120 GO:0032981 GO:0005747	O43674 P03923 O43676 O95139 P03886 O95178 O75489 P03891 Q9P0J0 Q8IUX1 P51970 Q9Y276 P28331	0.592%
G-protein coupled receptor signaling pathway	1.23	GO:0031683 GO:0005834 GO:0007188	Q14344 P38405 P30542 P29992 P11488 P19087 P09471 P63218 A8MTJ3 P21462 P50148	0.501%

Supplementary Table 20 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins	Proportion of HPV$^+$-exclusive antigens
Microtubule-mediated movement	1.23	GO:0003777 GO:0005871 GO:0007018	Q9NP97 Q9P2P6 Q9NYC9 Q96M86 Q9Y6G9 Q02224 O95235 Q8TD57 Q9UFH2 O00139 Q9H0B6 Q9BW19 Q9NSK0 Q07866 Q12840 P43034 O60282	0.774%
rRNA maturation	1.21	GO:0000460 GO:0000470 GO:0030687	Q99547 Q9NPD3 O95478 Q9BYD6 Q9H9Y2 P56182 Q9Y3B9 Q9UKD2 P56537	0.410%
mRNA catabolic process	1.18	GO:0035278 GO:0035194 GO:0035068 GO:0070578 GO:0006402 GO:0090625 GO:0035198 GO:0035279 GO:0003727 GO:0016442 GO:0035280 GO:0035196 GO:0031054 GO:0010501	P26651 Q8TF46 Q9H9G7 P48431 Q9UL18 O95931 Q8NDV7 Q53H82 Q9HCJ0 P36776 P56537 Q9NR97 Q9HCK5 Q8ND56 Q8TCS8 Q7L014 O77932 Q13206	0.820%
Autophagosome activity	1.12	GO:0034497 GO:0000422 GO:0000045	Q5MNZ6 Q13501 Q674R7 P42771 Q9Y484 A6NCE7 O95140 Q9Y3D6 Q7Z3C6 Q9GZQ8 Q7KZI7	0.501%
Lipid catabolic process	1.11	GO:0004629 GO:0004435 GO:0016042	Q00722 P06858 P30542 Q6PIU2 O75038 Q8NCG7 P19174 P68402 Q15111 P43034 Q4KWH8 P30041 Q9NQ66 Q13393 Q9UPR0	0.683%
Spindle assembly	1.07	GO:0051297 GO:0070652 GO:0051225	Q9BT25 Q5TZA2 Q7Z4H7 Q12982 P36404 Q9P2Y5 Q15004 Q9P2P6 Q9H6D7	0.410%

Supplementary Table 20 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁺-exclusive antigens
Rho protein-mediated signal transduction	1.01	GO:0005089 GO:0035023 GO:0005085 GO:0051056	Q8N1W1 Q96PE2 A8MVX0 Q9H6S3 Q9Y4F1 Q5VST9 Q9NZN5 Q9NZM3 A1L390 Q8IVF5 Q12979 Q7Z6J4 Q8TE68 O15085 Q96N96 P98171	P52565 Q9HCU5 Q9NR50 Q8NFG4 Q14232 Q8N103 Q0VGL1 Q9NPQ8 Q9UHV5 Q9BRG2 Q13017 Q9UNA1 O43295 A1A4S6 O00459 Q9HBH0	1.457%

Supplementary Table 21: Functional annotation clustering for HPV⁺-exclusive HLA class II ligands. GO term enrichment was performed with the functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11] on the basis of HPV⁺-exclusive source proteins of HLA class II-presented peptides. An enrichment score ≥ 1 was used as threshold. HPV – human papilloma virus; GO – gene ontology.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁺-exclusive antigens
Cell-cell adhesion	2.15	GO:0005913 GO:0098609 GO:0098641	Q9UM54 Q01813 Q53SF7 P19022 Q14160 Q9Y2D8 Q9UQB8 Q96NY8 P25685 Q9UH65 Q15762 Q9NQW6 O00515 Q08378 Q96PY5 Q7KZF4 O60716 P42166 P50990 O95433 Q9UNF0 P61978 Q9Y266 Q14258 P42566	P42167 P07948 Q9H4G0 Q8WWI1 P35613 Q9BXI6 Q15019 Q9H2G2 O15020 P41217 Q9Y371 Q13796 Q9NYQ8 Q99961 O00299 O75874 Q9NWB6 Q96HC4 P50914 Q8N3F8 P78344 P16422 Q9Y5R2 Q9NZU0	2.950%
Autophagosome activity	2.09	GO:0000045 GO:0000421 GO:0006995 GO:0000422	Q9H0R8 Q9H492 P49768 Q86Y82 Q2TAZ0 Q9GZQ8 Q9Y2L5	Q7Z3C6 Q9UMX0 A6NCE7 Q8TDY2 O15321 Q9Y371 O95166	0.843%

Supplementary Table 21 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins	Proportion of HPV$^+$-exclusive antigens
Calcium ion homeostasis	2.04	GO:0005388 GO:1903779 GO:0006874	Q01814 Q15413 P23634 P57103 P20020 Q13733 P16615 P05026 Q93084 P08246 Q16720 Q05586 P23582 Q9H7F0 Q13061 P28335	0.963%
Telomere maintenance	1.99	GO:1901998 GO:1904871 GO:0002199 GO:0005832 GO:0044297 GO:1904874 GO:0032212 GO:1904851 GO:0051082 GO:0007339	P58335 Q9HB71 Q53ET0 O95166 Q9UJY5 Q9NX24 P48643 Q13315 O75581 Q9Y6R4 P50990 P09651 P50991 P25685 O14497 Q58FG0 O14786 P61604 Q99832 Q99733 Q93050 P50502 P49368 Q9NU22 Q8TEY7 Q9UBS4 Q9BT88 Q9NQP4 P04899 Q9Y266 Q9H0R8 Q13111 Q9BXM7	1.987%
Microtubule-mediated movement	1.99	GO:0003777 GO:0005871 GO:0007018	Q9Y6G9 Q9P2D7 Q9BVG8 O43896 O95239 O14782 O75037 Q9HAQ2 Q12840 Q96DT5 Q96Q89 Q7Z494 Q02224 Q0VDD8 Q15058 Q99962 Q9NS87	1.023%
Viral DNA recombination	1.93	GO:0015074 GO:0003964 GO:0004523 GO:0006278 GO:0090502 GO:0019028 GO:0006310 GO:0004190 GO:0019031	P63136 P63135 Q9UQG0 P39748 P63133 P07998 Q6R2W3 O00584 Q9QC07 Q5D1E8 P10266 Q06330 Q53H47 Q9H981 Q8NDZ6 Q9NZ71 Q9BXR3 P49768 P63128	1.144%
Rho protein-mediated signal transduction	1.62	GO:0005089 GO:0035023 GO:0005085	Q8N1W1 Q9ULL1 O94887 Q6ZV73 Q96PE2 Q13009 Q58EX7 O15068 O75962 Q12802 Q5JSL3 P98171 Q9Y4F1 Q9BXI6 Q96M96 O95398 Q96PX9 Q8TF40 Q9NZM3 Q96BY6	1.204%

Supplementary Table 21 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins	Proportion of HPV$^+$-exclusive antigens
Canonical Wnt signaling pathway	1.37	GO:0060070 GO:0005109 GO:1990909 GO:1904886	P24385 Q92997 Q8N474 P54792 Q9HCK8 Q86UU0 P18827 P04628 P56706 O75581 P49768 O14640 A4QPB2 Q06945 P34925 Q01974 O75197	1.023%
Wnt signaling pathway in axis specification	1.30	GO:0042813 GO:0017147 GO:0044332	O60353 P34925 Q8N474 O75197 A4QPB2 Q01974 O75581	0.421%
Transmembrane receptor protein tyrosine kinase signaling pathway	1.13	GO:0004714 GO:0018108 GO:0007169	Q16620 P21781 P36888 P54753 P35968 P26951 P21802 P08631 P21860 P06241 Q08345 P29323 P06213 Q9P2A4 P34925 P54756 Q01974 P07948 P32927 P15509 P22607 P23470	1.325%
ATP biosynthesis	1.09	GO:0042776 GO:0046933 GO:0015078 GO:0000276 GO:0005753 GO:0015986 GO:0006754	P10606 P05496 P48201 Q16864 Q06055 P21281 O75947 P15313 P56381 Q93050	0.602%
Viral gene expression	1.06	GO:0019013 GO:0010467 GO:0030529	Q32P51 P25445 P51991 Q9H4A6 O75643 Q13151 P08621 Q92499 Q14103 P62314 P61978 Q4G0J3 P09651 Q07352 P35269	0.903%

Appendix

Supplementary Table 22: Functional annotation clustering for HPV⁻-exclusive HLA class I ligands. GO term enrichment was performed with the functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11] on the basis of HPV⁻-exclusive source proteins of HLA class I binders. An enrichment score ≥ 1 was used as threshold. HPV – human papilloma virus; GO – gene ontology.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins	Proportion of HPV⁻-exclusive antigens
Protein phosphorylation	2.08	GO:0004672 GO:0006468 GO:0004674	Q8TDC3 P54646 P08581 Q9Y2U5 Q16654 Q9UK32 P10398 Q8IVH8 Q8TEA7 P45985 Q6P2M8 P15735 Q09013 P21860 O15111 Q13554 O15075 O75582 Q92519 Q99759 Q13523 Q9NYY8 Q04771 Q5VT25 Q15119 P22612 Q9P2K8 P37023 Q9UQM7 P42685 Q8IW41 O75676 P42345 Q92585 Q6IQ55 Q13976 P07949 Q96NX5 Q8IWQ3 Q6P3R8 P49137 P11801 Q92759 Q13557 Q96S38 P46019 Q7L8L6 Q9C098 Q76MJ5 Q8N568 Q16566 Q13555 Q496M5 Q6XUX3	4.412%
RNA-dependent DNA biosynthesis	1.48	GO:0003964 GO:0006278 GO:0004523 GO:0015074 GO:0090502 GO:0006310 GO:0019028 GO:0004190	P63135 Q6ZW49 Q96PK6 P78346 Q9QC07 Q6PI98 Q9UQG0 P0C7P3 P10266 P63136 Q9BXR3 P55210 Q9WJR5	1.062%
Peptidyl-tyrosine phosphorylation	1.36	GO:0004714 GO:0007169 GO:0018108 GO:0004713	P10721 P42685 P36888 Q15303 P08069 P21860 P35590 P07949 P08581 Q7Z4S9 P01732 Q6XUX3 Q7Z7A1 Q12805 Q9H3Y6 Q8TEW6 P35916 P29322 P45985	1.552%

Supplementary Table 22 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁻-exclusive antigens
Metabolic process	1.23	GO:0051552 GO:1904224 GO:2001030 GO:0045922 GO:0052697 GO:0001972 GO:0052695 GO:0004857 GO:0052696 GO:0042573 GO:0015020 GO:0016758 GO:0008152	Q9HAW9 Q9HAW7 Q9NR63 Q5TBA9 O15382 Q4L235 P28161	Q9NRZ5 Q9HAW8 Q99798 P15289 O60656 Q92685	1.062%
Chloride transport	1.08	GO:1902476 GO:0007214 GO:1902711 GO:0004890 GO:0034707 GO:0006821 GO:0005230 GO:0045211 GO:0005254 GO:0006811	P55017 O14764 P47869 P48751 P14867 P31644 P55011 Q9NQ90 P34903 Q32M45 Q96S66 Q9NW15 P02730	O00555 Q9UPX8 O75970 P42263 O43815 Q9HCI5 Q9NSC5 Q9BVT8 Q7Z7J9 Q14118 Q13621 Q8IVB4	2.042%

Appendix

Supplementary Table 23: Functional annotation clustering for HPV⁻-exclusive HLA class II ligands. GO term enrichment was performed with the functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11] on the basis of HPV⁻-exclusive source proteins of HLA class II-presented peptides. An enrichment score ≥ 1 was used as threshold. HPV – human papilloma virus; GO – gene ontology.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁻-exclusive antigens
Protein phosphorylation	4.08	GO:0006468 GO:0004674 GO:0004672	Q8NB16 Q9NR09 P36897 Q8IXL6 P36896 Q13043 Q9NYL2 Q5VT25 Q99717 Q8IZX4 P11274 P25092 P53667 Q38SD2 Q6ZWH5 Q9UPZ9 Q96SB4 P21127 Q9UQ88 Q9BX84 Q9BUB5 Q76MJ5 Q9Y2K2 Q96S53 Q6P3R8 Q6P0Q8 P37173 O75131 Q9P1W9 P53671 P07332 P21675 P50750 O14965 P16298 Q96Q40 Q8NG66	Q92519 P42681 P41743 Q86V86 Q15118 P04626 P51617 P28482 P51948 O15198 Q92585 O15197 Q02846 P23458 P51813 P19525 Q15131 P27361 Q05655 O00141 Q15796 P18074 P54646 O95382 P42336 Q9UPE1 Q8NER5 P46019 P31749 P11309 P36507 Q16539 Q5TCY1 Q02750 P22392 Q13535 Q9UL54	4.917%

Supplementary Table 23 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁻-exclusive antigens
Proteasomal protein catabolic process	2.02	GO:0038095 GO:0010950 GO:0000209 GO:0060071 GO:0050852 GO:0043488 GO:0051437 GO:0061133 GO:0002223 GO:0031145 GO:0051436 GO:0038061 GO:0000502 GO:0008537 GO:0006521 GO:0002479 GO:0000165 GO:0061136 GO:2000045 GO:0022624 GO:0090263 GO:0033209 GO:0043161 GO:0090090	P01781 P01780 P43686 P48556 P62837 P25787 Q9Y297 Q08209 P06311 P16298 O00487 Q13094 P61289 P28482 P55036 P01609 Q9UL46 Q06323 P27361 Q07889 P01769 Q9UNM6 Q9BXL7 P42336 P04207 P51665 P01767 P01116 O00459 P01742 P01743 P01111 P01112 Q9Y255 Q53H80 P51965 Q9NX47 Q9HAU4 Q9UBN7 Q9UK99 P21675 P42771 Q14669 P61077 O95376 Q14139 Q6UWE0 P10415 Q9H2K2 O95271 Q9UKV5 Q9NZS9 Q9ULV1 Q9NPG1 Q93074 Q14332 P32121 Q9UP38	O95782 Q9UJZ1 P42681 O00481 P09693 P01850 Q9BZS1 Q13868 P47974 Q9Y5A9 Q05655 Q06787 Q92945 Q9Y3B2 P31749 Q13257 P60006 O14965 Q9UM73 Q05086 P12314 Q02750 Q15797 Q16690 O15399 P04626 Q99748 O14827 P36507 P23458 Q13813 P10721 P14784 P56159 O00255 P02549 P15923 P29350 P46937 Q70CQ2 Q38SD2 P26842 O14763 Q9BRA2 Q9Y5U5 P20333 Q16236 O00635 P35222 P04637 Q13362 Q15366 Q9Y283 Q13043 P10071 Q9GZV5 Q16539	7.641%

Appendix

Supplementary Table 23 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁻-exclusive antigens
Ubiquitination	1.70	GO:0070936 GO:0061631 GO:0070534 GO:0070979	P51965 O95376 Q05086 Q9UKV5 P62837 O76064 O00635 Q9NZS9	Q9Y2X8 Q969T4 Q96LR5 P61077 Q9NR09 P21675 P42771	0.997%
Protein quality control in endoplasmic reticulum	1.63	GO:1904153 GO:0030433 GO:0044322	Q96Q80 Q8NBM4 Q13438 Q96DZ1 Q9GZP9 Q8TCJ2	P43686 Q9UKV5 Q14139 Q96A33 Q9BV94	0.731%
Autophagosome activity	1.51	GO:0071203 GO:0034314 GO:0042147 GO:0055038 GO:0043014 GO:0016197 GO:0055037 GO:0031901 GO:0005776	A8MWX3 Q6VEQ5 C4AMC7 Q9NQA3 A8K0Z3 Q9Y4E1 O15144 O15145 O15511 Q13596 Q13190 Q641Q2 O15498 O94955 Q9UBQ0 O14653 Q6P5W5 Q96DZ5 Q9NZN4 Q9NZN3	Q9UBN7 Q9NVJ2 Q5T2P8 P42857 Q5VT79 Q8TCU4 P13928 Q9H9H4 Q9Y3R0 Q8IWF6 Q9Y6D5 P61018 Q9Y2H2 Q8TF17 O43747 P45844 Q96L93 P12314 Q96F24	2.591%
Translation initiation	1.54	GO:0001731 GO:0003743 GO:0006446 GO:0016281 GO:0006413 GO:0008135 GO:0071541 GO:0016282 GO:0033290 GO:0005852	Q15056 O15372 O15371 Q2VIR3 Q14152 P55010 Q9UBQ5 P41091 O43432 Q13144	Q14240 Q9UI10 P60842 P05386 P35268 P25398 P62273 P62888 P61247 Q92616	1.329%
Metabolic process	1.54	GO:0018636 GO:0071395 GO:0016655 GO:0044597 GO:0044598 GO:0047086 GO:0047718 GO:0047115 GO:0001758 GO:0004032 GO:0004033 GO:0042448 GO:0008202 GO:0001523 GO:0007586	P52895 P05181 P42330 Q04828 P17516 P16152	O60218 Q14849 P06858 P32239 Q8NHM4	0.731%

Supplementary Table 23 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins	Proportion of HPV⁻-exclusive antigens
Fatty acid biosynthesis	1.41	GO:0043651 GO:0036109 GO:0006636 GO:0035338 GO:0042761 GO:0102336 GO:0009922 GO:0102338 GO:0102337 GO:0034625 GO:0019367 GO:0034626 GO:0030148	P33897 O00767 Q9BW60 Q9NZ01 P08263 Q53H12 Q9NXB9 Q96CM8 O95864 O60488 P33121 O14495 Q9NYP7 Q6ZMG9 O15121 Q8IWX5	1.063%
Regulation of ERK cascade	1.41	GO:0060440 GO:0090170 GO:0032872 GO:2000641 GO:0060020 GO:0070371 GO:0060324 GO:0048538 GO:0030878 GO:0005770 GO:0070374	P37173 P84022 P28482 Q7Z2W4 P27361 Q6VEQ5 Q02750 Q9UBQ0 P35222 O75129 P36507 Q9GZP9 O14495 Q8TEB7 P05019 A8K0Z3 P47974 Q8NBP7 Q9NPJ1 P32121 O76080 Q99731 P61968 P55774 P36897 P01112 P10415	1.794%
Complement activation	1.39	GO:0003823 GO:0006958 GO:0006956	P01781 P04207 A6NJ16 Q9BZM6 P01609 P01767 P01780 P01880 P01591 P01742 Q8NHJ6 P01743 Q6H3X3 Q9BUD6 Q8TD07 P13612 P06311 P07357 O14817 P15529 P01769 P06681	1.462%
Telomere maintenance	1.38	GO:0060440 GO:0032212 GO:1904355 GO:0051973	P37173 P35222 P28482 O95271 P27361 Q13535 Q02750 Q9H2K2	0.532%
Cellular sphingolipid homeostasis	1.34	GO:0090156 GO:1900060 GO:0035339 GO:0006672	Q9P0S3 Q8N138 Q53FV1	0.199%
DNA damage response	1.18	GO:0010165 GO:0010332 GO:0051262 GO:0034644 GO:1900740 GO:0031571 GO:0042771 GO:0006978	P04637 Q86WG5 Q9H3D4 Q9BX84 P49917 P34897 P07992 P04424 P43246 Q16630 O15350 Q13535 P07204 P21675 P54098 Q06787 Q8IWE4 O60341 O00255 P30419 P10415 Q13362 P08069	1.528%

Appendix

Supplementary Table 23 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of HPV⁻-exclusive antigens
SMAD protein binding	1.17	GO:0046332 GO:0070411 GO:0010718	O95373 P36897 Q15796 P36896 P37173 Q9HAU4	Q99081 P35222 Q15797 Q9GZV5 P84022	0.731%
MAP kinase activity	1.17	GO:0051090 GO:0030278 GO:0031663 GO:0004707	O15360 P28482 P27361 Q16539 Q8N1H7 Q93009	O00141 P09603 P51617 P25105 P31749	0.731%
Regulation of muscle contraction	1.17	GO:0005861 GO:0030049 GO:0003009 GO:0006937	P48788 P19237 P45378 P13805 P20929	Q9UKX2 P14649 Q9Y623 Q96MF2	0.598%
SMAD protein signal transduction	1.17	GO:0030618 GO:0071141 GO:0007179 GO:0060395 GO:0007183 GO:0031053 GO:1901522 GO:0001657 GO:0070410 GO:0030509 GO:0009880 GO:0051216	Q15796 Q15797 Q99717 P84022 O15198 P55107 P36897 P37173 P32121 P33151 P49023 P43026	P09958 P08476 Q8N1F7 O94813 Q9HAU4 P27361 Q12841 A6ND36 P61077 P02458 Q9NPJ1 Q16665	1.595%
Protein processing	1.16	GO:0016486 GO:0016485 GO:0008233	Q92824 Q16549 P29122 P29120 P09958 Q13362 O15072 P10071	Q96SM3 Q86WA8 P15169 Q9UKP5 Q9UHP3 Q96FZ2 O96009	0.997%
Cellular response to heat	1.12	GO:0000189 GO:0070849 GO:1900034	P12270 P28482 P27361	Q13535 Q8WUM0 Q8N1F7	0.399%
Lipid antigen presentation by HLA	1.02	GO:0030884 GO:0030883 GO:0048007 GO:0071723 GO:0030881	P29017 P29016	P15813	0.199%
Regulation of protein phosphorylation	1.02	GO:0019888 GO:0008601 GO:0000159	Q13362 P19525 O75807 Q9UQ13	Q16537 Q15173 Q9NY27 P30154	0.532%
mRNA transcription termination and export from nucleus	1.01	GO:0006405 GO:0031124 GO:0006406 GO:0006369	P84103 P12270 Q9NZP6 Q08170 Q13247 O60508 P43487 Q96FV9 P38919	O94913 Q10570 Q8WUM0 Q8N1F7 O75494 Q96QD9 P63241 P62318	1.130%

Supplementary Table 24: Functional annotation clustering for up-regulated genes. GO term enrichment was performed with the functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11]. The analysis was based on the proteins that were encoded by the genes up-regulated in HPV^+ OPSCCs compared to HPV^- OPSCCs. An enrichment score ≥ 1 was used as threshold. HPV – human papilloma virus; GO – gene ontology.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of up-regulated genes
Chemokine-mediated signaling pathway	3.28	GO:0070098 GO:0004950 GO:0006935 GO:0016493 GO:0019957	O00421 P51684 P51681 P32302 O43927 O00590 O00574 P46092	Q9UBD3 P49238 P49682 Q92608 P16150 O00182 P55160	2.703%
Protein tyrosine kinase signaling pathway	2.58	GO:0007169 GO:0038083 GO:0004715 GO:0031234 GO:0004713	Q08881 P01732 P51451 P36888 P06239 Q9ULZ2 P43403 Q6PIZ9 P09564	Q06187 P07766 P52333 O60674 P34972 Q9Y463 Q96S53 P42229	2.872%
Melotic cohesin complex	2.38	GO:0000800 GO:0030893 GO:0034991	Q9BX26 Q8NDV3 Q9UJ98	O95072 Q8N6L0	0.845%

Appendix

Supplementary Table 24 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of up-regulated genes
DNA transcription	1.93	GO:0003700 GO:0006351 GO:0003677	Q9UL17 Q03112 Q12857 Q02556 Q15306 P51449 Q9BY31 Q96RE9 O95936 P42229 Q9UKT9 Q13342 Q3KNS6 Q13469 Q8N4W9 Q13422 P48382 P15863 Q9BZS1 Q99592 P51786 Q8NDQ6 Q14585 Q05066 Q9BX82 Q16520 Q96FA7 Q01101 Q6PF04 Q8TAQ5 P51522 Q8WXB4 Q96SK3 O75398 P15498 Q8NA19 P08235 Q7RTV3 Q8NHY6 Q92753 Q01892 Q9H0D2 Q02446 Q5JPB2 Q14188 P57071 P51531 Q8NE65 Q8N371	Q5HYK9 Q9C0A1 O94993 Q8TAU0 P42771 Q06187 Q8IZ20 O95931 Q9NV72 P02788 Q6ZN06 Q6BEB4 P12980 P05771 A6NNA5 P48023 Q9H165 Q15910 Q6ZNG0 O75475 Q8N726 Q8N5J4 Q86YH2 Q9UHJ3 Q460N3 P33076 O75346 Q8N228 Q3SXZ3 Q9BR84 Q9NXR8 Q14005 Q8IYN0 Q5U623 Q96GE5 Q9Y2A4 O14746 Q9NS87 Q9Y2T7 Q5H9K5 P05204 Q16633 Q3KP22 Q9NS91 Q9BX26 O00470 Q9H967 Q8NDV3 Q96E14	16.554%
Phosphatidylinositol-mediated signaling	1.40	GO:0046934 GO:0046854 GO:0014066 GO:0048015	P06239 Q5UE93 P15498 Q6PIZ9 P15391	P48736 P10747 P49441 Q96T49	1.520%

Supplementary Table 24 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of up-regulated genes
Cell-cell adhesion	1.29	GO:0050839 GO:0007157 GO:0008037 GO:0007156 GO:0005913	O95727 P19320 P29350 P20701 P26010 Q495A1 Q15762 P14151	P30203 P06127 Q9ULB5 Q9NYQ7 Q9Y6N7 Q9Y2D8 P16333	2.534%

Supplementary Table 25: Functional annotation clustering for down-regulated genes. GO term enrichment was performed with the functional annotation tool of DAVID Bioinformatics Resources 6.8 [10, 11]. The analysis was based on the proteins that were encoded by the genes down-regulated in HPV^+ OPSCCs compared to HPV^- OPSCCs. An enrichment score ≥ 1 was used as threshold. HPV – human papilloma virus; GO – gene ontology.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of down-regulated genes
Proteolysis	3.83	GO:0022617 GO:0004222 GO:0004252 GO:0006508	P02751 Q13753 Q16787 Q13751 Q9Y5K2 P51512 P45452 P35556 P09238 P07585 P16035 P08254 P08253 O43897	Q92743 P03956 Q12884 Q9ULZ9 Q9BXP8 O95450 O43184 P58397 P59510 Q8TE58 O95084 P00749 P56817	4.704%
Keratinization	3.30	GO:0018149 GO:0030216 GO:0031424 GO:0001533	P02751 O95932 Q5T5B0 Q9BYE4 P35325 P00488 Q9BYE3 P07996	Q5TA81 Q5TA82 Q86SJ6 O14944 P46937 P42330 Q07283	2.613%
Cytokine-mediated signaling pathway	2.07	GO:0046426 GO:0006469 GO:0019221 GO:0004860	Q6PEZ8 O43155 P07585 Q9NZU0 P52926 P21810 O75553 Q03135	P09211 P15090 Q8WVI7 Q8NI17 O14944 Q16849 Q8WWZ1	2.613%
Protein folding	1.98	GO:0005528 GO:0061077 GO:0003755 GO:0000413 GO:0006457	P68106 Q9NWM8 O95302 Q96AY3	Q75LS8 Q9Y680 P02511	1.220%

Appendix

Supplementary Table 25 continued.

Biological process of annotation cluster	Enrichment score	GO IDs	UniProtKB accessions of source proteins		Proportion of down-regulated genes
Metabolic process	1.91	GO:0044597 GO:0044598 GO:0047086 GO:0047718 GO:0047115 GO:0018636 GO:0071395 GO:0004032 GO:0016655 GO:0042448 GO:0001523 GO:0004033 GO:0008202 GO:0007586	P52895 O60218 P42330 Q04828	Q9Y625 P02753 Q16518	1.220%
Chondrocyte differentiation	1.78	GO:0002062 GO:0021983 GO:0007224	Q03431 P10070 Q5TA89 P52926 P12644	Q15582 Q9NSC2 P08151 A7MBM2	1.568%
Amino acid transport	1.76	GO:0006865 GO:0015175 GO:0015804	Q96QD8 P08195 Q9UHI5	Q969I6 Q01650 Q9UPY5	1.045%
Microfilament motor activity	1.72	GO:0000146 GO:0003774 GO:0016459 GO:0005516 GO:0030898 GO:0030048	Q9Y4I1 P35580 Q9UKX2 O43795 O94832 Q8NCM8	P12882 P35612 Q8TC29 Q13936 Q13033 P41220	2.091%
Bone morphogenetic protein signaling	1.52	GO:0071773 GO:0030501 GO:0030509	Q04771 O00238 P58397 P12644 Q9NQ87	P35556 Q5ZPR3 Q86SJ6 Q12841	1.568%
Glycosaminoglycan metabolic process	1.31	GO:0030206 GO:0043202 GO:0030208 GO:0030203 GO:0005796 GO:0030207 GO:0005539	P07585 P21810 P13611 Q8IZ52 Q9Y625	P51884 P09619 P34059 P51512 P01138	1.742%
Epidermal growth factor receptor signaling pathway	1.15	GO:0045740 GO:0005154 GO:0007173	P15514 P10070 O14944 P08151	P29353 P08648 P56945	1.220%
Regulation of cell adhesion and migration	1.00	GO:0045995 GO:0030155 GO:0030334	Q16787 Q16363 P24043	P42684 Q9NQC3 P09238	1.045%

8.4 References

[1] Krini M, Chouliaras G, Kanariou M, Varela I, Spanou K, Panayiotou J, Roma E, and Constantinidou N. HLA class II high-resolution genotyping in Greek children with celiac disease and impact on disease susceptibility. *Pediatr Res* 72(6): 625–630 (2012).

[2] Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, Cid-Pavon GMD, Ramsbottom K, Ghattaoraya G, Alfirevic A, Middleton D, and Jones AR. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 48(D1): D783-D788 (2020).

[3] Nelde A, Kowalewski DJ, and Stevanović S. Purification and Identification of Naturally Presented MHC Class I and II Ligands. *Methods Mol Biol* 1988: 123–136 (2019).

[4] Käll L, Canterbury JD, Weston J, Noble WS, and MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4(11): 923–925 (2007).

[5] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, and Nielsen M. NetMHCpan 4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* (2017).

[6] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45(6): 580–585 (2013).

[7] Thul PJ and Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci* 27(1): 233–244 (2018).

[8] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang G-D, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME,

McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, and Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218): 53–59 (2008).

[9] Michael Love SA. DESeq2: Bioconductor, 2017.

[10] Huang DW, Sherman BT, and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1): 1–13 (2009).

[11] Huang DW, Sherman BT, and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1): 44–57 (2009).

[12] Hulsen T, Vlieg J de, and Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 9: p. 488 (2008).

[13] Bardou P, Mariette J, Escudié F, Djemiel C, and Klopp C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15: p. 293 (2014).

[14] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, and Stevanović S. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics* 50(3-4): 213–219 (1999).

[15] Andreatta M and Nielsen M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* 32(4): 511–517 (2016).

[16] UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res* 35(Database issue): D193-7 (2007).

[17] Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, and Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res* 48(D1): D948-D955 (2020).

[18] Rapin N, Hoof I, Lund O, and Nielsen M. The MHC motif viewer: a visualization tool for MHC binding motifs. *Curr Protoc Immunol* Chapter 18: Unit 18.17 (2010).

[19] Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-T, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJG, Old LJ, Camargo AA, and Vasconcelos ATR. CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37(Database issue): D816-9 (2009).