Circulating Biomarkers for Cancer Immunoprofiling

by

Radwa Mohamed Emadeldin Mahmoud Ewaisha

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2018 by the
Graduate Supervisory Committee:

Karen S. Anderson, Chair
Joshua LaBaer
Douglas F. Lake
Valerie Stout

ARIZONA STATE UNIVERSITY

December 2018

ABSTRACT

Biomarkers find a wide variety of applications in oncology from risk assessment to diagnosis and predicting and monitoring recurrence and response to therapy. Developing clinically useful biomarkers for cancer is faced with several challenges, including cancer heterogeneity and factors related to assay development and biomarker performance. Circulating biomarkers offer a rapid, cost-effective, and minimally-invasive window to disease and are ideal for population-based screening. Circulating immune biomarkers are stable, measurable, and can betray the underlying antigen when present below detection levels or even no longer present. This dissertation aims to investigate potential circulating immune biomarkers with applications in cancer detection and novel therapies. Over 600,000 cancers each year are attributed to the human papillomavirus (HPV), including cervical, anogenital and oropharyngeal cancers. A key challenge in understanding HPV immunobiology and developing immune biomarkers is the diversity of HPV types and the need for multiplexed display of HPV antigens. In Project 1, nucleic acid programmable protein arrays displaying the proteomes of 12 HPV types were developed and used for serum immunoprofiling of women with cervical lesions or invasive cervical cancer. These arrays provide a valuable high-throughput tool for measuring the breadth, specificity, heterogeneity, and cross-reactivity of the serologic response to HPV. Project 2 investigates potential biomarkers of immunity to the bacterial CRISPR/Cas9 system that is currently in clinical trials for cancer. Pre-existing B cell and T cell immune responses to Cas9 were detected in humans and Cas9 was modified to eliminate immunodominant epitopes while preserving its function and specificity. This dissertation broadens our understanding of the immunobiology of cervical cancer and provides insights into the immune profiles that could serve as biomarkers of various applications in cancer.

To Mom,
for being the reason for every good thing that has ever happened to me;

To Dad,
for raising me to believe that everything is possible.

ACKNOWLEDGMENTS

I would first like to acknowledge my dissertation committee for their support and guidance. I would like to thank my advisor Dr. Karen Anderson for the truly transformative opportunity of working with her and being part of her incredible lab environment and for showing me how mentorship should be done. I would like to thank Dr. Josh LaBaer for his encouragement that made many days easier and for his outstanding leadership of the Biodesign Institute and the Center for Personalized Diagnositcs that I am fortunate to call my second home. I would like to thank Dr. Doug Lake for his honest advice and his support and Dr. Valerie Stout for always being there for me and her mentorship during both my research and teaching.

During this journey, I was blessed to have learned, studied, and grown alongside some of the best budding scientists and to have made invaluable lifelong friendships. I thank Shay Ferdosi for making my journey happier, easier, more productive, and more fun. I am so grateful I got to share the office and lab with a true friend who was always there for me, Peaches Ulrich. I thank Rowida Abdelgalel, Emma Chen, Sri Krishna, Diego Chowell, and Sandy Hou for their support and friendship. Special thanks to Benjamin Katchman, Marika Hopper, and Padhma Yuvaraj for their support in lab. I would also like to thank Justin Wolter, Heather Geissel, Kasuen Kotagama, Anasuya Pal, Jordan Yaron, Lusheng Song, and all members of BCPD for all their help and the wonderful memories. I am so thankful for my dear friends Maie Elkeshky, Youmna Ahmed, Nesreen Kandil, Aya El Nahas, and Rawa Awad for all their love and support. I am so lucky to have them in my life.

This research was only made possible through phenomenal collaborations over the years. I would like to thank Drs. Beth Unger, Gitika Panicker, Erich Sturgis, Jennifer

Blain Christen, Neerja Bhatla, Dean Brenner, Mack Ruffin, and Samira Kiani. I acknowledge Jen Van Duine, Lisa Miller, Eric Pacheco, and all members of the NAPPA and DNASU teams for their dedication and for always being willing to help. I thank Ian Meshay and Jack Resnik the former Barret's Honors students who worked on some of these projects before I started. I am also grateful to the undergraduate students that I was lucky to have either mentored or worked with over the years, Tirin Bharaj, Meredith Smith, and Meilin Zhu.

I would like to acknowledge the generous student and funding support from the School of Life Sciences, the Graduate College, the Graduate and Professional Students Association, the Harry Lowell Swift Award, the Edward and Linda Birge Award, and the Associated Students of ASU Council of Presidents.

I have been truly blessed with two loving and supportive brothers that mean the world to me. I could not have done any of this without Ahmed. He helped me with every step of this journey; from grad school application and moving to the US to sharing difficult moments, celebrating my achievements, being an amazing roommate, a caring friend, and an inspiring role model. I wish to thank Essam for keeping our close relation and friendship even though I am no longer home. Many thanks for our thought-provoking conversations, for inspiring me and teaching me many life lessons, and for his endless love and support.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

AAb	Autoantibody
Ab	Antibody
Ag	Antigen
ANN	Aritificial Neural Network
APC	Antigen Presenting Cell
BCM	B Cell Media
Cas9	CRISPR-Associated protein 9
CEA	Carcinoembryonic Antigen
CIN	Cervical Intraepithelial Neoplasia
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTC	Circulating Tumor Cell
DMSO	Dimethyl Sulfoxide
EBNA-1	Epstein-Barr Virus Nuclear Antigen-1
EBV	Epstein-Barr Virus
EDRN	Early Detection Research Network
EGFR	Epidermal Growth Factor Receptor
ELISpot	Enzyme-Linked Immunospot
GST	Glutathione S Transferase
HLA	Human Leukocyte Antigen
HPV	Human Papillomavirus
HPVOPC	Human Papillomavirus-Associated Oropharyngeal Cancer
HR	High Risk

hr	hour
ICC	Invasive Cervical Cancer
IEDB	Immune Epitope Database and Analysis Resource
IVTT	*In Vitro* Transcription/Translation
LMIC	Low and Middle-Income Country
MAb	Monoclonal Antibody
MHC	Major Histocompatability
NAPPA	Nucleic Acid Programmable Protein Arrays
NCI	National Cancer Institute
PBMC	Peripheral Blood Mononuclear Cell
PBS	Phosphate-Buffered Saline
POC	Point of Care
PTM	Post-Translational Modification
RAPID ELISA	Rapid Antigenic Protein In situ Display ELISA
RLU	Relative Light Units
S_b	Binding Score
SD	Standard Deviation
S_i	Immunogenicity Score
TNBC	Triple Negative Breast Cancer
WT	Wild Type

CHAPTER 1 : INTRODUCTION

The study of cellular pathways at the molecular level has revolutionized our understanding of mechanisms, markers, and classification of disease. Many conditions previously perceived as a single disease are now recognized to have distinct molecular patterns of perturbation. Personalized medicine envisions the delivery of treatments tailored to individual molecular disorders. This vision requires a transformation in the approaches for the discovery of reliable molecular markers of disease. The incorporation of the molecular changes observed in disease into clinical practice necessitates establishing validated correlations with clinical usefulness.

The topic of biomarkers has been one of the most exciting applications of the 'omics' technologies. It has raised hopes for the realization of precision medicine, thus improving healthcare quality and reducing treatment costs. For example, routinely screening colon cancer patients for K-RAS mutations will spare unresponsive patients costly and potentially toxic treatments with EGFR inhibitors and save at least US$600 million annually as estimated by the American Society of Clinical Oncology (Behl et al., 2012). Former US President Barack Obama announced the Precision Medicine Initiative during the 2015 State of the Union Address (Obama, 2015). The National Institutes of Health (NIH) leaders further explained the initiative goal – to combine large scale clinical data and biomarker measurements (Collins & Varmus, 2015; NIH, 2015). The number of large programs established by the NIH for biomarker discovery and validation has exceptionally expanded in the last decade. The excitement is reflected in the huge investments made in the field by both public and private companies, with the global cancer biomarkers market projected to reach US$27.63 billion by the end of 2025 from US$10.25 billion in 2016 ("Transparency Market Research," 2017).

This chapter provides background about biomarker applications in cancer and the process of biomarker discovery and validation, with a review of approaches for their identification. Focus is given to the current status of biomarkers of HPV-associated cancers, particularly cervical cancer, and their clinical utility. Finally, unanswered questions that constitute the topics of later chapters of this thesis as well as dissertation contributions are discussed.

1.1. WHAT ARE BIOMARKERS?

Biomarker is a portmanteau of biological marker and it is a biometric measurement that objectively indicates a medical state of the subject being tested. The state indicated could be related to disease incidence, outcome, response to interventions, and even unintended environmental exposure. The World Health Organization stated a broader definition that includes “almost any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical, or biological. The measured response may be functional and physiological, biochemical at the cellular level, or a molecular interaction” ("Biomarkers and Risk Assessment: Concepts and Principles," 1993). The National Cancer Institute defines biomarkers as “a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease.”

Biomarkers are biometric measurements that can range from pulse, blood pressure, and medical imaging techniques to quantitative analyses of genomic or proteomic analytes combined into mathematical models (Paik et al., 2004). The significance of this measurement is its capacity to distinguish between two biological states (LaBaer, 2005). The fundamental issue is to identify a reliable relationship between the biomarker measurement and a relevant clinical endpoint (Strimbu & Tavel, 2010). Often times, with enough variables, computational techniques may segregate two groups under

investigation only by random chance. Thus, it is important to recognize that establishing the distinction ability of the biomarker is not trivial.

Clinical endpoints are characteristics that indicate how the individual in a study “feels, functions, or survives” (Atkinson et al., 2001). Biomarkers, on the other hand, may or may not reflect what the subject feels or correlate with their clinical state. The ultimate goal of clinical practice and research is related to clinical endpoints (improving morbidity and mortality) rather than analytical measurements of biomarkers (Strimbu & Tavel, 2010). Thus, there needs to be enough evidence that a biomarker precisely and reliably predicts clinical outcome for it to be regarded as a surrogate endpoint, a substitute for a clinical endpoint (Atkinson et al., 2001). This brings up the question of what it takes to develop validated, reliable, and clinically useful biomarkers, which is discussed in detail in Chapter 2.

Some analyte measurements are designated biomarkers before their reliability and utility have been properly validated. A valid biomarker is “a biomarker that is measured in an analytical test system with well-established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the physiologic, toxicologic, pharmacologic, or clinical significance of the test results (FDA, 2005)”. Validation refers to the process of verifying the performance of a biomarker and the conditions required for reproducible and accurate prediction (J. W. Lee et al., 2006; Wagner, 2002). The term ‘evaluation’ has replaced validation to avoid the connotation that a relation between the biomarker and the underlying biological process or clinical endpoint needs to be established and fully understood (Atkinson et al., 2001; Strimbu & Tavel, 2010). ‘Qualification’ has recently replaced ‘evaluation’ and it refers to “the evidentiary process of linking a biomarker with biological processes and clinical end points (Hunter et

al., 2010; Wagner, Williams, & Webster, 2007)." It is not uncommon, however, for these three terms to be used interchangeably in the scientific literature.

1.2. APPLICATIONS OF BIOMARKERS IN CANCER

Biomarkers find a wide variety of applications in oncology from risk assessment and screening seemingly healthy individuals to diagnosis, determining prognosis, and predicting and monitoring recurrence and response to therapy. Some biomarkers have more than one clinical application and can be used for more than one type of cancer. Some of these potential applications are detailed below.

1.2.1. Risk Assessment

Primary prevention and early detection of cancer are regarded as the most successful approaches that can significantly reduce the overall burden of and mortality due to cancer (Etzioni et al., 2003; Jemal et al., 2004). One of the classical cancer risk models classifies the population into three groups: normal without cancer, asymptomatic cancer (can be detected and potentially treated), and symptomatic cancer (Zelen, 1993). Advances in the technologies for the identification of risk factors for each cancer type and even subtype has led to the development of more refined models for clinical use (NCI, 2018).

Inherited genetic abnormalities and environmental factors can change an individual's cancer risk over time. However, detection of markers of these factors is only useful if a high likelihood of developing cancer is known to occur in the presence of this marker (i.e. penetrance) and a potential intervention has been demonstrated to be effective (Calzone, 2012; X. Li, Blount, Vaughan, & Reid, 2011). Genetic biomarkers of cancer risk can be classified into high, moderate, and low penetrance according to the probability of the occurrence of cancer when the risk biomarker is detected (Calzone, 2012). The most well recognized high penetrance susceptibility genes are BRCA1 and BRCA2 mutations for

breast, ovarian, and other cancers (S. Chen & Parmigiani, 2007). A woman with this germline mutation and a family history of ovarian and/or breast cancer could benefit from more close monitoring, chemoprevention, or prophylactic surgery (Domchek et al., 2010). Moderate and low-penetrance markers are useful for research purposes, but more data are required, particularly from diverse ethnic populations, before these markers could be implemented in the clinical decision (Apostolou & Papasotiriou, 2017; Choi, Kipps, & Kurzrock, 2016; Sud, Kinnersley, & Houlston, 2017).

1.2.2. Early Detection

Disease diagnosis is the most explored application of biomarkers. Early detection is of particular interest in cancer given its potential for aiding in successful intervention before the patient succumbs to metastasis. There is an interest in developing minimally invasive biomarker tests that indicate disease long before the tumor is large enough to be observed for screening of seemingly healthy individuals. Biomarkers for cancer prognosis, progression, and response to therapy are currently in clinical use. However, finding biomarkers with enough sensitivity and specificity to find utility as population screens has been a challenge. One example is the commonly used but controversial prostate specific antigen (PSA) for the early detection of prostate cancer. Its implementation for screening of men over the age of 50 led to an increase in prostate cancer detection (Catalona, Smith, Ratliff, & Basler, 1993). However, recent analyses found insufficient evidence for the mortality benefit of early detection (Kim & Andriole, 2015).

For this purpose, the National Cancer Institute Early Detection Research Network (NCI EDRN) was developed for biomarker discovery and validation for cancer and cancer risk assessment (NCI, 2000). The EDRN allowed the collaboration of a network of scientists from both the academia and the industry for meticulous implementation of systematic evidence-based biomarker research for cancer screening and early detection. The result

is rigorous validation studies, prioritizing hundreds of biomarkers, and discontinuation of unpromising candidates from further development (Srivastava, 2013). The expectations of screening biomarkers are higher than those for biomarkers used in established patients and a narrower window of variation in measurements is allowed. Thus, combining a panel of biomarkers or combining a biomarker with other diagnostic tests could lead to optimal performance and certainty (LaBaer, 2005).

1.2.3. Disease Monitoring and Prognosis

Using biomarkers for monitoring of cancer progression has some advantages over other applications. Because most of these biomarkers are either normal molecules in the body that have abnormal changes or the immune response to these abnormalities and due to the heterogeneous nature of cancer, these biomarker measurements can vary greatly between individuals. This makes it difficult to establish a one-size-fits-all cutoff for population screening biomarkers. However, for patients already diagnosed with cancer, establishing a correlation between the biomarker level and disease severity for a given patient can help detect or predict disease progression, regardless of the absolute quantity of the biomarker.

A well-recognized example is the carcinoembryonic antigen (CEA) that correlates with disease progression in many but not all colon cancer patients (Fakih & Padmanabhan, 2006). If this correlation has been established for a specific patient through other means such as monitoring tumor size through imaging in relation with the biomarker level, then serial measurements of CEA over several visits can be informative of the cancer progression status. Beta-HCG, alpha feto-protein, and lactate dehydrogenase are similarly used for the early detection of nonseminomatous germ cell tumor recurrence through serial analysis (Gilligan et al., 2010). Additionally, a biomarker detected at higher

than normal levels but that has been stable in a given patient could indicate stable disease and success of therapy, despite being abnormally elevated (LaBaer, 2005).

1.2.4. Predicting Response to Therapy and Subtype Classification

Somatic mutations in KRAS are associated with poor response to the epidermal growth factor receptor (EGFR) inhibitors in colorectal cancers (Allegra et al., 2009). Estrogen receptor overexpression in breast cancer predicts response to the anti-endocrine drug Tamoxifen (Davies et al., 2011). Response to the anti-HER2 monoclonal antibody Trastuzumab could be predicted by HER2 overexpression in breast and gastric cancers (Bang et al., 2010). The overexpression of these proteins represents a basis for subtype classification within a given type of cancer that was once thought to be a single disease. This subtype classification influences therapeutic decisions.

Predicting success of a therapeutic agent is one of the costliest aspects of the drug development process. Pharmaceutical companies are interested in investing in biomarker tests that predict response to therapy before patients are put on drugs that they do not respond to, could lead to adverse events, or are expensive. Markers of ineffective or toxic lead compounds are useful for abandoning further research early on before huge investments are made. The field of pharmacogenomics focuses on studying how the genetic makeup of an individual dictates toxicity due to a certain therapeutic agent. For example, homozygosity for the UGT1A1*28 has been associated with increased risk of toxicity including severe neutropenia to the topoisomerase I inhibitor irinotecan used in colorectal and small cell lung cancers. This led the US Food and Drug Administration to change the labeling for this drug (Innocenti & Ratain, 2006).

Adverse reactions to a certain treatment can also be a function of the individual immune response and exposure history, rather than an inherited somatic gene. Pre-existing immune response to various components of gene therapies has been investigated as a

marker of a potential adverse immune reaction to these therapies (Brunetti-Pierri & Ng, 2009; Halbert, Standaert, Wilson, & Miller, 1998; H. Jiang et al., 2006; Kay, 2011; Nathwani et al., 2011; Nayak & Herzog, 2010). In this thesis, individual immune profiles to the CRISPR/Cas9 gene therapeutic is studied with the goal of exploring possible adverse reactions or therapy failure in certain individuals.

1.3. CIRCULATING BIOMARKERS: A WINDOW TO DISEASE

There has been increasing interest in developing circulating biomarkers for cancer, more so with the recognition of the concept of “liquid biopsy”. The blood has the potential to provide a rapid, cost-effective, and minimally invasive window to molecular changes both in the tumor and at distant metastatic sites. Circulating biomarkers may arise from direct shedding or secretion of tumor cells, proteins, nucleic acids, or subcellular components such as exosomes into the bloodstream (Bardelli & Pantel, 2017; O'Driscoll, 2015). The immune response to the tumor microenvironment also gives rise to potential biomarkers in circulation such as cytokines or autoantibodies (AAbs) (Anderson & LaBaer, 2005).

It is now better appreciated that tissue biopsies do not fully capture the full landscape of tumor heterogeneity and cancer evolution, the main cause of resistance to therapy (Y. Wang et al., 2014). On the other hand, circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA) shed by tumor cells into the blood give insights into the genomic landscape and overall disease including distinct metastatic sites (Bettegowda et al., 2014; Murtaza et al., 2015). Analyses of these changes can be used for cancer risk assessment, early detection, or monitoring minimal residual disease and response to treatment (Siravegna, Marsoni, Siena, & Bardelli, 2017).

1.4. HARNESSING THE IMMUNE SYSTEM

Besides abnormal molecules contained or released by tumors, the immune response to these molecules is of particular interest in the field of biomarkers. In comparison with other plasma proteins, antibodies (Abs) are highly stable, specific, and can be detected in minimal volume of patient serum using rigorous immune assays by probing with the specific protein. Detection of the cellular immune response is more labor intensive as discussed later in this chapter. The advances in epitope prediction algorithms and high throughput single cell technologies have expedited the process of T cell immunoprofiling (Fleri et al., 2017).

Our immune system is able to distinguish self from non-self antigens, although aberrant responses are sometimes manifested as autoimmune diseases. An immune response to self-proteins or glycans is often detected in cancer patients. This is because these Ags no longer look familiar due to one or more abnormal changes from the native form or the noncancerous setting. The alterations in self antigens could be mutations, misfolding, overexpression, aberrant location, aberrant glycosylation, or aberrant degradation in the protein. Alternatively, defects in tolerance could happen when self-reactive lymphocytes escape clonal deletion, which is believed to be a potential consequence of low level expression of the self-antigen (Goodnow, Sprent, de St Groth, & Vinuesa, 2005; Zaenker, Gray, & Ziman, 2016).

Aberrant post-translational modifications (PTMs) are one trigger of AAb production in cancer. AAbs against glycopeptides of the aberrantly glycosylated MUC-1 Ag have been detected in breast, ovarian, and prostate cancers (Wandall et al., 2010). Most tumor Ags, however, are the result of the overexpression of self-antigens. This is likely because of presentation of their epitopes on MHC molecules at levels higher than the threshold required for eliciting a T cell immune response and indirectly an Ab response as well

(Watanabe, Arase, Onodera, Ohashi, & Saito, 2000). There is a correlation between the expression levels of HER-2/neu in breast tumors and the levels of specific AAbs detected in sera (Goodell et al., 2008).

Other Ag alterations include neoantigen generation and their presentation to the immune system, which could elicit a T cell response against the neoepitope or the production of AAbs. These could result from somatic mutations that alter the protein immunogenicity or expose amino acid regions that were not previously exposed (Zaenker et al., 2016). This is further augmented by genetic instability, a hallmark of cancer that leads to more mutations and neoantigens. Anti-p53 AAbs are detected more frequently in serous than non-serous ovarian cancers, positively correlating with higher frequency of p53 mutations (K. S. Anderson, D. W. Cramer, et al., 2015). Two pieces of evidence suggest that p53-specific Ab production could be the result of p53 accumulation in the cell rather than its mutation. First, mutant p53 has a longer half-life than the wild-type protein. Additionally, most immunogenic epitopes of p53 are not located in the region that harbors mutations (Soussi, 2000).

Besides host-driven tumor Ags, viral proteins are of particular interest in the study of biomarkers of virus-related cancers. Because they are foreign proteins, their levels or the immune response to them are likely to be more specific than host-derived proteins. However, the existence of benign infections with oncogenic viruses complicates the use of their proteins as definitive detection biomarkers. This is in particular when the majority of infected people clears the infection without overt disease as in the case of HPV or when the virus causes chronic infections that only causes cancer in a proportion of patients such as hepatitis B and C viruses. It becomes useful in these cases to integrate a combination of test results in the clinical decision making. For example, hepatitis B patients with low viral load are more likely to develop cirrhosis and hepatocellular carcinoma if they do not

develop Abs against the hepatitis B e antigen (HBeAg) and have high serum levels of the hepatitis B surface antigen (HBsAg) (Lin & Kao, 2016). IgA Abs against Epstein-Barr Virus (EBV)-specific proteins have been evaluated in combination with EBV viral load for diagnosis and monitoring of undifferentiated nasopharyngeal carcinoma in endemic regions (Leung et al., 2004; Zhao et al., 2014). For HPV, only 50 – 70% of infected women seroconvert (Carter et al., 2000), but the immune profile could still be useful in combination with other tests or as an initial screening tool in low-resource settings.

1.5. ADVANCES IN HIGH-THROUGHPUT BIOMARKER DISCOVERY METHODS

The advent of the omics technologies and high-throughput methods for studying biochemical pathways and molecules has revolutionized our understanding of disease and medicine. In recent years, these methods have contributed rapid discoveries of candidate biomarkers and better recognition of the complexity of cancer. Even though they have accelerated the research process, only a handful of clinically useful biomarkers have made their way to the clinic. This highlights the importance of recognizing both the advantages and limitations of each of these methods for developing biomarkers that truly inform the clinical decision.

1.5.1. Assessment of Ab Biomarkers

One critical requirement for Ab-based assays is the efficient and reproducible expression, purification and display of proteins. Sera are typically screened for Abs to select Ags that are known to potentially be immunogenic or play a role in pathogenicity. This Ag selection does not measure the diversity of immune recognition (Ji Qiu & Anderson, 2013). To add complexity, proteome-wide immune monitoring requires the production of thousands of protein structures.

The need for tools to study proteins and the significant role they play in health and disease have led to the revolutionary advancements in the field of proteomics in the last twenty years. Effective targets of immunization and serological testing are best determined using a systems approach for monitoring the B cell immune response. Proteomic techniques that have been developed for epitope display are reviewed in [8, 11] and can be summarized as follows:

1.5.1.1. Phage display

Phage display was first described in 1985 (G. P. Smith, 1985). Candidate Ags are expressed in lambda phage from cDNA libraries constructed from a given pathogen or disease tissue. Phage expressing proteins of interest are subsequently replicated onto nitrocellulose membranes and probed with patient sera. Phage display has been applied in Ag discovery in various pathogens such as hepatitis C virus (Santini et al., 1998), human cytomegalovirus (Beghetto et al., 2008), *Mycoplasma pneumoniae* (Beghetto, De Paolis, Montagnani, Cellesi, & Gargano, 2009) and *Streptococcus pneumoniae* (Beghetto et al., 2006). Alternatively, solution-based phage display is used for autoantigen identification.

Phage-displayed peptide libraries are subjected to affinity purification to isolate phage carrying specific peptides. AAb biomarkers of several cancers such as the ovary and prostate have been identified using this technique (Chatterjee et al., 2006; G. Chen et al., 2007; X. Wang et al., 2005). However, because of the nature of cDNA cloned on the expression vectors, the major drawback of phage display is the expression of proteins with truncations, frame shifts and sometimes improper folding. In addition, PTMs are absent and abundant proteins are overrepresented.

1.5.1.2. Cellular fractionation and immunoblotting

In this strategy, candidate Ags from lysates of tissues or pathogens are separated by two-dimensional gel electrophoresis and serum reactivity is determined by immunoblotting or

mass spectrometry. Next, bands are excised, and proteins are identified by mass spectrometric analysis. This method has the advantage of using proteins with their relevant PTMs and it does not require cloning or expression procedures. However, proteins found in low concentrations may be masked by more abundant proteins.

1.5.1.3. Peptide arrays

Peptides are displayed on a solid surface such as a glass or plastic slide. Relevant peptides that have overlapping sequences are determined bioinformatically so as to cover the whole ORFeome or a portion of the proteome. This circumvents difficulty with expression of full-length proteins, but conformational epitopes and PTMs are not detected. Recently, peptide arrays have been used to determine individual immunosignatures that can predict the protective efficacy of a given vaccine in mice (Legutki & Johnston, 2013).

1.5.1.4. Protein arrays

Protein microarrays enable the display of thousands of proteins on the surface of a microscopic slide or in 96-well bead-array format. A wide variety of protein expression systems are used including *E. coli*, yeast or insect cells and then Ags are purified. However, these systems can be time-consuming and unsuitable for high-throughput proteomic methods. Additionally, bacteria often fail to express most proteins with intact tertiary structures or PTMs, particularly those with high molecular weights or multiple domains (Jackson, Boutell, Cooley, & He, 2004; Stevens, 2000). *In vitro* protein expression, on the other hand, diminishes the time required to obtain protein from DNA but adds the challenges of protein purity and reproducibility of expression (Ji Qiu & Anderson, 2013).

Protein microarrays are currently commercially available from several sources, and are provided either as purified, printed proteins, or as printed cDNA that can be expressed using *in vitro* transcription and translation. At this time, the antigenic display on protein

microarrays is primarily the protein backbone, so the diversity of displayed antigenic structures from PTMs is more limited. As the content of ORFeome collections content and the cost of protein expression improves, proteome-wide screening of sera for Ab responses is becoming feasible both for human Ags and pathogens. Here, three methods and overall strategies for using *in situ* protein display for detection of Ab responses in human sera or plasma are discussed.

Nucleic acid programmable protein array (NAPPA)

To improve both the cost of purification of recombinant proteins and the stability of displayed protein, the Nucleic Acid Programmable Protein Arrays (NAPPA) was developed using printed expression plasmids with an anti-tag Ab on microscopic glass slides (N. Ramachandran et al., 2004; Niroshan Ramachandran, Hainsworth, Demirkan, & LaBaer, 2006; N. Ramachandran et al., 2008). At the time of the assay, *in vitro* transcription/translation (IVTT) is used for *in situ* expression of tagged target proteins encoded by the arrayed plasmids (**Figure 1-1**). The use of a human coupled IVTT system derived from the human cell line HeLa results in ten times higher protein yields, more robust reproducibility, and less background than the previously used rabbit reticulocyte lysate system (Fernanda Festa et al., 2013). For immune monitoring, slides are incubated with subject sera or plasma to permit binding of Abs to their corresponding protein spot on the array. Signals are detected using either a fluorescently-labeled or HRP-labeled secondary Ab. NAPPA arrays are available from the Arizona State University protein array core, www.NAPPAproteinarray.org.

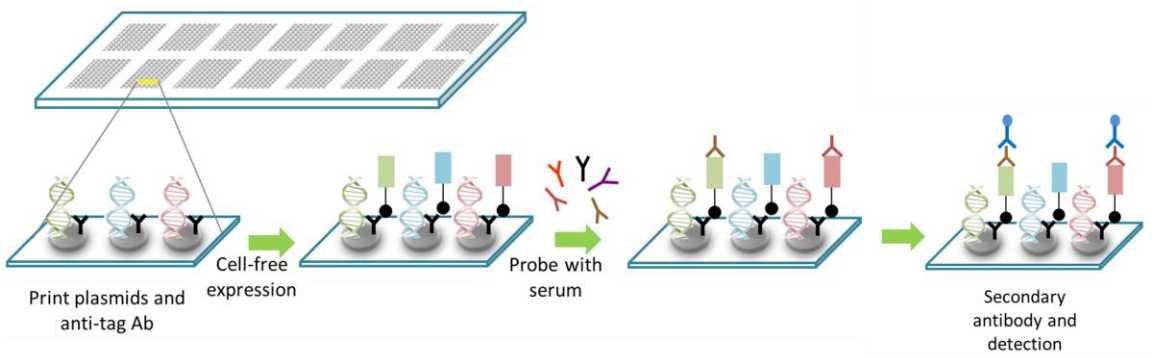

Figure 1-1. Nucleic Acid Programmable Protein Arrays (NAPPA). Expression plasmids with an anti-tag antibody are printed on microscopic glass slides. At the time of the assay, *in vitro* transcription/translation (IVTT) is used for *in situ* expression of tagged target proteins encoded by the arrayed plasmids. Slides are incubated with subject sera or plasma to permit binding of antibodies to their corresponding protein spot on the array. Signals are detected using either a fluorescently-labeled or HRP-labeled secondary antibody.

Printing DNA on the arrays has several advantages over printing proteins. Unlike protein arrays, printed plasmids retain their activity following months of storage of the arrays under arid conditions. Since the production and purification of thousands of proteins is expensive, time-consuming and liable to protein unfolding over the multistep process of protein array production, the on-demand IVTT avoids these issues (J.-R. Lee, Magee, Gaster, LaBaer, & Wang, 2013). However, IVTT-derived proteins are produced with limited PTMs that are a significant component of the immune response.

A key advance in the field of protein microarrays has been the steady improvement in relevant ORFeome collections. The DNASU Plasmid Repository is the source of plasmid DNA used for NAPPA array production (Seiler et al., 2013). This plasmid collection first started at the Harvard Institute of Proteomics in 2000 and is currently located at the Virginia G. Piper Center for Personalized Diagnostics in the Biodesign Institute (AZ, USA) (Seiler et al., 2013). The repository comprises and distributes a collection of over 200,000 plasmids containing the open reading frames (ORFs) of proteins from over 600 organisms,

including 12,000 full-length human genes. The DNASU website, database and physical repository (http://dnasu.asu.edu or http://dnasu.org) were designed to provide annotated and sequence-verified plasmids and online resources to the research community. All ORFs are cloned onto a master plasmid (pDONR), sequence verified and stored in the DNASU repository. ORFs in DONR plasmids can be moved to a wide array of expression vectors using Gateway recombinational cloning.

NAPPA arrays have been used for the discovery of AAbs in cancer patient sera, such as AAbs to p53 in breast and ovarian cancer, BCL2 in prostate cancer and ML-IAP in melanoma (Karen S. Anderson et al., 2015; Anderson et al., 2008; Niroshan Ramachandran et al., 2008). Examples of screening for infectious disease Ags include studies displaying *Pseudomonas aeruginosa* outer membrane proteins (Wagner R Montor et al., 2009) and *Mycobacterium tuberculosis* proteins (Prados-Rosales et al., 2014) on NAPPA to screen sera to identify immunogenic proteins.

Rapid antigenic protein in situ display (RAPID ELISA)

NAPPA protein microarrays are an excellent tool for Ag discovery. However, validation requires methods for the analysis of a few Ags but using thousands of sera. RAPID ELISA was developed as a robust tool that can be performed in most immunology laboratories using publicly available reagents. RAPID ELISA can be used to screen hundreds of sera rapidly and cost-effectively in order to confirm Ab biomarkers and immunogenicity of antigens discovered using protein microarrays (Anderson, 2011; Anderson et al., 2010; Niroshan Ramachandran et al., 2008). As with NAPPA assays, tagged proteins are expressed using an *in vitro* transcription and translation system, but then captured in a 96-well plate through an anti-tag Ab. Sera are then incubated with the displayed proteins, and bound immunoglobulins are detected using secondary Abs (**Figure 1-2**). To overcome the background problem encountered with human sera, an optimized serum blocking buffer

consisting of *E. coli* lysate diluted 1:10 in PBST and 5% milk was developed (J. Wang et al., 2013). An eight-fold increase in the Relative Light Unit (RLU) ratio of Ag-specific IgG compared with control GST protein was observed with the use of this serum blocking buffer. Additionally, human HeLa cell lysate IVTT system and automation have further enhanced the efficiency, rapidity, and reproducibility of this technique.

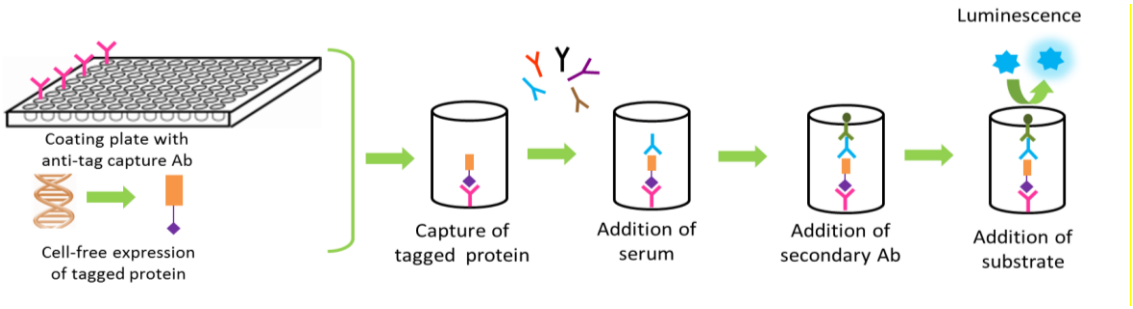

Figure 1-2. Rapid Antigenic Protein In Situ Display (RAPID ELISA). Tagged proteins are expressed using an *in vitro* transcription and translation system, but then captured in a 96-well plate through an anti-tag antibody. Sera are then incubated with the displayed proteins, and bound immunoglobulins are detected using secondary antibodies.

Magnetic programmable bead ELISA (MagProBE)

A similar technique as RAPID ELISA for high-throughput serum screening is the Magnetic Programmable Bead ELISA, MagProBE (Anderson, 2011). As with RAPID ELISA, tagged proteins are expressed by IVTT, but then expressed proteins are captured on anti-tag coupled fluorescent magnetic beads (such as Luminex beads) in a 96-well plate. This is followed by steps of incubation with sera and then with a secondary Ab (**Figure 1-3**). Beads are coupled with the anti-tag Ab in advance and they are stable for at least one year. Coupling efficiency is confirmed using anti-Ig secondary Abs. Chief among the advantages of MagProBE is the high reproducibility and automated washing. Additionally, bead-array ELISA can be used for multiplex assays by coupling of different Ags on beads of different colors and then pooling them. Multiplexing saves both time and volume of serum, but the cost per Ag is higher than with RAPID ELISA. This technique has been used for multiplex detection of immunity to a panel of EBV Ags in healthy donor sera

(Jessica Wong, Sahar Sibani, Naa Norkor Lokko, Joshua LaBaer, & Karen S Anderson, 2009) and to investigate potential biomarkers of HPV-associated oropharyngeal carcinoma (Anderson, Wong, et al., 2011).

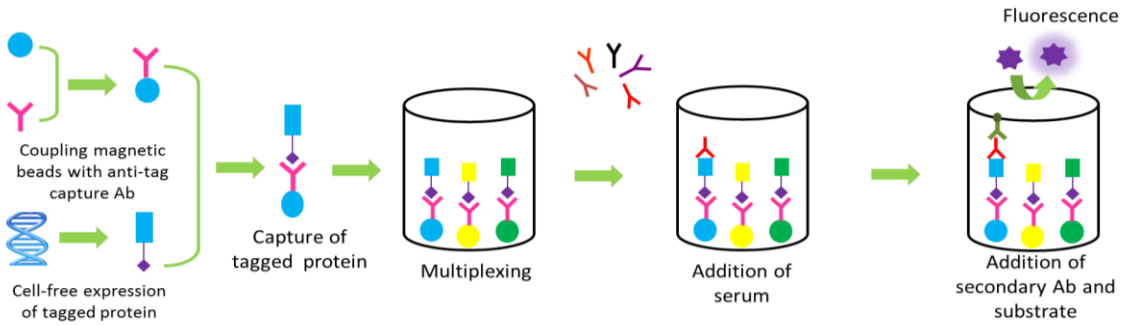

Figure 1-3. Magnetic Programmable Bead ELISA (MagProBE). Tagged proteins are expressed by IVTT and expressed proteins are captured on anti-tag coupled fluorescent magnetic beads in a 96-well plate. This is followed by steps of incubation with sera and then with a secondary antibody.

1.5.1.5. Recent advances in protein display and detection

Many immune-based biomarkers have clinical applications for early detection of disease. The applications require robust, reproducible, and cost-effective assays with improved limits of detection, multiplexing and automation, all of which have substantially improved in the last decade. For example, chromogenic enzyme substrates have been the traditional reporter molecules for ELISAs. The more sensitive chemiluminescent substrates can now detect analyte concentrations in the picomolar range (A. H. Wu, 2006). Ultrasensitive approaches, such as the Single Molecule Array technology may allow detection of femtomolar concentrations of Ab through digital measurements of immunocomplexes. Nanoparticle-based ELISAs are reported to detect attograms of analytes (de la Rica & Stevens, 2012).

Because routine laboratory diagnostics are costly and may not be accessible in resource-poor areas, point-of-care (POC) tests are emerging technologies for health screening,

much of which currently depends on detection of Abs. Affordable POC tests that give rapid and reliable results, require minimal training, and use no equipment are currently in use for HIV, syphilis and malaria (Ferguson et al., 2016). Integration of ELISA assays with microfluidics and molecular detection methods may transform vaccine monitoring and identification of at-risk individuals for clinical interventions.

1.5.2. Identification of T Cell Ags

As previously mentioned, the identification of targets of the T cell immune response is a longer and more complicated process compared to B cell Ags. Conventionally, T cells are screened for reactivity to peptides spanning candidate Ags, an approach limited by the prohibitive cost of synthesizing large numbers of peptides. The advances in the bioinformatics of immunogenic epitope prediction has made it possible to narrow down the number of screened peptides by prioritizing them according to MHC binding affinity or the probability of immunogenicity. Alternatively, the identification of antigenic T cell immune response targets depends on screening of peripheral blood or tumor-infiltrating T cells for reactivity to antigen presenting cells transfected with cDNA libraries. The tandem minigene (TGM) is a more recent approach that involves transfection of APCs with cDNA libraries encoding nonsynonymous mutation sequences identified by whole-exome sequencing of autologous tumors (Y.-C. Lu et al., 2014).

The identification of immunodominant T cell targets is useful for designing successful adoptive T cell immunotherapies in cancer. This can be achieved by identifying neoantigen-specific T cells that are found in the highest frequencies in tumor-infiltrating lymphocytes that result in tumor regression. T cell receptors are then identified using single-cell sequencing techniques and used for engineering chimeric antigen receptor (CAR) T cells (Yong et al., 2016). The detection of Ag-specific T cells can be useful as a biomarker to monitor and evaluate the success of adoptive T cell therapy. The presence,

homing, bioactivity, and counts of T cells following infusion can be evaluated by the detection of surface or intracellular markers by flow cytometry, ELISpot, or qPCR (Kalos, 2011).

1.6. DEVELOPING RELIABLE AND CLINICALLY USEFUL BIOMARKERS

The emergence of the omics technologies has raised hopes for delivering the promise of precision medicine to deliver therapies tailored to individual molecular phenotypes. Part of the reason why this has not been realized yet is the lag in the development of well-validated biomarkers of reliable clinical utility. These technologies have given rise to hundreds of thousands of research publications describing candidate biomarkers, but only a tiny fraction of this number is in clinical use (Poste, 2011). Many factors contribute to the candidate biomarkers not holding up in validation studies, ranging from specimen quality and pre-analytical variables to the complexity conferred by the heterogeneous nature of cancer. The first step to improve the validation pipeline is to recognize and control for these factors and to implement standardized procedures for every step along the way.

Breast cancer is the most common type of cancer in the US, accounting for about 15% of new cancer cases ("American Cancer Society: Cancer Facts and Figures 2018," 2018; NIH, 2017). Compared with other cancer types, breast cancer is widely studied and progress in research has made biomarkers an indispensable part of its management. Biomarkers exist for monitoring metastatic disease, such as CEA, CA27.29 and CA15-3, but there is a need for circulating biomarkers that are clinically available for early detection, prognosis, and monitoring for clinical relapse. Guidelines on the use of biomarkers in breast cancer were published by both the American Society of Clinical Oncology and the European Group on Tumor Markers (Duffy et al., 2017; Harris et al., 2016). As with much of biomarker research, despite the significant progress in the discovery of potential breast

cancer biomarkers, the vast majority of these biomarkers has not progressed beyond initial research discovery. In Chapter 2 of this thesis, breast cancer is used as a model to examine the reasons for the disconnect between the number of candidate biomarkers in research and those that make it to the clinic.

1.7. HPV-ASSOCIATED CANCERS

Human papillomaviruses (HPV) are a family of more than 200 closely related viruses with small circular double-stranded genomes. The virus is restricted to epithelial surfaces where it induces a non-lytic cellular proliferation and minimal immunologic response. The estimated number of new cases of cervical cancer worldwide was 528,000 in 2012, with an annual global mortality rate of 270,000 deaths (Ferlay et al., 2010; Ferlay et al., 2015). Vaccines preventing infection are highly effective for the prevention of type-specific cervical and anogenital cancer precursors (vaginal, vulvar, anal), may reduce relapse after conization (Kang, Choi, & Kim, 2013), but are not a treatment for pre-existing HPV infection. Vaccines are expected to substantially reduce the burden of HPV-associated cancers. In the US, even with low vaccine coverage, vaccination has resulted in a 64% reduction in the prevalence of types 6, 11, 16, and 18 among females aged 14 to 19 years and 34% decrease among those aged 20- 24 years (Lauri E. Markowitz et al., 2016).

1.7.1. HPV Genome Organization and Life Cycle

HPV is a circular double-stranded DNA virus, whose genome contains about 8000 base pairs. The virus has a limited repertoire of proteins, grouped as early (E1, E2, E4, E5, E6, E7) and late (L1, L2) proteins. The late proteins form the viral protein coat during productive infections. The early proteins interact with host and viral proteins to maintain viral replication and release, and they play a key role in transformation. The E1 protein is

a helicase enzyme that is expressed in the early stages of the infection and is important for viral DNA replication. E2 is also expressed in the early stages and is a key regulator, since it represses the expression of the E6 and E7 oncogenes (L. Wu et al., 2000). Viral integration occurs following a break in E2, which derepresses E6 and E7 expression, leading to cellular transformation (zur Hausen, 2009).

The E6 protein promotes cell proliferation by inhibiting p53, a key tumor suppressor and regulator of apoptosis (Boulet, Horvath, Broeck, Sahebali, & Bogers, 2007). It also promotes division through multiple other targets including telomerase and proteins involved in the regulation of the actin cytoskeleton (zur Hausen, 2002). E7 inhibits pRB, leading to the release of the transcription factor E2F and the transcription of genes involved in the cell cycle (Boulet et al., 2007). Thus, the expression of these two proteins is a potential hallmark of malignancy. However, there are cases of cellular transformation in the absence of E6 and E7 deregulation. This is the case when the virus has not integrated or when its genome breaks at sites other than E2 (Oyervides-Munoz et al., 2018).

HPV infects the basal layers of the stratified epithelium of the cervix and the expression of viral proteins is confined to keratinocytes. The virus infects the basal cells of the epithelium and the virus copy number reaches 50 – 100 copies/cell through a round of viral DNA replication that is independent of the cell cycle. The infected cell then moves to the proliferative compartment of the epithelium, where viral protein expression (particularly E6 and E7) is tightly regulated. As the cell then moves to the differentiating compartment of the epithelium and exits the cell cycle, differentiation is accompanied by amplification of the viral copy number to 1000 copies/cell and expression of the coat proteins. The expression of viral capsid proteins in basal epithelial cells is limited by the low availability

of the appropriate tRNAs (Zhou, Liu, Peng, Sun, & Frazer, 1999). The upper layers of the squamous epithelia are where viral proteins are expressed in high levels and where viral assembly occurs (Evans et al., 2001; Stanley, 2008).

1.7.2. HPV Integration

In cervical cells, the HPV virus is either episomal, integrated, or in mixed forms. Transcripts derived from integrated forms have higher stability than those from episomal virus. Cells with integrated HPV16 have been reported to have selective growth advantage (Jeon, Allen-Hoffmann, & Lambert, 1995). Integration is commonly thought of as a late event in cervical carcinogenesis, because it is rarely reported in earlier stages of the disease. However, technical difficulties associated with studying viral integration challenge this idea (Woodman, Collins, & Young, 2007). Integration often happens in the E1 or E2 region of the viral genome, thus disrupting the tight regulation on E6 and E7 expression coordinated by the E2 protein. Therefore, integration is commonly evaluated by the failure to amplify full-length E2 by PCR or by measuring the E2/E6 ratio. This method, however, has been shown to be able to distinguish integrated forms only when their frequency is at least 100 times that of the episomal forms (Arias-Pulido, Peyton, Joste, Vargas, & Wheeler, 2006).

In cervical cancer cells, integration is almost always detected at only one chromosomal site, which supports the proposal that cervical cancer is a clonal disease (Vinokurova et al., 2005). However, there is debate about whether integration is a necessary event for carcinogenesis to occur (Woodman et al., 2007). The host genetic alterations that result from viral integration are another mechanism that has been proposed for carcinogenesis, as opposed to the deregulation of E6 and E7 expression. Viral integration into oncogenes or tumor suppressor genes may disrupt their normal function, leading to cell proliferation, and integration into both types of genes has been reported (Ojesina et al., 2014; Parfenov

et al., 2014). Research is ongoing into whether other factors, such as viral load, HPV type, or host genetic or epigenetic factors could derive oncogenesis.

Whether integration can be a useful biomarker of progressive disease is also a matter of debate. The technical difficulties of detecting integrated forms in a background of mostly episomal forms have challenged answering this question. Additionally, the mere detection of integrated forms by PCR provides no insight into whether they are actively transcribed in these cells. Active transcription of integrants has been reported in only 37% of CIN3 women in which integration has been detected (Klaes et al., 1999; Melsheimer, Vinokurova, Wentzensen, Bastert, & von Knebel Doeberitz, 2004). Thus, integrant-derived HPV transcripts could be a more useful biomarker of progressive disease (Woodman et al., 2007).

1.7.3. Viral Load

Including HPV viral load measurement in HPV screening has been proposed based on the observation that women with high viral load of high-risk HPV are more likely to have cytological abnormalities (Heard et al., 2000; Lillo et al., 2005; Swan et al., 1999). However, several studies have challenged this conclusion. Some cross-sectional studies have reported higher viral load in low-grade than in high-grade CIN (Hall et al., 1996; Nindl et al., 1997); and longitudinal studies could not establish a relation between viral load and disease progression (Crum et al., 2004; Mark van Duin et al., 2002).

This inconsistency could be explained by the reduction in infectious viral particles and viral load after integration, an event associated with disease progression. Thus, women with high-grade CIN who also have low-grade preinvasive lesions could have higher viral load; but the vast majority of studies do not report multiple co-existing disease statuses when present (Sherman et al., 2003). Similarly, acquiring new HPV infections especially from

different HPV types may impact viral load and the development of cytologic lesions (Woodman et al., 2007). Therefore, the study of the usefulness of viral load as a marker of disease progression requires careful consideration of other factors that may be contributing to prognosis. These factors are important to consider when investigating the utility of the HPV-specific Ab immune response as a marker of disease, given that there is a reported increase in seropositivity in women with higher viral load (Viscidi et al., 1997).

1.7.4. Multiple HPV Infections

Detecting concurrent infections with multiple HPV types is common and is more frequent than to be attributed to random chance (Mendez et al., 2005; Thomas et al., 2000). There is no evidence that there is competition between the coinfecting HPV types (Liaw et al., 2001). Studies report that women who are already infected have a higher chance of acquiring an infection with a new HPV type than HPV negative women (Liaw et al., 2001; Mendez et al., 2005). Women with an HPV16 or HPV18 infection are seven times more likely to acquire an HPV58 infection than women who are negative for these types (Mendez et al., 2005). The viral load in women with HPV16 and multiple HPV infections was found to be higher than women with only HPV16, suggesting a possible interdependence between these viral types (Weissenborn et al., 2003).

Detecting an infection with a new HPV type that was not previously detected for a given individual could be due to the acquisition of a new sexual partner. However, the possibility of simultaneous transmission of these different HPV types with a lag in replication of one of them is not precluded (Woodman et al., 2007). Detecting an HPV-specific immune response to antigens from multiple HPV types is thus not surprising and could reflect multiple HPV infections, whether they are all concurrent and active infections or include past infections that have been cleared. This can also be explained by cross-reactivity

between antigens with a high degree of homology. With the technical challenges of studying the immune response to multiple HPV proteomes, this question has not been adequately addressed in the literature.

1.7.5. The Immune Response to HPV and Immune Evasion Strategies

The HPV replication strategy in the cervix allows viral replication without inducing inflammation, given the absence of virally-induced cytolysis or necrosis. E1 is the only DNA replication enzyme that the HPV genome encodes. Therefore, the virus depends for replication on the cellular machinery, which is active only in dividing cells. For amplification of the viral copy number to occur in cells that have exited the cell cycle and are differentiating, the virus uses its E6 and E7 proteins to reactivate the cellular DNA synthetic machinery in non-dividing cells, delay their differentiation, and inhibit apoptosis. A rare consequence of these events is the loss of growth control manifested as cancer (Munger et al., 2004).

Because the virus does not lyse the keratinocytes that it infects, the cervix acts as a protective niche for the virus, with minimal virus engulfment by APCs and antigen presentation to the immune system. There is also minimal release of proinflammatory cytokines that help induce APC migration and activation (Stanley, 2006). Outside the epithelium, viral antigens are not likely to be detected due to the absence of viremia. The low expression of capsid proteins in the basal layer gives less opportunity for the immune system to encounter them and initiate an immune response (Schwartz, 2000).

Several other factors contribute to the host immune evasion by the virus. IFN-β release results in clearance of episomal HPV, but this mechanism is not useful in cells with integrated viral DNA (Herdman et al., 2006; Pett et al., 2006). There is a reduction in the cell-mediated immune response directed against the E2 and E6 proteins in high-grade

pre-invasive cervical lesions and invasive carcinoma (de Jong et al., 2004). Regulatory T cells (Tregs) play an important role in counteracting the cytotoxic effect of HPV-specific T lymphocytes (Kobayashi et al., 2004). Tregs function to maintain the immune tolerance and prevent reactivity to self-antigens. Their activity in cancer patients has been associated with decreased immune reactivity to tumor antigens (R. F. Wang, 2006). Women with ICC and CIN were found to have increased frequencies of Treg cells in their peripheral blood. An increased T cell response against HPV16 E6 and E7 peptides in invasive cervical cancer was reported following *in vitro* depletion of these CD25+ T cells (Visser et al., 2007). Along with the localization of Treg cells, the expression of the immunoregulatory enzyme indoleamine 2, 3-dioxygenase (IDO) appears to also help cancer cells evade the immune response (Nakamura et al., 2007).

1.7.6. Clearance of Natural HPV Infection

Despite a high lifetime risk of acquiring an HPV infection (80%), most infections are cleared within 1-2 years (Rodriguez et al., 2008). Viral clearance means failure of detection of the DNA for a specific HPV type and it is the result of an effective adaptive immune response (Coleman et al., 1994). Failure of the immune response to clear the virus results in the establishment of a persistent infection, since the host can remain ignorant of the virus for a long time. There is an increased chance of progression to high-grade pre-invasive cervical lesions and invasive cancer in individuals with persistent infection (Liaw et al., 1999; Schlecht et al., 2001). The importance of cell-mediated immune response is demonstrated by the increased incidence and progression and delayed clearance of HPV infections in immunocompromised individuals. Several studies in HIV-infected individuals have reported increased incidence of genital warts, recurrences of pre-invasive cervical lesions, and progression of subclinical to clinical HPV infection

(Chirgwin, Feldman, Augenbraun, Landesman, & Minkoff, 1995; Fennema, van Ameijden, Coutinho, & van den Hoek, 1995; Fruchter et al., 1996).

Immunohistologic studies of naturally regressing genital warts have provided insights into the role of the cellular immune response in clearing HPV infections. Compared with non-regressing genital warts, regressing warts are characterized by a large number of infiltrating macrophages and CD4+ and CD8+ T cells in both the stroma and the epithelium of the wart. This response is directed against the E proteins, in particular E2 and E6 (de Jong et al., 2002; Welters et al., 2003). These lymphocytes release IL-12, TNF-α, and IFN-γ, which is characteristic of a Th1-based immune response (Coleman et al., 1994). Canine studies of oral warts corroborate these findings and report systemic T cell responses against peptides of the HPV E2 and E6 proteins that peak at wart regression, then decline rapidly (Ghim et al., 2000). During wart regression in both humans and animal models, seroconversion and neutralizing antibodies in serum against the major capsid protein L1 are maximal, which is accompanied by lifelong protection against infection (Ghim et al., 2000; Stanley, 2006). However, the antibody titers are low, and a significant proportion of women do not seroconvert (Carter et al., 2000; Dillner, 1999).

1.7.7. Diagnosis

The epidemiology and natural history of HPV infection has been best characterized in the cervix where precursor lesions are well recognized, using detection of HPV DNA or detection of Abs to HPV as biomarkers of disease pathogenesis (Crosbie, Einstein, Franceschi, & Kitchener, 2013). Genital HPV is usually acquired shortly after sexual debut, and prevalence is highest in adolescents and young adults (Dunne et al., 2007; Lauri E. Markowitz et al., 2016; L. E. Markowitz, Sternberg, Dunne, McQuillan, & Unger, 2009). Cervical cancer is a rare consequence of this common infection, with ~50% of the cases

worldwide caused by the HPV16 type. While high risk (HR) HPV infection is considered necessary for cervical carcinogenesis, additional factors are clearly involved. A small fraction of infected women gradually progress to invasive cancer, following a long, histologically well-defined pre-invasive phase (cervical intraepithelial neoplasia; CIN), ranging from low grade (CIN I) to high grade (CIN II and III) (Woodman et al., 2007). Cervical cancer is preventable because high grade lesions are detectable by clinical, histopathologic, or molecular alterations and can be surgically removed (Goodman, 2015). Current clinical practice in the US relies on regular screening with cytology (Pap test) often combined with HR HPV nucleic acid testing to refer women for colposcopy and biopsy. Recent data have documented that cytology screening is associated with a significantly reduced incidence and risk of death from cervical cancer, with odds ratios ranging from 0.28 to 0.60 (Vicus et al., 2014; Vicus et al., 2015), despite a reported high false negative rate (Soost, Lange, Lehmacher, & Ruffing-Kullmann, 1991). Cytologic screening remains subject to sampling errors, problems with cellular preservation, and reader subjectivity. Biomarkers are needed in particular to aid in the selection of patients for colposcopy screening in resource-limited settings in low and middle-income countries (LMICs), where nucleic acid and cytology testing are cost-prohibitive. Efforts are underway by the World Health Organization and the Program for Appropriate Technology in Health (PATH) to generate cost-effective HPV DNA testing (Qiao et al., 2008).

Measuring the humoral immune response to HPV Ags has been integral to understanding the natural history of infection and efficacy of vaccination (Doorbar et al., 2012; Villa et al., 2006; Woodman et al., 2007). Despite the potential of HPV serology in disease diagnosis and prognosis, its clinical application has been limited by HPV heterogeneity, assay variability, and viral immune evasion. The serologic response to genital HPV infection is

primarily directed at conformational epitopes on the viral major capsid protein L1. As the infection is non-lytic, the host Ab response to L1 is weak and may persist for years, as an indication of past infection but not malignancy (Luevano et al., 2010; Stanley, 2010). Although anti-L1 Abs are an indication of past infection, only 50 – 70% of infected women seroconvert (Carter et al., 2000; Dillner, 1999).

Abs to both HPV16 E6 and E7 proteins have been detected at low levels in both serum and cervical-vaginal secretions of invasive cervical cancer (ICC) patients (Bierl et al., 2005). Their levels increase with cervical disease progression, but they are not detectable in a subset of patients with cervical cancer (Gutierrez-Xicotencatl et al., 2016; Luevano et al., 2010; Reuschenbach et al., 2008; Stanley, 2003). They develop later in the course of ICC and are correlated with disease outcome (Gutierrez-Xicotencatl et al., 2016; Ravaggi et al., 2006; Silins et al., 2002). Studies of sera collected prior to the diagnosis of cervical cancer have shown that the presence of E6 and E7-specific Abs is associated with an increased relative risk (RR=2.7) for cervical cancer, and can be detected, albeit infrequently, up to 5 years prior to diagnosis (Lehtinen et al., 2003). The percentage of women with false negative serology is dependent on the method of Ab detection (Achour et al., 2009; Combes et al., 2014; Kontostathi et al., 2016; Luevano et al., 2010; Waterboer et al., 2005; Zumbach et al., 2000). There is a clinical need for circulating biomarkers that identify high-risk HPV infection for early detection and treatment of cervical disease. Thus, systematic investigation of proteome-wide HPV serology is much needed for this purpose.

1.8. BIOMARKERS OF IMMUNITY TO THE CRISPR/CAS9 GENE THERAPY SYSTEM

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/Cas9 technology has raised hopes for developing personalized gene therapies for complex diseases such as cancer. In October 2016, a Chinese group at Sichuan University in Chengdu was the first to inject CRISPR/Cas9-gene edited cells into a human to treat aggressive lung cancer (Cyranoski, 2016). Another trial by the University of Pennsylvania to treat sarcoma, multiple myeloma, and melanoma is currently under way. The trial will engineer T cells by CRISPR to delete PD-1 and create a tumor-specific T cell receptor (ClinicalTrials, 2018). Despite a not-so-distant history of disappointment with gene therapy, these trials reflect fast-paced strides toward the use of human gene editing in medicine.

The human experience with gene therapy has witnessed an excitement phase about the potential for curing diverse life-threatening diseases, especially genetic disorders, which was sparked by many success stories. However, the scientific community backtracked after the death of Jesse Gelsinger in 1999 at the age of 18 due to a massive immune reaction to the gene therapy he received (Marshall, 1999). Since then, several studies have evaluated the efficacy and safety of gene therapy in humans to treat genetic disorders (Cavazzana-Calvo et al., 2000; Gaspar et al., 2004; Hacein-Bey-Abina et al., 2002; Howe et al., 2008; Manno et al., 2006). One important finding is that healthy individuals have pre-existing immunity to adeno-associated viral vectors, some of the most widely used gene therapy vectors (F. Mingozzi et al., 2007; Thwaite, Pages, Chillon, & Bosch, 2015). AAV-specific CD8+ T cells detected were also shown to expand following gene delivery (F. Mingozzi et al., 2007).

The expression of the bacterial CRSIPR-associated protein 9 (Cas9) nuclease in humans for cancer gene therapy raises concerns over safety and potential adverse reactions. Both cellular and humoral immune responses were reported in response to the expression of *Streptococcus pyogenes* Cas9 protein (SpCas9) in mice (Chew et al., 2016; D. Wang et al., 2015). Given the ubiquity of the *S. pyogenes* bacteria, it is possible that healthy individuals have pre-existing immunity to the SpCas9 protein, which could have implications for using this system in humans. Biomarkers of immunity to this protein could be used to guide gene therapy decisions in patients, particularly as this system moves to cancer clinical trials and possibly to *in vivo* treatment approaches. Chapter 5 of this dissertation seeks to address this gap in knowledge and aims to identify these markers for the most common human HLA type in North America.

1.9. THESIS CONTRIBUTIONS

This dissertation addresses several questions in HPV immunology, cancer biomarkers, and immunity to the CRISPR/Cas9 system. The primary contributions of this dissertation are:

1. Generating custom HPV Nucleic Acid Programmable Protein Arrays (NAPPA) displaying 98 proteins, representing the proteomes of two low-risk and ten oncogenic high-risk HPV types for studying serology in HPV-associated cancers. A high correlation of HPV16-specific serum IgG detection with the previously described RAPID ELISA for HPV-associated oropharyngeal cancer was confirmed.
2. Profiling the HPV Ab response to the proteomes of 12 HPV types in serum samples from women with invasive cervical cancer, high-grade pre-invasive cervical lesions (CIN II/III) and low-grade pre-invasive cervical

lesions (CIN 0/I; no CIN to grade I CIN). The arrays allowed the systematic analysis of HPV serology in cervical disease and detection of the breadth, specificity, and changing levels of HPV-specific antibodies with disease progression.

3. Detecting pre-existing B cell and T cell immune responses to the *S. pyogenes* Cas9 protein in humans. Two immunodominant T cell epitopes for HLA-A*02:01 were identified and a single mutation in the anchor residue of one or both of these epitopes significantly reduced the protein immunogenicity while maintaining its function and specificity in the CRISPR/Cas9 system. The identified epitopes have the potential to serve as a biomarker of pre-existing immunity to SpCas9 in the studied HLA haplotype that can help guide gene therapy decision making.
4. Reviewing the crucial considerations of developing pipelines for the rapid evaluation of circulating cancer biomarkers, with a focus on breast cancer as a case study.

CHAPTER 2 : CRUCIAL CONSIDERATIONS FOR PIPELINES TO VALIDATE CIRCULATING BIOMARKERS: BREAST CANCER AS A CASE STUDY

This chapter has been published:

Radwa Ewaisha, Chelsea D. Gawryletz, and Karen S. Anderson (2016). Crucial considerations for pipelines to validate circulating biomarkers for breast cancer. *Expert Review of Proteomics, 13*(2), 201-211. doi: 10.1586/14789450.2016.1132170

ABSTRACT

Despite decades of progress in breast imaging, breast cancer remains the second most common cause of cancer mortality in women. The rapidly proliferative breast cancers that are associated with high relapse rates and mortality frequently present in younger women, in unscreened individuals, or in the intervals between screening mammography. Biomarkers exist for monitoring metastatic disease, such as CEA, CA27.29 and CA15-3, but there is a need for circulating biomarkers that are clinically available for early detection, prognosis, or monitoring for clinical relapse. There has been significant progress in the discovery of potential circulating biomarkers, including proteins, AAbs, nucleic acids, exosomes, and circulating tumor cells, but the vast majority of these biomarkers have not progressed beyond initial research discovery, and none have yet been approved for clinical use in early stage disease. Here, the crucial considerations of developing pipelines for the rapid evaluation of circulating biomarkers for breast cancer are reviewed.

2.1. INTRODUCTION AND BACKGROUND

There has been significant effort toward the development of circulating biomarkers for the diagnosis and management of cancers. Breast cancer, as with other solid malignancies, is associated with alterations in systemic proteomic (Pernikarova & Bouchal, 2015),

glycoproteomic (Boersema, Geiger, Wisniewski, & Mann, 2013), immune (Kroemer, Senovilla, Galluzzi, Andre, & Zitvogel, 2015), and nucleic acid biomarkers (Schwarzenbach, 2013) that can be measured in the blood. Many of these biomarkers are thought to arise from the shedding or secretion of proteins or nucleic acids directly from tumor cells, such as the glycoprotein MUC1 (Paoletti & Hayes, 2014), circulating tumor DNA (ctDNA) (Dawson et al., 2013; Garcia-Murillas et al., 2015; Gingras, Salgado, & Ignatiadis, 2015), or within subcellular components such as exosomes (O'Brien et al., 2013; O'Driscoll, 2015; Yu, Cao, Shen, & Feng, 2015), or circulating tumor cells (CTCs) (Z. F. Jiang et al., 2013; M. C. Liu, 2014). Other biomarkers, such as cytokines (Gunter et al., 2015; Patel et al., 2015) or AAbs (Anderson, Sibani, et al., 2011; Chapman et al., 2007; J. Wang et al., 2015), are measures of the systemic immune reaction to the local tumor microenvironment. The blood, or a “liquid biopsy”, has the potential to provide a rapid, cost-effective, and minimally invasive window to molecular changes both in the breast and at distant metastatic sites. The development of highly sensitive molecular diagnostics is now providing the opportunity to evaluate these biomarkers for clinical management.

As with much of biomarker science, thousands of potential breast cancer biomarkers have been identified by research laboratories, but few of these have reached clinical practice (Poste, 2011; Poste, Compton, & Barker, 2015). Only the circulating biomarkers CEA, CA27.29, and CA15-3 are in clinical use for the management of metastatic breast cancer (Van Poznak et al., 2015). There are several reasons for the delays in clinical translation of many of these biomarkers. First, unlike cancer therapeutics, there have been few established pipelines and standards for the rigorous selection and validation of biomarkers for breast cancer. Second, the risk of false discovery in the early stages of research is significant. The scientific literature contains many circulating biomarkers that fail in validation steps. This false discovery is due, in part, to recurrent biases in the research

design of discovery science, such as pre-analytic variables in sample collection and the routine use of small sets of convenience samples without appropriate controls. As a consequence, the time, effort, and materials needed to eliminate false biomarkers are significant. Third, the intra-patient and inter-patient heterogeneity of breast cancer contribute to the limited quantity of circulating biomolecules in early stage disease and false-negative elimination of potentially useful markers. Fourth, detection of secreted proteins that are diluted in plasma need is difficult due to the complexity and the wide dynamic range of the plasma proteome (Rifai, Gillette, & Carr, 2006). We need a systematic assessment of the target clinical applications for circulating biomarkers, recognizing that clinical care is a dynamic process with evolving needs, technologies, therapeutic outcomes, and economic feasibility.

This chapter will focus on the emerging scientific technologies that are used to detect circulating biomarkers for breast cancer, with an emphasis on protein and immune biomarkers. Critical elements in biomarker study design and assay development will be identified, both at the discovery and the validation stages, to increase the identification of clinically useful markers (**Figure 2-1**). Factors that are needed to establish pipelines for the rapid translation of these biomarkers to clinical practice across multiple clinical applications will also be identified.

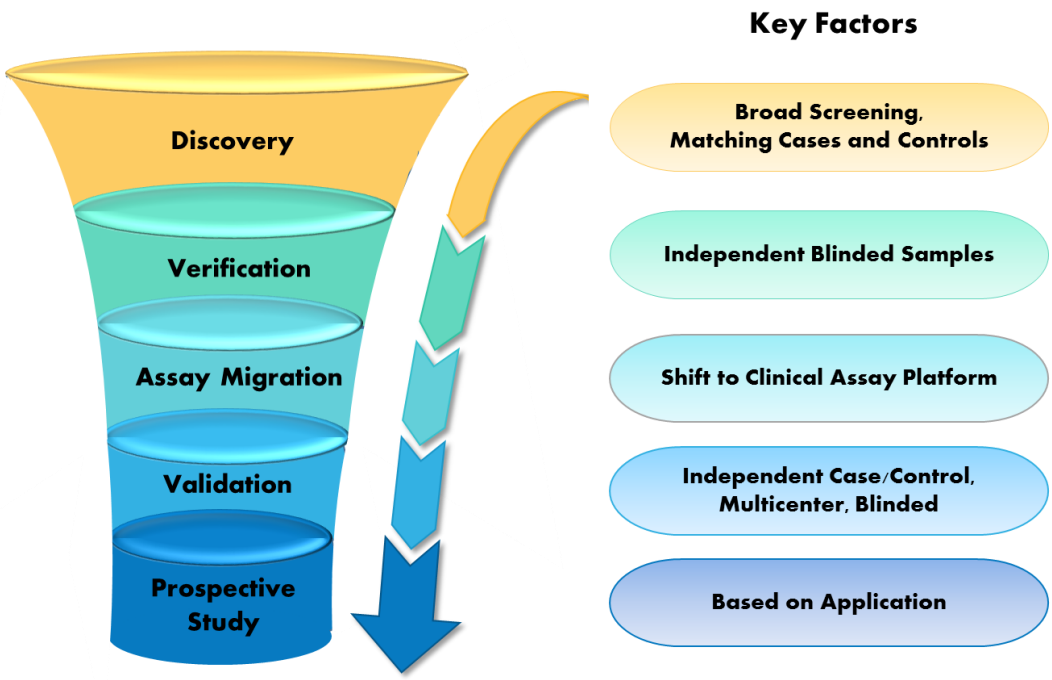

Figure 2-1. A Sample Pipeline of the Sequential Steps from Biomarker Discovery to Validation. The initial discovery usually involves screening many potential biomarkers with fewer samples but requires a strong statistical plan and matching of cases and controls to limit false discovery. Potential biomarkers should be verified using a larger number of samples from an independent, blinded sample set prior to publication. Prior to large-scale retrospective and prospective studies, migration of the assay to a highly reproducible, precise, and clinically-applicable platform is usually needed. Subsequent independent multicenter validation studies (both retrospective and prospective) require samples collected in the context of the intended clinical application.

2.2 THE CLINICAL APPLICATIONS OF BIOMARKERS FOR BREAST CANCER

Perhaps the most important process in biomarker discovery is the identification of the target clinical need (J. Chen et al., 2014; Zeidan, Townsend, Garbis, Copson, & Cutress, 2015). Circulating biomarkers have potential applications for risk assessment, early detection, prognosis, detection of early relapse, molecular profiling for treatment selection, and monitoring disease progression and evolution. The study design, selection of cases and controls, and validation strategy all depend on the intended use of the proposed biomarkers.

2.2.1. Biomarkers for Breast Cancer Risk Assessment

The current clinical practice in the United States is to provide population-based general breast cancer screening mammography and physical exams, in accordance with recommendations provided by the US Preventive Task Force guidelines ("US Preventive Task Force,"), the National Comprehensive Cancer Network (Bevers et al., 2009) and others. Patients at increased risk based on clinical, familial, and epidemiologic risk factors are routinely identified using the Breast Cancer Risk Assessment tool ("Breast Cancer Risk Assessment Tool,"), with additional risk factors (i.e. body mass index, breast density) being integrated into the current models (Tice et al., 2015). Patients with deleterious germline mutations in *BRCA1* and *BRCA2* are recommended to undergo more intensive screening, such as breast MRI (Narod, 2010). Emerging breast imaging tools, including tomosynthesis and contrast-enhanced digital mammography (CEDM), are rapidly changing the landscape of breast cancer screening (Chou et al., 2015). The role of additional genetic markers of risk (Garcia-Closas et al., 2013; Michailidou et al., 2013) identified from Genome-Wide Association Studies (GWAS) is being evaluated. As genetic screening becomes more prevalent and cost-effective, population-based genetic screening will become feasible (Manchanda et al., 2015).

In addition to genetic and epigenetic biomarkers of breast cancer risk, other circulating biomarkers of risk, such as protein, glycoprotein, and immune-based molecular changes may help inform targeted screening strategies. The systematic proteomic evaluation of blood from high risk women may lead to the discovery of biomarkers for risk stratification, to understand the molecular basis of tumor initiation and early development, and for selection of patients for targeted screening and prevention clinical trials. For both the discovery and validation of these biomarkers, early collaboration with epidemiologists and biostatisticians is necessary to design biorepositories that are technically compatible with

proteomic and immunomic analysis, and to identify appropriate samples (both circulating and tissue) with matching criteria of controls based on known and emerging breast cancer risk factors.

2.2.2. Biomarkers for Early Detection

Perhaps the greatest unmet need is for circulating biomarkers for early breast cancer detection. Screening mammography improves breast cancer survival (Harding et al., 2015), but the majority of breast cancers detected with mammography are ductal carcinomas *in situ* (DCIS) or low-grade, endocrine-responsive tumors where the timeline for early detection is less critical. High-grade, highly proliferative cancers, such as Her2+ breast cancer, triple-negative breast cancer (TNBC), and estrogen-receptor (ER) positive luminal B cancers frequently present as palpable masses in premenopausal women, women with increased breast density, or in the interval between screening mammography (Kirsh et al., 2011). There is a need for the specific detection of these cancers to reduce breast cancer mortality, in conjunction with radiographic imaging for tumor localization. In both the underserved and global health settings, mammography is rarely used as a primary screening tool, in part due to the cost and logistics of population-based mammographic screening. Circulating biomarkers, in conjunction with physical exams, could provide a rapid, cost-effective method for selection of patients for mammographic screening.

2.2.3. Biomarkers for Monitoring Breast Cancer

There is significant clinical interest in the use of liquid biopsies for monitoring breast cancer, either for the early detection of clinical relapse, for molecular profiling of metastatic tumors, or for monitoring metastatic disease response (Garcia-Murillas et al., 2015). Since the most common sites of metastatic spread are bone, liver, and lung, tissue-based diagnostics are associated with the morbidity of biopsies of relatively inaccessible

anatomic sites. The intra-patient heterogeneity of breast cancer evolution often results in mixed clinical responses that may be captured with deep molecular profiling. These circulating biomarkers could be used in conjunction with imaging for disease monitoring or for early detection of relapse. The clinical utility of detection of microscopic disease prior to relapse is not known but may identify patients for targeted clinical trials. In this setting, tissue-based genomic or proteomic alterations are a potential source of highly personalized marker profiles for tracking disease.

2.3. KEY FACTORS TO IMPROVE THE PIPELINE OF CIRCULATING BREAST CANCER BIOMARKERS

2.3.1. Incorporating PRoBE principles early in the biomarker discovery process

To improve the performance criteria of biomarkers in validation studies, it is important to limit hidden biases early in the design of discovery studies (Bohm et al., 2011; Marks et al., 2015; Opstal-van Winden et al., 2011; Pietrowska et al., 2010; Riley et al., 2011). The usual approach to the discovery of biomarkers is to measure markers in a small sample set of convenience, which usually contains serum or plasma samples of cases obtained from a single institution biorepository. Control samples often obtained from a separate collection from healthy people, with an inadequate sample size and unmatched for age, gender, race/ethnicity, location, or method of collection.

The prospective specimen collection, retrospective blinded evaluation (PRoBE) design, first described by Pepe et al. (Pepe, Feng, Janes, Bossuyt, & Potter, 2008) and updated in (Pepe, Li, & Feng, 2015), provides guidance for the strategic design of biomarker studies. PRoBE relies on well-designed prospective specimen collections that are fit-for-use, such as early diagnosis, prognosis, or serial samples for monitoring of early relapse.

The key element in the prospective study design is planned incorporation of relevant controls within the design of the biorepository, so that the comparison group is collected using standardized laboratory procedures in similar time frames and locations as the cases. For both discovery and validation studies, selection of the cases and controls from biorepositories should follow, as closely as possible, the biomarker development phase and the intended clinical use. The five phases of biomarker discovery progress from the preclinical exploratory phase, a clinical assay and validation phase, then onto retrospective longitudinal and prospective screening phases (Pepe et al., 2001). Following rigorous study design guidelines minimizes false discovery of biomarkers and streamlines the efficiency of the validation process.

2.3.2. Recognizing Breast Cancer Tumor Heterogeneity

Given the striking clinical and molecular differences between breast cancer subtypes, many circulating breast cancer biomarkers will have differential expression between the subtypes. Breast cancers are clinically divided into endocrine –responsive tumors that express the estrogen receptor (ER) and/or the progesterone receptor (PR), erbB2-expressing tumors (Her2+), and triple-negative breast cancers (TNBCs). Further molecular definitions based on RNA expression profiling have led to additional clinically relevant subtypes within these broad categories ("Comprehensive molecular portraits of human breast tumours," 2012). Use of a mixed collection of breast cancer sera in the discovery or validation phases of biomarker research is likely to over-represent the subset of low-grade ER+ Her2- breast cancers, for which survival rates are highest, mammographic screening is most effective, and circulating biomarkers for detection and relapse are less clinically needed. With this approach, biomarkers for the less common high-grade cancers are more likely to be eliminated early in discovery due to low overall sensitivity. There are two solutions to this challenge. One solution is to incorporate larger

numbers of samples within both the discovery and validation phases, to ensure adequate representation of the under-represented high-grade subtypes of breast cancer. This requires different statistical analysis strategies (Skates, 2014; Skates et al., 2013) that permit the initial selection of biomarkers with low clinical sensitivity (Wallstrom, Anderson, & LaBaer, 2013). The second solution is to focus the discovery (and validation) studies on a specific subtype of breast cancer. Our own strategy has evolved from broad, multiple-subtype screening for breast cancer (Anderson, Sibani, et al., 2011) to focused discovery and validation based on individual subtypes, such as basal-like breast cancers (J. Wang et al., 2015). As with all biomarker research, circulating biomarkers that are associated with changes in tumor biology (such as tumor mutation, gene expression, or alterations in the tumor microenvironment) may have a greater likelihood of validating in later phases of development.

2.3.3. Selection of Relevant Controls

The size, number, and selection of samples from cases and controls for both discovery and validation studies depends on the overall clinical use, so that early involvement of biostatisticians and epidemiologists throughout the discovery and validation pipeline is essential. For breast cancer, controls are usually matched by gender, age (+/- 5 years), which indirectly addresses menopausal status, and regional location of the collected samples, which limits confounding variables in sample collection (see below) and race/ethnicity variances. For early detection, incorporation of appropriate benign breast disease controls and follow up are needed to limit both false positive and false negative discovery. Ideally, both cases and controls are collected in similar clinical settings (screening mammography or diagnostic mammography), prior to treatment, at the time of biopsy or resection (for prognostic markers), or at defined time intervals post-surgery (for early relapse markers). The impact of stress, diet, concomitant medications, and systemic

therapy on biomarker levels are hidden potential confounders that can be minimized with a planned study design. Even with careful selection, a certain number of healthy controls may be positive for a biomarker, and it may be difficult to determine an analytical false positive from an asymptomatic early cancer diagnosis.

2.3.4. Minimizing Pre-analytic Variability

Standardization of specimen collection procedures and clinical annotation for biomarker research is an essential component of a reliable pipeline. The current significant variation in quality and suitability of samples collected impacts research quality and the ultimate clinical utility of identified biomarkers. Protocols that detail procedures (Dash et al., 2012) with defined clinical annotation (Robb et al., 2014) are available and can be generalized across the biomarker pipeline. Funding agencies and peer-reviewed publications have an emerging recognition of the value of high quality samples prepared using standard protocols and the need for detailed description of sample collection methods. This includes protocols for collection, processing and storage, time delays to collection and processing, batch numbers and reference sets, validation of tissue pathology data linked to blood samples, inventory management, sample distribution and standardized equipment (Dash et al., 2012). Hundreds of pre-analytic variables may be considered for annotation in biorepositories (Robb et al., 2014).

2.3.5. Assay Migration to a Clinically-Compatible Platform

Once the early discovery and verification studies have identified a discrete number of potential biomarkers for an intended clinical use, proteomics, immunomics, and genomics assays usually need to be migrated to a clinically-compatible platform prior to further validation studies. This process may take 1-2 years to establish standardized procedures, precision and variability, and to replicate the prior data obtained from the discovery

platform. Typically, at this stage the numbers of biomarkers are limited, and the assay becomes cost-effective for large-scale serum or plasma screening.

2.3.6. Leveraging Existing Biorepositories

Once the single-institution discovery and initial independent blinded verification study is completed, the next phase is to perform further validation studies using larger cohorts of multi-institutional, blinded specimens. There are multiple biorepositories already available for select clinical applications. To meet the needs for early detection, the NCI Early Detection Research Network (EDRN, ("The Early Detection Research Network,")) has created the Breast Cancer Reference Set (BCRS), a prospective, multicenter annotated collection of over 700 sera and plasma from women undergoing screening mammography and diagnostic mammography (Marks et al., 2015). This is an invaluable resource collected in compliance with PRoBE principles that are available for distribution to researchers upon request (Feng et al., 2013; Marks et al., 2015). In a recent evaluation of 90 potential protein biomarkers using the BCRS, only CA 125 was found to have potential utility for the discrimination of ER- breast cancers (Marks et al., 2015).

A key question for any early detection biomarker is the lead time of biomarker detection prior to clinical diagnosis. Several large multicenter biorepositories have blood collections from healthy individuals, annotated for subsequent cancer diagnosis. This includes the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), the Women's Health Initiative (WHI), the Health Professionals Study, Janus, ROCA (Risk of Ovarian Cancer Algorithm), EPIC (European Prospective Investigation into Cancer), the Carotene and Retinol Efficacy Trial (CARET) and the Nurses' Health Study (NHS), all of which have blood samples obtained prior to clinical diagnosis (Bohm et al., 2011; Chan, Bandera, Greenwood, & Norat, 2015; Feng et al., 2013; Gonzalez, Daly, Tan, Marks, & Zangar, 2011; Gunter et al., 2015; J. Wang et al., 2015). Fewer biorepositories exist for evaluating

biomarkers for the early detection of clinical relapse, which requires prospective collection of serial samples after clinical diagnosis. Overall, other than early detection, there is a need for rapid, adaptable, multicenter, annotated specimens for targeted validation studies available to the research community.

2.4. SPECIFIC CLASSES OF BIOMARKERS

2.4.1. Proteins

A number of proteomic platforms have been used to identify novel circulating protein biomarkers for breast cancer (**Table 2-1**). Targeted sandwich ELISA microarrays have been used to identify novel proteins in pre-diagnostic plasmas for early detection of ER+/PR+ ductal breast carcinoma (Buas et al., 2015), and a separate panel of proteins were detected for triple-negative breast cancer (C. I. Li et al., 2012). In an independent microarray platform, ten proteins were altered in at least one breast cancer subtype, including elevated expression of RANTES (Gonzalez et al., 2011). Plasma proteomic profiling has also been used to identify breast cancer biomarkers that are differentially detected in patient plasma. Elevated EGFR levels have been detected in plasma samples collected up to 17 months prior to ER+ breast cancer diagnosis in post-menopausal hormone therapy users (Pitteri et al., 2010). Elevated glycolysis pathway proteins have also been detected in pre-diagnostic plasma compared with controls (Amon et al., 2012). A panel of 14 biomarkers that distinguish primary non-metastatic breast cancer and healthy controls have been reported using mass spectrometry of serum samples (Bohm et al., 2011). Serum levels of 20 peptides were found to change one year after the end of chemotherapy in breast cancer patients subjected to tumor surgical resection (Pietrowska et al., 2010). A large LC-MS proteomics dataset has been generated from plasma samples

prospectively collected from breast cancer patients and controls (Riley et al., 2011), providing a valuable resource for comparative analysis.

Table 2-1. Examples of Protein Biomarkers for Breast Cancer.

Protein Marker	Breast Cancer Subtype	Clinical Application	PRoBE Phase	References
CCL27, CCL28	TNBC	Early detection	Phase 3	(C. I. Li et al., 2012)
CSF2, RYBP, TFRC, ITGB4	ER+/PR+	Risk assessment and early detection	Phase 3	(Buas et al., 2015)
Glycolysis proteins	ER+	Early detection	Phase 3	(Amon et al., 2012)
EGFR	ER+	Early detection	Phase 3	(Pitteri et al., 2010)
Multiple proteins	Unselected	Early detection	Phase 3	(Opstal-van Winden et al., 2011)
Specific PTMs *	Unselected	Diagnosis	Phase 1/2	(Jin, Daly, Marks, & Zangar, 2013)
Dermcidin (DCD)	Intrinsic	Diagnosis	Rat model	(Brauer et al., 2014)
14 proteins	Unselected	Diagnosis	Phase 1/2	(Bohm et al., 2011)
20 peptides	Unselected	Monitoring response to therapy	Phase 1/2	(Pietrowska et al., 2010)

***Glutathione (GSH)-modified ceruloplasmin and 4-hydroxynonenal (HNE)-modified PDGF**

Two-dimensional nano-liquid chromatography coupled with tandem mass spectrometry (2D-Nano LC-MS/MS) has been used to detect elevated levels of afamin, apolipoprotein E and the isoform 1 of inter-alpha trypsin inhibitor heavy chain H4 (ITIH4), as well as decreased levels of alpha-2-macroglobulin and ceruloplasmin in sera collected prior to breast cancer diagnosis (Opstal-van Winden et al., 2011). Subcellular fractionation of plasma membranes (Leth-Larsen, Lund, & Ditzel, 2010), and secretome analysis

(Zawadzka et al., 2014) have also been applied for biomarker selection. There has been increasing interest in exploring the tumor interstitial fluid (TIF) as an enriched source of the tumor secretome (Gromov et al., 2010; Turtoi et al., 2011). TIF has the potential advantage of local detection of disease-specific biomarkers in high concentrations before they are diluted in the blood. (Haslene-Hox, Tenstad, & Wiig, 2013; Rifai et al., 2006).

Other potential biomarkers include angiogenic factors, cytokines, metalloproteases (Gonzalez et al., 2011), and circulating C-reactive protein (Chan et al., 2015; Cheng, Liu, & Zhang, 2015; Gunter et al., 2015). Dermcidin and the related host defense peptide psoriasin (S100A7 (Anderson, Wong, Polyak, Aronzon, & Enerback, 2009)) function in mammary carcinogenesis, and dermcidin levels are increased prior to breast cancer diagnosis in a Japanese study (Brauer et al., 2014), but S100A7 levels are not elevated (Anderson et al., 2009). PTMs have also been evaluated as potential circulating biomarkers. Alterations in mucin glycoproteins are common, functional molecular changes in breast cancer (Vester-Christensen et al., 2013), and MUC1 glycoproteins are currently used in clinical practice for monitoring advanced disease (Paoletti & Hayes, 2014). Additional PTMs have been evaluated, such as oxidized ceruloplasmin and platelet-derived growth factor (PDGF) which are differentially detected in plasma from breast cancer patients and benign controls (Jin et al., 2013). Overall, the majority of these circulating proteins have not yet been validated in blinded, retrospective or prospective multi-center clinical studies.

2.4.2. Autoantibodies (AAbs)

Abs to tumor-associated antigens are induced in cancer patients in response to altered structure or expression of proteins or glycans. Because of their stability, high specificity and ease of detection in cancer patient sera, they represent a class of circulating biomarkers that can be rapidly adapted to clinical diagnostic platforms and may reflect

underlying immune surveillance of tumors. A number of these AAbs have been identified (**Table 2-2**). Since individual AAbs often have low clinical sensitivity, AAb combinations may improve the sensitivity of detection (Anderson, Sibani, et al., 2011; Chapman et al., 2007; Katayama et al., 2015; Lacombe, Mange, Bougnoux, Prassas, & Solassol, 2014; Ladd et al., 2013; H. Lu et al., 2012; J. Wang et al., 2015). For example, a three-phase screening approach was used to identify an immunosignature of 28 AAbs in early stage breast cancer (IBC) for potential early diagnosis (Anderson, Sibani, et al., 2011). In a subsequent study focused on basal-like breast cancer, 10,000 antigens were screened using protein microarrays and identified a panel of 13 AAbs that can distinguish between cancer cases and controls (J. Wang et al., 2015). A separate phase 3 study using plasma derived from the Women's Health Initiative cohort identified AAbs to proteins involved in BRCA1, TP53 and cytokeratin networks prior to TNBC diagnosis (Katayama et al., 2015). Additional autoantigens include plasminogen (Goufman, Iakovlev, Tikhonova, & Lokshin, 2015), centrosomal antigens (Maroun et al., 2014), nuclear antigens (Mohammed & Abdelhafiz, 2015), and Alpha 1-antitrypsin (Lopez-Arias et al., 2012). Using sera from primary breast cancer patients, DCIS and healthy controls for Abs against P53, c-myc, NY-ESO-1, BRCA1, BRCA2, HER2 and MUC1, AAbs against at least one of these 6 proteins were associated with early-stage breast cancer (Chapman et al., 2007).

Table 2-2. Examples of Autoantibody Biomarkers for Breast Cancer.

Autoantibody Marker	Breast Cancer Subtype	Clinical Application	Probe Phase	References
13-AAb signature: Tp53*, MN1*, CTAG2, CTAG1B, RNF216, PPHLN1, PIP4K2C, ZBTB16, TAS2R8, WBP2NL, DOK2, PSRC1, TRIM21	Basal-like	Diagnosis	Phase 1/2	(J. Wang et al., 2015)
28-AAb panel	Unselected	Early detection	Phase 1/2	(Anderson, Sibani, et al., 2011)
BRCA1, Tp53, cytokeratin	TNBC	Early detection	Phase 3	(Katayama et al., 2015)
Glu-plasminogen (Pg)	Unselected	Diagnosis	Phase 1/2	(Goufman et al., 2015)
Anti-nuclear Ab (ANA)	Unselected	Diagnosis	Phase 1/2	(Mohammed & Abdelhafiz, 2015)
Anti-centrosome Ab	Unselected	Diagnosis	Phase 1/2	(Maroun et al., 2014)
Alpha 1-antitrypsin Ab	Unselected	Diagnosis	Phase 1/2	(Lopez-Arias et al., 2012)
6-AAb Panel: P53, c-myc, NY-ESO-1, BRCA1, BRCA2, HER2, MUC1	Unselected	Diagnosis	Phase 1/2	(Chapman et al., 2007)
CCNB1, FKBP52, GAL3, PAK2, PRDX2, PPIA, P53, and MUC1	Unselected	Diagnosis	Phase 1/2	(Lacombe et al., 2014)
HER2/neu and p53	Unselected	Early detection	Phase 3	(H. Lu et al., 2012)

*Associated with worse survival

2.4.3. Circulating Tumor Cells (CTCs)

Circulating tumor cells are cancer cells that have disseminated into the peripheral blood, but are rare in healthy individuals and patients with local disease (Bidard et al., 2013). Their identification and enumeration in blood is minimally invasive and have been

evaluated for prognosis and treatment selection. Detection of 5 or more CTCs per 7.5 mL of blood is associated with increased metastatogenesis, decreased progression-free survival and overall survival (Bidard et al., 2010; Giuliano et al., 2014; Z. F. Jiang et al., 2013). CTCs have been detected from multiple molecular subtypes, but at low frequencies in early stage disease (Bidard et al., 2014). Very high counts were observed more frequently in Luminal A and TNBC subtypes (Peeters et al., 2014). At this time, their primary use is to identify molecular changes in metastatic tumors (M. C. Liu, 2014). Ongoing clinical trials to investigate CTC clinical utility are reviewed in (Castle, Shaker, Morris, Tugwood, & Kirwan, 2014).

2.4.4. Exosomes

Exosomes are nano-sized vesicles released from all types of cells, present in almost all body fluids and are involved in intercellular communication (O'Driscoll, 2015). They are rich in proteins, DNA, mRNA, lncRNA and miRNA and their constituents provide insights into the contents of their cells of origin (Yu et al., 2015). Tumor-derived exosomes transfer information to other cells and can thus actively enhance cancer progression and metastasis, which has been demonstrated for TNBC (O'Brien et al., 2013). Sera from cancer patients may have both increased quantity and altered content of exosomes compared with exosomes in healthy control sera. Unlike normal exosomes, breast cancer-derived exosomes were found to convert pre-miRNA into miRNA including the two breast cancer-related miRs, miR-10b and miR-21 (Melo et al., 2014), and to contain unique tRNAs (Guzman et al., 2015).

2.4.5. MicroRNA (miRNA)

miRNAs function to modulate post-transcriptional regulation of gene expression in cancers, including breast cancers, and can be classified as oncogenes (oncomiRs) or tumor suppressor genes (oncosuppressor-miRs) depending on the genomic context (Andorfer, Necela, Thompson, & Perez, 2011; Jansson & Lund, 2012). Many circulating miRNAs are differentially detected in the blood of breast cancer patients (reviewed in (Bertoli, Cava, & Castiglioni, 2015)). These miRNAs are potential biomarkers for breast cancer diagnosis (Hu et al., 2012; E. J. Jung et al., 2012; Kleivi Sahlberg et al., 2015; Q. Wu et al., 2011), prognosis (Hu et al., 2012; Joosse, Muller, Steinbach, Pantel, & Schwarzenbach, 2014; Kleivi Sahlberg et al., 2015; Q. Wu et al., 2011) and prediction of response to therapy (E. J. Jung et al., 2012; Sun et al., 2012). A recent study identified upregulation of miR-106a-5p and miR-454-3p and downregulation of miR-195-5p and miR-495 ($p < 0.05$) in PBMCs from patients with early stage breast cancer (Mishra, Srivastava, Suman, Kumar, & Shukla, 2015). In addition to the widely used RT-qPCR, exploring other highly sensitive techniques such as small RNA sequencing (Kelly et al., 2015) and droplet digital PCR (Mangolini et al., 2015) will help detect specific, low abundance analytes within complex biologic fluids.

2.4.6. Circulating Tumor DNA (ctDNA)

Circulating cell-free DNA was originally identified in the 1970s in the blood of advanced cancer patients (Chapman et al., 2007; Goufman et al., 2015). Because of low plasma concentrations, reliable detection required the later development of highly sensitive fluorescent dye and PCR assays (K. Jung, Fleischhacker, & Rabien, 2010). Initially, cell-free DNA was used to measure copy number alterations in tumors (Skates, 2014) as well as acquired resistance to drugs (Pepe et al., 2001; Tripathy et al., 2014; Wallstrom et al., 2013). However, total circulating DNA content has not been consistent between studies

(K. Jung et al., 2010; Mohammed & Abdelhafiz, 2015), and the association with tumor stage and metastasis in breast cancer has been variable.

More recently, methods have evolved to detect tumor-specific aberrations within total DNA, termed circulating tumor DNA (ctDNA). ctDNA provides a unique insight into the tumor genomic profile, including mutations, loss of heterozygosity, gene amplification, chromosomal aberrations and epigenetic changes. High levels (95-100%) of similarity of PIK3CA mutation between tumor and plasma DNA have been reported in breast cancer (Board et al., 2010; Higgins et al., 2012). HER2 amplification in circulating DNA was detected in sera from patients with HER2 negative tumors at diagnosis (Pepe et al., 2008). ctDNA has been detected in the majority of blood samples from patients with advanced breast cancer (Bettegowda et al., 2014) and may have applications as liquid biopsies for monitoring minimal residual disease (MRD) and predicting relapse (Garcia-Murillas et al., 2015). In one study, the sensitivity of ctDNA was superior to both CA 15-3 (85% vs 59%) and circulating tumor cells (90% vs 67%) (Dawson et al., 2013).

In addition to somatic mutations and copy number alterations, tumor-specific epigenetic modifications within ctDNA have also been observed. *De novo* DNA methylation is often conserved through disease progression and are highly stable analytes (Wittenberger et al., 2014). Methylation of the tumor suppressor RASSF1 gene was detected in 80% of tumors and 40% of sera from 52 breast cancer patients using a methylation sensitive PCR-based method, and may change with response to therapy (Avraham et al., 2012). Recent advances in the detection of whole genome methylated DNA (reviewed in (Wittenberger et al., 2014)) may broaden the scope of potential circulating epigenetic biomarkers for breast cancer. Overall, rigorous prospective validation of ctDNA biomarkers will be required, and protocols for processing, storage and enrichment of ctDNA will need

standardization to create biorepositories for their evaluation (Mohammed & Abdelhafiz, 2015; Skates et al., 2013).

CHAPTER 3 : PROGRAMMABLE PROTEIN ARRAYS FOR IMMUNOPROFILING HPV-ASSOCIATED CANCERS

This chapter has been published:

Radwa Ewaisha, Ian Meshay, Jack Resnik, Benjamin A. Katchman, and Karen S. Anderson (2016). Programmable protein arrays for immunoprofiling HPV-associated cancers. *PROTEOMICS, 16*(8), 1215-1224. doi: 10.1002/pmic.201500376

ABSTRACT

Over 600,000 cancers each year are attributed to the human papillomavirus (HPV), including cervical, anogenital and oropharyngeal cancers (OPC). A key challenge in understanding HPV immunobiology is the diversity of oncogenic HPV types and the need for multiplexed display of HPV antigens to measure Ab responses. Custom HPV protein microarrays displaying 98 proteins as C-terminal GST fusion proteins, representing 8 antigens of two low-risk HPV types (HPV6 and 11) and ten oncogenic high-risk HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52 and 58) were generated. Robust and reproducible protein expression of 96/98 of the antigens was demonstrated using a human cell lysate expression system. The target epitopes and specificity of four monoclonal Abs were identified. Using sera from 10 patients with newly diagnosed OPC and 10 controls, specific IgG seroreactivity to HPV16 E1, E2, and E7 (a fold increase of 1.52, 2.19 and 1.35 in cases vs controls, respectively, all $p<0.005$) was demonstrated, confirming our prior data on an ELISA platform. HPV52 E7 Abs were also detected in serum from a patient with cervical cancer. The HPV protein array has potential for rapid identification of serologic responses to 12 HPV types.

3.1. INTRODUCTION AND BACKGROUND

Approximately 79 million Americans are currently infected with human papillomavirus (HPV) ("HPV Factsheet,"). While most infections are subclinical or result in benign neoplastic growth, HPV infection is a necessary event for the development of cervical cancer (Bosch & de Sanjose, 2003) and is strongly associated with anogenital and oropharyngeal cancers. Cervical cancer is the most common cancer among women in Eastern and Middle Africa and fourth in women worldwide (Ferlay et al., 2013), with an estimated 528,000 new cases in 2012 and an annual mortality rate of 270,000 deaths (Ferlay et al., 2010; Ferlay et al., 2015). The majority of these cases occurs in less developed countries due to limited resources for cytologic screening and HPV vaccination (Jemal, Ward, & Thun, 2010). In the US, there has been a recorded rise in incidence of HPV-positive oropharyngeal cancers among men in the last decade (Marur, D'Souza, Westra, & Forastiere, 2010).

Although HPV is a small double-stranded DNA virus, a major challenge to the detection of specific immune responses is the diversity of over 200 HPV types. These vary from non-oncogenic low-risk types such as HPV 6 and 11 that cause anogenital warts, to high-risk types that are oncogenic. HPV16 is responsible for 85-90% of HPV-associated OPCs (D'Souza et al., 2007; Marur et al., 2010), but only 50-55% of cervical cancers (Bosch et al., 2008). Eight high-risk types (HPV16, 18, 31, 33, 35, 45, 52 and 58) are responsible for 90% of invasive cervical cancer (Bosch et al., 2008; Woodman et al., 2007).

The humoral immune response to HPV plays a significant role in the settings of natural infection, vaccination and cancer. Type-specific IgG Abs to the L1 coat protein are induced in response to acute HPV infection, and the HPV vaccines are designed to induce high levels of protective anti-L1 IgG Abs (Villa et al., 2006). In contrast to HPV infection, the development of HPV cancers is associated with IgG Abs, primarily to the oncogenic

antigens E6 and E7, best studied for the most common subtype, HPV16. Serum Abs to HPV16 E6 and E7 proteins have been detected in sera of 30-40% of patients with invasive cervical cancer (Achour et al., 2009; Reuschenbach et al., 2008). Abs to HPV16 E6 and E7 were similarly detected in 30-50% of HPV-positive OPC patients (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015; E. M. Smith et al., 2010) and Abs to E6 have been detected 10 years before OPC diagnosis (Kreimer et al., 2013). Using a mammalian-based method of antigen display of the full HPV16 proteome (8 antigens), our laboratory has observed that the immune response to HPV16 is heterogeneous in OPC and also includes strong immune responses to HPV16 E1, E2, and E4 antigens (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015; Anderson, Wong, et al., 2011), highlighting the significance of exploring the immune response across the HPV proteome.

The primary methods for measuring HPV-specific IgG across multiple HPV types have used multiplexed bead arrays coupled with HPV fusion proteins expressed in *E. coli* (Combes et al., 2014; Kreimer et al., 2013). One extensive study of an HPV protein slide-based array displaying multiple HPV proteomes expressed in *E. coli* identified IgG Abs to HPV16 E7 in the sera of 60% of patients with HPV16+ invasive cervical cancer, but limited detection of other Abs (Luevano et al., 2010). Our experience with mammalian expression of HPV16 proteins suggest that the method of protein display markedly impacts both yield and epitope presentation (F. Festa et al., 2013). However, when used in bead- and ELISA-formats, this method is cost-ineffective for detecting Abs across a wide range of viral proteins. We predicted that custom protein arrays displaying a broad spectrum of HPV antigens in a slide-based format may improve the detection of serologic responses to HPV.

Here, we have generated the first HPV programmable protein microarrays (termed NAPPA) displaying eight antigens from 2 low-risk and 10 high-risk HPV types. Expression plasmids were co-printed on glass slides with anti-GST Abs using human cell lysate (N. Ramachandran et al., 2004; N. Ramachandran et al., 2008). We demonstrate robust and reproducible protein expression across 96 of 98 viral antigens, and rapid mapping of the specificity of four HPV-specific monoclonal Abs (MAbs). The arrays permitted rapid mapping of Ab responses in sera from 11 patients with HPV-associated malignancies. These high-risk HPV protein arrays are designed to rapidly and specifically detect a wide array of serum Abs to multiple HPV types and may be useful to identify biomarkers for the detection and prognosis of HPV-associated malignancies.

3.2. MATERIALS AND METHODS

3.2.1. Patient Sera

The HPVOPC sera used in this study are a subset of the serum collection used in the HOTSPOT study previously described (K. S. Anderson, J. E. Gerber, et al., 2015; D'Souza et al., 2014). Samples were collected from newly diagnosed, histopathologically confirmed oropharyngeal cancer (OPC) patients (n = 10) from four study sites prior to initiation of treatment between October 2009 and May 2013. Healthy control sera (n = 10) used in this study, previously described (K. S. Anderson, K. R. Dahlstrom, et al., 2015), are a subset of a molecular epidemiology study of head and neck cancer at the MD Anderson Cancer Center, collected between January 2006 and September 2008. Serum from one patient with invasive cervical cancer serum was obtained from Biorepository for Molecular Signatures of Cervical Cancer developed by the Centers for Disease Control and Prevention as a Clinical Epidemiology and Validation Center for NCI's Early Detection Research Network (EDRN). Samples were collected using a standardized sample

collection protocol and stored at -80°C until use. Written informed consent was obtained from all participants under institutional review board approval. The HPVOPC sera used in this study are a subset of the serum collection used in the HOTSPOT study previously described (K. S. Anderson, J. E. Gerber, et al., 2015; D'Souza et al., 2014). Samples were collected from newly diagnosed, histopathologically confirmed oropharyngeal cancer (OPC) patients (n = 10) from four study sites prior to initiation of treatment between October 2009 and May 2013. Healthy control sera (n = 10) used in this study, previously described (K. S. Anderson, K. R. Dahlstrom, et al., 2015), are a subset of a molecular epidemiology study of head and neck cancer at the MD Anderson Cancer Center, collected between January 2006 and September 2008. Serum from one patient with invasive cervical cancer serum was obtained from Biorepository for Molecular Signatures of Cervical Cancer developed by the Centers for Disease Control and Prevention as a Clinical Epidemiology and Validation Center for NCI's Early Detection Research Network (EDRN). Samples were collected using a standardized sample collection protocol and stored at -80°C until use. Written informed consent was obtained from all participants under institutional review board approval.

3.2.2. Gene Design, Codon Optimization, and Cloning

DNA constructs of all genes for 10 high-risk (16, 18, 31, 33, 35, 39, 45, 51, 52, and 58) and 2 low-risk (6 and 11) HPV types were commercially prepared by GenScript (Piscataway, NJ) in the pDONR221 vector. For HPV types 6 and 11, both variations of E5 (denoted E5a and E5b) found naturally due to alternative reading frames, were generated. To use the Gateway cloning system, attB1 (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACC–3') and attB2 (3'–GACCCAGCTTTCTTGTACAAAGTGGTCCCC–5') linkers were added to flank the genes. The ORFs were transferred by recombination cloning into the pANT7-cGST vector, which

is optimized for maximal *in vitro* expression of proteins with a C-terminal glutathione S-transferase (GST) fusion tag (N. Ramachandran et al., 2004). DNA sequences were obtained from the Papillomavirus Episteme (PaVE), an online database maintained by the National Institute of Allergy and Infectious Disease (NIAID) (http://pave.niaid.nih.gov). The sequences were codon optimized by the manufacturer using the proprietary OptimumGene Algorithm that optimized codon usage, as codon adaptability, mRNA structure, and cis-elements in transcription and translation (X. LIU et al., 2012). Plasmids were sequence verified and deposited in the DNASU Plasmid Repository (Seiler et al., 2013) and will be made available upon request.

3.2.3. HPV Microarray Generation

Highly purified plasmid DNA was prepared using a high-throughput DNA factory robotic system as previously described (J. Qiu & LaBaer, 2011) and DNA concentration was normalized to 1200 ng/µl prior to printing. Protein arrays were generated as previously described (Anderson, Sibani, et al., 2011; J. Qiu & LaBaer, 2011). Briefly, plasmid DNA was incubated overnight at 4°C with a printing mix of capture anti-GST Ab (50 µg/mL, GE Healthcare Biosciences, Piscataway, NJ), BS3 protein cross-linker (2 mM, Pierce, Rockford, IL) and Bovine Serum Albumin (BSA) (3 mg/mL, Sigma-Aldrich) before arraying onto aminosilane-coated glass slides using a Genetix QArray2 with 300 µm solid tungsten pins. Positive controls included the highly immunogenic EBV-derived antigen EBNA-1 (Rickinson & Kieff, 2001) and purified human IgG protein. Negative controls included no spotted material (nonspots) and all material except DNA (no DNA). Each sample was spotted in duplicate and in non-adjacent positions. Three subarrays, each printed with 98 genes in duplicate, were printed on each slide. Two sequential printing batches were run on the same day using the same source plate of plasmid DNA preps. Arrays were stored in an air-tight container at room temperature and protected from light. Array quality control

was performed using both picogreen DNA staining and anti-GST staining of IVTT-expressed proteins. A 1:200 dilution of anti-GST MAb (Cell Signaling Technology, MA, USA) was used to confirm expression of proteins from the plasmids printed on the arrays. The correlation of protein expression was compared within (intra-batch) and between (inter-batch) printing runs and between subarrays on the same slide (intra-array) to determine the reproducibility of slide production. Highly purified plasmid DNA was prepared using a high-throughput DNA factory robotic system as previously described (J. Qiu & LaBaer, 2011) and DNA concentration was normalized to 1200 ng/µl prior to printing. Protein arrays were generated as previously described (Anderson, Sibani, et al., 2011; J. Qiu & LaBaer, 2011). Briefly, plasmid DNA was incubated overnight at 4°C with a printing mix of capture anti-GST Ab (50 µg/mL, GE Healthcare Biosciences, Piscataway, NJ), BS3 protein cross-linker (2 mM, Pierce, Rockford, IL) and Bovine Serum Albumin (BSA) (3 mg/mL, Sigma-Aldrich) before arraying onto aminosilane-coated glass slides using a Genetix QArray2 with 300 µm solid tungsten pins. Positive controls included the highly immunogenic EBV-derived antigen EBNA-1 (Rickinson & Kieff, 2001) and purified human IgG protein. Negative controls included no spotted material (nonspots) and all material except DNA (no DNA). Each sample was spotted in duplicate and in non-adjacent positions. Three subarrays, each printed with 98 genes in duplicate, were printed on each slide. Two sequential printing batches were run on the same day using the same source plate of plasmid DNA preps. Arrays were stored in an air-tight container at room temperature and protected from light. Array quality control was performed using both picogreen DNA staining and anti-GST staining of IVTT-expressed proteins. A 1:200 dilution of anti-GST MAb (Cell Signaling Technology, MA, USA) was used to confirm expression of proteins from the plasmids printed on the arrays. The correlation of protein expression was compared within (intra-batch) and between (inter-batch) printing runs and

between subarrays on the same slide (intra-array) to determine the reproducibility of slide production.

3.2.4. Detection of Serum Abs using HPV Arrays

Serum Ab detection was performed as previously described (K. S. Anderson, D. W. Cramer, et al., 2015; Anderson, Sibani, et al., 2011; J. Qiu & LaBaer, 2011). Slides were incubated with rocking for 1 h at room temperature with SuperBlock (Thermo Fisher Scientific, Waltham, MA). Proteins were expressed by injecting 150 µl 1-Step Human Coupled *in vitro* Expression system (Thermo Fisher Scientific) into slides sealed with HybriWell (Grace BIO-LABS, OR) and incubating for 1.5 h at 30°C for protein expression followed by 30 min at 15°C for protein capture. Slides were rinsed twice with PBST (0.2% tween), washed 3 times for 5 min each with 5% w/v milk powder (MP Biomedicals, LLC, CA, USA) in 0.2% PBS-Tween (5% milk-PBST) and then blocked in milk for 1 h. Serum was diluted 1:80 in 25% *E. coli* lysate prepared in 5% milk-PBST (0.2% tween) (J. Wang et al., 2013) and rotated for 2 h at room temperature. The serum was added to individual subarrays separated by a multi-well gasket (Grace BIO-LABS). Slides were incubated with serum for 16 h at 4°C with rocking, washed 3 times with 5% milk-PBST (0.2% tween) then incubated for 1 h at room temperature with Alexa Fluor 647 goat anti-human IgG (#109-605-008, Jackson Immunoresearch Laboratories, PA, USA). Slides were washed, dried, and scanned by Tecan PowerScanner (Tecan Group, Männedorf, Switzerland).

3.2.5. Protein Array Image Analysis and Quantification

Signal intensity of individual spots on the scanned slides was measured using ArrayPro Analyzer version 6.3 (MediaCybernetics, Bethesda, MD). Normalization of raw intensity values was done by subtracting the slide background signal and dividing by the background signal subtracted from intensity of noncontrol spots. The slide background signal was determined by the first quartile of signal intensity of nonspots (spots with no

printed DNA). In addition, array images were qualitatively examined to identify and confirm positive responses by adjusting raw images to extreme brightness and contrast using ArrayPro Analyzer and visual inspection of diffused signal (ring) as described previously (W. R. Montor et al., 2009).

3.2.6. Detection of Expressed Proteins using Anti-HPV Monoclonal Antibodies

RAPID ELISA was performed as described in (D'Souza et al., 2014) to evaluate the binding of specific MAbs raised against HPV16 E2 (Abcam, UK), HPV16/18 E6, HPV16 E7 or HPV16 E1/E4 (Santa Cruz Biotechnology, TX). GST-tagged E1, E2, E6 or E7, proteins from 12 HPV types were expressed by IVTT from 200 ng cDNAs in pANT7-cGST vectors using human HeLa lysate. A plasmid expressing GST was used as a positive control. Expressed proteins were diluted 1:100 with 5% Milk PBST 0.2% and 100 µL of diluted antigen were added to the specified wells. MAbs were diluted 1:3000 in blocking buffer and bound Abs were detected using a 1:6250 dilution of secondary HRP goat anti-mouse IgG Ab (Life Technologies, Carlsbad, CA). Luminescence detection was done using Glomax 96 Microplate Luminometer (Promega, WI).

3.2.7. Statistical Analysis

For the RAPID programmable ELISA, proteins were expressed in duplicate and RLU measurements were plotted as mean values. Each protein was expressed from DNA spotted in duplicate and signal intensities were plotted as mean values. The correlation of raw signal intensities of protein expression between the arrays randomly selected for quality control was determined with scatter plots and the correlation coefficient (R) was calculated to determine consistency. Levels of protein expression on the arrays were determined by calculating the mean values of raw signal intensities of duplicate spots from 2 arrays. Mean values of normalized signal intensity for serum Ab reactivity on protein

arrays were plotted for cases and controls and compared using unpaired t-test (Graphpad Prism version 5.0c, San Diego, CA).

3.3. RESULTS

3.3.1. Gene Design and Codon Optimization

A total of 98 proteins derived from 12 HPV PaVE reference sequence genomes (Van Doorslaer et al., 2013), including 10 high risk (16, 18, 31, 33, 35, 39, 45, 51, 52, 58) and 2 low risk (6, 11) were generated for this study. All genes were codon optimized, commercially synthesized and sequence confirmed. All reference and codon optimized DNA sequences, the percentage of nucleotide change as well as translated amino acid sequences are provided in **Appendix A**. We evaluated the difference between codon optimized or non-codon optimized sequences for the HPV16 and HPV18 genomes. Overall, protein expression levels of antigens from these 2 HPV types were similar following codon optimization (**Figure 3-1**), and three antigens (HPV16 E4, HPV16 L2 and HPV18 E5) had a fold increase of 1.19, 1.48 and 37.62, respectively (all $p<0.05$) after codon optimization. Non-codon-optimized HPV16 E2 has been difficult to express and was previously used as N-terminal (NE2) and C-terminal (CE2) fragments (K. S. Anderson, J. E. Gerber, et al., 2015); the codon-optimized gene is strongly expressed. In contrast, HPV16 L1 expression decreased after codon optimization.

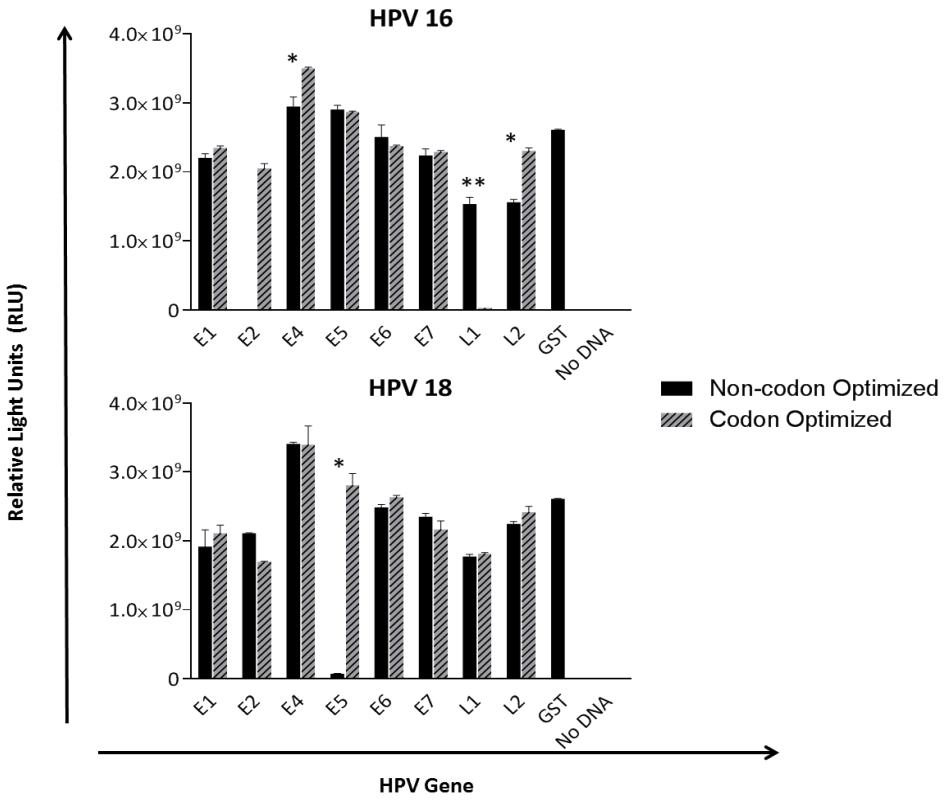

Figure 3-1. Comparison of Levels of Protein Expression Before and After Codon Optimization for Antigens from HPV16 (Top) and HPV18 (Bottom). Protein expression levels of antigens from HPV16 and HPV18 were similar following codon optimization. Non-codon optimized HPV16 E2 (not shown) was previously difficult to express but expressed at high levels after codon optimization. Protein expression levels markedly improved after codon optimization of HPV16 E4 ($p<0.03$), HPV16 L2 ($p<0.0031$) and HPV18 E5 ($p<0.0019$) (*). HPV16 L1 expression was lower after codon optimization (**).

3.3.2. Generation and Quality Control of Custom HPV Protein Microarrays

HPV protein microarrays comprising all 98 HPV proteins from 12 HPV types (**Appendix A**) were generated from printed cDNAs with IVTT and *in situ* protein purification using the C-terminal GST tags. The quality of the array printing was evaluated with picogreen DNA staining and the protein expressed was measured with anti-GST Abs. The reproducibility of array printing and protein display are shown in **Figure 3-2A**. Correlation coefficients of anti-GST signal intensities were determined for intra-array, intra-batch, and inter-batch replicate arrays (R>0.90) from two subarrays on the same slide or two randomly selected slides within a print run or between two print runs (**Figure 3-2B**). There were two independent printing batches and from each, one slide for DNA staining and two for protein staining were randomly selected. Scatter plots showing the variability in the correlation of protein expression among these controls are shown in **Figure 3-3**.

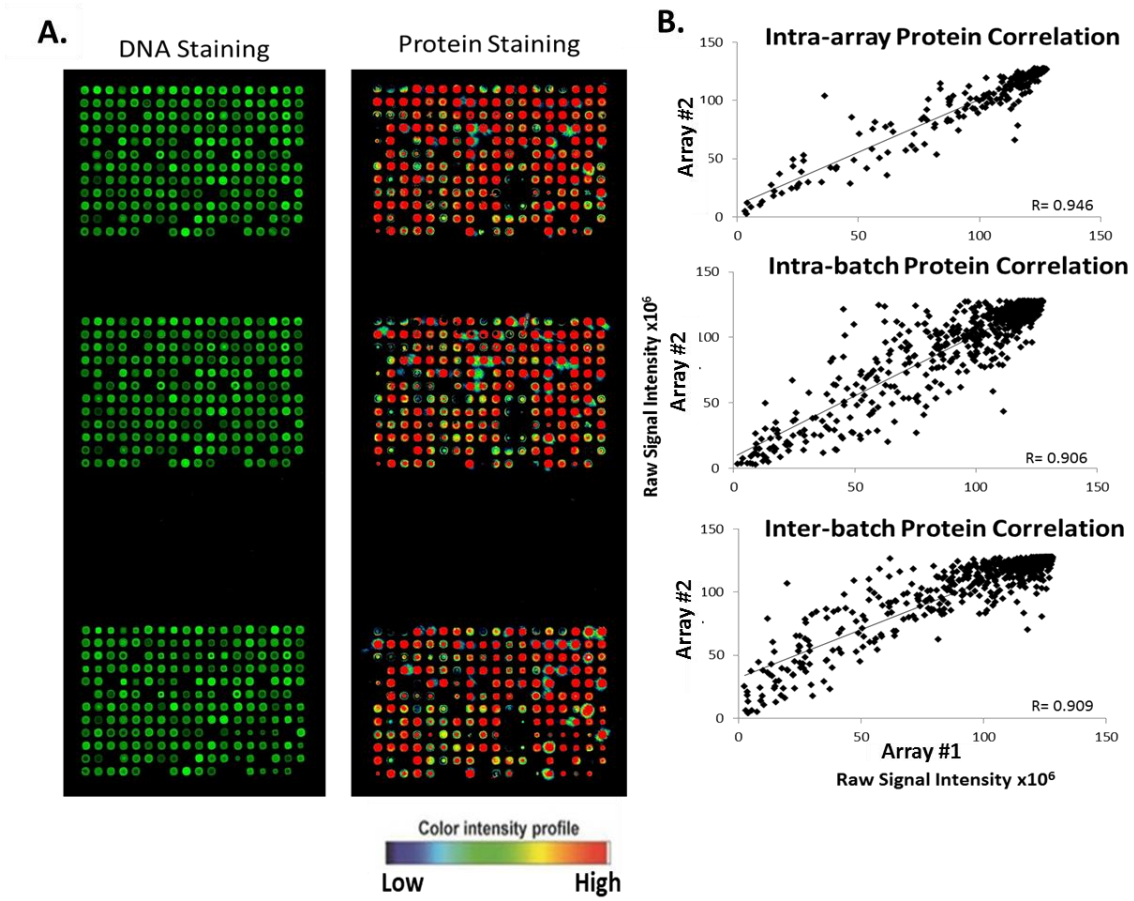

Figure 3-2. Reproducibility of Printing and Protein Expression on NAPPA. (A) Left: Picogreen to detect content of DNA printed on the microarrays. Right: Image of anti-GST binding to measure the level of protein display. (B) Plots of protein signal intensities from all spots from two subarrays on the same slide or two randomly selected slides within a print run or between two print runs.

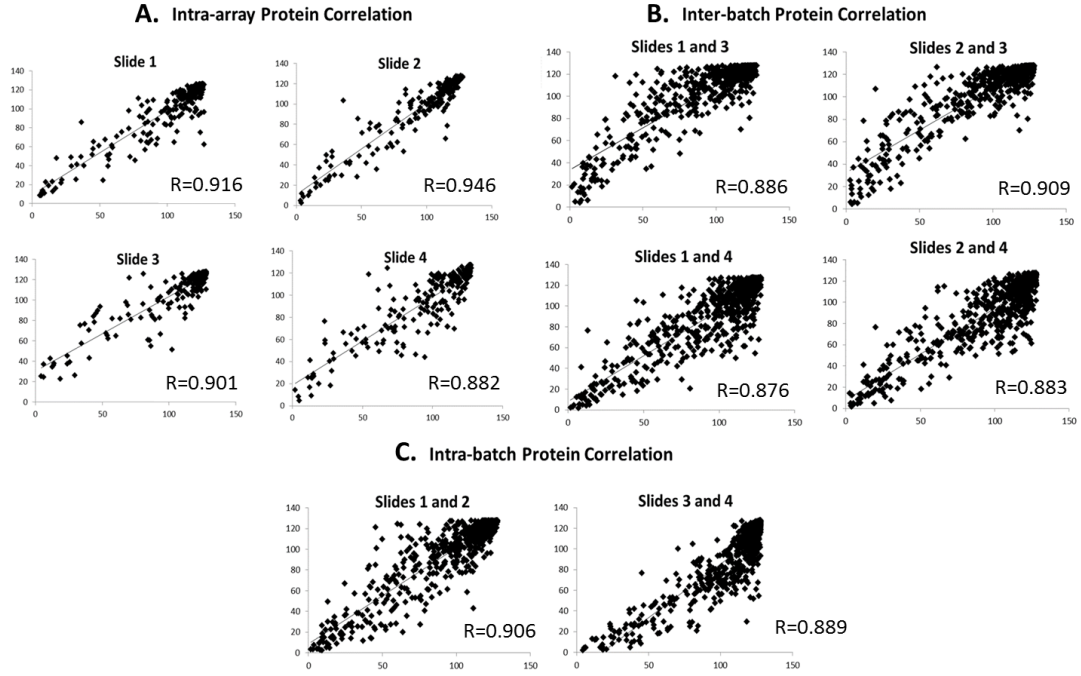

Figure 3-3. Plots of Protein Signal Intensities from All Spots on All Four Protein Staining Controls Used. Slides 1 and 2 were randomly selected from batch 1 while 3 and 4 from batch 2. These plots show signal intensities from two subarrays on the same slide (A) or two randomly selected slides within a print run (C) or between two print runs (B).

3.3.3. High-level HPV Protein Expression across Viral Types and Antigens

Expression of C-terminal GST-tagged HPV Ags on the arrays was confirmed by IVTT expression followed by anti-GST detection (**Figures 3-4**, **3-5**). We determined that 96/98 antigens had detectable protein signal over background, defined as five times the average signal of blank spots (dotted line, **Figure 3-5**). The C-terminal portion of the EBV-derived protein EBNA-1 was expressed as a positive control. Only HPV16 L1 and HPV35 E2 antigens had low expression signals, and IgG Abs to these proteins were not detected in patient sera (data not shown). Since the non-codon-optimized HPV16 L1 gene is well-expressed, future arrays will contain this version.

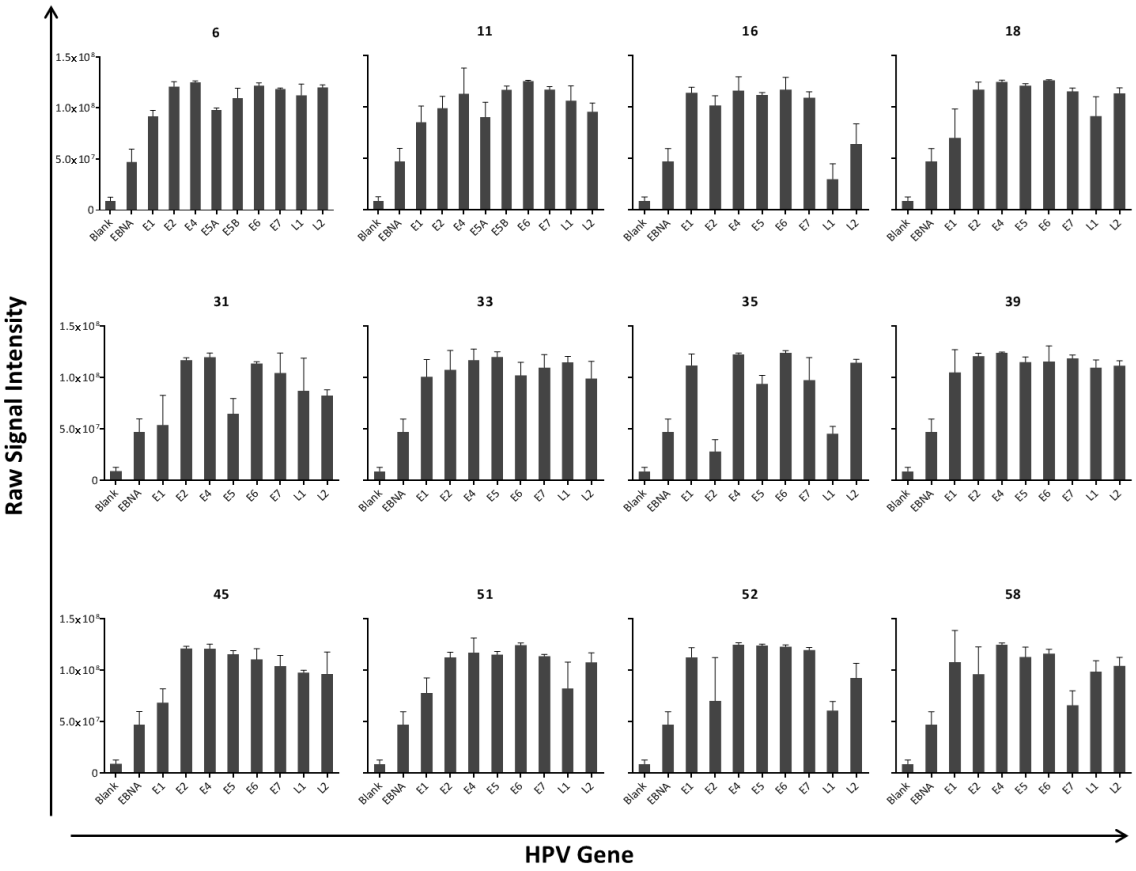

Figure 3-4. Detection of GST-Tagged HPV Antigens on the Arrays. Mean values of raw signal intensity from individual spots on the arrays are shown for all antigens from each HPV type.

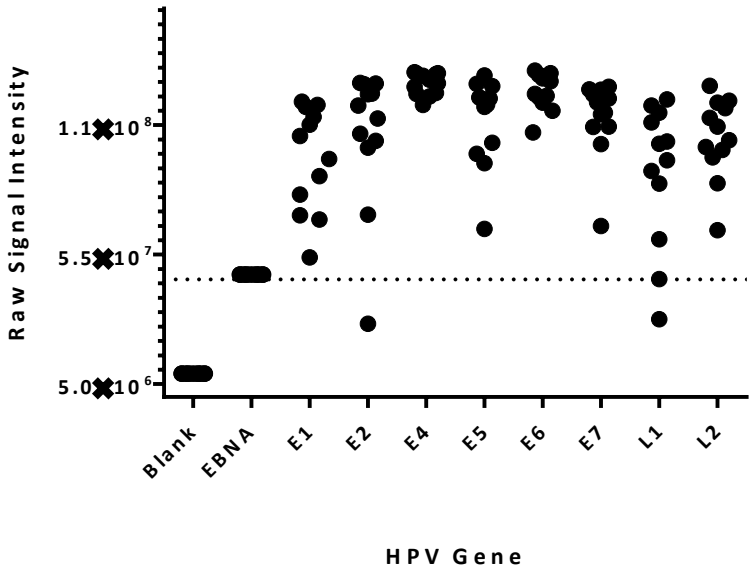

Figure 3-5. Detection of GST-tagged HPV Antigens on the Arrays. Protein signal was detected over background for 98% of antigens, defined as five times the average signal of blank spots (dotted line).

3.3.4. HPV Viral Proteins Retain Antigenic Specificity

To evaluate whether the displayed proteins retain their antigenic specificity, the immunoreactivity was evaluated using four commercially-available mouse monoclonal IgG antibodies (MAbs) that were raised against specific HPV Ags (HPV16 E1/E4, HPV16/18 E6, HPV16 E2 and HPV16 E7). HPV antigens were displayed in RAPID ELISA format, which uses the same expression system in a 96-well format for focused antigen display. The MAbs were added and mouse IgG was detected using secondary Abs and luminescence (relative light units, RLU, **Figure 3-6**). All expressed proteins were detected using anti-GST MAb, confirming expression of full-length proteins with C-terminal GST tags (data not shown). The anti-HPV16 E2 MAb was raised to amino acids 2-17 (ETLCQRLNVCQDKILT) of HPV16 E2 and was specific for HPV16 E2 with cross-reactivity to HPV51 E2. HPV51 E2 has a similar epitope as the parent HPV16 E2 protein (ETLCHRLNVCQEKILD) (**Figure 3-6A**). The E4 ORF is the most divergent between HPV

types (Bell, Martin, & Roberts, 2007), and the anti-HPV16 E4 MAb reacted specifically with HPV16 E4 with no cross-reactivity with other E4 proteins (**Figure 3-6B**). The anti-HPV E6 MAb was raised for cross-specificity with HPV16 and HPV18 E6 proteins (**Figure 3-6C**). As expected, the HPV E6 MAb was specific, but weakly reactive, for both HPV16 E6 and HPV18 E6. The HPV16 E7 MAb showed strong cross-reactivity with HPV35 E7, which is highly homologous to HPV16 E7 (**Figure 3-6D**). Overall, these data demonstrate the display of seroreactive epitopes from the proteins expressed using human cell lysate, and demonstrate the utility of the protein display for rapid mapping of the specificity of MAbs to viral antigens.

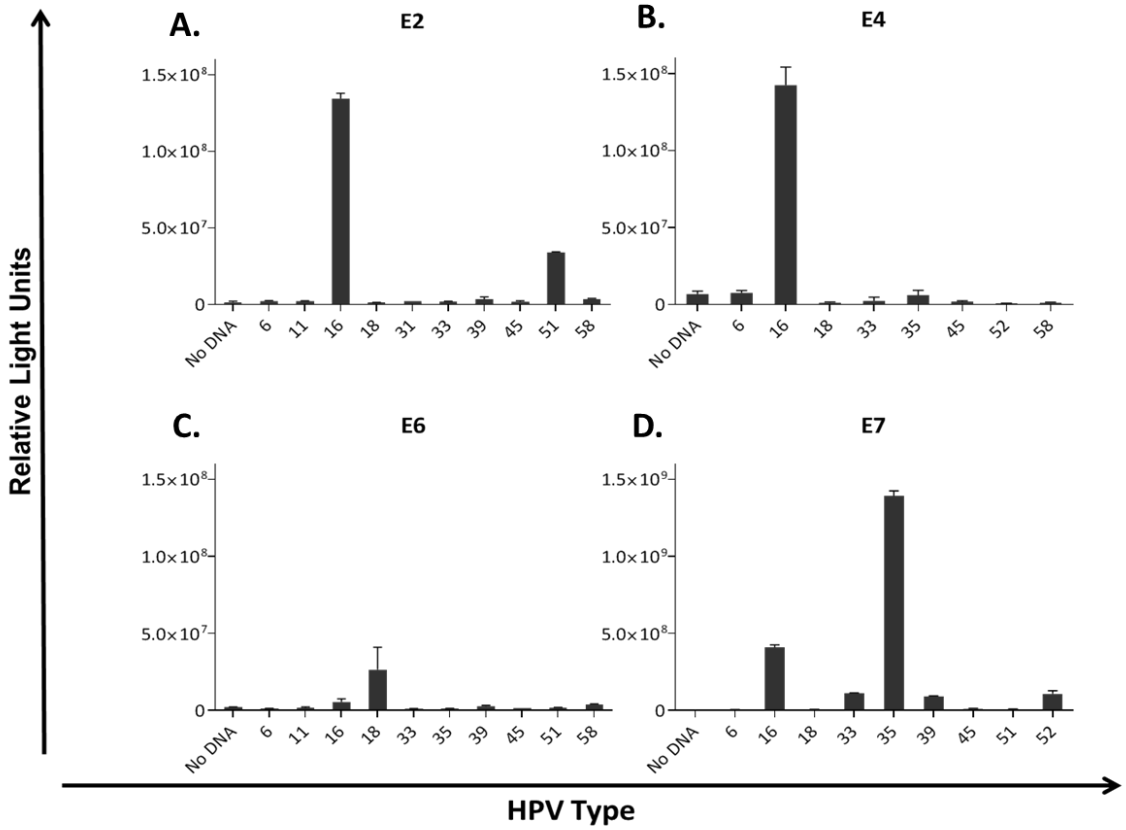

Figure 3-6. Mapping Monoclonal Antibody Reactivity Using Arrays of Displayed HPV Antigens. Antibody specificity to GST-tagged HPV antigens were detected using the 96-well RAPID ELISA format. Monoclonal antibodies specific for HPV16 E2 (A), HPV16 E4 (B) HPV16/18 E6 (C) and HPV16 E7 (D) were detected with anti-mouse IgG secondary antibodies.

3.3.5. Detection of HPV-Specific Abs in Patient Serum Using HPV Protein Arrays

The primary goal of generating HPV protein arrays is to develop methods for the rapid detection of a broad-spectrum of Abs in sera to multiple HPV types. We displayed all 98 HPV proteins from 12 HPV types in duplicate in protein microarray format, with the C-terminal portion of EBNA-1 included as a positive control. Each individual subarray on microarray slides was incubated with selected serum from patients with HPV-associated malignancies, either oropharyngeal cancer, invasive cervical cancer, or healthy controls,

to demonstrate the detection of HPV-specific Abs. In **Figure 3-7A**, reactivity of serum from a patient with known HPV16 positive oropharyngeal cancer is shown. In this serum, we observed IgG Abs to the HPV16 E1 and E2 proteins, which is found in 56% of newly-diagnosed OPC patients [14]. We also observed seroreactivity to HPV31 E2 protein, which likely represents epitope cross-reactivity, as infection with HPV31 is present in <5% of OPC. Serum from a healthy control subject only showed reaction with the EBNA-1 positive control protein (**Figure 3-7B**). We also identified a patient with HPV16-negative invasive cervical cancer who had seroreactivity to the HPV52 E7 protein, (**Figure 3-7C**), suggesting that Abs to different HPV types may be detectable with the arrays.

Serum Ab reactivity to four early HPV16 antigens (E1, E2, E4 and E7) was determined for serum from 10 HPVOPC cases and 10 healthy controls by measuring normalized signal intensity values from protein arrays probed with serum. These HPVOPC cases were previously shown via a RAPID ELISA assay to have Abs to at least one early HPV16 protein, while the negative healthy control sera did not (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015). The geometric mean signal intensities of cases vs controls and fold change were E1 (1.82 vs 1.20, $p<0.0016$), E2 (1.50 vs 0.69, $p<0.0001$), E4 (1.42 vs 1.26, not significant) and E7 (1.23 vs 0.91, $p<0.0083$), respectively (**Figure 3-8** and **Table 3-1**). On visual inspection, patients who had Abs to HPV16 E7 showed no cross-reactivity to E7 proteins from other HPV types, while all patients who had anti-HPV16 E4 Abs cross-reacted with E4 from HPV types 31 and 35. Four of the eight sera that had Abs to HPV16 E1 had Abs to HPV 11 E1 and one had Abs to HPV39 E1. All patients who had Abs to HPV16 E2 except one showed cross-reactivity to E2 Ag from at least two of the HPV types 18, 31, 39 or 58.

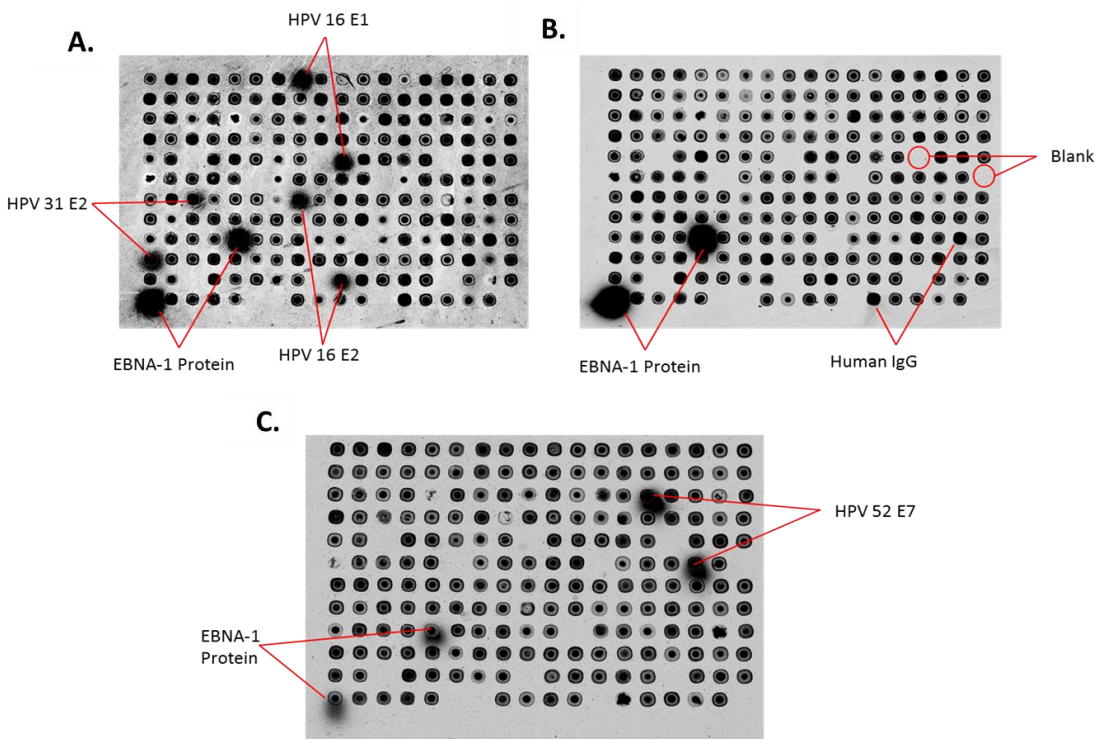

Figure 3-7. Detection of IgG in Human Serum Using HPV Protein Arrays. (A) Detection of IgG Abs in sera from a patient with HPV16 OPC. Antibodies to EBNA-1 protein, HPV16 E1 and E2 and HPV35 E2 proteins are detected. (B) Healthy control serum with specificity for EBNA-1 protein. (C) Detection of IgG Abs in sera from a patient with HPV16-negative invasive cervical cancer. Strong immunoreactivity to the HPV52 E7 protein, as well as EBNA-1 protein, is shown.

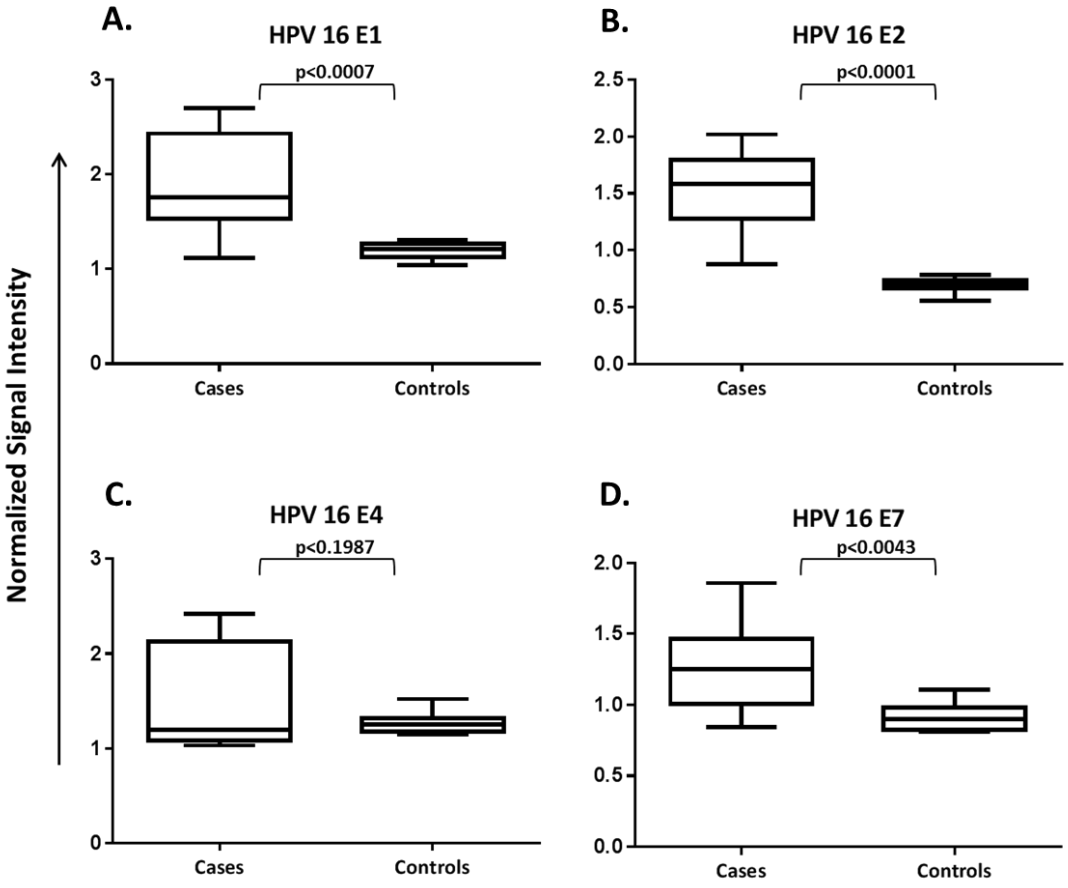

Figure 3-8. Normalized Proteome Microarray Antibody Reactivity Against Four Early hpv16 Antigens: E1 (A), E2 (B), E4 (C) and E7 (D) in Sera from 10 HPVOPC Cases and 10 Healthy Controls. Box-and-whisker plots of antibody reactivity are presented as normalized signal intensity value on the vertical axis. The top and bottom of the box indicate the 75th and 25th percentiles, respectively, while the line within the box indicates the median. The whiskers mark the interquartile range.

Table 3-1. Normalized Signal Intensities of HPVOPC Sera on ELISA Compared with NAPPA Protein Arrays.

		RAPID ELISA					*NAPPA*			
	HPV16 Antigen	E1	NE2[a]	CE2[a]	E4	E7	E1	E2	E4	E7
	Case 1	19.14	4.45	6.70	1.13	70.84	1.78	1.28	2.42	1.86
	Case 2	7.99	8.15	11.52	1.25	1.22	1.33	1.75	1.22	0.84
	Case 3	36.50	2.21	1.88	1.14	42.43	1.74	0.88	1.08	1.16
	Case 4	0.84	17.74	31.80	25.34	0.78	1.12	1.79	2.08	1.03
	Case 5	26.44	13.29	15.12	0.42	46.64	2.70	1.45	1.51	1.68
	Case 6	7.05	1.94	8.44	1.89	0.98	2.61	1.68	2.26	0.94
	Case 7	42.27	18.63	37.61	1.33	41.25	2.37	2.02	1.08	1.39
	Case 8	25.94	2.16	19.46	1.05	39.23	1.74	1.50	1.15	1.37
	Case 9	38.86	4.91	26.75	3.83	87.85	1.60	1.28	1.17	1.34
	Case 10	25.61	4.01	25.57	1.66	1.43	1.82	1.80	1.03	1.04
	Control 1	1.79	1.55	1.51	1.33	0.90	1.27	0.70	1.31	0.82
	Control 2	0.94	0.95	1.06	0.88	0.99	1.21	0.73	1.52	0.90
	Control 3	1.08	0.98	1.74	0.81	0.83	1.31	0.69	1.26	0.90
	Control 4	1.39	1.09	0.85	1.00	1.08	1.13	0.78	1.14	0.81
	Control 5	0.67	0.95	0.89	0.77	0.71	1.22	0.70	1.27	0.98
	Control 6	1.01	0.99	1.21	1.09	1.05	1.27	0.70	1.22	1.11
	Control 7	1.43	1.16	1.36	1.26	1.35	1.22	0.73	1.24	0.82
	Control 8	1.27	1.16	1.41	1.21	1.09	1.12	0.56	1.37	1.00
	Control 9	0.92	0.81	0.94	1.00	1.16	1.19	0.74	1.19	0.94
	Control 10	0.76	0.82	1.01	1.17	1.01	1.04	0.58	1.15	0.89
Median	Cases	25.77	4.68	17.29	1.29	40.24	1.76	1.59	1.20	1.25
	Controls	1.04	0.99	1.14	1.05	1.03	1.21	0.70	1.25	0.90
Mean	Cases	23.06	7.75	18.49	3.91	33.27	1.88	1.54	1.50	1.27
	Controls	1.13	1.05	1.20	1.05	1.02	1.20	0.69	1.27	0.92
Geometric mean	Cases	15.76	5.58	14.03	1.76	11.03	1.82	1.50	1.42	1.23
	Controls	1.08	1.03	1.17	1.04	1.00	1.20	0.69	1.26	0.91
Standard deviation	Cases	14.21	6.48	11.73	7.59	31.43	0.52	0.34	0.54	0.33
	Controls	0.34	0.21	0.30	0.19	0.18	0.08	0.07	0.11	0.09
Cut-off		1.81	1.47	1.80	1.44	1.37	1.36	0.83	1.49	1.10
Fold increase		14.60	5.42	12.03	1.70	11.02	1.52	2.19	1.13	1.35
p value		0.0043	0.0396	0.0028	0.768	0.3263	0.0016	0.0001	0.3719	0.0083

Highlighted numbers are above the established cut-off values. Fold increase was calculated for the corresponding geometric means of cases and controls. [a]E2 was expressed as N-terminal and C-terminal fragments.

3.4. DISCUSSION

The marked advances in proteomic technologies for display of antigenic structures have led to novel tools for proteome-wide immune monitoring. Protein microarrays have enabled profiling of the Ab immune response to the proteomes of pathogens and self-antigens in human sera. These advances are rapidly changing our understanding of the disease biology and the heterogeneity of the immune response. Programmable protein display (NAPPA) permits rapid flexibility and high-throughput protein production for custom pathogen protein arrays which retain activity following months of storage (N. Ramachandran et al., 2004; N. Ramachandran et al., 2008). Our current method of using human cell lysate results in high levels and reproducible protein expression for multiplexed antigen display.

We have generated custom programmable HPV protein microarrays displaying 98 antigens of two low-risk HPV types that are associated with anogenital warts, and ten oncogenic high-risk HPV types that are the most prevalent types in cervical and HPV-associated oropharyngeal cancers. We demonstrate robust protein expression for all the protein antigens except HPV16 L1 and HPV35 E2. The variation in protein expression was not associated with antigen size or hydrophobicity (data not shown). The displayed antigens retained antigenic specificity of target epitopes, as measured by four HPV-specific MAbs. Ab reactivity to multiple early HPV proteins were detected in sera from patients with HPV-associated malignancies using the arrays (**Figures 3-7** and **3-8**).

The serologic response to HPV has been well-characterized for HPV16. Abs to the major capsid protein HPV16 L1 are an indication of past HPV infection but are not a reliable marker of HPV-associated tumors (Dillner, 1999). Since HPV early antigen expression, especially E6 and E7, is restricted to the later stages of viral progression, there is high frequency of seropositivity to HPV16 E6 and E7 in patients with an underlying HPV-

associated malignancy but not in controls (Combes et al., 2014; D'Souza et al., 2014; Zumbach et al., 2000). We confirmed a high correlation of HPV16-specific serum IgG detection between our ELISA (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015) and the custom HPV protein microarrays for Abs to E1, E2 and E7 proteins (R=0.66-0.86, Supplemental Table 2 and Figures 4 and 5). Abs to HPV16 E2 were detected in all 10 cases while only 4 cases showed reactivity to HPV16 E4 (Figure 5). This heterogeneous immune response is similar to our prior ELISA-based data using these sera (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015; Anderson, Wong, et al., 2011), but antigenic specificity in patient sera will need to be confirmed using large well-annotated and matched serum biorepositories. As compared with OPC, immunoreactivity to E1, E2 and E4 is uncommon in cervical disease (unpublished observations). The biology and clinical significance underlying the immune response within the same HPV type and between different anatomic sites is unknown at this time.

A key challenge in HPV serology is the diversity of oncogenic HPV types and the potential for serologic cross-reactivity on the arrays. We observed immunoreactivity to non-HPV16 Ags in HPV16+ OPC (**Figure 3-7A**), which could be due to epitope cross-reactivity between HPV types that have close phylogenetic linkage (Combes et al., 2014). Approximately half of invasive cervical cancers are caused by non-16 HPV types (Bosch et al., 2008). In a multiplex bead assay, Abs to HPV52 E6 and HPV 58 E7 were associated with invasive cervical cancer (Waterboer et al., 2005). Here, we detected seroreactivity of an HPV16 negative invasive cervical cancer patient to HPV52 E7 protein with no cross-reactivity with HPV16 Ags (**Figure 3-7C**). These data support the need for multiplexed detection of multiple high-risk HPV types for biomarker discovery.

To our knowledge, only one prior study displayed the proteomes of multiple HPV types on slide-based protein microarrays, and Abs to E7 were the immunodominant response in cervical cancer patient sera (Luevano et al., 2010). In that study, Abs to E6 and L1, although previously reported (Achour et al., 2009; Reuschenbach et al., 2008; Waterboer et al., 2005), were not strongly detected, which may be attributable to the larger size of these proteins compared with E7, or the difficulties associated with expression, folding, and stability using *E. coli* for protein array generation. No significant difference was observed between seroreactivity of patients with precancerous lesions (HSIL) and healthy controls, but this needs to be confirmed using independent approaches. Based on our experience with autoantigens (K. S. Anderson, D. W. Cramer, et al., 2015; Anderson, Sibani, et al., 2011), we predict that the high protein expression levels, antigenic specificity and the on-demand cell-free expression system used here may improve detection of HPV antigen-specific Abs.

Programmable protein arrays have also been used to investigate the humoral immune response to proteomes of other infectious agents including *Pseudomonas aeruginosa* (W. R. Montor et al., 2009) and *Mycobacterium tuberculosis* (Prados-Rosales et al., 2014). With the advancement in ORFeome collections and consequent reduction in the cost per antigen studied, display of full proteomes of pathogens is becoming more feasible. The DNASU Plasmid Repository (http://dnasu.asu.edu), comprising a collection of over 200,000 plasmids encoding proteins from over 600 organisms, was the source of plasmid DNA used for these arrays and is widely available for researchers (Seiler et al., 2013). Once in the expression vectors, these methods permit rapid conversion from gene panels to displayed proteins for immune monitoring within two weeks. In our data, batch-to-batch variation was limited and known HPV16 antigenic epitopes were detected. Proteins can be displayed in native or denatured forms, which may result in the display of unique

epitopes (J. Wang et al., 2013). At this time, PTMs of the *in vitro* expressed proteins are limited. In our experiment, careful attention to quantitative printing of cDNA is important to minimize variation.

The clinical importance of understanding the immunobiology of HPV infection is striking. Despite decades of research, the biologic basis by which the majority, but not all, of infected individuals clear HPV infections prior to cancer development remains largely unknown (M. van Duin et al., 2002; Woodman et al., 2007). In 2012, approximately 270,000 women died from cervical cancer (Ferlay et al., 2013), and the incidence of oropharyngeal cancer is rapidly rising in the US and Europe (Marur et al., 2010; Nasman et al., 2009). Biomarkers for the rapid detection of HPV-associated cancers are needed for targeted health care delivery on a global scale. Proteome-wide immune monitoring of HPV has the potential to identify novel biomarker of diagnosis and prognosis and facilitate studies of the dynamic interaction of HPV virology and host immunity.

CHAPTER 4 : SERUM IMMUNE PROFILING FOR EARLY DETECTION OF CERVICAL DISEASE

This chapter has been published:

Radwa Ewaisha, Gitika Panicker, Paul Maranian, Elizabeth R. Unger, and Karen S. Anderson (2017). Serum Immune Profiling for Early Detection of Cervical Disease. *Theranostics, 7*(16), 3814-3823. doi: 10.7150/thno.21098

ABSTRACT

The most recent (2012) worldwide estimates from International Agency for Research on Cancer indicate that approximately 528,000 new cases and 270,000 deaths per year are attributed to cervical cancer worldwide. The disease is preventable with HPV vaccination and with early detection and treatment of pre-invasive cervical intraepithelial neoplasia, CIN. Abs to HPV proteins are under investigation as potential biomarkers for early detection. To detect circulating HPV-specific IgG Abs, we developed programmable protein arrays (NAPPA) that display the proteomes of two low-risk HPV types (HPV6 and 11) and ten oncogenic high-risk HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52 and 58). Arrays were probed with sera from women with CIN 0/I (n=78), CIN II/III (n=84), or invasive cervical cancer (ICC, n=83). Abs to any early (E) HPV protein were detected less frequently in women with CIN 0/I (23.7%) than women with CIN II/III (39.0%) and ICC (46.1%, $p<0.04$). Of the E Abs, anti-E7 Abs were the most frequently detected (6.6%, 19.5%, and 30.3%, respectively). The least frequently detected Abs were E1 and E2-Abs in CIN 0/I (1.3%) and E1-Abs in CIN II/III (1.2%) and ICC (7.9%). HPV16-specific Abs correlated with HPV16 DNA detected in the cervix in 0% of CIN 0/I, 21.2% of CIN II/III, and 45.5% of ICC. A significant number (29 – 73%) of E4, E7, L1, and L2 Abs had cross-reactivity between HPV types. HPV protein arrays provide a valuable high-throughput tool

for measuring the breadth, specificity, and heterogeneity of the serologic response to HPV in cervical disease.

4.1. INTRODUCTION AND BACKGROUND

Measuring the humoral immune response to HPV Ags has been integral to understanding the natural history of infection and efficacy of vaccination (Doorbar et al., 2012; Villa et al., 2006; Woodman et al., 2007). Despite the potential of HPV serology in disease diagnosis and prognosis, its clinical application has been limited by HPV heterogeneity, assay variability, and viral immune evasion. HPV has a limited repertoire of proteins, grouped as early (E1, E2, E4, E5, E6, E7) and late (L1, L2) proteins. The late proteins form the viral protein coat during productive infections. The early proteins interact with host and viral proteins to maintain viral replication. The serologic response to genital HPV infection is primarily directed at conformational epitopes on the viral major capsid protein L1. As the infection is non-lytic, the host Ab response to L1 is weak and may persist for years, as an indication of past infection but not malignancy (Luevano et al., 2010; Stanley, 2010). Although anti-L1 Abs are an indication of past infection, only 50 – 70% of infected women seroconvert (Carter et al., 2000; Dillner, 1999).

Abs to both HPV16 E6 and E7 proteins have been detected at low levels in both serum and cervical-vaginal secretions of invasive cervical cancer (ICC) patients (Bierl et al., 2005). Their levels increase with cervical disease progression but they are not detectable in a subset of patients with cervical cancer (Gutierrez-Xicotencatl et al., 2016; Luevano et al., 2010; Reuschenbach et al., 2008; Stanley, 2003). They develop later in the course of ICC and are correlated with disease outcome (Gutierrez-Xicotencatl et al., 2016; Ravaggi et al., 2006; Silins et al., 2002). Studies of sera collected prior to the diagnosis of cervical cancer have shown that the presence of E6 and E7-specific Abs is associated with an

increased relative risk (RR=2.7) for cervical cancer, and can be detected, albeit infrequently, up to 5 years prior to diagnosis (Lehtinen et al., 2003). The percentage of women with false negative serology is dependent on the method of Ab detection (Achour et al., 2009; Combes et al., 2014; Kontostathi et al., 2016; Luevano et al., 2010; Waterboer et al., 2005; Zumbach et al., 2000).

The diverse array of oncogenic HPV types and the technical limitations of high throughput protein expression and display have been impediments to HPV immune profiling and most research has focused on select Ags from the most common viral types (Combes et al., 2014; Gutierrez-Xicotencatl et al., 2016; Luevano et al., 2010). Nucleic Acid Programmable Protein Arrays (NAPPA) (N. Ramachandran et al., 2004; N. Ramachandran et al., 2008) have enabled rapid profiling of the serum Ab response in the settings of infections (W. R. Montor et al., 2009; Prados-Rosales et al., 2014), autoimmune diseases (Bian et al., 2016; Miersch et al., 2013) and cancer (K. S. Anderson, D. W. Cramer, et al., 2015; Katchman et al., 2016; J. Wang et al., 2015). To measure the serologic responses across multiple HPV types, we adapted the arrays described in Chapter 3(Ewaisha, Meshay, Resnik, Katchman, & Anderson, 2016) for the detection of HPV-specific IgG Abs in sera. Full length cDNAs encoding the proteomes of 12 HPV types are expressed as C-terminal GST fusion proteins using mammalian *in vitro* transcription/translation and captured onto a glass slide surface (J. Wong, S. Sibani, N. N. Lokko, J. LaBaer, & K. S. Anderson, 2009). In a pilot study, we demonstrated that HPV protein arrays display immunogenic epitopes that can be detected using HPV-specific monoclonal Abs (MAbs) and with select sera from HPV-specific malignancies (Ewaisha et al., 2016). The purpose of this study was to systematically investigate the serologic immune profile to HPV in women with high-grade pre-invasive cervical lesions and ICC, and to identify serologic biomarkers for diagnosis and early detection of cervical cancer.

4.2. MATERIALS AND METHODS

4.2.1. Sample Selection

We used the Biorepository for Molecular Signatures of Cervical Cancer developed by the Centers for Disease Control and Prevention as a Clinical Epidemiology and Validation Center for NCI's Early Detection Research Network (EDRN). Samples in the biorepository were collected from women attending colposcopy clinics at urban public hospitals in Atlanta, GA, Detroit, MI, or Galveston, TX between 2000 and 2010 and linked to epidemiologic and clinical data, including HPV detected in exfoliated cervical cells, age, race, and tissue confirmation of cervical disease status (Rajeevan et al., 2005). For this study, 162 samples from women with cervical intraepithelial neoplasia (CIN) grade 0 (no CIN), I, II, III were selected, of which 78 were CIN 0/I and 84 were CIN II/III. We used 83 archived anonymized plasma samples from women with ICC collected in Atlanta, GA prior to 1997. For convenience, the term serum is used throughout the manuscript. While HPV vaccine history was not collected, HPV vaccination was not introduced before 2006, and it is unlikely that any study participants were vaccinated. Only a subset (n=51) of the ICC samples had information on the HPV DNA status of the tumor, of which 24 (47.1%) were HPV16+. Samples were collected using a standardized sample collection protocol and stored at -80°C until use. Written informed consent was obtained from all subjects under institutional review board approval.

4.2.2. HPV Microarray Generation and Detection of Serum Abs

Production of custom HPV protein arrays and array quality control experiments were performed as previously reported (Ewaisha et al., 2016) with modifications described here. In brief, arrays displaying codon-optimized proteomes of 2 low risk (HPV6 and 11) and 10 high risk (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, and 58) HPV types were generated. The codon-optimized HPV16 L1 gene previously used (Ewaisha et al., 2016) was replaced

with the non-codon-optimized version, which had higher protein expression. Both codon-optimized and non-optimized HPV16 E6 and E7 were printed on the arrays for direct comparison of Ab reactivity. Since only the non-codon optimized HPV16 E6 and HPV16 E7 (Ewaisha et al., 2016) showed immunoreactivity, only those results are shown. All non-codon-optimized genes were obtained by nested PCR using gene-specific primers from HPV16 plasmid DNA (American Type Culture Collection, Manassas, VA) as described (Anderson, Sibani, et al., 2011). All genes were inserted into pDONR221 vector per manufacturer's instructions (Invitrogen, Carlsbad, CA), and were converted to the pANT7_cGST vector (http://dnasu.asu.edu/DNASU/Home.jsp) (N. Ramachandran et al., 2004). Human IgG and the C-terminal portion of the Epstein-Barr virus-derived Ag EBNA-1 were used as positive controls. A set of non-HPV related negative control proteins (n=93) were printed on the arrays and used for array signal intensity normalization and establishment of cut-off values. Arrays were incubated with serum samples diluted 1:50 in 10% *E. coli* lysate prepared in 5% milk-PBST (0.2% tween) (J. Wang et al., 2013) and serum Ab detection was performed as previously described (K. S. Anderson, D. W. Cramer, et al., 2015; Anderson, Sibani, et al., 2011; Ewaisha et al., 2016; J. Qiu & LaBaer, 2011).

4.2.3. Protein Array Image Analysis and Quantification

After serum binding and IgG detection, arrays were scanned by Tecan PowerScanner (Tecan Group, Männedorf, Switzerland). ArrayPro Analyzer version 6.3 (MediaCybernetics, Bethesda, MD) was used to measure the signal intensity of individual spots on the scanned slides. Normalization of raw intensity values was performed by subtracting the slide background signal and dividing by the background signal subtracted from the median intensity of all spots. The slide background signal was determined by the first quartile of signal intensity of the no-DNA control spots (all material except DNA). In

addition, array images were qualitatively inspected to identify and confirm positive responses by adjusting raw images to extreme brightness and contrast using ArrayPro Analyzer and visual analysis of diffused signal (ring) as described previously (W. R. Montor et al., 2009; J. Wang et al., 2015). Each spot was scored based on the intensity and morphology of the ring on a scale of 0 to 5.

4.2.4. HPV DNA Detection by L1 Consensus PCR

For all the samples from the biorepository, HPV DNA was detected in extracts of exfoliated cervical cells collected in PreservCyt media as previously described (Rajeevan et al., 2005). Briefly, 16 ml of the PreservCyt collection media was extracted using MasterPure Complete DNA and RNA purification kit (Epicentre, Madison, WI). HPV detection and typing was performed using the Roche linear array that detects 37 types. HPV results for the anonymized archived cervical cancer cases (n=83) were based on combined results of colorimetric ISH for HPV16, 18, 31, 33, 35 on formalin-fixed paraffin embedded tissue sections and L1 consensus PCR with MY09/11 primers and type-specific hybridization to 6 HR types (16, 18, 31, 33, 35, 45) on DNA extracts from the same tissues (methods in use at the time of archiving (Unger, Vernon, Lee, Miller, & Reeves, 1998)).

4.2.5. Statistical Analysis

The correlation of raw signal intensities of protein expression between the arrays randomly selected for quality control was determined with scatter plots and the Pearson correlation coefficient (R) was calculated to assess consistency. Levels of protein expression on the arrays were measured by calculating the mean values of raw signal intensities of duplicate spots from two arrays. For serum Ab reactivity on protein arrays, mean values (of duplicate spots for a given Ag) of normalized signal intensity were compared for different disease groups using Fisher's exact test (Graphpad Prism version 5.0c, San Diego, CA). A p-value of <0.05 was considered significant. Seropositivity for any given Ag was defined as the

median of normalized signal intensity values of all negative control proteins (n=93) in all sera (n=234) +3 standard deviations or spots that were positive by visual analysis. A total of 245 serum samples were tested on the arrays, of which 11 (4.5%; n=7 ICC, n=4 CIN) were excluded from the analysis due to high background. High array background was defined as an array with normalized signal intensity values for ≥14 out of 93 negative control spots exceeding the 75th percentile + 1.5*interquartile range of this negative control protein across all arrays.

4.3. RESULTS

4.3.1. Characteristics of Study Samples

Our primary goal was to determine the prevalence and specificity of HPV-specific Ab responses in women with cervical cancer precursors and with ICC. Age, race, and HPV DNA status of patients contributing samples to the study are shown in **Table 4-1**. Ab levels were compared in women with CIN 0/I (n=78) and CIN II/III (n=84) who were referred to colposcopy because of abnormalities in cervical cancer screening, and in women with ICC (n=83). Women with CIN 0/I were chosen as the relevant control population to determine the utility of these biomarkers within a high-risk population. As expected, women with CIN 0/I had a lower frequency of cervical high-risk (HR) HPV than women with CIN II/III (57.7% vs. 97.6%, $p<0.0001$). Infection with 2 or more HPV types was detected in more than 35% of women in both CIN 0/I and CIN II/III (**Table 4-1**). Women with CIN II/III were as expected significantly younger than women with ICC (mean 30.0 yrs vs. 52.0 yrs, $p<0.0001$). There was also a lower frequency ($p<0.0015$) of HPV16 in CIN 0/I (19.2%) than CIN II/III (63.1%) and ICC (47.1%). The clinics participating in the EDRN study had a high proportion of minority and Hispanic white patients. The racial distribution of the samples occurred by chance.

Table 4-1. Characteristics of Study Samples.

Characteristics	Disease Status		
	CIN 0/I N = 78 N (%)	**CIN II/III N = 84 N (%)**	**ICC N = 83 N (%)***
Age in yrs, Mean	**28.7**	**30.0**	**52.0**
< 30	51 (65.3)	45 (53.6)	3/79 (3.8)
≥ 30	27 (34.6)	39 (46.4)	76/79 (96.2)
Race			
Black	71 (91.0)	55 (65.5)	64/79 (81.0)
Other	7 (9.0)	29 (34.5)	15/79 (19.0)
HPV16 DNA status†			
HPV16+	15 (19.2)	53 (63.1)	24/51 (47.1)
HPV DNA status overall†			
Negative	23 (29.5)	2 (2.4)	12/51 (23.5)
1 HPV type	29 (37.2)	46 (54.8)	39/51 (76.5)
2 HPV types	10 (12.8)	19 (22.6)	0/51 (0)
≥ 3 HPV types	16 (20.5)	17 (20.2)	0/51 (0)
Any HR HPV‡	45 (57.7)	82 (97.6)	39/51 (76.5)

*N varies for each category because of missing information. The numbers of samples are shown.
†HPV testing methods used for anonymized archived samples differed from those used in biorepository, so results are not directly comparable.
‡The following HPV types were considered as high-risk types for this analysis -HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68

4.3.2. Production and Reproducibility of NAPPA HPV Protein Arrays

The quality and reproducibility of the array printing were evaluated by picogreen staining of DNA and measuring protein expression with anti-GST Abs (**Figure 4-1A**). Three arrays were printed on each slide and the correlation coefficients of anti-GST signal intensities were determined for intra-array (R=0.98) and intra-batch replicate arrays (R=0.90) from two subarrays on the same slide or two randomly selected slides within the print batch (**Figure 4-1B**).

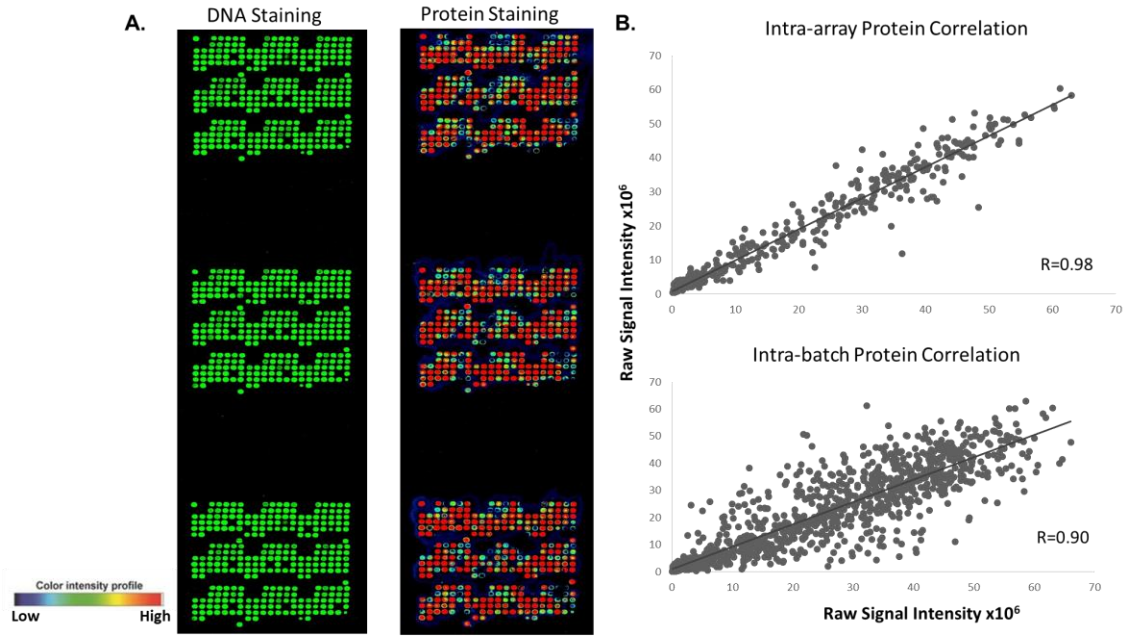

Figure 4-1. Reproducibility of Printing and Protein Expression on NAPPA. (A) Left: Picogreen staining to detect DNA content printed on the microarrays. Right: Image of anti-GST binding to measure the level of protein display. (B) Plots of protein signal intensities from all spots from two subarrays on the same slide or two randomly selected slides within the print run.

4.3.3. HPV-Specific Ab Prevalence

There was no significant difference (p=0.46) in the percentages of negative control spots (displaying the non-HPV related proteins; n=93) that exceeded the cut-off value between arrays probed with CIN 0/I (0.71%), CIN II/III (0.81%), and ICC (1.77%) sera.

The prevalence of HPV-specific serum IgG Abs among women with CIN 0/I, CIN II/III, and ICC is summarized in **Table 4-2**. At least one of the HPV-specific Abs was detected in serum from women with CIN 0/I (46.1%), CIN II/III (59.8%), and ICC (68.4%). Abs to any early (E) HPV protein were detected more frequently in women with ICC (46.1%) and CIN II/III (39.0%) than women with CIN 0/I (23.7%, p<0.04). Abs to any L1 protein had the highest prevalence (28.9%, 34.1%, and 44.7% in CIN 0/I, CIN II/III, and ICC, respectively). Of the E Abs, anti-E7 Abs were the most frequently detected (CIN 0/I, 6.6%; CIN II/III,

19.5%; and ICC, 30.3%). The least frequently detected Abs were E1 and E2-Abs in CIN 0/I (1.3%) and E1-Abs in CIN II/III (1.2%) and ICC (7.9%). The sensitivity [proportion of cases with at least one HPV Ag-specific Ab detected] was comparable when restricted to cases known to have HPV16 (as opposed to any other oncogenic HPV) DNA detected; 59.6% vs 59.8% for CIN II/III and 81.8% vs 68.4% for ICC, (p-value N.S.). Among all women (irrespective of HPV DNA status), Abs to HPV16 Ags were detected in only 6.6%, 19.5%, and 35.5% in CIN 0/I, CIN II/III, and ICC, demonstrating the importance of multi-antigenic immunoprofiling.

Table 4-2. Prevalence of Positive Antibody Response [(1)] to Each HPV Protein From Any HPV Type [(2)].

	No. + (%)						
	CIN 0/I		CIN II/III		ICC		Total
HPV Antibodies	Total N=76	HPV16+[(3)] N=14	Total N=82	HPV16+[(3)] N=52	Total N=76	HPV16+[(3)] N=22[(4)]	
E1	1 (1.3)	0 (0)	1 (1.2)	0 (0)	6 (7.9)*	1 (4.5)	8 (3.4)
E2	1 (1.3)	0 (0)	4 (4.9)	1 (1.9)	11 (14.5)	5 (22.7)	16 (6.8)
E4	5 (6.6)	0 (0)	14 (17.1)	5 (9.6)	12 (15.8)	4 (18.2)	31 (13.2)
E5	5 (6.6)	1 (7.1)	4 (4.9)	3 (5.8)	12 (15.8)	6 (27.3)	21 (9.0)
E6	3 (3.9)	0 (0)	8 (9.8)	5 (9.6)	12 (15.8)*	5 (22.7)	23 (9.8)
E7	5 (6.6)	0 (0)	16 (19.5)*	10 (19.2)	23(30.3)*	8 (36.4)*	44 (18.8)
Any E [(2)]	18 (23.7)	2 (14.3)	32 (39.0)*	19 (36.5)	35 (46.1)*	15 (68.2)*	85 (36.3)
L1	22 (28.9)	4 (28.6)	28 (34.1)	16 (30.8)	34 (44.7)	12 (54.5)	84 (35.9)
L2	8 (10.5)	0 (0)	5 (6.1)	5 (9.6)	13 (17.1)	6 (27.3)	26 (11.1)
Any L [(2)]	25 (32.9)	4 (28.6)	32 (39.0)	20 (38.5)	38 (50.0)	14 (63.6)*	95 (40.6)
Any E and/or L[(2)]	35 (46.1)	4 (28.6)	49 (59.8)	31 (59.6)	52 (68.4)*	18 (81.8)*	136 (58.1)
Any HPV16 Ag	5 (6.6)	0 (0)	16 (19.5)*	11 (21.2)	27 (35.5)*	10 (45.5)*	48 (20.5)

[(1)] Cut-off values defined as the median of normalized signal intensity values of all negative control proteins +3 standard deviations in all sera (n=234) or spots that were positive by visual analysis.
[(2)] Any positive vs. all negative from any of the 12 HPV types tested.
[(3)] HPV16 DNA detected in cervix. HPV testing methods used for anonymized archived samples differed from those used in biorepository.
[(4)] HPV DNA status was known for only a subset (n=51) of the ICC samples.
* $p<0.05$, compared with CIN 0/I

4.3.4. Type-Specific Ab Response

To determine whether patients with a specific HPV infection develop type-specific Abs, there are multiple challenges. First, a significant number of women with CIN have multiple HPV types detected (33.3% of CIN 0/I and 42.8% of CIN II/III, **Table 4-1**), and past exposure to other HPV types cannot be excluded. Second, there is likely serologic cross-reactivity across HPV types. **Figure 4-2A** shows Abs from an ICC patient reacting with E4 protein from 4 different HPV types (16, 31, 35, and 45). As examples, **Figures 4-2B** and **4-2C** show serum from two women with CIN II/III with HPV16 DNA and Abs against HPV16 E4 and HPV52 E4 (B) and HPV 58 E4 (C).

To determine whether there is any correlation between HPV DNA types detected in the patient and type-specific serum Abs, we analyzed the data in two ways. In **Table 4-3**, subjects were stratified based on cervical HPV DNA status. For example, among women with successful Ab testing, cervical HPV16 DNA was detected in 18.4% of those with CIN 0/I; 63.4% of CIN II/III; and 46.8% of ICC. The type-specific Ab detection rate in this group was 0%, 21.2%, and 45.5% for CIN 0/I, CIN II/III and ICC. For the most common HPV DNA detected in cervical samples in CIN II/III, HPV16, 31, 35, and 52, the range of detection of type-specific Abs was 8.3 – 25.0%. In **Table 4-4**, subjects were stratified by type-specific seropositivity to evaluate the proportion with detection of type-specific HPV DNA in the cervix. Women with HPV16 Abs and CIN II/III had the highest type-specific DNA detection, 78.6%, followed by HPV31 (36.4%), 45 (25.0%), and 52 (14.2%).

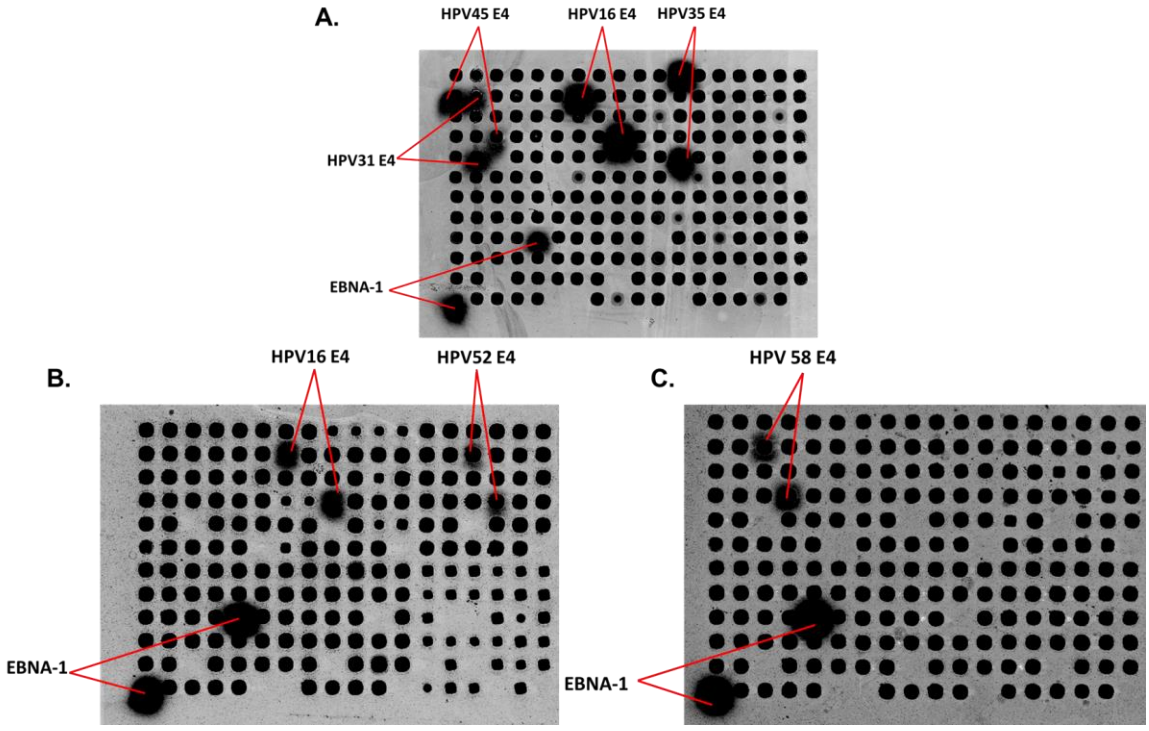

Figure 4-2. Detection of IgG in Human Serum Using HPV Protein Arrays. (A) Detection of IgG Abs in serum from a patient with ICC. Immunoreactivity to the positive control EBV EBNA-1 protein, and HPV E4 protein from 4 different HPV types (16, 31, 35, and 45) are detected. (B) and (C) Detection of IgG Abs in sera from two women with CIN II/III. Immunoreactivity to HPV16 E4, HPV52 E4 (B) and HPV58E4 (C) as well as EBNA-1 protein, is shown. Dark spots represent the individual proteins (HPV Ags and non HPV-related controls in random order) displayed on the arrays after adjusting the raw images to extreme brightness and contrast. Positive spots (with diffused signal) are labeled.

Table 4-3. Prevalence of HPV Type-Specific IgG Abs in Women with Known Cervical HPV DNA Status.

HPV DNA in cervix	Total Number[1]	No. + (%) CIN 0/I N=76		CIN II/III N=82		ICC N=47[2]	
		Total	Ab+[3]	Total	Ab+[3]	Total	Ab+[3]
HPV6	3	1 (1.3)	0 (0)	2 (2.4)	0 (0)	0 (0)	0 (0)
11	2	1 (1.3)	0 (0)	1 (1.2)	0 (0)	0 (0)	0 (0)
16	88	14 (18.4)	0 (0)	52 (63.4)	11 (21.2)	22 (46.8)	10 (45.5)
18	16	3 (3.9)	1 (33.3)	5 (6.1)	1 (20.0)	8 (17.0)	2 (25.0)
31	20	2 (2.6)	0 (0)	16 (19.5)	4 (25.0)	2 (4.3)	0 (0)
33	10	2 (2.6)	0 (0)	5 (6.1)	0 (0)	3 (6.4)	1 (33.3)
35	16	5 (6.6)	2 (40.0)	10 (12.2)	2 (20.0)	1 (2.1)	1 (100.0)
39	5	3 (3.9)	1 (33.3)	2 (2.4)	1 (50.0)	0 (0)	0 (0)
45	10	5 (6.6)	1 (20.0)	5 (6.1)	1 (20.0)	0 (0)	0 (0)
51	7	4 (5.3)	1 (25.0)	2 (2.4)	0 (0)	1 (2.1)	0 (0)
52	21	9 (11.8)	1 (11.1)	12 (14.6)	1 (8.3)	0 (0)	0 (0)
58	14	10 (13.2)	1 (10.0)	3 (3.7)	0 (0)	1 (2.1)	0 (0)

Table 4-4. Prevalence of Type-Specific HPV DNA in Women with Known Seropositivity.

HPV - specific Abs	Total Number[4]	No. + (%) CIN 0/I N=76		CIN II/III N=82		ICC N=76[2]	
		Total	DNA+[5]	Total	DNA+[5]	Total	DNA+[5]
HPV6	24	4 (5.3)	0 (0)	10 (12.2)	0 (0)	10 (13.2)	0 (0)
11	25	5 (6.6)	0 (0)	8 (9.8)	0 (0)	12 (15.8)	0 (0)
16	46	5 (6.6)	0 (0)	14 (17.1)	11 (78.6)	27 (35.5)	10 (37.0)
18	50	11 (14.5)	1 (9.0)	15 (18.3)	1 (6.7)	24 (31.6)	2 (8.3)
31	37	5 (6.6)	0 (0)	11 (13.4)	4 (36.4)	21 (27.6)	0 (0)
33	58	14 (18.4)	0 (0)	18 (22.0)	0 (0)	26 (34.2)	1 (3.8)
35	78	21 (27.6)	2 (9.5)	19 (23.2)	2 (10.5)	38 (50.0)	1 (2.6)
39	51	10 (13.2)	1 (10.0)	20 (24.4)	1 (5.0)	21 (27.6)	0 (0)
45	26	5 (6.6)	1 (20.0)	4 (4.9)	1 (25.0)	17 (22.4)	0 (0)
51	24	5 (6.6)	1 (20.0)	6 (7.3)	0 (0)	13 (17.1)	0 (0)
52	31	5 (6.6)	1 (20.0)	7 (8.5)	1 (14.2)	19 (25.0)	0 (0)
58	18	4 (5.3)	1 (25.0)	5 (6.1)	0 (0)	9 (11.8)	0 (0)

[1] of women with the corresponding HPV DNA type in the cervix.

[2] The ICC samples with unknown tumor DNA status were excluded from the analysis in table 3A.

[3] Positive for any Ab specific to the given HPV type.

[4] of women with serum Abs to any Ag of the corresponding HPV type.

[5] Type-specific HPV DNA positive.

4.3.5. Cross-Reactivity of Serologic Responses

We determined the prevalence of Abs against homologous Ags (i.e. all E7 Ags) from more than one HPV type in all sera (n=234) from the three cervical disease groups under investigation (**Table 4-5**). Abs against L2 were the most cross-reactive, while anti-E1 Abs were the least cross-reactive. Of sera that had Abs against any L2 Ag, 57.7% were positive for L2 from at least 6 HPV types. For E1, all 8 women who had specific Abs were positive for E1 from only one HPV type. 8.1% of all women had Abs to E7 from at least 2 HPV types. The percentages of sera with cross-reactive Abs to at least one other HPV type were as follows: E2 (6.3%), E4 (29.0%), E5 (14.3%), E6 (17.4%), E7 (43.2%), L1 (39.3%), and L2 (73.1%).

Table 4-5. Prevalence of Serum Abs to Homologous Ags from Different HPV Types.

Homologous Abs[(1)]		**E1**	**E2**	**E4**	**E5**	**E6**	**E7**	**L1**	**L2**
=2	**No.**	0	1	7	3	1	10	17	3
	% of total[(2)]	0.0	0.4	3.0	1.3	0.4	4.3	7.3	1.3
	% of positive[(3)]	0.0	6.3	22.6	14.3	4.3	22.7	20.2	11.5
=3	**No.**	0	0	1	0	3	9	9	1
	% of total[(2)]	0.0	0.0	0.4	0.0	1.3	3.8	3.8	0.4
	% of positive[(3)]	0.0	0.0	3.2	0.0	13.0	20.5	10.7	3.8
=4	**No.**	0	0	1	0	0	3	3	0
	% of total[(2)]	0.0	0.0	0.4	0.0	0.0	1.3	1.3	0.0
	% of positive[(3)]	0.0	0.0	3.2	0.0	0.0	6.8	3.6	0.0
=5	**No.**	0	0	0	0	0	0	2	0
	% of total[(2)]	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0
	% of positive[(3)]	0.0	0.0	0.0	0.0	0.0	0.0	2.4	0.0
≥6	**No.**	0	0	0	0	0	1	2	15
	% of total[(2)]	0.0	0.0	0.0	0.0	0.0	0.4	0.9	6.4
	% of positive[(3)]	0.0	0.0	0.0	0.0	0.0	2.3	2.4	57.7

[(1)] Number of homologous Ags from different HPV types to which Abs were detected in sera from all three disease groups (n=234).
[(2)] Percentage of women who had Abs to a given Ag from one or more HPV types over the total number of women.
[(3)] Percentage of women who had Abs to a given Ag from multiple HPV types over the number of women who had Abs to this Ag from at least one HPV type.

4.4. DISCUSSION

There is a clinical need for circulating biomarkers that identify high-risk HPV infection for early detection and treatment of cervical disease. Here, we have used our custom HPV protein microarrays, displaying the proteomes of two low-risk and ten high-risk HPV types, to characterize the diversity of the immune response in cervical cancer and in pre-invasive cervical disease. We find that 20 – 46% of patients with CIN and ICC have a broad range of Abs to HPV early proteins in their sera and these biomarkers correlate with disease severity.

Up to 80% of patients with HPV-associated oropharyngeal cancer (HPVOPC) have detectable serum HPV16 Abs to E Ags (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015; Kreimer et al., 2013). Abs against the oncogenic proteins E6 and E7 are also highly specific to ICC. As a result, most cervical cancer studies have focused on E6, E7, and L1 Abs from HPV16 and 18 (Achour et al., 2009; Gutierrez-Xicotencatl et al., 2016; Reuschenbach et al., 2008; Waterboer et al., 2005). In the only previous study of Abs to entire HPV proteomes using slide-based protein microarrays, high levels of Abs were detected against E7 but not E6 or L1 in ICC sera, possibly due to the difficulty of expressing larger-sized proteins (Luevano et al., 2010). In that study and others (Gutierrez-Xicotencatl et al., 2016; Lehtinen et al., 2003; Stanley, 2003), no significant difference in Ab response between CIN II/III and asymptomatic controls was detected.

While L-specific Abs were detected in serum from women in all three groups of cervical disease in our study, Abs against any E protein were, as expected, more prevalent in ICC and CIN II/III than women with CIN 0/I. Anti-E7 Abs were the most frequently detected E-Abs, and our data of E7-specific Abs in ICC is consistent with previous studies. Using ELISA (Achour et al., 2009; Silins et al., 2002; Tjiong et al., 2001) and Luminex bead

arrays (Castellsague et al., 2014; Lang Kuhs et al., 2015; Reuschenbach et al., 2008), anti-E7 Abs were detected in 13 – 53% of women with ICC and in ~60% of ICC (and 10% of healthy controls) using protein microarrays (Luevano et al., 2010).

As cervical disease progresses towards malignancy, infectious viral particle production becomes limited to a small area near the surface of the cervical epithelium (Doorbar, 2006; Griffin et al., 2012). E4 plays a role in viral synthesis and possibly viral release (Doorbar, 2013). The expression of E4 in CIN II and III is restricted to this subset of cells and is generally lost in ICC (Doorbar, 2006). Expression of E4 protein in tissue has been proposed as an early detection marker (Griffin et al., 2012) but specific serum Abs may provide a more convenient detection method. Our data show that E4-specific Abs develop early in disease progression, with 16 – 17% prevalence in both CIN II/III and ICC. Anti-E4 Abs have been reported in sera from women with CIN II/III (34%) and ICC (29%) (Pedroza-Saavedra et al., 2000), which is consistent with our findings and the predicted level of E4 expression especially in CIN II/III.

Viral integration into the host genome, with loss of E2 expression, is a frequent hallmark of HPV-associated cancers, leading to derepression of E6 and E7 expression. Since E2 Ag is expressed in CIN II/III (Xue et al., 2010), we predicted that E2 Abs would also be detected early in CIN II/III. E2-specific Abs were detected in both CIN II/III (4.9%) and at higher ($p<0.05$) frequency in ICC (14.5%) but were at low prevalence. E2-specific IgG Abs have been reported in 24% of women with CIN I-III (compared to 13% of healthy controls) (Marais, Rose, & Williamson, 1997) and in 12% of women with ICC (compared to 2% of healthy controls) (Combes et al., 2014), consistent with our data. Since the majority (64%) of patients with HPVOPC have Abs against HPV16 E2 (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015), our data suggest that these two tumor sites may have significant differences in viral integration and expression of E2 Ag,

or exposure of E2 Ag to B cells. In cervical disease, E1 Abs have been reported to have a low prevalence (10% and 0.3% in ICC and healthy controls, respectively) (Combes et al., 2014). We also found an overall low prevalence of anti-E1 Abs in cervical disease, as opposed to a 60% prevalence in HPVOPCs (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015).

Only a few studies have investigated the immune response against multiple HPV types (Combes et al., 2014; Gutierrez-Xicotencatl et al., 2016; Luevano et al., 2010; Waterboer et al., 2005). Only one study displayed multiple HPV proteomes on protein arrays (Luevano et al., 2010) and the others have used glutathione S-transferase-based multiplex serology to evaluate serum Abs against only E6, E7, and L1 Ags from multiple HPV types (Combes et al., 2014; Waterboer et al., 2005). The question of cross-reactivity of Abs against Ags from closely-related viral types has therefore not been adequately addressed. We and others have previously demonstrated that MAbs raised against specific HPV proteins may cross-react with homologous proteins from different HPV types due to sequence similarity (Ewaisha et al., 2016; Luevano et al., 2010). Here, we detected cross-reactive Abs including against the E4 Ag. The E4 ORF is the most divergent between HPV types (Bell et al., 2007). The range of amino acid sequence similarity between E4 from the non-HPV16 types detected in the 3 sample sera illustrated in **Figure 4-2** and HPV16 E4 is 42 – 59%. It is therefore not known if Abs to multiple homologous E4 proteins reflect cross-reactivity with conserved epitopes or prior multiple HPV infections. We also observed Abs against L2 from at least 6 HPV types in 57.7% of women who had L2-specific Abs (**Table 4-4**), which likely indicates cross-reactivity, given the high (46 – 63%) sequence conservation of the L2 protein, and the interest in developing it as a vaccine (Karanam, Jagu, Huh, & Roden, 2009). Abs against L1 from three or more HPV types were also detected in 19% of women positive for L1 Abs. This is consistent with previously

reported L1 cross-reactivity detected by an HPV16 L1-specific MAb (Luevano et al., 2010). Overall, these data suggest that these HPV arrays will have limited utility as surrogate markers for HPV typing.

While not directly compared in this study, the signal intensity of Ab binding on the arrays (Ewaisha et al., 2016) and on RAPID ELISA (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015) are consistently weaker in cervical disease than in HPVOPC (unpublished observations). Additionally, despite tissue expression of the oncoproteins E6 and E7 in ICC, we and others (Luevano et al., 2010; Reuschenbach et al., 2008; Stanley, 2003) have reported specific serum Abs in less than half of the patients, while they are detected in up to 75% of HPVOPC cases (K. S. Anderson, K. R. Dahlstrom, et al., 2015). Anti-E6 Abs have been detected in HPVOPC and cervical cancer cases years prior to the establishment of a clinical diagnosis (Kreimer et al., 2013; Lehtinen et al., 2003), suggesting these may be useful for early detection. The Ab response to the E proteins in high grade pre-invasive cervical lesions, however, has been difficult to detect in previous studies (Lehtinen et al., 2003; Luevano et al., 2010; Stanley, 2003). Even though we detect E-specific Abs in a subset of CIN II/III, the frequencies are low. The presence of both viral DNA and viral oncoproteins in HPVOPC tumors suggest that both cancer sites have similar pathogenetic mechanisms (Gillison et al., 2008). Therefore, it is likely that the close proximity of lymphoid tissue in the tonsils results in a more potent immune induction in HPVOPC compared with cervical disease.

The lack of infrastructure and resources in LMICs hampers large-scale implementation of Pap test screening (Kontostathi et al., 2016; Wentzensen & von Knebel Doeberitz, 2007; Wright, 2006). In low-resource environments, visual examination with acetic acid (VIA) is an inexpensive alternative (Sankaranarayanan et al., 2007). It results in a 25% reduction in cervical cancer incidence and a 35% reduction in cervical cancer mortality after a single

screen (Sankaranarayanan et al., 2007; Shastri et al., 2014), with significant downstaging of cervical cancers (Shastri et al., 2014). However, VIA has low sensitivity in women older than 50 years, poor reproducibility between operators and it requires continuous training and supervision. The absence of HPV nucleic acid in the cervix is a good negative predictor of cervical disease but HPV testing is not recommended for women <30 years old because transient infection reduces specificity (Arbyn et al., 2012; Goodman, 2015). In pooled analyses, HPV testing is more sensitive (90 – 95% for CIN II/III) than cervical cytology alone or VIA but lacks the specificity (89%) for a reliable biomarker (Arbyn et al., 2012). To date, there are no established tissue, blood, or vaginal biomarkers other than HPV nucleic acid and cytology for CIN II/III in high risk patients. Biomarkers such as serology that identify high-risk HPV infection and invasive cervical cancers (ICC) could have an impact on the screening, detection, and treatment of cervical disease.

CHAPTER 5 : IDENTIFICATION OF CIRCULATING MARKERS OF PRE-EXISTING IMMUNITY TO *STREPTOCOCCUS PYOGENES* CRISPR/CAS9 IN HUMANS

This chapter is in revisions in Nature Communications and has been published as a preprint on BioRxiv:

Ferdosi, S. R., Ewaisha, R., Moghadam, F., Krishna, S., Park, J. G., Ebrahimkhani, M. R., and Anderson, K. S. (2018). Multifunctional CRISPR/Cas9 with engineered immunosilenced human T cell epitopes. *bioRxiv*.

ABSTRACT

The application of Cas9 protein for genetic and epigenetic therapies in humans raises concerns over immunogenicity of this foreign protein. This chapter investigates potential biomarkers of immunity to the CRISPR/Cas9 system that is currently in clinical trials for cancer. Pre-existing B cell and T cell immune responses to the *Streptococcus pyogenes* Cas9 protein were detected in humans. Two immunodominant T cell epitopes for HLA-A*02:01 were identified and a single mutation in the anchor residue of one or both of these epitopes significantly reduced the protein immunogenicity while maintaining its function and specificity in the CRISPR/Cas9 system. The identified immune responses could serve as immune biomarkers that help guide decisions of using this therapeutic in humans.

5.1. INTRODUCTION AND BACKGROUND

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/Cas9 technology has raised hopes for developing personalized gene therapies for complex diseases such as cancer as well as genetic disorders, and is currently entering clinical trials (Cyranoski, 2016; Reardon, 2016). The history of gene therapy has included both

impressive success stories and serious immunologic adverse events (Cavazzana-Calvo et al., 2000; Gaspar et al., 2004; Hacein-Bey-Abina et al., 2002; Howe et al., 2008; Manno et al., 2006; Marshall, 1999). The expression of *Streptococcus pyogenes* Cas9 protein (SpCas9) in mice has evoked both cellular and humoral immune responses (Chew et al., 2016; D. Wang et al., 2015), which raises concerns regarding its safety and efficacy as a gene or epi-gene therapy in humans. These pre-clinical models and host immune reactions to other exogenous gene delivery systems (Mays & Wilson, 2011; Federico Mingozzi & High, 2013; Yin et al., 2014) suggest that the pathogenic “non-self” origin of Cas9 may be immunogenic in humans.

Both B cell and T cell host responses specific to either the transgene or the viral components of adenoviral (Ahi, Bangari, & Mittal, 2011; Aldhamen & Amalfitano, 2016) and adeno-associated viral (AAV) (Mays & Wilson, 2011; Federico Mingozzi & High, 2013) vectors have been detected, despite relatively low immunogenicity of AAV vectors. In the case of AAV, specific neutralizing Abs and T cells are frequently detected in healthy donors (Boutin et al., 2010; F. Mingozzi et al., 2007; Scallan et al., 2006; Thwaite et al., 2015) and, specific CD8+ T cells have been shown to expand following gene delivery (F. Mingozzi et al., 2007). There has been recent progress in developing strategies to overcome this problem, such as capsid engineering and transient immunosuppression (Bartel, Schaffer, & Buning, 2011; Martino et al., 2013; F. Mingozzi et al., 2013). The potential consequences of immune responses to expressed proteins from viral vectors or transgenes include neutralization of the gene product; destruction of the cells expressing it, leading to loss of therapeutic activity or tissue destruction; induction of immune memory that prevents re-administration; and fulminant innate inflammatory responses (Brunetti-Pierri & Ng, 2009; Halbert et al., 1998; H. Jiang et al., 2006; Kay, 2011; Nathwani et al., 2011; Nayak & Herzog, 2010). More potent immune responses to gene therapies have

been observed in humans and non-human primate models compared to mice (Gao et al., 2009; Manno et al., 2006).

Of the Cas9 orthologs derived from bacterial species (Esvelt et al., 2013; Hirano et al., 2016; Jinek et al., 2014; Ran et al., 2015; Zetsche et al., 2015), the SpCas9 is the best characterized. *S. pyogenes* is a ubiquitous pathogen, with an annual incidence of 700 million worldwide (Carapetis, Steer, Mulholland, & Weber, 2005), but immunity to SpCas9 in humans has not been reported. Here, we sought to characterize the pre-existing immune response to SpCas9 in healthy individuals and to identify the immunodominant T cell epitopes with the aim of developing SpCas9 proteins that have diminished capacity to invoke human adaptive response.

5.2. METHODS

5.2.1. Detection of Cas9-Specific Serum Abs in Healthy Controls

Healthy control sera (n = 183) used in this study, and previously described 47, are a subset of a molecular epidemiology study of head and neck cancer at the MD Anderson Cancer Center, collected between January 2006 and September 2008. *S. pyogenes* lysate was prepared by sonication of bacterial pellets from overnight cultures of *S. pyogenes* ATCC 19615 in the presence of 1 pill of cOmplete Protease Inhibitor (Sigma-Aldrich) after 3 cycles of freezing and thawing. Serum Ab detection was performed using ELISA. 96-well plates were coated with 20 µg/mL of recombinant *S. pyogenes* Cas9 nuclease (New England Biolabs, Ipswich, MA) or *S. pyogenes* lysate. Sera were diluted 1:50 in 10% *E. coli* lysate prepared in 5% milk-PBST (0.2% tween) 48, incubated with shaking for 2 hrs at room temperature, and added to the specified wells in duplicate. Horseradish peroxidase (HRP) anti-human IgG Abs (Jackson ImmunoResearch Laboratories, West Grove, PA) were added at 1:10,000, and detected using Supersignal ELISA Femto

Chemiluminescent substrate (Thermo Fisher Scientific, Waltham, MA). Luminescence was detected as relative light units (RLU) on a Glomax 96 Microplate Luminometer (Promega, Madison, WI) at 425 nm. To establish cut-off values, a RLU ratio > (the mean + 3 standard deviations) of 125 randomly chosen control samples was designated positive (**Fig. 5-1**, dotted and dashed lines for bacterial lysate and Cas9 protein, respectively).

5.2.2. Cas9 Candidate T cell Epitope Prediction

We used our previously described prediction strategies (Chowell et al., 2015; Krishna & Anderson, 2016) to predict candidate Cas9 T cell epitopes. Briefly, we predicted MHC class I restricted 9-mer and 10-mer candidate epitopes derived from the Cas9 protein (Uniprot - Q99ZW2) for HLA A*02:01. The protein reference sequence was entered into 5 different prediction algorithms; 3 MHC-binding: IEDB-consensus binding (Moutaftsi et al., 2006), NetMHCpan binding (Hoof et al., 2009), Syfpeithi (Rammensee, Bachmann, Emmerich, Bachor, & Stevanović, 1999) and 2 antigen-processing algorithms: IEDB-consensus processing, ANN processing (Tenzer et al., 2005). The individual scores from each of the prediction algorithms were then normalized within the pool of predicted peptides after exclusion of poor binders as previously detailed (Chowell et al., 2015; Krishna & Anderson, 2016), and the average normalized binding scores were used to re-rank the candidate peptides. The top 38 candidate peptides (**Table 5-1**) were selected for experimental testing.

In brief, the IEDB consensus MHC-binding prediction algorithm (http://www.iedb.org/) was applied to obtain a list of high binding Cas9 peptides, each of which was assigned a normalized binding score (S_b). The immunogenicity score (S_i) was calculated for each peptide based on its amino acid hydrophobicity (ANN-Hydro) (Chowell et al., 2015).

5.2.3. Ex vivo Stimulation and Epitope Mapping of Cas9 by ELISpot

All peripheral blood mononuclear cells (PBMCs) were obtained from healthy individuals using informed consent, protocol MOD00006783 under ASU's Institutional Review Board. PBMCs were isolated from fresh heparinized blood by Ficoll–Hypaque (GE Healthcare, UK) density gradient centrifugation and stimulated as previously described (Krishna & Anderson, 2016). Briefly, predicted Cas9 peptides with S_b < 0.148 (N=38) were synthesized (> 80% purity) by Proimmune, UK. Each peptide was reconstituted at 1mg/mL in sterile PBS and pools were created by mixing 3-4 candidate peptides. Sterile multiscreen ELISpot plates (Merck Millipore, Billerica, MA, USA) were coated overnight with 5µg/well of anti-IFN-γ capture Ab (clone D1K, Mabtech, USA) diluted in sterile PBS. Frozen PBMCs were thawed rapidly and recombinant human IL-2 (20U/mL, R&D Systems) was added. They were then stimulated in triplicates with 10µg/mL Cas9 peptide pools (or individual peptides), pre-mixed CEF pool as a positive control (ProImmune, UK), or DMSO as a negative control in the anti-IFN-γ-coated ELISpot plates, (Merck Millipore, Billerica, MA, USA) and incubated in a 37°C, 5% CO_2 incubator for 48 hrs. Plates were washed three times for 5 min each with ELISpot buffer (PBS + 0.5% FBS) and incubated with 1µg/mL anti-IFN-γ secondary detection Ab (clone 7-B6-1, Mabtech, USA) for 2 hrs at room temperature, washed and incubated with 1µg/mL Streptavidin ALP conjugate for 1 hr at room temperature. The wells were washed again with ELISpot buffer and spots were developed by incubating for 8-10 min with detection buffer (33µL NBT, 16.5µL BCIP, in 100mM Tris-HCl pH 9, 1mM MgCl2, 150mM NaCl). Plates were left to dry for 2 days and spots were read using the AID ELISpot reader (Autoimmun Diagnostika GmbH, Germany). The average number of spot forming units for each triplicate was calculated for each test peptide or peptide pool and subtracted from the background signal.

5.2.4. Autologous APC Generation from Healthy Individual PBMCs

Autologous CD40L-activated B cell APCs were generated from healthy donors by incubating whole PBMCs with irradiated (32 Gy) K562-cell line expressing human CD40L (KCD40L) at a ratio of 4:1 (800,000 PBMCs to 200,000 irradiated KCD40Ls) in each well. The cells were maintained in B cell media (BCM) consisting of IMDM (Gibco, USA), 10% heat-inactivated human serum (Gemini Bio Products, CA, USA), and Antibiotic-Antimycotic (Anti-Anti, Gibco, USA). BCM was supplemented with 10 ng/mL recombinant human IL-4 (R&D Systems, MN, USA), 2µg/mL Cyclosporin A (Sigma-Aldrich, CA, USA), and insulin transferrin supplement (ITES, Lonza, MD, USA). APCs were re-stimulated with fresh irradiated KCD40Ls on days 5 and 10, after washing with PBS and expanding into a whole 24-well plate. After two weeks, APC purity was assessed by CD19+ CD86+ expressing cells using flow cytometry and were used for T cell stimulation after >90% purity. APCs were either restimulated up to 4 weeks or cryopreserved for re-expansion as necessary.

5.2.5. T cell Stimulation by Autologous APCs

Antigen-specific T cells were generated by stimulating healthy donor B cell APCs bypeptide pulsing of specific Cas9 epitopes. Peptide pulsing of APCs was done under BCM 5% human serum, with recombinant IL-4. Twenty-four hours later, on day 1, APCs were washed and incubated with thawed whole PBMCs at a ratio of 1:2 (200,000 APCs : 400,000 PBMCs) in a 24-well plate in BCM supplemented with 20U/mL recombinant human IL-2 (R&D Systems, MN, USA) and 5ng/mL IL-7 (R&D Systems, MN, USA). On day 5, partial media exchange was performed by replacing half the well with fresh BCM and IL-2. On day 10, fresh APCs were peptide pulsed in a new 24-well plate. On day 11, expanded T cells were restimulated with peptide-pulsed APCs similar to day 1. T cells were used for T cell assays or immunophenotyped after day 18.

5.2.6. Flow Cytometry Staining for T cells

Cells were washed once in MACS buffer (containing PBS, 1% BSA, 0.5mM EDTA), centrifuged at 550g for 5 min and re-suspended in 200µL MACS buffer. Cells were stained in 100µL of staining buffer containing anti-CD137, conjugated with phycoerythrin (PE, clone 4B4-1; BD Biosciences, USA), anti-CD8-PC5 (clone B9.11; Beckman Coulter 1:100), anti-CD4 (clone SK3; BioLegend, 1:200), anti-CD14 (clone 63D3; BioLegend, 1:200), and anti-CD19 (clone HIB19; BioLegend,1:200), all conjugated to Fluorescein isothiocyanate (FITC) for exclusion gates, for 30 min on ice. Samples were covered and incubated for 30 min on ice, washed twice in PBS, and resuspended in 1mL PBS prior to analysis.

5.2.7. Pentamer Staining for T cell Immunophenotyping

The following HLA-A*02:01 PE-conjugated Cas9 pentamers were obtained from ProImmune: F2A-D-CUS-A*02:01-ILEDIVLTL-Pentamer, 007-Influenza A MP 58-66-GILGFVFTL-Pentamer. T cells were washed twice in MACS buffer with 5% human serum and centrifuged at 550g for 5 min each time. They were then re-suspended in 100µL staining buffer (MACS buffer, with 5% human serum and 1mM Dasatanib (ThermoFisher Scientific, MA, USA). Each of the pentamers was added to resuspended T cells, stimulated with the respective peptide or APCs at a concentration of 1:100. Samples were incubated at room temperature for 30 min in the dark, then washed twice in MACS buffer. Cells were stained in 100µL MACS buffer with anti-CD8-PC5, anti-CD4-FITC, anti-CD14-FITC, and anti-CD19-FITC for exclusion gates. Samples were then washed twice with PBS and analyzed by flow cytometry. For flow cytometric analysis, all samples were acquired with Attune flow cytometer (ThermoFisher Scientific, MA, USA) and analyzed using the Attune software. Gates for expression of different markers and pentamers were determined based on flow minus one (FMO) samples for each color after doublet

discrimination. Percentages from each of the gated populations were used for the analysis.

5.2.8. Vector Design and Construction

Modified Cas9 plasmids - Human codon-optimized *Streptococcus pyogenes* Cas9 sequence was amplified from pSpCas9 (pX330; Addgene plasmid ID: 42230), using forward and reverse primers and inserted within gateway entry vectors using golden gate reaction. Desired mutations were designed within gBlocks (Integrated DNA Technologies). The gblocks and amplicons were then cloned into entry vectors using golden gate reaction. All the primers and gblocks sequences are listed in **Appendix B**. Next, the Cas9 vectors and CAG promoter cassettes were cloned into an appropriate gateway destination vector via LR reaction (Invitrogen).

U6-sgRNA-MS2 plasmids - These plasmids were constructed by inserting either 14bp or 20bp spacers of gRNAs (**Appendix B**) into sgRNA (MS2) cloning backbone (Addgene plasmid ID: 61424) at BbsI site. All the gRNA sequences are listed in **Appendix B**.

5.2.9. Cell Culture for Endogenous Target Mutation and Activation

HEK293FT cell line was purchased from ATCC and maintained in Dulbecco's modified Eagle's medium (DMEM - Life Technologies) containing 10% fetal bovine serum (FBS - Life Technologies), 2mM glutamine, 1.0 mM sodium pyruvate (Life Technologies) and 1% penicillin-streptomycin (Life Technologies) in incubators at 37 °C and 5% CO_2. Polyethylenimine (PEI) was used to transfect HEK293FT cells seeded into 24-well plates. Transfection complexes were prepared according to manufacturer's instructions.

5.2.10. Fluorescent Reporter Assay for Quantifying Cas9 Function

HEK293FT cells were co-transfected with 10 ng gRNA, 200 ng Cas9 constructs, 100 ng reporter plasmid and 25 ng EBFP2 expressing plasmid as the transfection control. Fluorescent reporter experiments were performed 48 hrs after transfection. Flow

cytometry data were analyzed using FlowJo. Cells were gated for positive EBFP expression to remove the un-transfected cells from the analysis. Un-transfected controls were included in each experiment.

5.2.11. Quantitative RT-PCR Analysis

HEK293FT cells were co-transfected with 10 ng gRNA, 200 ng Cas9 constructs, 100 ng MS2-P65-HSF1 (Addgene plasmid ID: 61423) and 25 ng transfection control. Cells were lysed, and RNA was extracted using RNeasy Plus mini kit (Qiagen) 72 hrs post transfection, followed by cDNA synthesis using the High-Capacity RNA-to-cDNA Kit (Thermo fisher). qRT-PCR was performed using SYBR Green PCR Master Mix (Thermo fisher). All analyses were normalized to 18s rRNA (ΔCt) and fold-changes were calculated against un-transfected controls ($2^{-\Delta\Delta Ct}$). Primer sequences for qPCR are listed in **Appendix B**.

5.2.12. Endogenous Indel Analysis

HEK293FT cells were co-transfected with 200ng of Cas9 plasmids, 10ng of gRNA coding cassette and and 25 ng transfection control. 72 hrs later, transfected cells were dissociated and spun down at 200 g for 5 min at room temperature. Genomic DNA was extracted using 50 µl of QuickExtract DNA extraction solution (Epicentre) according to the manufacturer's instructions. Genomic DNA was amplified by PCR using primers flanking the targeted region. Illumina Tru-Seq library was created by ligating partial adaptors and a unique barcode to the DNA samples. Next, a small number of PCR cycles were performed to complete the partial adaptors. Equal amounts of each sample were then pooled and sequenced on Illumina Tru-Seq platform with 2x150 run parameters, which yielded approximately 80,000 reads per sample. Sequencing was performed using a 2x150 paired-end (PE) configuration by CCIB DNA Core Facility at Massachusetts General Hospital (Cambridge, MA, USA). The reads were aligned to the target gene

reference in Mus musculus genome using Geneious software, 9-1-5. To detect the indels (insertions and deletions of nucleic acid sequence at the site of double-strand break), each mutation was evaluated carefully in order to exclude the ones that are caused by sequencing error or any off-target mutation. The variant frequencies (percentage to total) assigned to each read containing indels were summed up. i.e. indel percentage = total number of indel containing reads/ total number of reads. The minimum number of analyzed reads per sample was 70,000.

5.2.13. RNA Sequencing for Quantifying Activator Specificity

HEK293FT cells were co-transfected with 10 ng gRNA for MIAT locus, 200 ng Cas9 constructs, 100 ng MS2-P65-HSF1 (Addgene plasmid ID: 61423) and 25 ng transfection control. Total RNA was extracted 72 hrs post transfection using RNeasy Plus mini kit (Qiagen) and sent to UCLA TCGB core on dry ice. Ribosomal RNA depletion, and single read library preparation were performed at UCLA core followed by RNA sequencing using NextSeq500. Coverage was 14 million reads per sample. FASTQ files with single-ended 75 bp reads were then aligned to the human GRCh38 reference genome sequence (Ensembl release 90) with STAR (Dobin et al., 2013), and uniquely-mapped read counts (an average of 14.8 million reads per sample) were obtained with Cufflink (Trapnell et al., 2012). The read counts for each sample were then normalized for the library size to CPM (counts per million reads) with edgeR (Robinson, McCarthy, & Smyth, 2010). Custom R scripts were then used to generate plots.

5.3. RESULTS

5.3.1. Detection of Cas9-Specific Serum Abs in Healthy Controls

We first determined whether healthy individuals have detectable IgG Abs to SpCas9. Of 143 healthy control sera screened, 70 (49.0%) had detectable Abs against *S. pyogenes*

lysate using ELISA (**Figure 5-1**). This positive subset along with 12 sera that were borderline negative for Abs to *S. pyogenes* lysate were screened for Abs against recombinant SpCas9 (**Figure 5-1**), of which 36.6% were positive. At least 21.0% (n=30) of healthy individuals in this study had Cas9-specific Abs (**Figure 5-1**).

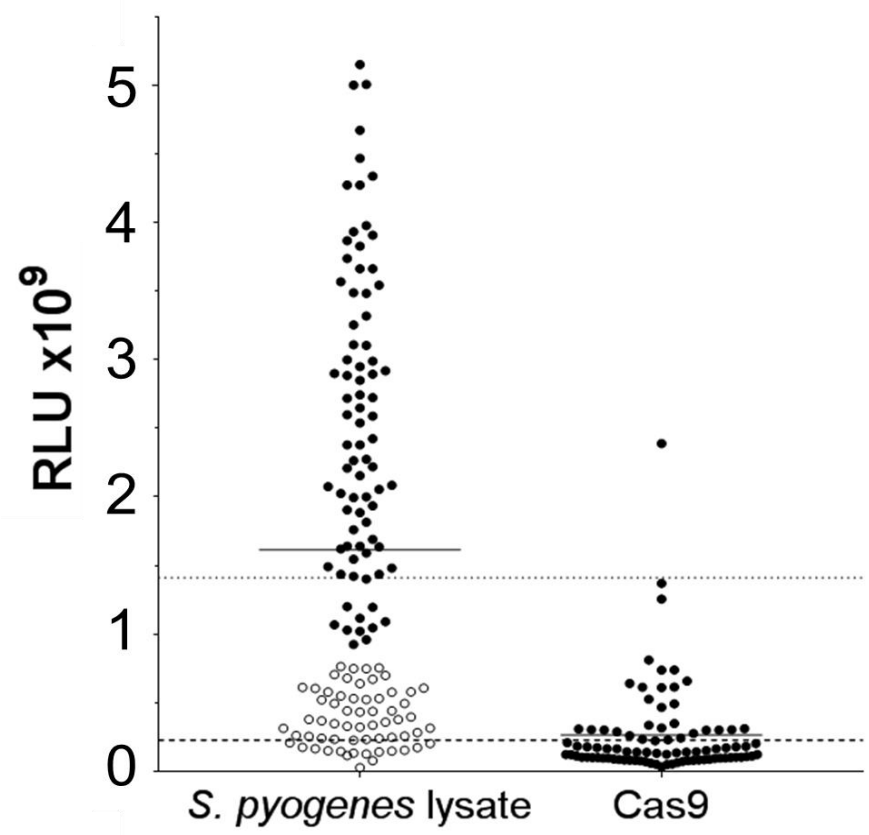

Figure 5-1. Detection of Pre-existing SpCas9-Specific Antibodies in Healthy Individuals. Specific serum Abs were detected against *S. pyogenes* lysate in 49.0% (above the dotted line) of 143 healthy controls (left). The subset shown in black circles was screened for Abs against recombinant Cas9 protein (right), of which 36.6% (21.0% of total samples screened) were positive (above the dashed line). Nucleic Acid Programmable Protein Arrays (NAPPA).

5.3.2. Cas9 Candidate T cell Epitope Prediction

We predicted HLA-A*02:01-restricted T cell epitopes derived from SpCas9 using a model based on both HLA binding and biochemical properties of immunogenicity (Chowell et al., 2015) (**Table 5-1**). We plotted the calculated normalized binding (S_b) and immunogenicity

(S_i) scores for each peptide (**Figure 5-2**) to predict the more immunogenic epitopes, which are expected to have both high HLA binding (low S_b) and more hydrophobicity (high S_i).

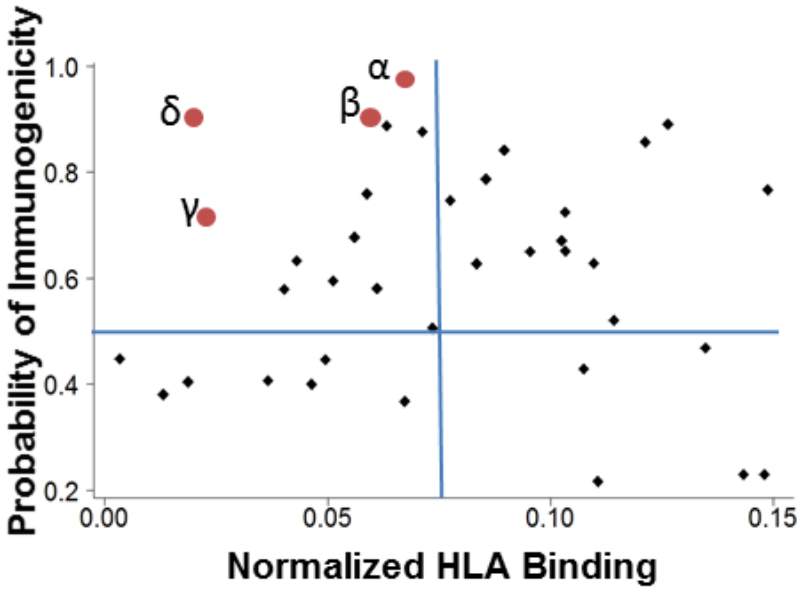

Figure 5-2. Plot of S_b and S_i of Predicted HLA-A*02:01 Epitopes for the SpCas9 Protein. Red dots represent the immunodominant and subdominant epitopes as found by IFN-γ ELISpot.

Table 5-1. Predicted Cas9 Immunogenic T Cell Epitopes.

				Binding			Protein Processing				
Rank	**Position**	**Sequence**	**Code**	**IEDB**	**NetMHC**	**Syfpeithi**	**IEDB**	**ANN**	**S_b**	**S_i**	**S_b.S_i**
1	988-997	YLNAVVGTAL	γ	1.25	21.5	24	0.27	0.02	0.068	0.975	0.002
2	1281-1290	ILADANLDKV		1.25	11.37	31	-0.06	-0.49	0.003	0.447	0.002
3	236-244	GLFGNLIAL	δ	0.6	10.12	29	1.15	1.04	0.020	0.900	0.002
4	240-248	NLIALSLGL	α	1.7	61.18	25	0.15	0.22	0.061	0.903	0.006
5	615-623	ILEDIVLTL	β	1.5	53.29	29	0.28	0.56	0.023	0.710	0.007
6	614-623	DILEDIVLTL		4.6	3105.79	28	-1.53	-1.02	0.063	0.888	0.007
7	719-727	SLHEHIANL		1.4	9.14	30	0.93	0.82	0.013	0.380	0.008
8	415-423	HLGELHAIL		4.4	276.73	25	-0.75	-0.81	0.071	0.876	0.009
9	300-308	ILLSDILRV		0.3	6.51	29	0.67	0.7	0.019	0.404	0.011
10	1086-1095	VLSMPQVNIV		3.65	178.87	26	-1.05	-1.43	0.059	0.758	0.014
11	719-728	SLHEHIANLA		4.7	60.17	19	-0.98	-1.74	0.126	0.890	0.014
12	1194-1203	LIIKLPKYSL		8.5	966.31	25	-0.97	-1.04	0.090	0.841	0.014
13	1346-1355	TLIHQSITGL		1.95	57.8	27	0.12	-0.06	0.043	0.632	0.016
14	1197-1207	KLPKYSLFEL		1.2	10.93	27	0.9	0.5	0.040	0.579	0.017
15	1041-1050	NIMNFFKTEI		2.65	314.8	19	-1.03	-0.9	0.121	0.857	0.017
16	512-520	SLLYEYFTV		0.4	4.56	25	0.67	0.55	0.056	0.678	0.018
17	1309-1318	IIHLFTLTNL		4.25	1083.6	24	-1.04	-0.78	0.085	0.787	0.018
18	661-670	RLSRKLINGI		3.5	278.03	24	-0.82	-1.05	0.078	0.746	0.020
19	1227-1236	ALPSKYVNFL		4.3	111.14	27	0.05	-0.26	0.051	0.594	0.021
20	996-1004	ALIKKYPKL		2.6	154.09	28	-0.27	0	0.037	0.407	0.022
21	221-229	RLENLIAQL		4.2	242.87	26	-0.46	-0.46	0.061	0.581	0.026
22	1237-1245	YLASHYEKL		1.2	10.3	26	0.9	0.84	0.050	0.446	0.027
23	1265-1273	YLDEIIEQI		0.3	4.8	26	0.62	0.6	0.046	0.399	0.028
24	1042-1050	IMNFFKTEI		3.2	131.4	21	-0.69	-0.87	0.103	0.724	0.028
25	815-824	YLQNGRDMYV		0.25	13.01	22	-0.18	-0.07	0.083	0.627	0.031
26	1212-1220	RMLASAGEL		3.2	333.2	22	-0.64	-0.51	0.095	0.650	0.033
27	1020-1029	KMIAKSEQEI		3.1	64.01	21	-0.36	-0.9	0.103	0.671	0.034
28	793-801	SQILKEHPV		2.8	191.23	16	-1.4	-1.36	0.149	0.766	0.035
29	742-750	KVVDELVKV		2.8	44.75	24	-0.06	-0.26	0.074	0.505	0.036
30	1181-1190	FLEAKGYKEV		3.25	105.27	21	-1.08	-1.42	0.103	0.651	0.036
31	160-169	HMIKFRGHFL		4.75	324.13	21	-0.59	-0.73	0.110	0.628	0.041
32	551-559	LLFKTNRKV		3	381.3	25	-1.52	-1.25	0.067	0.368	0.043
33	141-149	KLVDSTDKA		3.4	274.05	20	-1.48	-1.17	0.114	0.520	0.055
34	472-481	TITPWNFEEV		4.45	124.55	21	-0.84	-1.21	0.107	0.429	0.061
35	194-203	QLFEENPINA		1.65	67.94	17	-0.71	-0.79	0.135	0.469	0.072
36	518-527	FTVYNELTKV		2.55	169.93	20	-1.12	-1.15	0.111	0.216	0.087
37	473-481	ITPWNFEEV		6.4	351.14	18	-1.25	-1.65	0.143	0.229	0.110
38	970-978	FQFYKVREI		2.7	135.61	16	-0.66	-0.39	0.148	0.229	0.114

The table shows Cas9 HLA-A*02:01 epitopes predicted using an integrative prediction model and ranked according to their S_b.S_i score (the lower the more immunogenic). The immunodominant and subdominant epitopes as confirmed by ELISpot are highlighted in dark gray and light gray, respectively. S_b, binding score; S_i, immunogenicity score.

5.3.3. Ex vivo Stimulation and Epitope Mapping of Cas9 by ELISpot

We then investigated whether peripheral blood mononuclear cells (PBMCs) derived from healthy individuals had measurable T cell reactivity against the predicted SpCas9 MHC class I epitopes. We synthesized 38 peptides (**Table 5-1**) and grouped them into 10 pools of 3-4 peptides each. We measured peptide-specific T cell immunity using IFN-γ secretion ELISpot assays in PBMCs derived from 12 healthy individuals (HLA-A*02:01, n=10; non-HLA-A*02:01, n=2) and identified immunoreactive epitopes within pools 3 or 5 in 83.0% of the donors tested (90% of the HLA-A*02:01 donors; **Figure 5-3**). The seven individual peptides from pools 3 and 5 were evaluated by IFN-γ ELISpot and the dominant immunogenic epitopes were SpCas9_240-248 and SpCas9_615-623, designated peptides α and β, from pools 5 and 3, respectively. The subdominant epitopes were found to be γ and δ from pools 3 and 5, respectively. Both peptides α and β are located in the REC lobe of the Cas9 protein (**Figure 5-5**) that binds the sgRNA and the target DNA heteroduplex (Nishimasu et al., 2014) .The position of peptides α and β within the protein structure is shown in **Figure 5-4**. The individual peptides within pools that were positive for any donor were evaluated for this donor by IFN-γ ELISpot. The immunoreactivity and position of the 38 predicted peptides (a few of which are overlapping) within the SpCas9 protein are shown in **Figure 5-5**.

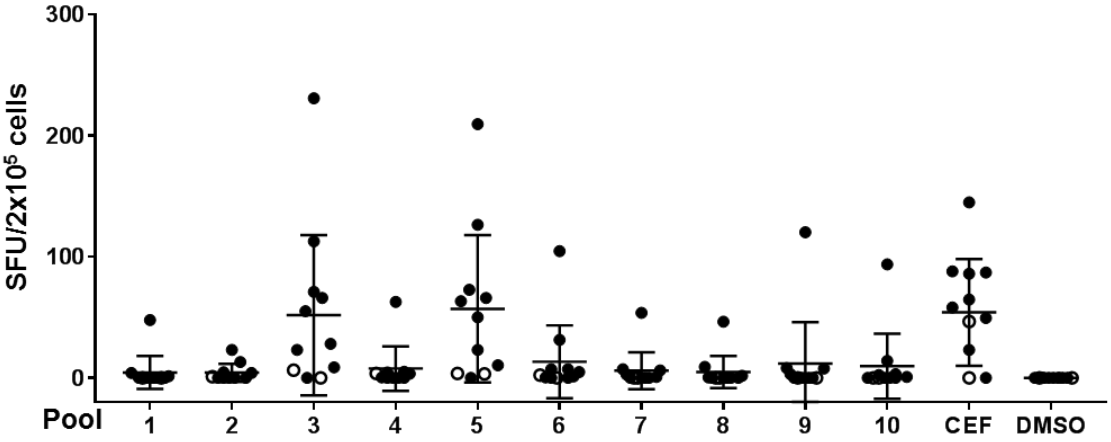

Figure 5-3 Detection of Pre-existing T Cell Immune Response to SpCas9 in Healthy Donors. IFN-γ ELISpot assay of T cell reactivity of 12 healthy donors (non-HLA-A*02:01 are shown as open circles; n=2) to 38 predicted epitopes grouped in 10 pools, CEF (positive control), and DMSO (negative control). Peptides α and δ were in pool 5 while β and γ were in pool 3.

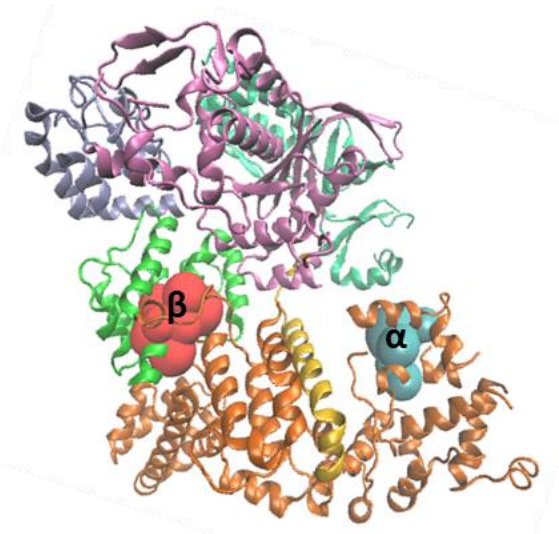

Figure 5-4. 3D Structure of the SpCas9 Protein. The location of the identified immunodominant epitopes α and β is shown.

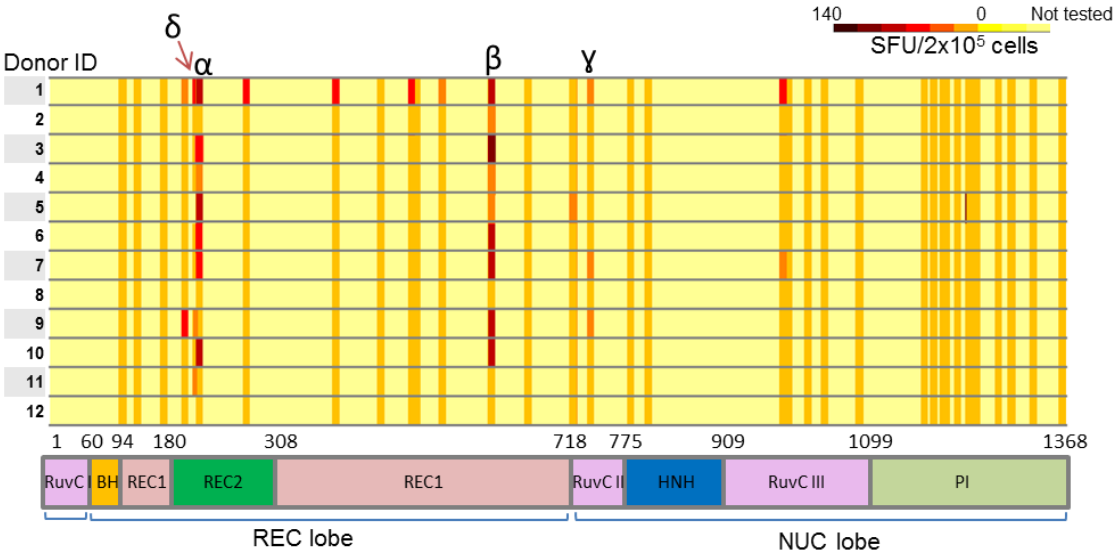

Figure 5-5. IFN-γ ELISpot Reactivity of Healthy Donor T Cells (N=12) to Epitopes Across the Different Domains of the *S. pyogenes* Cas9 Protein. Donors 1-10 were HLA-A*02:01, while 11 and 12 were not. Peptides α and δ overlap in 5 amino acid residues.

5.3.4. Sequence Similarity of Identified T cell Epitopes

Peptides α and β are shown as red dots on the epitope prediction plot (**Fig. 5-2**) and their sequences and predicted ranking are shown in **Table 5-1**. As predicted, these peptides had low S_b and high S_i values. Both the immunodominant (α and β) and subdominant (γ and δ) T cell epitopes identified by IFN-γ ELISpot were within the top 5 most immunogenic epitopes predicted by the previously described immunogenicity model (Chowell et al., 2015). Their ranking as predicted by the consensus method hosted on the IEDB server using default settings was 14, 5, 18, and 4, respectively. Sequence similarity of peptides α and β to amino acid sequences in known proteins was investigated using Protein BLAST and the IEDB epitope database (Vita et al., 2015). This was done to investigate whether there is any chance that the T cell immune response that we are detecting in healthy individuals could be due to previous exposure to another protein of similar sequence. A

peptide was considered ‘similar’ to α or β if at least 7 of 9 (78%) amino acid residues (that are not the second or ninth) were matching. None of these two peptides resembled known epitopes in the IEDB database, but similarity to other Cas9 orthologs and other bacterial proteins was detected (**Tables 5-2** and **5-3**). Epitope β has sequence similarity to a peptide derived from the *Neisseria meningitidis* peptide chain release factor 2 protein (ILEDIVLTL versus ILE**G**IVLTL).

Table 5-2. Sequence Homology of Epitope α to Amino Acid Sequences from Known Proteins.

	Sequence	*Similarity (%)*	*Protein*	*Sequence ID*	*Source*
1	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_014612333.1	*Streptococcus dysgalactiae*
2	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_054279288.1	*Streptococcus phocae*
3	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_067062573.1	*Streptococcus pantholopis*
4	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_048800889.1	*Streptococcus constellatus*
5	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_002304487.1	*Streptococcus mutans*
6	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_049516684.1	*Streptococcus anginosus*
7	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_003079701.1	*Streptococcus macacae*
8	NLIALSLGL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	GAD40915.1	*Streptococcus intermedius* SK54
9	NLIA**F**SLGL	8/9 (89%)	Full=RNA polymerase-associated protein RapA; AltName: Full=ATP-dependent helicase HepA	Q6LV34.1	*Photobacterium profundum* SS9
10	NLI**S**LSLGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_096633625.1	*Streptococcus parauberis*
11	NLIAL**A**LGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_075103982.1	*Streptococcus cuniculi*

	Sequence	*Similarity (%)*	*Protein*	*Sequence ID*	*Source*
12	NLIAL**A**LGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_058692367.1	*Streptococcus gallolyticus*
13	NLIAL**A**LGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_061100419.1	*Streptococcus pasteurianus*
14	NLIAL**A**LGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_018363470.1	*Streptococcus caballi*
15	NLIAL**A**LGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_099412266.1	*Streptococcus macedonicus*
16	NLIAL**A**LGL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_014334983.1	*Streptococcus infantarius*
17	**D**LIAL**Y**LGL	7/9 (78%)	Full=NADH-quinone oxidoreductase subunit N; AltName: Full=NADH dehydrogenase I subunit N; AltName: Full=NDH-1 subunit N	A8I421.1	*Azorhizobium caulinodans* ORS 571
18	NL**L**AL**A**LGL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_007896501.1	*Streptococcus pseudoporcinus*
19	NLI**G**L**A**LGL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_061587801.1	*Streptococcus oralis*
20	NL**V**AL**A**LGL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_074862269.1	*Streptococcus equinus*
21	NL**V**AL**V**LGL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_020917064.1	*Streptococcus lutetiensis*
22	**S**LIA**F**SLGL	7/9 (78%)	ectoine/hydroxyectoine ABC transporter permease subunit EhuD	WP_086160327.1	*Streptomyces* sp. SCSIO 03032
23	**Y**LIAL**A**LGL	7/9 (78%)	ectoine/hydroxyectoine ABC transporter permease subunit EhuD	WP_026413155.1	*Actinomadura oligospora*

Table 5-3. Sequence Homology of Epitope β to Amino Acid Sequences from Known Proteins.

	Sequence	*Similarity (%)*	*Protein*	*Sequence ID*	*Source*
1	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_084916602.1	*Streptococcus dysgalactiae*
2	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_074484960.1	*Streptococcus henryi*
3	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_003088697.1	*Streptococcus ratti*
4	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_044681799.1	*Streptococcus suis*
5	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_024786433.1	*Streptococcus mutans*
6	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_057491067.1	*Streptococcus orisasini*
7	ILEDIVLTL	9/9 (100%)	type II CRISPR RNA-guided endonuclease Cas9	WP_082312238.1	*Streptococcus intermedius*
8	ILE**G**IVLTL	8/9 (89%)	peptide chain release factor 2	NP_275123.1	*Neisseria meningitidis* MC58
9	ILEDIV**Q**TL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	EAO61901.1	*Streptococcus agalactiae*
10	ILEDIV**Q**TL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_070454905.1	*Streptococcus* sp. HMSC063D10
11	**V**LEDIVLTL	8/9 (89%)	type II CRISPR RNA-guided endonuclease Cas9	WP_075346866.1	*Streptococcus* sp. '*caviae*'
12	**V**LEDIVL**S**L	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_093650272.1	*Streptococcus varani*

	Sequence	*Similarity (%)*	*Protein*	*Sequence ID*	*Source*
13	ILE**N**IV**H**TL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	KYF37509.1	*Streptococcus mitis*
14	ILE**N**IV**H**TL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_084972088.1	*Streptococcus oralis*
15	ILE**N**IV**H**TL	7/9 (78%)	type II CRISPR RNA-guided endonuclease Cas9	WP_045635197.1	*Streptococcus gordonii*

5.3.5. Detection of Cas9-Specific T Cell Immune Response Against the Identified Immunodominant Epitopes

Antigen-specific T cells were expanded for 18 days *in vitro* by coculturing healthy donor PBMCs with peptide β-pulsed autologous antigen presenting cells (APCs). Cas9-specific CD8+ T cell responses were assessed by flow cytometry. CD8+ T cells specific for the HLA-A*0201/β pentamer were detected after stimulation (3.09%; **Fig. 5-6A**).

We next hypothesized that mutation of the MHC-binding anchor residues of the identified immunogenic epitopes would abolish specific T cell recognition (**Fig. 5-6A**). The epitope anchor residues (2nd and 9th) are not only necessary for peptide binding to the MHC groove but are also crucial for recognition by the T cell receptor (Chowell et al., 2015). The percentage of CD8+ pentamerβ+ T cells dropped to 0.3% when APCs were pulsed with the mutated peptide (β2; **Fig. 5-6B**) compared with 3.09% with the wild type peptide (β; **Fig. 5-6A**). We then examined the reactivity of healthy donor T cells to modified peptides α or β with mutations in residues 2, 9, or both (sequences are shown in **Table 5-4**) using IFN-γ ELISpot assay. The epitope-specific T cell reactivity was markedly reduced with the mutant peptides (**Fig. 5-7**, **5-8**). The average reduction for the responsive HLA-A*02:01 donors was 25-fold from α to α29 ($n=7$, $p<0.03$) and 30-fold from β to β29 ($n=8$; $p<0.03$; **Fig. 5-7**).

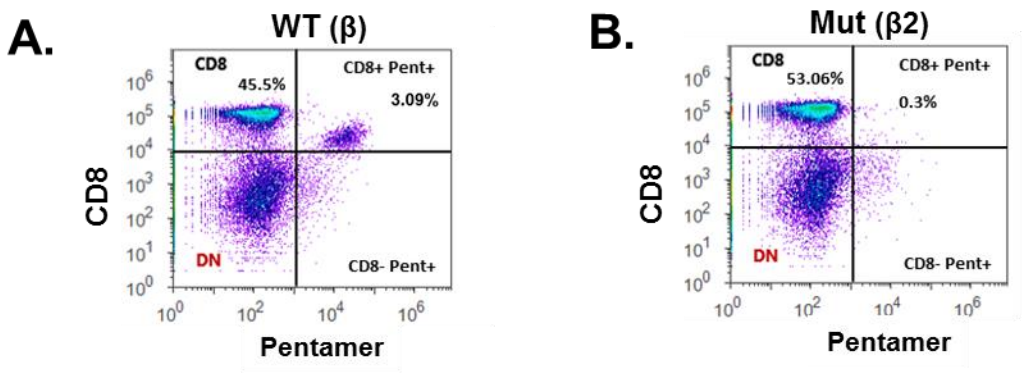

Figure 5-6. Detection of Cas9-Specific T Cells Recognizing the Immunodominant Epitope β in a Healthy Individual and Reduction of the Immune Recognition by Mutating One Anchor Residue of the Epitope. A. Epitope β-specific CD8+ T cell response detected using β-specific pentamer in PBMCs stimulated with peptide β-pulsed antigen presenting cells. **B.** The percentage of CD8+ pentamerβ+ T cells was reduced to 0.3% when APCs were pulsed with the mutated peptide β2.

Table 5-4. Positions, Sequences and IEDB HLA Binding Percentile Rank of Epitopes α and β Before and After Mutation of the Anchor (2nd and/or 9th) Residues.

Peptide Code (Position)	Peptide Sequence	HLA binding (percentile rank)	Peptide Code (Position)	Peptide Sequence	HLA binding (percentile rank)
α (240-248)	NLIALSLGL	1.7	β (615-623)	ILEDIVLTL	1.5
α2	NGIALSLGL	26	β2	IGEDIVLTL	23
α9	NLIALSLGG	14	β9	ILEDIVLTG	12
α29	NGIALSLGG	62	β29	IGEDIVLTG	49

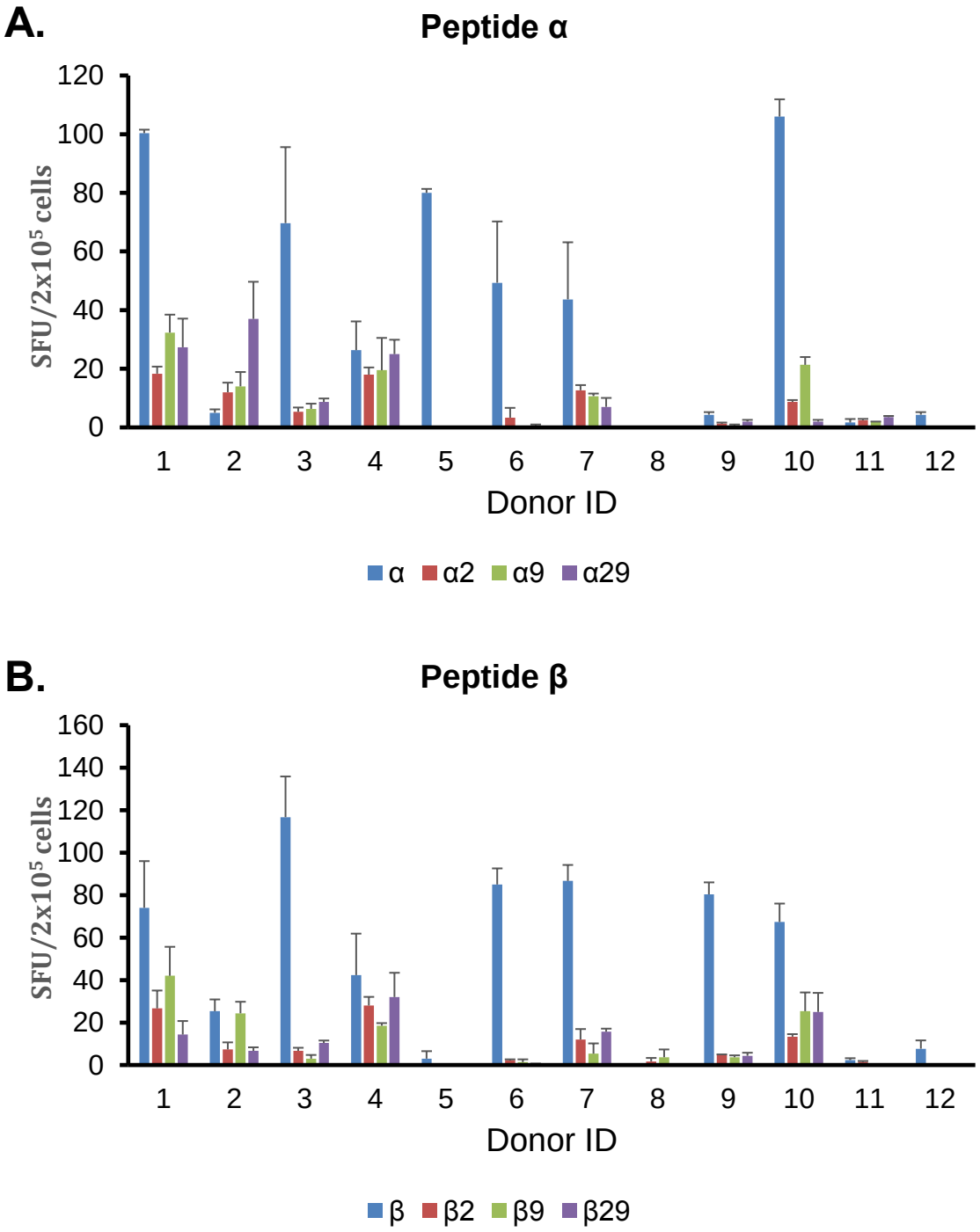

Figure 5-7. Reduced T Cell Response to Epitopes α and β After Mutation of the Anchor Residues. IFN-γ ELISpot for 12 healthy donor PBMCs stimulated with wild type or mutated peptide α (A) or β (B). Data represent mean+/- SEM.

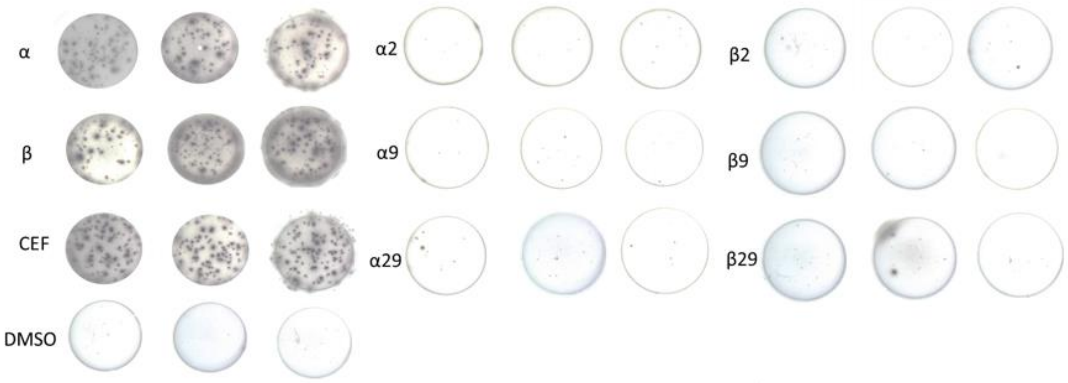

Figure 5-8. SpCas9 Immunodominant Epitope-Specific CD8+ T Cell Recognition Is Abolished After Anchor Residue Mutation. IFN-γ ELISpot assay in triplicate wells comparing T cell reactivity to wild type or mutated epitopes α and β. These results are representative of 12 donors and two independent replicates.

5.3.6. Mutated Cas9 Proteins Have Lower Immune Recognition and Maintain their Function and Specificity

We then generated modified Cas9 constructs by mutating the second residue of peptide α (L241G; Cas9-α2), peptide β (L616G; Cas9-β2), or both (Cas9-α2β2). To measure the effect of mutating the anchor residue of the immunogenic epitopes on T cell recognition of the Cas9 protein, we transiently transfected healthy donor B cell APCs with mRNA encoding WT Cas9, Cas9-α2, Cas9-β2, or Cas9-α2β2. Protein expression was confirmed by Western blot and the levels were comparable for all four constructs (data not shown). The T cell response measured by IFN-γ ELISpot after coculturing of transfected APCs with autologous PBMCs was significantly decreased for the modified Cas9 proteins (**Figure 5-9**). These results demonstrate that mutating the anchor amino acid residue at a highly immunogenic epitope can influence the overall immunogenicity of Cas9. Hence, engineering Cas9 variants with reduced immunogenicity potential can be used in conjunction with other strategies for safer CRISPR therapies and even possibly reduce the dosage of systemic immunosuppression needed for patients. Introduction of the β2

mutation was the most effective in reducing T cell immunogenicity (5.5-fold, p<0.0001). This mutation in the REC1 domain (**Figure 5-4** and **5-5**) is not located in any of the two regions that are absolutely essential for DNA cleavage, the repeat-interacting (97–150) and the anti-repeat-interacting (312–409) regions (Nishimasu et al., 2014).

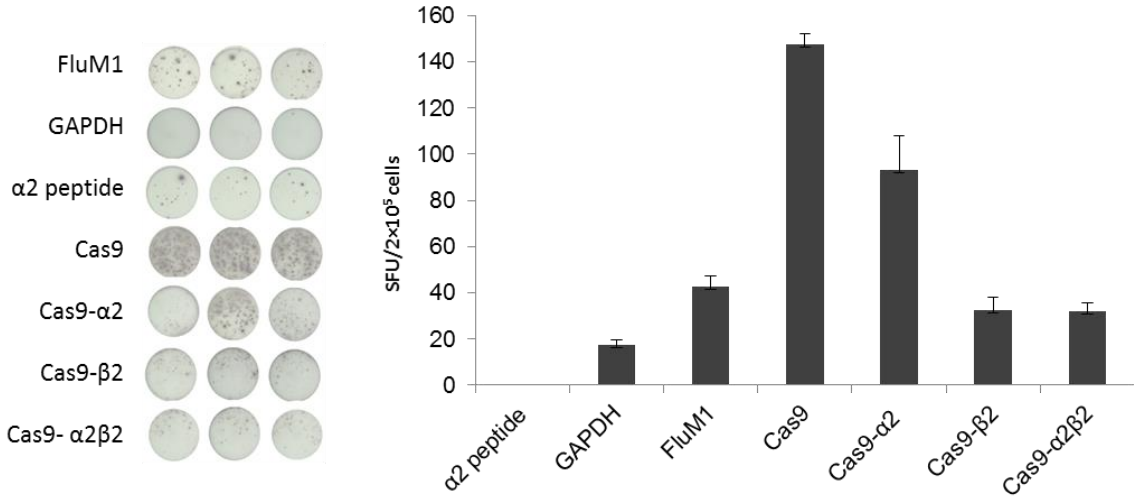

Figure 5-9. Mutated Cas9 Proteins Have Lower Immune Recognition. IFN-γ ELISpot comparing T cell reactivity to APCs expressing WT or modified Cas9 proteins. APCs expressing FluM1 were used as a positive control. APCs expressing GAPDH or spiked with peptide α2 were used as negative controls. Data represent mean +/- SEM of 5 replicates (right).

We then tested the function of Cas9-β2 in comparison with wild type Cas9 (WT-Cas9) in the context of DNA cleavage and transcriptional modulation. To examine the nuclease activity of Cas9-β2 and compare with WT-Cas9, we targeted Cas9-β2 or WT-Cas9 to an endogenous locus (*EMX-1*) and measured percent indel formation (**Figure 5-10**). Our data demonstrate that Cas9-β2 retains nuclease capacity in the locus we studied (**Fig. 5-10B**) as well as on a synthetic promoter (**Fig. 5-11**).

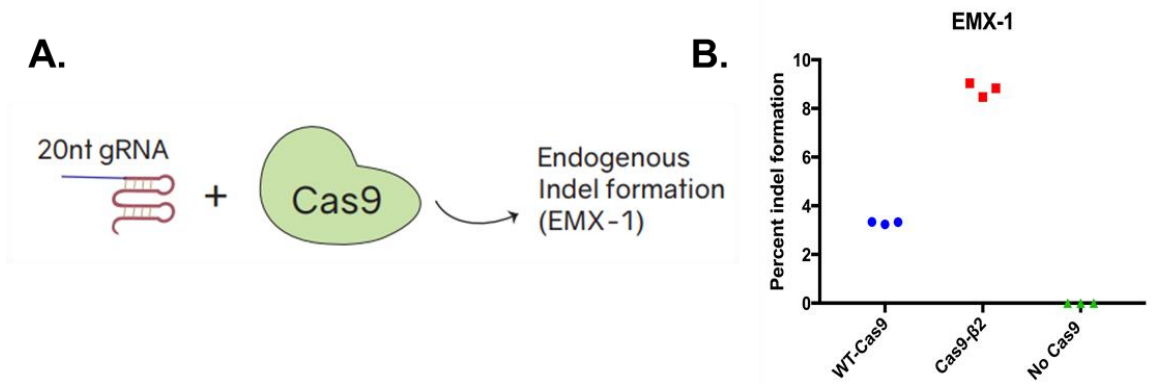

Figure 5-10. Mutated Cas9 Retains Its Nuclease Function at an Endogenous Locus. **A.** Schematic of the experiment assessing mutagenesis capacity of Cas9-β2. Cells were transfected with either WT-Cas9, Cas9-β2, or an empty plasmid as well as 20nt gRNA targeting *EMX-1* locus. 72 hrs after the transfection, percent cleavage was assessed by DNA extraction and illumina sequencing. **B.** Percentage of indel formation in *EMX-1* locus. Data represent mean +/-SD of three individual transfections.

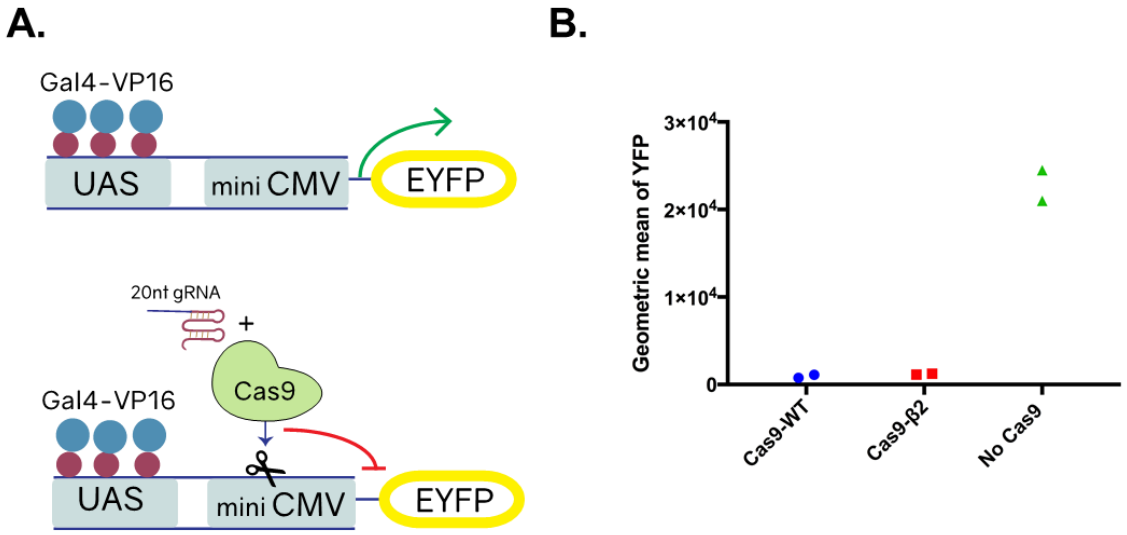

Figure 5-11. Mutated Cas9 Retains Its Nuclease Function at a Synthetic Promoter. **A.** Schematic of the experiment assessing Cas9-β2 cleavage capacity at a synthetic promoter. Cells were transfected with either WT-Cas9, Cas9- β2 or an empty plasmid as well as 20nt gRNA targeting a synthetic CRISPR promoter that harbors two gRNA target sites flanking a mini-CMV promoter. The targeting and cleavage at the promoter should disrupt the promoter and decrease EYFP expression. **B.** Each individual dot EYFP expression 48 hours after the transfection in cells expressing >2x10^2 A.U. of a transfection marker measured by flow cytometry (n=2 individual transfections represented by individual dots).

Next, we determined whether Cas9-β2 can successfully recognize and bind its target DNA leading to transcriptional modulation. We first tested this in the context of enhanced transgene expression from a synthetic CRISPR responsive promoter in HEK293 cells using 14nt gRNAs and aptamer-mediated recruitment of transcriptional modulators similar to what we had shown before (**Fig. 5-12**). Having shown successful transgene activation, we then investigated whether this variant retains such capacity within the chromosomal contexts of endogenous genes. We transfected the cells with plasmids encoding Cas9-β2 or WT-Cas9 and 14nt gRNAs against two different endogenous genes (*TTN* and *MIAT*). qRT-PCR analysis showed that this variant successfully led to target gene expression (**Fig. 5-13A-C**). To further characterize Cas9-β2 specificity, we performed genome-wide RNA sequencing after targeting Cas9-β2 or WT-Cas9 to the *MIAT* locus for transcriptional activation. The results demonstrated no significant increase in undesired off-target activity by Cas9-β2 as compared to WT-Cas9 (**Fig. 5-13D**).

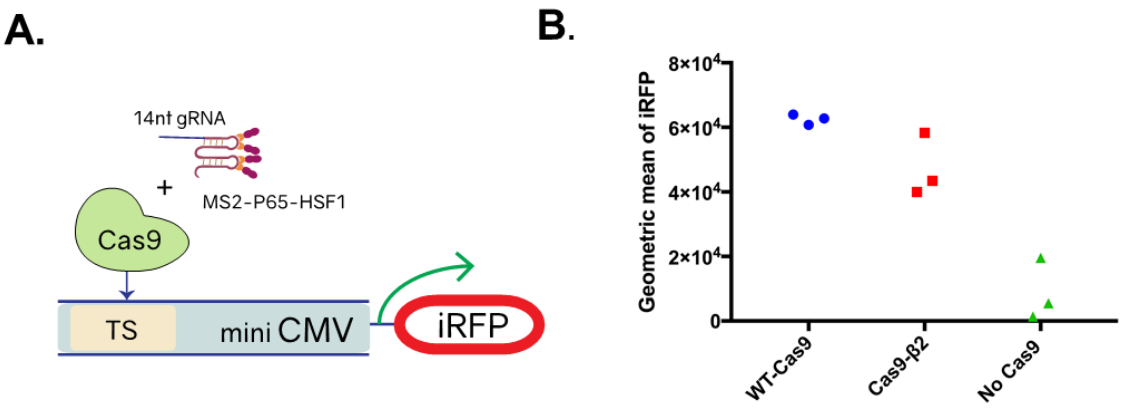

Figure 5-12. Mutated Cas9 Retains Its Transcriptional Modulation Capacity at a Synthetic Promoter. A. Schematic of the experiment assessing Cas9-β2 transcriptional activation capacity at a synthetic promoter. Cells were transfected with either WT-Cas9, Cas9- β2 or an empty plasmid as well as aptamer binding transcriptional activation domains, and14nt gRNA targeting a synthetic CRISPR promoter that harbors multiple target sites upstream of a mini-CMV promoter. The targeting at the promoter should should enable iRFP expression. **B.** Data shows mean+/-SD of geometric mean of iRFP expression 48 hrs after the transfection in cells expressing >2x10^2 A.U of a transfection marker measured by flow cytometry (n=3 individual transfections).

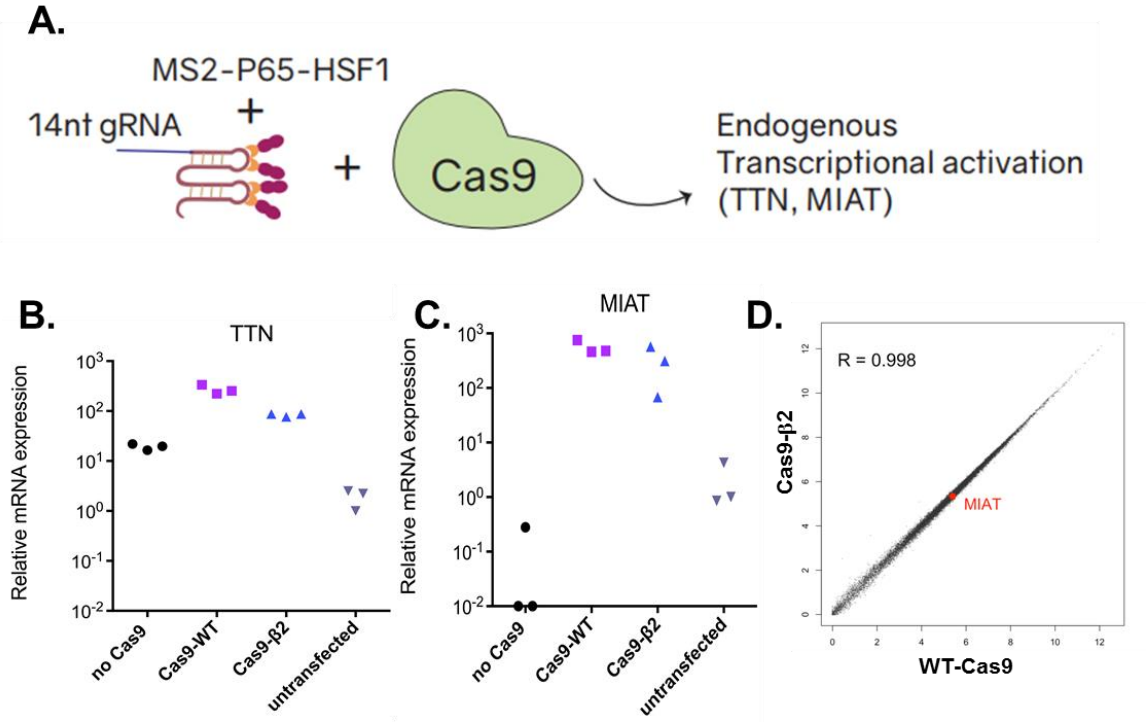

Figure 5-13. Mutated Cas9 Retains its Specificity and Transcriptional Modulation Capacity at an Endogenous Chromosomal Locus. A. Schematic of the experiment assessing gRNA binding, DNA targeting and transcriptional modulation with Cas9-β2. Cells were transfected with either WT-Cas9, Cas9-β2, or an empty plasmid as well as 14nt gRNA targeting *TTN* or *MIAT* in the presence of MS2-P65-HSF1 (transcriptional modulation). 72 hrs after the transfection, mRNA was assessed by qRT-PCR. **B, C.** Shown is the mRNA level relative to an untransfected control experiment. Each individual dot represents an individual transfection. **D.** Mean expression levels of 24,078 protein-coding and non-coding RNA genes for WT-Cas9 and Cas9-β2 (each in duplicate) are shown. For visualization purposes, the values were transformed to a log2(CPM+1) scale. *MIAT*, the gRNA target gene, is highlighted in red, and R denotes Pearson correlation coefficient between two groups.

To show the extensibility of our approach, we tested the function of Cas9-α2, with a mutation located in the REC2 domain (**Figures 5-4** and **5-5**). Cas9-α2 also demonstrated DNA cleavage and transcriptional modulation functionality comparable with WT-Cas9 (**Fig. 5-14**). This is consistent with a previous study which showed that Cas9 with a deleted REC2 domain retains its nuclease activity (Nishimasu et al., 2014). When T cells were

stimulated with APCs spiked with peptide α2, the percentage of CD8+ CD137+ T cells (a marker of T cell activation (Wolfl et al., 2007)) was decreased by 2.3 fold as compared to WT peptide α stimulation (**Fig. 5-15**).

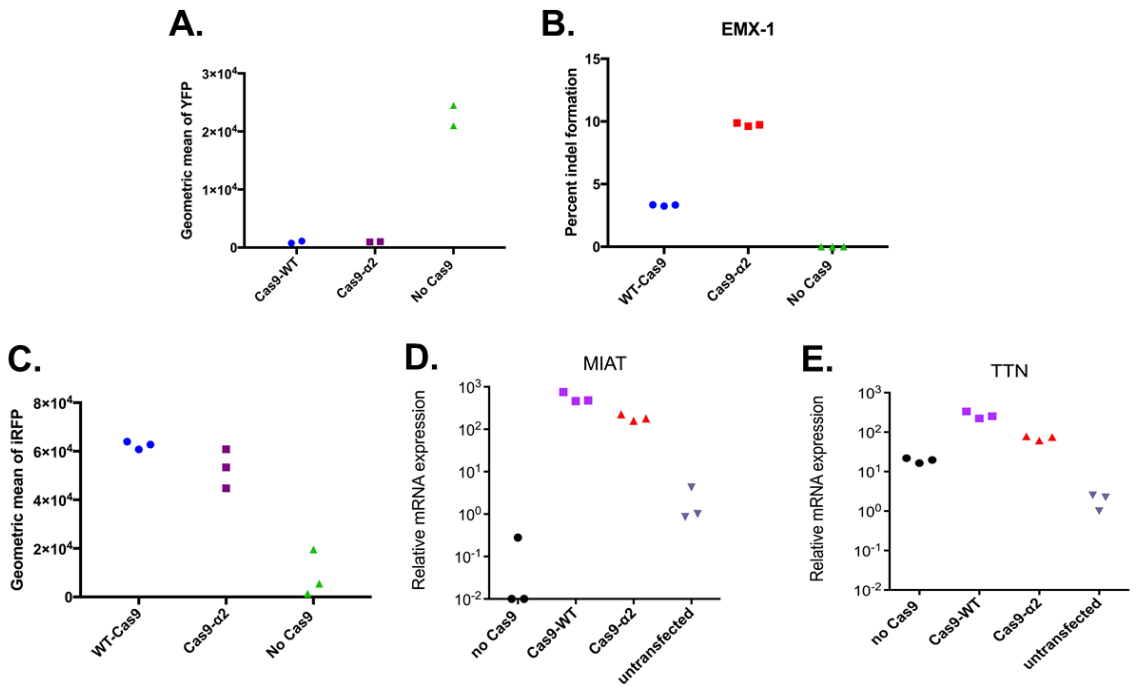

Figure 5-14. Cas9-α2 Retains Its Nuclease and Transcriptional Modulation Activities. A. Analysis of mutagenesis capacity of Cas9-α2 as compared to WT-Cas9 in a synthetic promoter. Data shows mean+/-SD of geometric mean of EYFP expression 72 hrs after the transfection in cells expressing >2x10^2 A.U of a transfection marker measured by flow cytometry (n=2 individual transfections). **B.** Mutagenesis in endogenous *EMX-1* locus. Percentage of indel formation in *EMX-1* locus. Data is mean +SD of three individual transfections. **C.** Transcriptional modulation by Cas9-α2 at a synthetic promoter. Data is mean+/-SD of geometric mean of iRFP expression 72 hrs after the transfection in cells expressing >2x10^2 A.U of a transfection marker measured by flow cytometry (n=3 individual transfections). **D, E.** Shown is mRNA level relative to an untransfected control experiment. Each individual dot represents individual transfections. Note for **A** and **C** WT-Cas9 and no Cas9 data is also reported in Fig.5-10 and 5-11. For **B-E** WT-Cas9 and no Cas9 data is also reported in Fig.5-9 and 5-12.

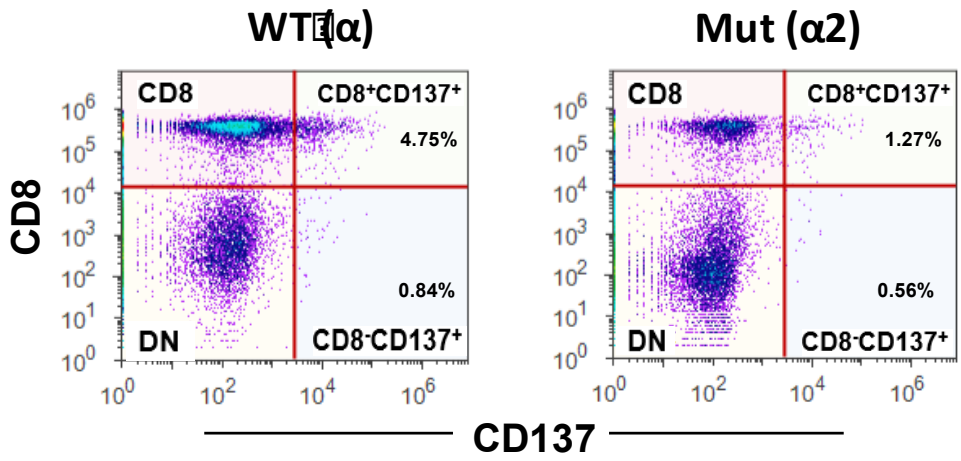

Figure 5-15. Mutation of the Second Residue of Epitope α Reduces Its Immunogenicity. Activated CD8+CD137+ T cells detected in PBMCs stimulated with peptide α were reduced in PBMCs stimulated with peptide α2.

5.4. DISCUSSION

The detection of pre-existing B cell and T cell immunity to the most widely used nuclease ortholog of the CRISPR/Cas9 tool in a significant proportion of healthy humans confirms previous studies in mice (Chew et al., 2016; D. Wang et al., 2015) and sheds light on the need for more studies of the immunological risks of this system. The CD8+ T cell immunity we observed is likely memory responses, as they are observed without ex vivo stimulation. Following 18 days of T cell stimulation by peptides α or β, expansion of naïve T cells is not precluded. This suggests that the expression of Cas9 in naïve individuals may trigger a T cell response that could prevent subsequent administration. This could be avoided by switching to Cas9 orthologs from other bacterial species, but attention needs to be given to individual and distinct immune repertoires. This practice can be difficult given the epitope conservation across Cas9 proteins from multiple *Streptococcus* species and resemblance to sequences from other bacterial proteins such as the common pathogen *N. meningitidis*, that asymptomatically colonizes the nasopharynx in 10% of the population (Pollard & Maiden, 2001). Therefore, selective deimmunization (immunosilencing) of Cas9

can represent an attractive alternative. Selective deimmunization can be an effective alternative for reduction of immune response to CRISPR in patients where high-dose systemic immunosuppression is contraindicated, such as in patients with chronic infectious diseases. This strategy can be important particularly when longer expression of Cas9 will be desired for epigenetic therapy.

CRISPR application for human therapies will span its use both for gene editing (through DNA DSB) or epigenetic therapies (without DNA DSB). In fact, recent reports shed light on CRISPR's ability to activate or repress gene expression in mice (Ibrahim & Robertson, 2018; Liao et al., 2017; Zheng et al., 2018), which open the door to a variety of new therapeutic applications such as activating silent genes, compensating for disrupted genes, cell fate reprogramming or silencing disrupted genes, without the concern over permanent change in DNA sequence. However, unlike the use of Cas9 for gene editing, which may only require Cas9 presence in cells for a few hours, current techniques for CRISPR-based epigenetic therapies require longer term expression of Cas9 *in vivo*, possibly for weeks and months (Liao et al., 2017; Zheng et al., 2018). The requirement for longer term expression of Cas9 *in vivo* poses the challenge of combating pre-existing immune response towards Cas9. This challenge will need to be addressed before CRISPR application for human therapies, especially for epigenetic therapies, can be fully implemented.

Delivery of CRISPR *in vivo* by incorporating its expression cassette in adeno-associated virus (AAV), will most likely shape many of initial clinical trials as AAV-based gene delivery is one of the safest and most prevalent form of gene therapies in human. AAV will enable longer term expression of Cas9, desirable for epigenetic therapies. Therefore, unlike Cas9 delivery in the form of ribonucleoprotein complexes (which are short term), it is highly likely

that CRISPR delivery through AAV and its expression within target cells will engage CD8+ T cell immunity.

Conventional methods of deimmunizing non-human therapeutic proteins rely on trial-and-error mutagenesis, machine learning, and often includes deletion of whole regions of the protein (Cantor et al., 2011; King et al., 2014; Mazor et al., 2017; Salvat et al., 2017; Tangri et al., 2005). Here, as a general principle, we show that alteration of one of the anchor residues of an immunodominant epitope abolished specific T cell recognition. However, HLA allotype diversity and the existence of numerous epitopes in the large Cas9 protein complicate the process of complete deimmunization. The overall impact of removal of select immunodominant epitopes remains to be seen; both reduction (Yeung et al., 2004) and enhancement (Mok, Lee, Wright, & Crowe, 2008) of the immunogenicity of subdominant epitopes have been reported with similar approaches for other proteins.

CHAPTER 6 : SUMMARY AND FUTURE PERSPECTIVES

6.1. SUMMARY

The immune response provides a valuable tool that can be harnessed for developing biomarkers with various applications in cancer. These range from screening seemingly healthy individuals to diagnosis, prognosis, monitoring recurrence and prediction of response to therapy. This dissertation employs classical and novel techniques to identify immune profiles for use as biomarkers for cancer detection and novel therapies.

Despite hundreds of thousands of research publications reporting claimed biomarkers, the number in clinical use is astonishingly small. This has been a well-recognized but poorly addressed problem in biomarker research, that results in wasted financial and intellectual resources. In Chapter 2, the crucial considerations of developing pipelines for the rapid evaluation of circulating biomarkers are reviewed, with a focus on breast cancer as a case study. I discuss critical elements in biomarker study design and assay development, both at the discovery and the validation stages, to increase the identification of clinically useful markers. Factors that are needed to establish pipelines for the rapid translation of these biomarkers to clinical practice across multiple clinical applications were also identified.

In Chapter 3 of this dissertation, I describe the generation and validation of custom HPV NAPPA arrays displaying the proteomes of 12 HPV types for immunoprofiling HPV-associated cancers. We demonstrate robust protein expression for 98% of the Ags expressed. The displayed Ags retained antigenic specificity of target epitopes, as measured by four HPV-specific MAbs. Ab reactivity to multiple early HPV proteins were detected in sera from patients with HPV-associated malignancies using the arrays and high correlation of HPV16 IgG detection with RAPID ELISA was confirmed.

These arrays were used in Chapter 4 to identify the immune response in cervical cancer and pre-invasive cervical lesions. Abs to any early (E) HPV protein were detected less

frequently in women with CIN 0/I (23.7%) than women with CIN II/III (39.0%) and ICC (46.1%, $p<0.04$). Of the E Abs, anti-E7 Abs were the most frequently detected (6.6%, 19.5%, and 30.3%, respectively). The least frequently detected Abs were E1 and E2-Abs in CIN 0/I (1.3%) and E1-Abs in CIN II/III (1.2%) and ICC (7.9%). HPV16-specific Abs correlated with HPV16 DNA detected in the cervix in 0% of CIN 0/I, 21.2% of CIN II/III, and 45.5% of ICC. A significant number (29 – 73%) of E4, E7, L1, and L2 Abs had cross-reactivity between HPV types. HPV protein arrays provide a valuable high-throughput tool for measuring the breadth, specificity, and heterogeneity of the serologic response to HPV in cervical disease.

The CRISPR/Cas9 system has raised hopes for developing personalized gene therapies for cancer. However, the expression of the bacterial Cas9 nuclease in humans raises concerns over safety and potential immune adverse reactions. In Chapter 5, biomarkers of Cas9-specific B cell and T cell immunity are identified. Two immunodominant T cell epitopes for HLA-A*02:01 were identified and a single mutation in the anchor residue of one or both of these epitopes significantly reduced the protein immunogenicity while maintaining its function and specificity in the CRISPR/Cas9 system. The identified immune response may find application in informing *in vivo* gene therapy decisions for cancer as well as other diseases.

6.2. FUTURE PERSPECTIVES

To date, there are no established tissue, blood, or vaginal biomarkers other than HPV nucleic acid and cytology for CIN II/III in high risk patients. Biomarkers such as serology that identify high-risk HPV infection and invasive cervical cancers (ICC) could have an impact on the screening, detection, and treatment of cervical disease. In low-resource settings, a low-cost blood-based point-of-care (POC) test for HPV serology is ideal for population screening. Development of such test has been challenged by the difficulties of

biomarker development for cervical disease and of achieving an analytical sensitivity high enough for their detection in a few drops of blood in a POC setting.

The discovery of cervical disease immune biomarkers has been limited by several factors. These include the diversity of HPV types, the technical challenges of high-throughput study of the immune response to multiple proteins, and the heterogeneity of the immune response to viral proteins in HPV-associated cancers, as reported here. Since HPV16 is responsible for only 50-55% of cervical cancers (Bosch et al., 2008), we predicted that custom protein arrays displaying a broad spectrum of HPV antigens in a slide-based format may improve the detection of serologic responses to HPV. While we were able to detect an HPV-specific immune response that increases with disease severity, an improved signal-to-noise ratio is required for developing clinically useful biomarkers. More studies are needed to reach the ultimate goal of developing a panel of a limited number of select biomarkers that can be produced at low cost for display in a POC test with optimal performance for the detection of women at risk of developing cancer. This study highlights the importance of investigating proteins from HPV types other than the most widely studied HPV16 for improved sensitivity.

One challenge with slide-based protein arrays is the possibility of diffusion of the desired product from one feature on the array to another, resulting in signal cross-talk. This could happen during the *in vitro* protein synthesis step in NAPPA, before the GST-tagged protein is fully synthesized and captured on the array surface. An innovative technology that addresses this problem is high-density nucleic acid programmable protein arrays (HD-NAPPA) (Takulapalli et al., 2012). In this technique, the printing mix is deposited in physically isolated nanowells etched on a silicon surface on the slide. This allows both the protein expression and display and the immune reaction with serum antibodies to occur in these nanowells without diffusion to adjacent spots. This significantly enhances the signal-

to-noise ratio and may allow detection of weak signals over background. Future studies evaluating potential biomarkers for cervical disease may need to use this platform as a discovery tool.

Another limitation, as demonstrated here, is that the HPV-specific antibody response in cervical disease is weak and even absent in some women (Carter et al., 2000), especially compared with that detected in HPVOPC (K. S. Anderson, K. R. Dahlstrom, et al., 2015; K. S. Anderson, J. E. Gerber, et al., 2015). This is not surprising since the cervix acts as a protective niche for the virus and the host can remain ignorant of its presence for years. Thus, for discovery, nanowell arrays could help detect weak signals with the improved signal-to-noise ratio. Additionally, an important requirement for a POC test with a select panel of biomarkers for cervical disease is high analytical sensitivity to ensure reproducibility of the biomarker performance in a POC setting. This could be achieved by the use of fluorescent-based biorecognition instead of colorimetry (L. Lee, Nordman, Johnson, & Oldham, 2013; J. Smith et al., 2015), where a bright light source is used to actively interrogate the biorecognition site and the emitted fluorescent signal is detected electronically. There is a need for the development and optimization of such platform for use at low cost in a LMIC setting. Eventually, the serologic immune response could be useful in combination with other tests or as an initial screening tool in low-resource environments.

The detection of pre-existing B cell and T cell immunity to the most widely used nuclease ortholog of the CRISPR/Cas9 tool in a significant proportion of healthy humans confirms previous studies in mice (Chew et al., 2016; D. Wang et al., 2015) and sheds light on the need for more studies of the immunological risks of this system. More studies are needed to explore the immune response in other HLA haplotypes and other Cas9 orthologs. Whether Cas9-specific B cell or T cell immunity impact the efficacy or safety of Cas9 gene

delivery remains to be seen. Unlike the use of Cas9 for gene editing, which may only require Cas9 presence in cells for a few hours, current techniques for CRISPR-based epigenetic therapies require longer term expression of Cas9 *in vivo*, possibly for weeks and months (Liao et al., 2017; Zheng et al., 2018). This will likely engage memory T cell responses which could have implications for safety and efficacy. Studies that find associations with the identified biomarkers of Cas9-specific immunity and clinical consequences are needed for the biomarkers to have clinical utility. The use of CRISPR/Cas9 in humans may eventually necessitate creating HLA type-specific Cas9 variants, particularly for applications that require long-term Cas9 expression.

The top binding T cell epitopes within Cas9 that are most promiscuous for common HLA class I and class II alleles have been recently predicted *in silico* using IEDB (Chew, 2018). However, this is the first study that experimentally validates predicted immunodominant epitopes. None of the epitopes we report overlap with the peptides previously predicted (Chew, 2018). This is not unsurprising since we restricted our analysis to one HLA haplotype. Additionally, improved algorithms are needed to predict epitopes that hold up in experimental validation, as we show here. The use of CRISPR/Cas9 in humans may eventually necessitate creating HLA type-specific Cas9 variants, particularly for applications that require long-term Cas9 expression.

Generating partially immune-silenced Cas9 can be an attractive strategy to reduce the immune response to Cas9 particularly in patients where systemic immunosuppression will be contraindicated. Non-specific localized immune suppressive approaches, such as those used by tumor cells and some viruses may complement these strategies for complete deimmunization. One attractive strategy is the transient and inducible co-expression of programmed death-ligand 1 (PD-L1) or Indoleamine 2,3-Dioxygenase 1 (IDO1) activating gRNAs inside cells that express Cas9 to protect them against attack by

T cells. Alternatively, antigen presentation can be blocked by viral proteins interfering with antigen presentation (VIPRs), such as the adenoviral E319K or US2 and US11 from the human cytomegalovirus (Yewdell & Hill, 2002) or molecules that inhibit proteasomal antigen processing such as the Epstein-Barr virus Gly-Ala repeat (Levitskaya et al., 1995). We anticipate that deimmunized Cas9 will be useful in reduction of the dosage of the other immunomodulatory measures needed to be co-administered in patients, which could thus facilitate therapeutic CRISPR applications as we develop better understanding of the immunological consequences of this system.

REFERENCES

Achour, M., Zeghal, D., Kochbati, L., Kahla, S., Zouari, F., Maalej, M., & Oueslati, R. (2009). Antibody response for L1, E6, E7 HPV 16, and HPV 18 antigens in Tunisian women with cervical cancer. *J Immunoassay Immunochem, 30*(1), 82-96. doi:10.1080/15321810802569543

Ahi, Y. S., Bangari, D. S., & Mittal, S. K. (2011). Adenoviral Vector Immunity: Its Implications and circumvention strategies. *Current gene therapy, 11*(4), 307-320.

Aldhamen, Y. A., & Amalfitano, A. (2016). 16 - Methods to Mitigate Immune Responses to Adenoviral Vectors A2 - Curiel, David T *Adenoviral Vectors for Gene Therapy (Second Edition)* (pp. 391-422). San Diego: Academic Press.

Allegra, C. J., Jessup, J. M., Somerfield, M. R., Hamilton, S. R., Hammond, E. H., Hayes, D. F., . . . Schilsky, R. L. (2009). American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clin Oncol, 27*(12), 2091-2096. doi:10.1200/JCO.2009.21.9170

American Cancer Society: Cancer Facts and Figures 2018. (2018). Retrieved from https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html

Amon, L. M., Pitteri, S. J., Li, C. I., McIntosh, M., Ladd, J. J., Disis, M., . . . Hanash, S. M. (2012). Concordant release of glycolysis proteins into the plasma preceding a diagnosis of ER+ breast cancer. *Cancer Res, 72*(8), 1935-1942. doi:10.1158/0008-5472.CAN-11-3266

Anderson, K. S. (2011). Multiplexed Detection of Antibodies using Programmable Bead Arrays *Protein Microarray for Disease Analysis* (pp. 227-238): Springer.

Anderson, K. S., Cramer, D. W., Sibani, S., Wallstrom, G., Wong, J., Park, J., . . . LaBaer, J. (2015). Autoantibody signature for the serologic detection of ovarian cancer. *J Proteome Res, 14*(1), 578-586. doi:10.1021/pr500908n

Anderson, K. S., Cramer, D. W., Sibani, S., Wallstrom, G., Wong, J., Park, J., . . . LaBaer, J. (2015). Autoantibody Signature for the Serologic Detection of Ovarian Cancer. *Journal of proteome research, 14*(1), 578-586. doi:10.1021/pr500908n

Anderson, K. S., Dahlstrom, K. R., Cheng, J. N., Alam, R., Li, G., Wei, Q., . . . Sturgis, E. M. (2015). HPV16 antibodies as risk factors for oropharyngeal cancer and their association with tumor HPV and smoking status. *Oral Oncol, 51*(7), 662-667. doi:10.1016/j.oraloncology.2015.04.011

Anderson, K. S., Gerber, J. E., D'Souza, G., Pai, S. I., Cheng, J. N., Alam, R., . . . Posner, M. (2015). Biologic predictors of serologic responses to HPV in

oropharyngeal cancer: The HOTSPOT study. *Oral Oncol, 51*(8), 751-758. doi:10.1016/j.oraloncology.2015.05.007

Anderson, K. S., & LaBaer, J. (2005). The sentinel within: exploiting the immune system for cancer biomarkers. *J Proteome Res, 4*(4), 1123-1133. doi:10.1021/pr0500814

Anderson, K. S., Ramachandran, N., Wong, J., Raphael, J. V., Hainsworth, E., Demirkan, G., . . . Harris, L. (2008). Application of protein microarrays for multiplexed detection of antibodies to tumor antigens in breast cancer. *Journal of proteome research, 7*(4), 1490-1499.

Anderson, K. S., Sibani, S., Wallstrom, G., Qiu, J., Mendoza, E. A., Raphael, J., . . . Labaer, J. (2011). Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *Journal of proteome research, 10*(1), 85-96. doi:10.1021/pr100686b

Anderson, K. S., Wong, J., D'Souza, G., Riemer, A. B., Lorch, J., Haddad, R., . . . Posner, M. (2011). Serum antibodies to the HPV16 proteome as biomarkers for head and neck cancer. *Br J Cancer, 104*(12), 1896-1905. doi:10.1038/bjc.2011.171

Anderson, K. S., Wong, J., Polyak, K., Aronzon, D., & Enerback, C. (2009). Detection of psoriasin/S100A7 in the sera of patients with psoriasis. *The British journal of dermatology, 160*(2), 325-332. doi:10.1111/j.1365-2133.2008.08904.x

Anderson, K. S., Wong, J., Vitonis, A., Crum, C. P., Sluss, P. M., LaBaer, J., & Cramer, D. (2010). p53 autoantibodies as potential detection and prognostic biomarkers in serous ovarian cancer. *Cancer Epidemiology Biomarkers & Prevention, 19*(3), 859-868.

Andorfer, C. A., Necela, B. M., Thompson, E. A., & Perez, E. A. (2011). MicroRNA signatures: clinical biomarkers for the diagnosis and treatment of breast cancer. *Trends Mol Med, 17*(6), 313-319. doi:10.1016/j.molmed.2011.01.006

Apostolou, P., & Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. *Breast Cancer : Targets and Therapy, 9*, 331-335. doi:10.2147/bctt.s111394

Arbyn, M., Ronco, G., Anttila, A., Meijer, C. J., Poljak, M., Ogilvie, G., . . . Peto, J. (2012). Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. *Vaccine, 30 Suppl 5*, F88-99. doi:10.1016/j.vaccine.2012.06.095

Arias-Pulido, H., Peyton, C. L., Joste, N. E., Vargas, H., & Wheeler, C. M. (2006). Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. *J Clin Microbiol, 44*(5), 1755-1762. doi:10.1128/JCM.44.5.1755-1762.2006

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., . . . Zeger, S. L. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics, 69*(3), 89-95. doi:doi:10.1067/mcp.2001.113989

Avraham, A., Uhlmann, R., Shperber, A., Birnbaum, M., Sandbank, J., Sella, A., . . . Evron, E. (2012). Serum DNA methylation for monitoring response to neoadjuvant chemotherapy in breast cancer patients. *Int J Cancer, 131*(7), E1166-1172. doi:10.1002/ijc.27526

Bang, Y. J., Van Cutsem, E., Feyereislova, A., Chung, H. C., Shen, L., Sawaki, A., . . . Kang, Y. K. (2010). Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet, 376*(9742), 687-697. doi:10.1016/S0140-6736(10)61121-X

Bardelli, A., & Pantel, K. (2017). Liquid Biopsies, What We Do Not Know (Yet). *Cancer Cell, 31*(2), 172-179. doi:10.1016/j.ccell.2017.01.002

Bartel, M., Schaffer, D., & Buning, H. (2011). Enhancing the Clinical Potential of AAV Vectors by Capsid Engineering to Evade Pre-Existing Immunity. *Front Microbiol, 2*, 204. doi:10.3389/fmicb.2011.00204

Beghetto, E., De Paolis, F., Montagnani, F., Cellesi, C., & Gargano, N. (2009). Discovery of new Mycoplasma pneumoniae antigens by use of a whole-genome lambda display library. *Microbes and Infection, 11*(1), 66-73.

Beghetto, E., De Paolis, F., Spadoni, A., Del Porto, P., Buffolano, W., & Gargano, N. (2008). Molecular dissection of the human B cell response against cytomegalovirus infection by lambda display. *Journal of virological methods, 151*(1), 7-14.

Beghetto, E., Gargano, N., Ricci, S., Garufi, G., Peppoloni, S., Montagnani, F., . . . Felici, F. (2006). Discovery of novel Streptococcus pneumoniae antigens by screening a whole-genome λ-display library. *FEMS microbiology letters, 262*(1), 14-21.

Behl, A. S., Goddard, K. A., Flottemesch, T. J., Veenstra, D., Meenan, R. T., Lin, J. S., & Maciosek, M. V. (2012). Cost-effectiveness analysis of screening for KRAS and BRAF mutations in metastatic colorectal cancer. *JNCI: Journal of the National Cancer Institute, 104*(23), 1785-1795.

Bell, I., Martin, A., & Roberts, S. (2007). The E1^E4 protein of human papillomavirus interacts with the serine-arginine-specific protein kinase SRPK1. *J Virol, 81*(11), 5437-5448. doi:10.1128/JVI.02609-06

Bertoli, G., Cava, C., & Castiglioni, I. (2015). MicroRNAs: New Biomarkers for Diagnosis, Prognosis, Therapy Prediction and Therapeutic Tools for Breast Cancer. *Theranostics, 5*(10), 1122-1143. doi:10.7150/thno.11543

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., . . . Diaz, L. A., Jr. (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med, 6*(224), 224ra224. doi:10.1126/scitranslmed.3007094

Bevers, T. B., Anderson, B. O., Bonaccio, E., Buys, S., Daly, M. B., Dempsey, P. J., . . . Yankeelov, T. (2009). NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. *Journal of the National Comprehensive Cancer Network : JNCCN, 7*(10), 1060-1096.

Bian, X., Wasserfall, C., Wallstrom, G., Wang, J., Wang, H., Barker, K., . . . LaBaer, J. (2016). Tracking the Antibody Immunome in Type 1 Diabetes Using Protein Arrays. *J Proteome Res*. doi:10.1021/acs.jproteome.6b00354

Bidard, F. C., Fehm, T., Ignatiadis, M., Smerage, J. B., Alix-Panabieres, C., Janni, W., . . . Pierga, J. Y. (2013). Clinical application of circulating tumor cells in breast cancer: overview of the current interventional trials. *Cancer metastasis reviews, 32*(1-2), 179-188. doi:10.1007/s10555-012-9398-0

Bidard, F. C., Mathiot, C., Degeorges, A., Etienne-Grimaldi, M. C., Delva, R., Pivot, X., . . . Pierga, J. Y. (2010). Clinical value of circulating endothelial cells and circulating tumor cells in metastatic breast cancer patients treated first line with bevacizumab and chemotherapy. *Ann Oncol, 21*(9), 1765-1771. doi:10.1093/annonc/mdq052

Bidard, F. C., Peeters, D. J., Fehm, T., Nole, F., Gisbert-Criado, R., Mavroudis, D., . . . Michiels, S. (2014). Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. *Lancet Oncol, 15*(4), 406-414. doi:10.1016/S1470-2045(14)70069-5

Bierl, C., Karem, K., Poon, A. C., Swan, D., Tortolero-Luna, G., Follen, M., . . . Reeves, W. C. (2005). Correlates of cervical mucosal antibodies to human papillomavirus 16: results from a case control study. *Gynecol Oncol, 99*(3 Suppl 1), S262-268. doi:10.1016/j.ygyno.2005.07.100

Biomarkers and Risk Assessment: Concepts and Principles. (1993). Retrieved from http://www.inchem.org/documents/ehc/ehc/ehc155.htm

Board, R. E., Wardley, A. M., Dixon, J. M., Armstrong, A. C., Howell, S., Renshaw, L., . . . Dive, C. (2010). Detection of PIK3CA mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Res Treat, 120*(2), 461-467. doi:10.1007/s10549-010-0747-9

Boersema, P. J., Geiger, T., Wisniewski, J. R., & Mann, M. (2013). Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Molecular & cellular proteomics : MCP, 12*(1), 158-171. doi:10.1074/mcp.M112.023614

Bohm, D., Keller, K., Wehrwein, N., Lebrecht, A., Schmidt, M., Kolbl, H., & Grus, F. H. (2011). Serum proteome profiling of primary breast cancer indicates a specific biomarker profile. *Oncol Rep, 26*(5), 1051-1056. doi:10.3892/or.2011.1420

Bosch, F. X., Burchell, A. N., Schiffman, M., Giuliano, A. R., de Sanjose, S., Bruni, L., . . . Munoz, N. (2008). Epidemiology and natural history of human papillomavirus infections and type-specific implications in cervical neoplasia. *Vaccine, 26 Suppl 10*, K1-16. doi:10.1016/j.vaccine.2008.05.064

Bosch, F. X., & de Sanjose, S. (2003). Chapter 1: Human papillomavirus and cervical cancer--burden and assessment of causality. *J Natl Cancer Inst Monogr*(31), 3-13.

Boulet, G., Horvath, C., Broeck, D. V., Sahebali, S., & Bogers, J. (2007). Human papillomavirus: E6 and E7 oncogenes. *The International Journal of Biochemistry & Cell Biology, 39*(11), 2006-2011. doi:https://doi.org/10.1016/j.biocel.2007.07.004

Boutin, S., Monteilhet, V., Veron, P., Leborgne, C., Benveniste, O., Montus, M. F., & Masurier, C. (2010). Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Hum Gene Ther, 21*(6), 704-712. doi:10.1089/hum.2009.182

Brauer, H. A., D'Arcy, M., Libby, T. E., Thompson, H. J., Yasui, Y. Y., Hamajima, N., . . . Lampe, P. D. (2014). Dermcidin expression is associated with disease progression and survival among breast cancer patients. *Breast Cancer Res Treat, 144*(2), 299-306. doi:10.1007/s10549-014-2880-3

Breast Cancer Risk Assessment Tool. Retrieved from http://www.cancer.gov/bcrisktool/

Brunetti-Pierri, N., & Ng, P. (2009). Progress towards liver and lung-directed gene therapy with helper-dependent adenoviral vectors. *Curr Gene Ther, 9*(5), 329-340.

Buas, M. F., Rho, J. H., Chai, X., Zhang, Y., Lampe, P. D., & Li, C. I. (2015). Candidate early detection protein biomarkers for ER+/PR+ invasive ductal breast carcinoma identified using pre-clinical plasma from the WHI observational study. *Breast Cancer Res Treat, 153*(2), 445-454. doi:10.1007/s10549-015-3554-5

Calzone, K. A. (2012). Genetic biomarkers of cancer risk. *Semin Oncol Nurs, 28*(2), 122-128. doi:10.1016/j.soncn.2012.03.007

Cantor, J. R., Yoo, T. H., Dixit, A., Iverson, B. L., Forsthuber, T. G., & Georgiou, G. (2011). Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. *Proc Natl Acad Sci U S A, 108*(4), 1272-1277. doi:10.1073/pnas.1014739108

Carapetis, J. R., Steer, A. C., Mulholland, E. K., & Weber, M. (2005). The global burden of group A streptococcal diseases. *Lancet Infect Dis, 5*(11), 685-694. doi:10.1016/S1473-3099(05)70267-X

Carter, J. J., Koutsky, L. A., Hughes, J. P., Lee, S. K., Kuypers, J., Kiviat, N., & Galloway, D. A. (2000). Comparison of human papillomavirus types 16, 18, and 6 capsid antibody responses following incident infection. *J Infect Dis, 181*(6), 1911-1919. doi:10.1086/315498

Castellsague, X., Pawlita, M., Roura, E., Margall, N., Waterboer, T., Bosch, F. X., . . . Riboli, E. (2014). Prospective seroepidemiologic study on the role of Human Papillomavirus and other infections in cervical carcinogenesis: evidence from the EPIC cohort. *Int J Cancer, 135*(2), 440-452. doi:10.1002/ijc.28665

Castle, J., Shaker, H., Morris, K., Tugwood, J. D., & Kirwan, C. C. (2014). The significance of circulating tumour cells in breast cancer: A review. *The Breast, 23*(5), 552-560. doi:http://dx.doi.org/10.1016/j.breast.2014.07.002

Catalona, W. J., Smith, D. S., Ratliff, T. L., & Basler, J. W. (1993). Detection of organ-confined prostate cancer is increased through prostate-specific antigen—based screening. *JAMA, 270*(8), 948-954.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., . . . Fischer, A. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science, 288*(5466), 669-672.

Chan, D. S., Bandera, E. V., Greenwood, D. C., & Norat, T. (2015). Circulating C-Reactive Protein and Breast Cancer Risk-Systematic Literature Review and Meta-analysis of Prospective Cohort Studies. *Cancer Epidemiol Biomarkers Prev, 24*(10), 1439-1449. doi:10.1158/1055-9965.EPI-15-0324

Chapman, C., Murray, A., Chakrabarti, J., Thorpe, A., Woolston, C., Sahin, U., . . . Robertson, J. (2007). Autoantibodies in breast cancer: their use as an aid to early diagnosis. *Ann Oncol, 18*(5), 868-873. doi:10.1093/annonc/mdm007

Chatterjee, M., Mohapatra, S., Ionan, A., Bawa, G., Ali-Fehmi, R., Wang, X., . . . Tainsky, M. A. (2006). Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. *Cancer Res, 66*(2), 1181-1190. doi:10.1158/0008-5472.CAN-04-2962

Chen, G., Wang, X., Yu, J., Varambally, S., Thomas, D. G., Lin, M. Y., . . . Chinnaiyan, A. M. (2007). Autoantibody profiles reveal ubiquilin 1 as a humoral immune response target in lung adenocarcinoma. *Cancer research, 67*(7), 3461-3467. doi:10.1158/0008-5472.CAN-06-4475

Chen, J., Liu, Q. J., Wang, D., Zhou, X. Y., Xiong, D., Li, H. J., & Li, C. L. (2014). Hepatoma upregulated protein expression is involved in the pathogenesis of human breast carcinogenesis. *Oncol Lett, 8*(6), 2543-2548. doi:10.3892/ol.2014.2614

Chen, S., & Parmigiani, G. (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol, 25*(11), 1329-1333. doi:10.1200/JCO.2006.09.1066

Cheng, J., Liu, S., & Zhang, F. F. (2015). Circulating Biomarkers in Breast Cancer Prognosis and Survival: A Systematic Review. *The FASEB Journal, 29*(1 Supplement).

Chew, W. L. (2018). Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdiscip Rev Syst Biol Med, 10*(1). doi:10.1002/wsbm.1408

Chew, W. L., Tabebordbar, M., Cheng, J. K., Mali, P., Wu, E. Y., Ng, A. H., . . . Church, G. M. (2016). A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods, 13*(10), 868-874. doi:10.1038/nmeth.3993

Chirgwin, K. D., Feldman, J., Augenbraun, M., Landesman, S., & Minkoff, H. (1995). Incidence of venereal warts in human immunodeficiency virus-infected and uninfected women. *J Infect Dis, 172*(1), 235-238.

Choi, M., Kipps, T., & Kurzrock, R. (2016). ATM Mutations in Cancer: Therapeutic Implications. *Molecular Cancer Therapeutics, 15*(8), 1781-1791. doi:10.1158/1535-7163.mct-15-0945

Chou, C. P., Lewin, J. M., Chiang, C. L., Hung, B. H., Yang, T. L., Huang, J. S., . . . Pan, H. B. (2015). Clinical evaluation of contrast-enhanced digital mammography and contrast enhanced tomosynthesis-Comparison to contrast-enhanced breast MRI. *European journal of radiology*. doi:10.1016/j.ejrad.2015.09.019

Chowell, D., Krishna, S., Becker, P. D., Cocita, C., Shu, J., Tan, X., . . . Anderson, K. S. (2015). TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes. *Proc Natl Acad Sci U S A, 112*(14), E1754-1762. doi:10.1073/pnas.1500973112

ClinicalTrials. (2018). NY-ESO-1-redirected CRISPR (TCRendo and PD1) Edited T Cells (NYCE T Cells). Retrieved from https://clinicaltrials.gov/ct2/show/NCT03399448

Coleman, N., Birley, H. D., Renton, A. M., Hanna, N. F., Ryait, B. K., Byrne, M., . . . Stanley, M. A. (1994). Immunological events in regressing genital warts. *Am J Clin Pathol, 102*(6), 768-774.

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793-795.

Combes, J. D., Pawlita, M., Waterboer, T., Hammouda, D., Rajkumar, T., Vanhems, P., . . . Clifford, G. (2014). Antibodies against high-risk human papillomavirus proteins as markers for invasive cervical cancer. *Int J Cancer, 135*(10), 2453-2461. doi:10.1002/ijc.28888

Comprehensive molecular portraits of human breast tumours. (2012, Oct 4). *Nature,* pp. 61-70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23000897

Crosbie, E. J., Einstein, M. H., Franceschi, S., & Kitchener, H. C. (2013). Human papillomavirus and cervical cancer. *Lancet, 382*(9895), 889-899. doi:10.1016/S0140-6736(13)60022-7

Crum, C. P., Beach, K. J., Hedley, M. L., Yuan, L., Lee, K. R., Wright, T. C., & Urban, R. G. (2004). Dynamics of human papillomavirus infection between biopsy and excision of cervical intraepithelial neoplasia: results from the ZYC101a protocol. *J Infect Dis, 189*(8), 1348-1354. doi:10.1086/382956

Cyranoski, D. (2016). Chinese scientists to pioneer first human CRISPR trial. *Nature News, 535*(7613), 476.

D'Souza, G., Gross, N. D., Pai, S. I., Haddad, R., Anderson, K. S., Rajan, S., . . . Posner, M. R. (2014). Oral human papillomavirus (HPV) infection in HPV-positive patients with oropharyngeal cancer and their partners. *J Clin Oncol, 32*(23), 2408-2415. doi:10.1200/JCO.2014.55.1341

D'Souza, G., Kreimer, A. R., Viscidi, R., Pawlita, M., Fakhry, C., Koch, W. M., . . . Gillison, M. L. (2007). Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med, 356*(19), 1944-1956. doi:10.1056/NEJMoa065497

Dash, R. C., Robb, J. A., Booker, D. L., Foo, W. C., Witte, D. L., & Bry, L. (2012). Biospecimens and biorepositories for the community pathologist. *Arch Pathol Lab Med, 136*(6), 668-678. doi:10.5858/arpa.2011-0274-SO

Davies, C., Godwin, J., Gray, R., Clarke, M., Cutter, D., Darby, S., . . . Peto, R. (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet, 378*(9793), 771-784. doi:10.1016/S0140-6736(11)60993-8

Dawson, S. J., Tsui, D. W., Murtaza, M., Biggs, H., Rueda, O. M., Chin, S. F., . . . Rosenfeld, N. (2013). Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med, 368*(13), 1199-1209. doi:10.1056/NEJMoa1213261

de Jong, A., van der Burg, S. H., Kwappenberg, K. M., van der Hulst, J. M., Franken, K. L., Geluk, A., . . . Offringa, R. (2002). Frequent detection of human papillomavirus 16 E2-specific T-helper immunity in healthy subjects. *Cancer Res, 62*(2), 472-479.

de Jong, A., van Poelgeest, M. I., van der Hulst, J. M., Drijfhout, J. W., Fleuren, G. J., Melief, C. J., . . . van der Burg, S. H. (2004). Human papillomavirus type 16-positive cervical cancer is associated with impaired CD4+ T-cell immunity against early antigens E2 and E6. *Cancer Res, 64*(15), 5449-5455. doi:10.1158/0008-5472.CAN-04-0831

de la Rica, R., & Stevens, M. M. (2012). Plasmonic ELISA for the ultrasensitive detection of disease biomarkers with the naked eye. *Nat Nano, 7*(12), 821-824.

doi:http://www.nature.com/nnano/journal/v7/n12/abs/nnano.2012.186.html#supplementary-information

Dillner, J. (1999). The serological response to papillomaviruses. *Semin Cancer Biol, 9*(6), 423-430. doi:10.1006/scbi.1999.0146

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Domchek, S. M., Friebel, T. M., Singer, C. F., Evans, D. G., Lynch, H. T., Isaacs, C., . . . Rebbeck, T. R. (2010). Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA, 304*(9), 967-975. doi:10.1001/jama.2010.1237

Doorbar, J. (2006). Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond), 110*(5), 525-541. doi:10.1042/CS20050369

Doorbar, J. (2013). The E4 protein; structure, function and patterns of expression. *Virology, 445*(1-2), 80-98. doi:10.1016/j.virol.2013.07.008

Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., & Stanley, M. A. (2012). The biology and life-cycle of human papillomaviruses. *Vaccine, 30 Suppl 5*, F55-70. doi:10.1016/j.vaccine.2012.06.083

Duffy, M. J., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., & Cardoso, F. (2017). Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *European Journal of Cancer, 75*, 284-298. doi:https://doi.org/10.1016/j.ejca.2017.01.017

Dunne, E. F., Unger, E. R., Sternberg, M., McQuillan, G., Swan, D. C., Patel, S. S., & Markowitz, L. E. (2007). Prevalence of HPV infection among females in the United States. *JAMA, 297*(8), 813-819. doi:10.1001/jama.297.8.813

The Early Detection Research Network. Retrieved from http://edrn.nci.nih.gov/

Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., & Church, G. M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods, 10*(11), 1116-1121. doi:10.1038/nmeth.2681

Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., . . . Hartwell, L. (2003). The case for early detection. *Nat Rev Cancer, 3*(4), 243-252. doi:10.1038/nrc1041

Evans, M., Borysiewicz, L. K., Evans, A. S., Rowe, M., Jones, M., Gileadi, U., . . . Man, S. (2001). Antigen processing defects in cervical carcinomas limit the presentation of a CTL epitope from human papillomavirus 16 E6. *J Immunol, 167*(9), 5420-5428.

Ewaisha, R., Meshay, I., Resnik, J., Katchman, B. A., & Anderson, K. S. (2016). Programmable protein arrays for immunoprofiling HPV-associated cancers. *Proteomics, 16*(8), 1215-1224. doi:10.1002/pmic.201500376

Fakih, M. G., & Padmanabhan, A. (2006). CEA monitoring in colorectal cancer. *ONCOLOGY-WILLISTON PARK THEN HUNTINGTON THE MELVILLE NEW YORK-, 20*(6), 579.

FDA, U. (2005). Guidance for industry - pharmacogenomic data submissions. Retrieved from https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm079849.pdf

Feng, Z., Kagan, J., Pepe, M., Thornquist, M., Ann Rinaudo, J., Dahlgren, J., . . . Srivastava, S. (2013). The Early Detection Research Network's Specimen reference sets: paving the way for rapid evaluation of potential biomarkers. *Clin Chem, 59*(1), 68-74. doi:10.1373/clinchem.2012.185140

Fennema, J. S., van Ameijden, E. J., Coutinho, R. A., & van den Hoek, A. A. (1995). HIV, sexually transmitted diseases and gynaecologic disorders in women: increased risk for genital herpes and warts among HIV-infected prostitutes in Amsterdam. *AIDS, 9*(9), 1071-1078.

Ferguson, N. M., Rodríguez-Barraquer, I., Dorigatti, I., Mier-y-Teran-Romero, L., Laydon, D. J., & Cummings, D. A. (2016). Benefits and risks of the Sanofi-Pasteur dengue vaccine: Modeling optimal deployment. *Science, 353*(6303), 1033-1036.

Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer, 127*(12), 2893-2917. doi:10.1002/ijc.25516

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., . . . Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer, 136*(5), E359-386. doi:10.1002/ijc.29210

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., . . . Bray, F. (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Retrieved from http://globocan.iarc.fr

Festa, F., Rollins, S. M., Vattem, K., Hathaway, M., Lorenz, P., Mendoza, E. A., . . . Jensen, P. (2013). Robust microarray production of freshly expressed proteins in a human milieu. *PROTEOMICS-Clinical Applications, 7*(5-6), 372-377.

Festa, F., Rollins, S. M., Vattem, K., Hathaway, M., Lorenz, P., Mendoza, E. A., . . . LaBaer, J. (2013). Robust microarray production of freshly expressed proteins in a human milieu. *Proteomics Clin Appl, 7*(5-6), 372-377. doi:10.1002/prca.201200063

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Frontiers in Immunology, 8*, 278. doi:10.3389/fimmu.2017.00278

Fruchter, R. G., Maiman, M., Sedlis, A., Bartley, L., Camilien, L., & Arrastia, C. D. (1996). Multiple recurrences of cervical intraepithelial neoplasia in women with the human immunodeficiency virus. *Obstet Gynecol, 87*(3), 338-344.

Gao, G., Wang, Q., Calcedo, R., Mays, L., Bell, P., Wang, L., . . . Wilson, J. M. (2009). Adeno-associated virus-mediated gene transfer to nonhuman primate liver can elicit destructive transgene-specific T cell responses. *Hum Gene Ther, 20*(9), 930-942. doi:10.1089/hum.2009.060

Garcia-Closas, M., Couch, F. J., Lindstrom, S., Michailidou, K., Schmidt, M. K., Brook, M. N., . . . Kraft, P. (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature genetics, 45*(4), 392-398, 398e391-392. doi:10.1038/ng.2561

Garcia-Murillas, I., Schiavon, G., Weigelt, B., Ng, C., Hrebien, S., Cutts, R. J., . . . Turner, N. C. (2015). Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med, 7*(302), 302ra133. doi:10.1126/scitranslmed.aab0021

Gaspar, H. B., Parsley, K. L., Howe, S., King, D., Gilmour, K. C., Sinclair, J., . . . Thrasher, A. J. (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet, 364*(9452), 2181-2187. doi:10.1016/S0140-6736(04)17590-9

Ghim, S., Newsome, J., Bell, J., Sundberg, J. P., Schlegel, R., & Jenson, A. B. (2000). Spontaneously regressing oral papillomas induce systemic antibodies that neutralize canine oral papillomavirus. *Exp Mol Pathol, 68*(3), 147-151. doi:10.1006/exmp.1999.2298

Gilligan, T. D., Seidenfeld, J., Basch, E. M., Einhorn, L. H., Fancher, T., Smith, D. C., . . . Hayes, D. F. (2010). American Society of Clinical Oncology Clinical Practice Guideline on Uses of Serum Tumor Markers in Adult Males With Germ Cell Tumors. *Journal of Clinical Oncology, 28*(20), 3388-3404. doi:10.1200/jco.2009.26.4481

Gillison, M. L., D'Souza, G., Westra, W., Sugar, E., Xiao, W., Begum, S., & Viscidi, R. (2008). Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst, 100*(6), 407-420. doi:10.1093/jnci/djn025

Gingras, I., Salgado, R., & Ignatiadis, M. (2015). Liquid biopsy: will it be the 'magic tool' for monitoring response of solid tumors to anticancer therapies? *Curr Opin Oncol, 27*(6), 560-567. doi:10.1097/CCO.0000000000000223

Giuliano, M., Giordano, A., Jackson, S., De Giorgi, U., Mego, M., Cohen, E. N., . . . Cristofanilli, M. (2014). Circulating tumor cells as early predictors of metastatic spread in breast cancer patients with limited metastatic dissemination. *Breast Cancer Res, 16*(5), 440. doi:10.1186/s13058-014-0440-8

Gonzalez, R. M., Daly, D. S., Tan, R., Marks, J. R., & Zangar, R. C. (2011). Plasma biomarker profiles differ depending on breast cancer subtype but RANTES is consistently increased. *Cancer Epidemiol Biomarkers Prev, 20*(7), 1543-1551. doi:10.1158/1055-9965.EPI-10-1248

Goodell, V., Waisman, J., Salazar, L. G., de la Rosa, C., Link, J., Coveler, A. L., . . . Disis, M. L. (2008). Level of HER-2/neu protein expression in breast cancer may affect the development of endogenous HER-2/neu-specific immunity. *Mol Cancer Ther, 7*(3), 449-454. doi:10.1158/1535-7163.MCT-07-0386

Goodman, A. (2015). HPV testing as a screen for cervical cancer. *BMJ, 350*, h2372. doi:10.1136/bmj.h2372

Goodnow, C. C., Sprent, J., de St Groth, B. F., & Vinuesa, C. G. (2005). Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature, 435*, 590. doi:10.1038/nature03724

Goufman, E. I., Iakovlev, V. N., Tikhonova, N. B., & Lokshin, A. E. (2015). Quantification of autoantibodies to plasminogen in plasma of patients with cancer. *Cancer Biomark, 15*(3), 281-287. doi:10.3233/CBM-150469

Griffin, H., Wu, Z., Marnane, R., Dewar, V., Molijn, A., Quint, W., . . . Doorbar, J. (2012). E4 antibodies facilitate detection and type-assignment of active HPV infection in cervical disease. *PLoS One, 7*(12), e49974. doi:10.1371/journal.pone.0049974

Gromov, P., Gromova, I., Bunkenborg, J., Cabezon, T., Moreira, J. M., Timmermans-Wielenga, V., . . . Celis, J. E. (2010). Up-regulated proteins in the fluid bathing the tumour cell microenvironment as potential serological markers for early detection of cancer of the breast. *Mol Oncol, 4*(1), 65-89. doi:10.1016/j.molonc.2009.11.003

Gunter, M. J., Wang, T., Cushman, M., Xue, X., Wassertheil-Smoller, S., Strickler, H. D., . . . Ho, G. Y. (2015). Circulating Adipokines and Inflammatory Markers and Postmenopausal Breast Cancer Risk. *J Natl Cancer Inst, 107*(9). doi:10.1093/jnci/djv169

Gutierrez-Xicotencatl, L., Salazar-Pina, D. A., Pedroza-Saavedra, A., Chihu-Amparan, L., Rodriguez-Ocampo, A. N., Maldonado-Gama, M., & Esquivel-Guadarrama, F. R. (2016). Humoral Immune Response Against Human Papillomavirus as Source of Biomarkers for the Prediction and Detection of Cervical Cancer. *Viral Immunol, 29*(2), 83-94. doi:10.1089/vim.2015.0087

Guzman, N., Agarwal, K., Asthagiri, D., Yu, L., Saji, M., Ringel, M. D., & Paulaitis, M. E. (2015). Breast Cancer-Specific miR Signature Unique to Extracellular Vesicles

Includes "microRNA-like" tRNA Fragments. *Mol Cancer Res, 13*(5), 891-901. doi:10.1158/1541-7786.MCR-14-0533

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J. P., . . . Cavazzana-Calvo, M. (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med, 346*(16), 1185-1193. doi:10.1056/NEJMoa012616

Halbert, C. L., Standaert, T. A., Wilson, C. B., & Miller, A. D. (1998). Successful readministration of adeno-associated virus vectors to the mouse lung requires transient immunosuppression during the initial exposure. *J Virol, 72*(12), 9795-9805.

Hall, S., Lorincz, A., Shah, F., Sherman, M. E., Abbas, F., Paull, G., . . . Shah, K. V. (1996). Human papillomavirus DNA detection in cervical specimens by hybrid capture: correlation with cytologic and histologic diagnoses of squamous intraepithelial lesions of the cervix. *Gynecol Oncol, 62*(3), 353-359. doi:10.1006/gyno.1996.0248

Harding, C., Pompei, F., Burmistrov, D., Welch, H. G., Abebe, R., & Wilson, R. (2015). Breast Cancer Screening, Incidence, and Mortality Across US Counties. *JAMA internal medicine*. doi:10.1001/jamainternmed.2015.3043

Harris, L. N., Ismaila, N., McShane, L. M., Andre, F., Collyar, D. E., Gonzalez-Angulo, A. M., . . . Hayes, D. F. (2016). Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *Journal of Clinical Oncology, 34*(10), 1134-1150. doi:10.1200/jco.2015.65.2289

Haslene-Hox, H., Tenstad, O., & Wiig, H. (2013). Interstitial fluid-a reflection of the tumor cell microenvironment and secretome. *Biochim Biophys Acta, 1834*(11), 2336-2346. doi:10.1016/j.bbapap.2013.01.028

Heard, I., Tassie, J. M., Schmitz, V., Mandelbrot, L., Kazatchkine, M. D., & Orth, G. (2000). Increased risk of cervical disease among human immunodeficiency virus-infected women with severe immunosuppression and high human papillomavirus load(1). *Obstet Gynecol, 96*(3), 403-409.

Herdman, M. T., Pett, M. R., Roberts, I., Alazawi, W. O., Teschendorff, A. E., Zhang, X. Y., . . . Coleman, N. (2006). Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. *Carcinogenesis, 27*(11), 2341-2353. doi:10.1093/carcin/bgl172

Higgins, M. J., Jelovac, D., Barnathan, E., Blair, B., Slater, S., Powers, P., . . . Park, B. H. (2012). Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. *Clin Cancer Res, 18*(12), 3462-3469. doi:10.1158/1078-0432.CCR-11-2696

Hirano, H., Gootenberg, J. S., Horii, T., Abudayyeh, O. O., Kimura, M., Hsu, P. D., . . . Nureki, O. (2016). Structure and Engineering of Francisella novicida Cas9. *Cell, 164*(5), 950-961. doi:10.1016/j.cell.2016.01.039

Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., . . . Nielsen, M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics, 61*(1), 1.

Howe, S. J., Mansour, M. R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., . . . Thrasher, A. J. (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest, 118*(9), 3143-3150. doi:10.1172/JCI35798

HPV Factsheet. (Jan 23 2014). *Centers for Disease Control and Prevention.* Retrieved from http://www.cdc.gov/std/hpv/hpv-factsheet-march-2014.pdf

Hu, Z., Dong, J., Wang, L. E., Ma, H., Liu, J., Zhao, Y., . . . Shen, H. (2012). Serum microRNA profiling and breast cancer risk: the use of miR-484/191 as endogenous controls. *Carcinogenesis, 33*(4), 828-834. doi:10.1093/carcin/bgs030

Hunter, D. J., Losina, E., Guermazi, A., Burstein, D., Lassere, M. N., & Kraus, V. (2010). A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. *Curr Drug Targets, 11*(5), 536-545.

Ibrahim, S. H., & Robertson, K. D. (2018). Use of the CRISPR/Cas9-Based Epigenetic Gene Activation System in Vivo: a New Potential Therapeutic Modality. *Hepatology*. doi:10.1002/hep.29860

Innocenti, F., & Ratain, M. J. (2006). Pharmacogenetics of irinotecan: clinical perspectives on the utility of genotyping. *Pharmacogenomics, 7*(8), 1211-1221. doi:10.2217/14622416.7.8.1211

Jackson, A. M., Boutell, J., Cooley, N., & He, M. (2004). Cell-free protein synthesis for proteomics. *Briefings in functional genomics & proteomics, 2*(4), 308-319.

Jansson, M. D., & Lund, A. H. (2012). MicroRNA and cancer. *Mol Oncol, 6*(6), 590-610. doi:10.1016/j.molonc.2012.09.006

Jemal, A., Clegg, L. X., Ward, E., Ries, L. A., Wu, X., Jamison, P. M., . . . Edwards, B. K. (2004). Annual report to the nation on the status of cancer, 1975-2001, with a special feature regarding survival. *Cancer, 101*(1), 3-27. doi:10.1002/cncr.20288

Jemal, A., Ward, E., & Thun, M. (2010). Declining death rates reflect progress against cancer. *PLoS One, 5*(3), e9584. doi:10.1371/journal.pone.0009584

Jeon, S., Allen-Hoffmann, B. L., & Lambert, P. F. (1995). Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol, 69*(5), 2989-2997.

Jiang, H., Couto, L. B., Patarroyo-White, S., Liu, T., Nagy, D., Vargas, J. A., . . . Pierce, G. F. (2006). Effects of transient immunosuppression on adenoassociated, virus-mediated, liver-directed gene transfer in rhesus macaques and implications for human gene therapy. *Blood, 108*(10), 3321-3328. doi:10.1182/blood-2006-04-017913

Jiang, Z. F., Cristofanilli, M., Shao, Z. M., Tong, Z. S., Song, E. W., Wang, X. J., . . . Zhang, M. (2013). Circulating tumor cells predict progression-free and overall survival in Chinese patients with metastatic breast cancer, HER2-positive or triple-negative (CBCSG004): a multicenter, double-blind, prospective trial. *Ann Oncol, 24*(11), 2766-2772. doi:10.1093/annonc/mdt246

Jin, H., Daly, D. S., Marks, J. R., & Zangar, R. C. (2013). Oxidatively modified proteins as plasma biomarkers in breast cancer. *Cancer Biomark, 13*(3), 193-200. doi:10.3233/CBM-130349

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., . . . Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science, 343*(6176), 1247997. doi:10.1126/science.1247997

Joosse, S. A., Muller, V., Steinbach, B., Pantel, K., & Schwarzenbach, H. (2014). Circulating cell-free cancer-testis MAGE-A RNA, BORIS RNA, let-7b and miR-202 in the blood of patients with breast cancer and benign breast diseases. *Br J Cancer, 111*(5), 909-917. doi:10.1038/bjc.2014.360

Jung, E. J., Santarpia, L., Kim, J., Esteva, F. J., Moretti, E., Buzdar, A. U., . . . Calin, G. A. (2012). Plasma microRNA 210 levels correlate with sensitivity to trastuzumab and tumor presence in breast cancer patients. *Cancer, 118*(10), 2603-2614. doi:10.1002/cncr.26565

Jung, K., Fleischhacker, M., & Rabien, A. (2010). Cell-free DNA in the blood as a solid tumor biomarker--a critical appraisal of the literature. *Clin Chim Acta, 411*(21-22), 1611-1624. doi:10.1016/j.cca.2010.07.032

Kalos, M. (2011). Biomarkers in T cell therapy clinical trials. *Journal of Translational Medicine, 9*, 138-138. doi:10.1186/1479-5876-9-138

Kang, W. D., Choi, H. S., & Kim, S. M. (2013). Is vaccination with quadrivalent HPV vaccine after loop electrosurgical excision procedure effective in preventing recurrence in patients with high-grade cervical intraepithelial neoplasia (CIN2–3)? *Gynecol Oncol, 130*(2), 264-268. doi:http://dx.doi.org/10.1016/j.ygyno.2013.04.050

Karanam, B., Jagu, S., Huh, W. K., & Roden, R. B. (2009). Developing vaccines against minor capsid antigen L2 to prevent papillomavirus infection. *Immunol Cell Biol, 87*(4), 287-299. doi:10.1038/icb.2009.13

Katayama, H., Boldt, C., Ladd, J. J., Johnson, M. M., Chao, T., Capello, M., . . . Hanash, S. (2015). An Autoimmune Response Signature Associated with the

Development of Triple-Negative Breast Cancer Reflects Disease Pathogenesis. *Cancer Res, 75*(16), 3246-3254. doi:10.1158/0008-5472.CAN-15-0248

Katchman, B. A., Barderas, R., Alam, R., Chowell, D., Field, M. S., Esserman, L. J., . . . Anderson, K. S. (2016). Proteomic mapping of p53 immunogenicity in pancreatic, ovarian, and breast cancers. *PROTEOMICS – Clinical Applications, 10*(7), 720-731. doi:10.1002/prca.201500096

Kay, M. A. (2011). State-of-the-art gene-based therapies: the road ahead. *Nat Rev Genet, 12*(5), 316-328. doi:10.1038/nrg2971

Kelly, H., Downing, T., Tuite, N. L., Smith, T. J., Kerin, M. J., Dwyer, R. M., . . . Reddington, K. (2015). Cross Platform Standardisation of an Experimental Pipeline for Use in the Identification of Dysregulated Human Circulating MiRNAs. *PLoS One, 10*(9), e0137389. doi:10.1371/journal.pone.0137389

Kim, E. H., & Andriole, G. L. (2015). Prostate-specific antigen-based screening: controversy and guidelines. *BMC Medicine, 13*(1), 61. doi:10.1186/s12916-015-0296-5

King, C., Garza, E. N., Mazor, R., Linehan, J. L., Pastan, I., Pepper, M., & Baker, D. (2014). Removing T-cell epitopes with computational protein design. *Proc Natl Acad Sci U S A, 111*(23), 8577-8582. doi:10.1073/pnas.1321126111

Kirsh, V. A., Chiarelli, A. M., Edwards, S. A., O'Malley, F. P., Shumak, R. S., Yaffe, M. J., & Boyd, N. F. (2011). Tumor characteristics associated with mammographic detection of breast cancer in the Ontario breast screening program. *Journal of the National Cancer Institute, 103*(12), 942-950. doi:10.1093/jnci/djr138

Klaes, R., Woerner, S. M., Ridder, R., Wentzensen, N., Duerst, M., Schneider, A., . . . von Knebel Doeberitz, M. (1999). Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res, 59*(24), 6132-6136.

Kleivi Sahlberg, K., Bottai, G., Naume, B., Burwinkel, B., Calin, G. A., Borresen-Dale, A. L., & Santarpia, L. (2015). A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. *Clin Cancer Res, 21*(5), 1207-1214. doi:10.1158/1078-0432.CCR-14-2011

Kobayashi, A., Greenblatt, R. M., Anastos, K., Minkoff, H., Massad, L. S., Young, M., . . . Smith-McCune, K. K. (2004). Functional attributes of mucosal immunity in cervical intraepithelial neoplasia and effects of HIV infection. *Cancer Res, 64*(18), 6766-6774. doi:10.1158/0008-5472.CAN-04-1091

Kontostathi, G., Zoidakis, J., Anagnou, N. P., Pappa, K. I., Vlahou, A., & Makridakis, M. (2016). Proteomics approaches in cervical cancer: focus on the discovery of biomarkers for diagnosis and drug treatment monitoring. *Expert Rev Proteomics, 13*(8), 731-745. doi:10.1080/14789450.2016.1210514

Kreimer, A. R., Johansson, M., Waterboer, T., Kaaks, R., Chang-Claude, J., Drogen, D., . . . Brennan, P. (2013). Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer. *J Clin Oncol, 31*(21), 2708-2715. doi:10.1200/JCO.2012.47.2738

Krishna, S., & Anderson, K. S. (2016). T-Cell Epitope Discovery for Therapeutic Cancer Vaccines. *Vaccine Design: Methods and Protocols: Volume 1: Vaccines for Human Diseases*, 779-796.

Kroemer, G., Senovilla, L., Galluzzi, L., Andre, F., & Zitvogel, L. (2015). Natural and therapy-induced immunosurveillance in breast cancer. *Nature medicine, 21*(10), 1128-1138. doi:10.1038/nm.3944

LaBaer, J. (2005). So, you want to look for biomarkers (introduction to the special biomarkers issue). *J Proteome Res, 4*(4), 1053-1059. doi:10.1021/pr0501259

Lacombe, J., Mange, A., Bougnoux, A. C., Prassas, I., & Solassol, J. (2014). A multiparametric serum marker panel as a complementary test to mammography for the diagnosis of node-negative early-stage breast cancer and DCIS in young women. *Cancer Epidemiol Biomarkers Prev, 23*(9), 1834-1842. doi:10.1158/1055-9965.EPI-14-0267

Ladd, J. J., Chao, T., Johnson, M. M., Qiu, J., Chin, A., Israel, R., . . . Hanash, S. (2013). Autoantibody signatures involving glycolysis and splicesome proteins precede a diagnosis of breast cancer among postmenopausal women. *Cancer research, 73*(5), 1502-1513. doi:10.1158/0008-5472.CAN-12-2560

Lang Kuhs, K. A., Anantharaman, D., Waterboer, T., Johansson, M., Brennan, P., Michel, A., . . . Kreimer, A. R. (2015). Human Papillomavirus 16 E6 Antibodies in Individuals without Diagnosed Cancer: A Pooled Analysis. *Cancer Epidemiol Biomarkers Prev, 24*(4), 683-689. doi:10.1158/1055-9965.EPI-14-1217

Lee, J.-R., Magee, D. M., Gaster, R. S., LaBaer, J., & Wang, S. X. (2013). Emerging protein array technologies for proteomics.

Lee, J. W., Devanarayan, V., Barrett, Y. C., Weiner, R., Allinson, J., Fountain, S., . . . Wagner, J. A. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res, 23*(2), 312-328. doi:10.1007/s11095-005-9045-3

Lee, L., Nordman, E., Johnson, M., & Oldham, M. (2013). A Low-Cost, High-Performance System for Fluorescence Lateral Flow Assays. *Biosensors, 3*(4), 360-373.

Legutki, J. B., & Johnston, S. A. (2013). Immunosignatures can predict vaccine efficacy. *Proceedings of the National Academy of Sciences, 110*(46), 18614-18619.

Lehtinen, M., Pawlita, M., Zumbach, K., Lie, K., Hakama, M., Jellum, E., . . . Dillner, J. (2003). Evaluation of antibody response to human papillomavirus early proteins

in women in whom cervical cancer developed 1 to 20 years later. *Am J Obstet Gynecol, 188*(1), 49-55.

Leth-Larsen, R., Lund, R. R., & Ditzel, H. J. (2010). Plasma membrane proteomics and its application in clinical cancer biomarker discovery. *Molecular & cellular proteomics : MCP, 9*(7), 1369-1382. doi:10.1074/mcp.R900006-MCP200

Leung, S. F., Tam, J. S., Chan, A. T., Zee, B., Chan, L. Y., Huang, D. P., . . . Lo, Y. M. (2004). Improved accuracy of detection of nasopharyngeal carcinoma by combined application of circulating Epstein-Barr virus DNA and anti-Epstein-Barr viral capsid antigen IgA antibody. *Clin Chem, 50*(2), 339-345. doi:10.1373/clinchem.2003.022426

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P. M., Klein, G., . . . Masucci, M. G. (1995). Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature, 375*(6533), 685-688. doi:10.1038/375685a0

Li, C. I., Mirus, J. E., Zhang, Y., Ramirez, A. B., Ladd, J. J., Prentice, R. L., . . . Lampe, P. D. (2012). Discovery and preliminary confirmation of novel early detection biomarkers for triple-negative breast cancer using preclinical plasma samples from the Women's Health Initiative observational study. *Breast Cancer Res Treat, 135*(2), 611-618. doi:10.1007/s10549-012-2204-4

Li, X., Blount, P. L., Vaughan, T. L., & Reid, B. J. (2011). Application of biomarkers in cancer risk management: evaluation from stochastic clonal evolutionary and dynamic system optimization points of view. *PLoS Comput Biol, 7*(2), e1001087. doi:10.1371/journal.pcbi.1001087

Liao, H. K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M. Z., Sui, Y., . . . Izpisua Belmonte, J. C. (2017). In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell, 171*(7), 1495-1507 e1415. doi:10.1016/j.cell.2017.10.025

Liaw, K. L., Glass, A. G., Manos, M. M., Greer, C. E., Scott, D. R., Sherman, M., . . . Schiffman, M. (1999). Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. *J Natl Cancer Inst, 91*(11), 954-960.

Liaw, K. L., Hildesheim, A., Burk, R. D., Gravitt, P., Wacholder, S., Manos, M. M., . . . Schiffman, M. (2001). A prospective study of human papillomavirus (HPV) type 16 DNA detection by polymerase chain reaction and its association with acquisition and persistence of other HPV types. *J Infect Dis, 183*(1), 8-15. doi:10.1086/317638

Lillo, F. B., Lodini, S., Ferrari, D., Stayton, C., Taccagni, G., Galli, L., . . . Uberti-Foppa, C. (2005). Determination of human papillomavirus (HPV) load and type in high-grade cervical lesions surgically resected from HIV-infected women during follow-up of HPV infection. *Clin Infect Dis, 40*(3), 451-457. doi:10.1086/427032

Lin, C.-L., & Kao, J.-H. (2016). New perspectives of biomarkers for the management of chronic hepatitis B. *Clinical and Molecular Hepatology, 22*(4), 423-431. doi:10.3350/cmh.2016.0069

Liu, M. C. (2014). By the numbers: does circulating tumor cell enumeration have a role in metastatic breast cancer? *J Clin Oncol, 32*(31), 3479-3482. doi:10.1200/JCO.2014.56.6851

LIU, X., He, Y., Wang, Z., WANG, C., Liu, Z., XIA, T., . . . Zhang, F. L. (2012). Method of sequence optimization for improved recombinant protein expression using a particle swarm optimization algorithm: Google Patents.

Lopez-Arias, E., Aguilar-Lemarroy, A., Felipe Jave-Suarez, L., Morgan-Villela, G., Mariscal-Ramirez, I., Martinez-Velazquez, M., . . . Hernandez-Gutierrez, R. (2012). Alpha 1-antitrypsin: a novel tumor-associated antigen identified in patients with early-stage breast cancer. *Electrophoresis, 33*(14), 2130-2137. doi:10.1002/elps.201100491

Lu, H., Ladd, J., Feng, Z., Wu, M., Goodell, V., Pitteri, S. J., . . . Disis, M. L. (2012). Evaluation of known oncoantibodies, HER2, p53, and cyclin B1, in prediagnostic breast cancer sera. *Cancer Prev Res (Phila), 5*(8), 1036-1043. doi:10.1158/1940-6207.CAPR-11-0558

Lu, Y.-C., Yao, X., Crystal, J. S., Li, Y. F., El-Gamil, M., Gross, C., . . . Robbins, P. F. (2014). Efficient Identification of Mutated Cancer Antigens Recognized by T Cells Associated with Durable Tumor Regressions. *Clinical Cancer Research, 20*(13), 3401-3410. doi:10.1158/1078-0432.ccr-14-0433

Luevano, M., Bernard, H. U., Barrera-Saldana, H. A., Trevino, V., Garcia-Carranca, A., Villa, L. L., . . . Kalantari, M. (2010). High-throughput profiling of the humoral immune responses against thirteen human papillomavirus types by proteome microarrays. *Virology, 405*(1), 31-40. doi:10.1016/j.virol.2010.05.011

Manchanda, R., Legood, R., Burnell, M., McGuire, A., Raikou, M., Loggenberg, K., . . . Jacobs, I. (2015). Cost-effectiveness of population screening for BRCA mutations in Ashkenazi jewish women compared with family history-based testing. *Journal of the National Cancer Institute, 107*(1), 380. doi:10.1093/jnci/dju380

Mangolini, A., Ferracin, M., Zanzi, M. V., Saccenti, E., Ebnaof, S. O., Poma, V. V., . . . Negrini, M. (2015). Diagnostic and prognostic microRNAs in the serum of breast cancer patients measured by droplet digital PCR. *Biomark Res, 3*, 12. doi:10.1186/s40364-015-0037-0

Manno, C. S., Pierce, G. F., Arruda, V. R., Glader, B., Ragni, M., Rasko, J. J., . . . Kay, M. A. (2006). Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med, 12*(3), 342-347. doi:10.1038/nm1358

Marais, D., Rose, R. C., & Williamson, A. L. (1997). Age distribution of antibodies to human papillomavirus in children, women with cervical intraepithelial neoplasia and blood donors from South Africa. *J Med Virol, 51*(2), 126-131.

Markowitz, L. E., Liu, G., Hariri, S., Steinau, M., Dunne, E. F., & Unger, E. R. (2016). Prevalence of HPV After Introduction of the Vaccination Program in the United States. *Pediatrics, 137*(3), e20151968. doi:10.1542/peds.2015-1968

Markowitz, L. E., Sternberg, M., Dunne, E. F., McQuillan, G., & Unger, E. R. (2009). Seroprevalence of human papillomavirus types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003-2004. *J Infect Dis, 200*(7), 1059-1067. doi:10.1086/604729

Marks, J. R., Anderson, K. S., Engstrom, P., Godwin, A. K., Esserman, L. J., Longton, G., . . . Pepe, M. S. (2015). Construction and analysis of the NCI-EDRN breast cancer reference set for circulating markers of disease. *Cancer Epidemiol Biomarkers Prev, 24*(2), 435-441. doi:10.1158/1055-9965.EPI-14-1178

Maroun, M. C., Olivero, O., Lipovich, L., Stark, A., Tait, L., Bandyopadhyay, S., . . . Madrid, F. F. (2014). Anti-centrosome antibodies in breast cancer are the expression of autoimmunity. *Immunol Res, 60*(2-3), 339-347. doi:10.1007/s12026-014-8582-4

Marshall, E. (1999). Gene Therapy Death Prompts Review of Adenovirus Vector. *Science, 286*(5448), 2244-2245. doi:10.1126/science.286.5448.2244

Martino, A. T., Basner-Tschakarjan, E., Markusic, D. M., Finn, J. D., Hinderer, C., Zhou, S., . . . Herzog, R. W. (2013). Engineered AAV vector minimizes in vivo targeting of transduced hepatocytes by capsid-specific CD8+ T cells. *Blood, 121*(12), 2224-2233. doi:10.1182/blood-2012-10-460733

Marur, S., D'Souza, G., Westra, W. H., & Forastiere, A. A. (2010). HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol, 11*(8), 781-789. doi:10.1016/S1470-2045(10)70017-6

Mays, L. E., & Wilson, J. M. (2011). The complex and evolving story of T cell activation to AAV vector-encoded transgene products. *Mol Ther, 19*(1), 16-27. doi:10.1038/mt.2010.250

Mazor, R., Kaplan, G., Park, D., Jang, Y., Lee, F., Kreitman, R., & Pastan, I. (2017). Rational design of low immunogenic anti CD25 recombinant immunotoxin for T cell malignancies by elimination of T cell epitopes in PE38. *Cell Immunol, 313*, 59-66. doi:10.1016/j.cellimm.2017.01.003

Melo, S. A., Sugimoto, H., O'Connell, J. T., Kato, N., Villanueva, A., Vidal, A., . . . Kalluri, R. (2014). Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell, 26*(5), 707-721. doi:10.1016/j.ccell.2014.09.005

Melsheimer, P., Vinokurova, S., Wentzensen, N., Bastert, G., & von Knebel Doeberitz, M. (2004). DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. *Clin Cancer Res, 10*(9), 3059-3063.

Mendez, F., Munoz, N., Posso, H., Molano, M., Moreno, V., van den Brule, A. J., . . . Instituto Nacional de Cancerologia Human Papillomavirus Study, G. (2005). Cervical coinfection with human papillomavirus (HPV) types and possible implications for the prevention of cervical cancer by HPV vaccines. *J Infect Dis, 192*(7), 1158-1165. doi:10.1086/444391

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., . . . Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature genetics, 45*(4), 353-361, 361e351-352. doi:10.1038/ng.2563

Miersch, S., Bian, X., Wallstrom, G., Sibani, S., Logvinenko, T., Wasserfall, C. H., . . . LaBaer, J. (2013). Serological autoantibody profiling of type 1 diabetes by protein arrays. *J Proteomics, 94*, 486-496. doi:10.1016/j.jprot.2013.10.018

Mingozzi, F., Anguela, X. M., Pavani, G., Chen, Y., Davidson, R. J., Hui, D. J., . . . High, K. A. (2013). Overcoming preexisting humoral immunity to AAV using capsid decoys. *Sci Transl Med, 5*(194), 194ra192. doi:10.1126/scitranslmed.3005795

Mingozzi, F., & High, K. A. (2013). Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood, 122*(1), 23-36. doi:10.1182/blood-2013-01-306647

Mingozzi, F., Maus, M. V., Hui, D. J., Sabatino, D. E., Murphy, S. L., Rasko, J. E., . . . High, K. A. (2007). CD8(+) T-cell responses to adeno-associated virus capsid in humans. *Nat Med, 13*(4), 419-422. doi:10.1038/nm1549

Mishra, S., Srivastava, A. K., Suman, S., Kumar, V., & Shukla, Y. (2015). Circulating miRNAs revealed as surrogate molecular signatures for the early detection of breast cancer. *Cancer Lett, 369*(1), 67-75. doi:10.1016/j.canlet.2015.07.045

Mohammed, M. E., & Abdelhafiz, K. (2015). Autoantibodies in the sera of breast cancer patients: Antinuclear and anti-double stranded DNA antibodies as example. *J Cancer Res Ther, 11*(2), 341-344. doi:10.4103/0973-1482.157314

Mok, H., Lee, S., Wright, D. W., & Crowe, J. E. (2008). Enhancement of the CD8+ T cell response to a subdominant epitope of respiratory syncytial virus by deletion of an immunodominant epitope. *Vaccine, 26*(37), 4775-4782. doi:https://doi.org/10.1016/j.vaccine.2008.07.012

Montor, W. R., Huang, J., Hu, Y., Hainsworth, E., Lynch, S., Kronish, J.-W., . . . LaBaer, J. (2009). Genome-wide study of Pseudomonas aeruginosa outer membrane protein immunogenicity using self-assembling protein microarrays. *Infection and immunity, 77*(11), 4877-4886.

Montor, W. R., Huang, J., Hu, Y., Hainsworth, E., Lynch, S., Kronish, J. W., . . . LaBaer, J. (2009). Genome-wide study of Pseudomonas aeruginosa outer membrane protein immunogenicity using self-assembling protein microarrays. *Infect Immun, 77*(11), 4877-4886. doi:10.1128/IAI.00698-09

Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D. C., Sidney, J., Bui, H.-H., . . . Sette, A. (2006). A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus. *Nat Biotechnol, 24*(7), 817.

Munger, K., Baldwin, A., Edwards, K. M., Hayakawa, H., Nguyen, C. L., Owens, M., . . . Huh, K. (2004). Mechanisms of human papillomavirus-induced oncogenesis. *J Virol, 78*(21), 11451-11460. doi:10.1128/JVI.78.21.11451-11460.2004

Murtaza, M., Dawson, S. J., Pogrebniak, K., Rueda, O. M., Provenzano, E., Grant, J., . . . Caldas, C. (2015). Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun, 6*, 8760. doi:10.1038/ncomms9760

Nakamura, T., Shima, T., Saeki, A., Hidaka, T., Nakashima, A., Takikawa, O., & Saito, S. (2007). Expression of indoleamine 2, 3-dioxygenase and the recruitment of Foxp3-expressing regulatory T cells in the development and progression of uterine cervical cancer. *Cancer Sci, 98*(6), 874-881. doi:10.1111/j.1349-7006.2007.00470.x

Narod, S. A. (2010). BRCA mutations in the management of breast cancer: the state of the art. *Nature reviews. Clinical oncology, 7*(12), 702-707. doi:10.1038/nrclinonc.2010.166

Nasman, A., Attner, P., Hammarstedt, L., Du, J., Eriksson, M., Giraud, G., . . . Dalianis, T. (2009). Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? *Int J Cancer, 125*(2), 362-366. doi:10.1002/ijc.24339

Nathwani, A. C., Tuddenham, E. G., Rangarajan, S., Rosales, C., McIntosh, J., Linch, D. C., . . . Davidoff, A. M. (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med, 365*(25), 2357-2365. doi:10.1056/NEJMoa1108046

Nayak, S., & Herzog, R. W. (2010). Progress and prospects: immune responses to viral vectors. *Gene Ther, 17*(3), 295-304. doi:10.1038/gt.2009.148

NCI. (2000). EDRN initial report: translational research to identify early cancer and cancer risk. Retrieved from https://edrn.nci.nih.gov/docs

NCI. (2018). Cancer Risk Prediction and Assessment. Retrieved from https://epi.grants.cancer.gov/cancer_risk_prediction/

NIH. (2015). NIH workshop on building a precision medicine research cohort. Retrieved from http://www.nih.gov/precisionmedicine/workshop.htm

NIH. (2017). National Cancer Institute: Surveillence, Epidemiology, and End Results Program (SEER). Retrieved from https://seer.cancer.gov/statfacts/html/breast.html

Nindl, I., Greinke, C., Zahm, D. M., Stockfleth, E., Hoyer, H., & Schneider, A. (1997). Human papillomavirus distribution in cervical tissues of different morphology as determined by hybrid capture assay and PCR. *Int J Gynecol Pathol, 16*(3), 197-204.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S., Dohmae, N., . . . Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell, 156*(5), 935-949. doi:10.1016/j.cell.2014.02.001

O'Brien, K., Rani, S., Corcoran, C., Wallace, R., Hughes, L., Friel, A. M., . . . O'Driscoll, L. (2013). Exosomes from triple-negative breast cancer cells can transfer phenotypic traits representing their cells of origin to secondary cells. *Eur J Cancer, 49*(8), 1845-1859. doi:10.1016/j.ejca.2013.01.017

O'Driscoll, L. (2015). Expanding on exosomes and ectosomes in cancer. *N Engl J Med, 372*(24), 2359-2362. doi:10.1056/NEJMcibr1503100

Obama, B. (2015). Remarks by the President in State of the Union Address. *The White House*. Retrieved from https://www.whitehouse.gov/the-pressoffice/2015/01/20/remarks-president-state-unionaddress-january-20-2015

Ojesina, A. I., Lichtenstein, L., Freeman, S. S., Pedamallu, C. S., Imaz-Rosshandler, I., Pugh, T. J., . . . Meyerson, M. (2014). Landscape of genomic alterations in cervical carcinomas. *Nature, 506*(7488), 371-375. doi:10.1038/nature12881

Opstal-van Winden, A. W., Krop, E. J., Karedal, M. H., Gast, M. C., Lindh, C. H., Jeppsson, M. C., . . . Vermeulen, R. C. (2011). Searching for early breast cancer biomarkers by serum protein profiling of pre-diagnostic serum; a nested case-control study. *BMC Cancer, 11*, 381. doi:10.1186/1471-2407-11-381

Oyervides-Munoz, M. A., Perez-Maya, A. A., Rodriguez-Gutierrez, H. F., Gomez-Macias, G. S., Fajardo-Ramirez, O. R., Trevino, V., . . . Garza-Rodriguez, M. L. (2018). Understanding the HPV integration and its progression to cervical cancer. *Infect Genet Evol, 61*, 134-144. doi:10.1016/j.meegid.2018.03.003

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., . . . Wolmark, N. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med, 351*(27), 2817-2826. doi:10.1056/NEJMoa041588

Paoletti, C., & Hayes, D. F. (2014). Molecular testing in breast cancer. *Annual review of medicine, 65*, 95-110. doi:10.1146/annurev-med-070912-143853

Parfenov, M., Pedamallu, C. S., Gehlenborg, N., Freeman, S. S., Danilova, L., Bristow, C. A., . . . Cancer Genome Atlas, N. (2014). Characterization of HPV and host

genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A, 111*(43), 15544-15549. doi:10.1073/pnas.1416074111

Patel, S. K., Wong, A. L., Wong, F. L., Breen, E. C., Hurria, A., Smith, M., . . . Bhatia, S. (2015). Inflammatory Biomarkers, Comorbidity, and Neurocognition in Women With Newly Diagnosed Breast Cancer. *J Natl Cancer Inst, 107*(8). doi:10.1093/jnci/djv131

Pedroza-Saavedra, A., Cruz, A., Esquivel, F., De La Torre, F., Berumen, J., Gariglio, P., & Gutierrez, L. (2000). High prevalence of serum antibodies to Ras and type 16 E4 proteins of human papillomavirus in patients with precancerous lesions of the uterine cervix. *Arch Virol, 145*(3), 603-623.

Peeters, D. J. E., van Dam, P. J., Van den Eynden, G. G. M., Rutten, A., Wuyts, H., Pouillon, L., . . . Dirix, L. Y. (2014). Detection and prognostic significance of circulating tumour cells in patients with metastatic breast cancer according to immunohistochemical subtypes. *Br J Cancer, 110*(2), 375-383. doi:10.1038/bjc.2013.743

Pepe, M. S., Etzioni, R., Feng, Z., Potter, J. D., Thompson, M. L., Thornquist, M., . . . Yasui, Y. (2001). Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst, 93*(14), 1054-1061.

Pepe, M. S., Feng, Z., Janes, H., Bossuyt, P. M., & Potter, J. D. (2008). Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst, 100*(20), 1432-1438. doi:10.1093/jnci/djn326

Pepe, M. S., Li, C. I., & Feng, Z. (2015). Improving the quality of biomarker discovery research: the right samples and enough of them. *Cancer Epidemiol Biomarkers Prev, 24*(6), 944-950. doi:10.1158/1055-9965.EPI-14-1227

Pernikarova, V., & Bouchal, P. (2015). Targeted proteomics of solid cancers: from quantification of known biomarkers towards reading the digital proteome maps. *Expert review of proteomics*, 1-17. doi:10.1586/14789450.2015.1094381

Pett, M. R., Herdman, M. T., Palmer, R. D., Yeo, G. S., Shivji, M. K., Stanley, M. A., & Coleman, N. (2006). Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. *Proc Natl Acad Sci U S A, 103*(10), 3822-3827. doi:10.1073/pnas.0600078103

Pietrowska, M., Polanska, J., Marczak, L., Behrendt, K., Nowicka, E., Stobiecki, M., . . . Widlak, P. (2010). Mass spectrometry-based analysis of therapy-related changes in serum proteome patterns of patients with early-stage breast cancer. *J Transl Med, 8*, 66. doi:10.1186/1479-5876-8-66

Pitteri, S. J., Amon, L. M., Busald Buson, T., Zhang, Y., Johnson, M. M., Chin, A., . . . Li, C. I. (2010). Detection of elevated plasma levels of epidermal growth factor

receptor before breast cancer diagnosis among hormone therapy users. *Cancer Res, 70*(21), 8598-8606. doi:10.1158/0008-5472.CAN-10-1676

Pollard, A. J., & Maiden, M. C. J. (2001). *Meningococcal Vaccines*: Humana Press.

Poste, G. (2011). Bring on the biomarkers. *Nature, 469*(7329), 156-157. doi:10.1038/469156a

Poste, G., Compton, C. C., & Barker, A. D. (2015). The national biomarker development alliance: confronting the poor productivity of biomarker research and development. *Expert Rev Mol Diagn, 15*(2), 211-218. doi:10.1586/14737159.2015.974561

Prados-Rosales, R., Carreno, L. J., Batista-Gonzalez, A., Baena, A., Venkataswamy, M. M., Xu, J., . . . Casadevall, A. (2014). Mycobacterial membrane vesicles administered systemically in mice induce a protective immune response to surface compartments of Mycobacterium tuberculosis. *MBio, 5*(5), e01921-01914. doi:10.1128/mBio.01921-14

Qiao, Y.-l., Sellors, J. W., Eder, P. S., Bao, Y.-p., Lim, J. M., Zhao, F.-h., . . . Lorincz, A. T. (2008). A new HPV-DNA test for cervical-cancer screening in developing regions: a cross-sectional study of clinical accuracy in rural China. *Lancet Oncol, 9*(10), 929-936. doi:10.1016/S1470-2045(08)70210-9

Qiu, J., & Anderson, K. S. (2013). Autoantibodies and Biomarker Discovery. In H. J. Issaq & T. D. Veenstra (Eds.), *Proteomic and Metabolomic to Biomarker Discovery*: Elsevier Inc.

Qiu, J., & LaBaer, J. (2011). Nucleic acid programmable protein array a just-in-time multiplexed protein expression and purification platform. *Methods Enzymol, 500*, 151-163. doi:10.1016/B978-0-12-385118-5.00009-8

Rajeevan, M. S., Swan, D. C., Nisenbaum, R., Lee, D. R., Vernon, S. D., Ruffin, M. T., . . . Unger, E. R. (2005). Epidemiologic and viral factors associated with cervical neoplasia in HPV-16-positive women. *Int J Cancer, 115*(1), 114-120. doi:10.1002/ijc.20894

Ramachandran, N., Anderson, K. S., Raphael, J. v., Hainsworth, E., Sibani, S., Montor, W. R., . . . Sanda, M. G. (2008). Tracking humoral responses using self assembling protein microarrays. *PROTEOMICS-Clinical Applications, 2*(10‐11), 1518-1527.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., . . . LaBaer, J. (2004). Self-assembling protein microarrays. *Science, 305*(5680), 86-90. doi:10.1126/science.1097639

Ramachandran, N., Hainsworth, E., Demirkan, G., & LaBaer, J. (2006). On-chip protein synthesis for making microarrays *New and Emerging Proteomic Techniques* (pp. 1-14): Springer.

Ramachandran, N., Raphael, J. V., Hainsworth, E., Demirkan, G., Fuentes, M. G., Rolfs, A., . . . LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. *Nat Methods, 5*(6), 535-538. doi:10.1038/nmeth.1210

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

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., . . . Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature, 520*(7546), 186-191. doi:10.1038/nature14299

Ravaggi, A., Romani, C., Pasinetti, B., Tassi, R. A., Bignotti, E., Bandiera, E., . . . Pecorelli, S. (2006). Correlation between serological immune response analyzed by a new ELISA for HPV-16/18 E7 oncoprotein and clinical characteristics of cervical cancer patients. *Arch Virol, 151*(10), 1899-1916. doi:10.1007/s00705-006-0787-y

Reardon, S. (2016). First CRISPR clinical trial gets green light from US panel. *Nature, 531*.

Reuschenbach, M., Waterboer, T., Wallin, K. L., Einenkel, J., Dillner, J., Hamsikova, E., . . . Wentzensen, N. (2008). Characterization of humoral immune responses against p16, p53, HPV16 E6 and HPV16 E7 in patients with HPV-associated cancers. *Int J Cancer, 123*(11), 2626-2631. doi:10.1002/ijc.23837

Rickinson, A., & Kieff, E. (2001). Epstein-Barr virus. In D. Knipe, P. Howley, D. Griffin, M. Martin, & R. Lamb (Eds.), *Fields Virology* (pp. 2575–2627). Philadelphia: Lippincott Williams and Wilkins.

Rifai, N., Gillette, M. A., & Carr, S. A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol, 24*(8), 971-983. doi:10.1038/nbt1235

Riley, C. P., Zhang, X., Nakshatri, H., Schneider, B., Regnier, F. E., Adamec, J., & Buck, C. (2011). A large, consistent plasma proteomics data set from prospectively collected breast cancer patient and healthy volunteer samples. *J Transl Med, 9*, 80. doi:10.1186/1479-5876-9-80

Robb, J. A., Gulley, M. L., Fitzgibbons, P. L., Kennedy, M. F., Cosentino, L. M., Washington, K., . . . Lapham, R. L. (2014). A call to standardize preanalytic data elements for biospecimens. *Arch Pathol Lab Med, 138*(4), 526-537. doi:10.5858/arpa.2013-0250-CP

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Rodriguez, A. C., Schiffman, M., Herrero, R., Wacholder, S., Hildesheim, A., Castle, P. E., . . . Proyecto Epidemiologico Guanacaste, G. (2008). Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst, 100*(7), 513-517. doi:10.1093/jnci/djn044

Salvat, R. S., Verma, D., Parker, A. S., Kirsch, J. R., Brooks, S. A., Bailey-Kellogg, C., & Griswold, K. E. (2017). Computationally optimized deimmunization libraries yield highly mutated enzymes with low immunogenicity and enhanced activity. *Proc Natl Acad Sci U S A, 114*(26), E5085-E5093. doi:10.1073/pnas.1621233114

Sankaranarayanan, R., Esmy, P. O., Rajkumar, R., Muwonge, R., Swaminathan, R., Shanthakumari, S., . . . Cherian, J. (2007). Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial. *Lancet, 370*(9585), 398-406. doi:10.1016/S0140-6736(07)61195-7

Santini, C., Brennan, D., Mennuni, C., Hoess, R. H., Nicosia, A., Cortese, R., & Luzzago, A. (1998). Eficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. *Journal of molecular biology, 282*(1), 125-135.

Scallan, C. D., Jiang, H., Liu, T., Patarroyo-White, S., Sommer, J. M., Zhou, S., . . . Pierce, G. F. (2006). Human immunoglobulin inhibits liver transduction by AAV vectors at low AAV2 neutralizing titers in SCID mice. *Blood, 107*(5), 1810-1817. doi:10.1182/blood-2005-08-3229

Schlecht, N. F., Kulaga, S., Robitaille, J., Ferreira, S., Santos, M., Miyamura, R. A., . . . Franco, E. L. (2001). Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia. *JAMA, 286*(24), 3106-3114.

Schwartz, S. (2000). Regulation of human papillomavirus late gene expression. *Ups J Med Sci, 105*(3), 171-192.

Schwarzenbach, H. (2013). Circulating nucleic acids as biomarkers in breast cancer. *Breast cancer research : BCR, 15*(5), 211. doi:10.1186/bcr3446

Seiler, C. Y., Park, J. G., Sharma, A., Hunter, P., Surapaneni, P., Sedillo, C., . . . Steel, J. (2013). DNASU plasmid and PSI: Biology-Materials repositories: resources to accelerate biological research. *Nucleic acids research*, gkt1060.

Shastri, S. S., Mittra, I., Mishra, G. A., Gupta, S., Dikshit, R., Singh, S., & Badwe, R. A. (2014). Effect of VIA screening by primary health workers: randomized controlled study in Mumbai, India. *J Natl Cancer Inst, 106*(3), dju009. doi:10.1093/jnci/dju009

Sherman, M. E., Wang, S. S., Wheeler, C. M., Rich, L., Gravitt, P. E., Tarone, R., & Schiffman, M. (2003). Determinants of human papillomavirus load among women with histological cervical intraepithelial neoplasia 3: dominant impact of surrounding low-grade lesions. *Cancer Epidemiol Biomarkers Prev, 12*(10), 1038-1044.

Silins, I., Avall-Lundqvist, E., Tadesse, A., Jansen, K. U., Stendahl, U., Lenner, P., . . . Frankendal, B. (2002). Evaluation of antibodies to human papillomavirus as prognostic markers in cervical cancer patients. *Gynecol Oncol, 85*(2), 333-338. doi:10.1006/gyno.2002.6628

Siravegna, G., Marsoni, S., Siena, S., & Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol, 14*(9), 531-548. doi:10.1038/nrclinonc.2017.14

Skates, S. J. (2014). A statistical challenge: developing tests for biomarker utility specific to the intended use. *J Natl Cancer Inst, 106*(4), dju076. doi:10.1093/jnci/dju076

Skates, S. J., Gillette, M. A., LaBaer, J., Carr, S. A., Anderson, L., Liebler, D. C., . . . Boja, E. S. (2013). Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J Proteome Res, 12*(12), 5383-5394. doi:10.1021/pr400132j

Smith, E. M., Pawlita, M., Rubenstein, L. M., Haugen, T. H., Hamsikova, E., & Turek, L. P. (2010). Risk factors and survival by HPV-16 E6 and E7 antibody status in human papillomavirus positive head and neck cancer. *Int J Cancer, 127*(1), 111-117. doi:10.1002/ijc.25015

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science, 228*(4705), 1315-1317.

Smith, J., Katchman, B., Kullman, D., Obahiagbon, U., Lee, Y.-K., O'Brien, B., . . . Blain Christen, J. (2015). Application of Flexible OLED Display Technology to Point-of-Care Medical Diagnostic Testing. *IEEE*.

Soost, H. J., Lange, H. J., Lehmacher, W., & Ruffing-Kullmann, B. (1991). The validation of cervical cytology. Sensitivity, specificity and predictive values. *Acta Cytol, 35*(1), 8-14.

Soussi, T. (2000). p53 Antibodies in the sera of patients with various types of cancer: a review. *Cancer Res, 60*(7), 1777-1788.

Srivastava, S. (2013). The early detection research network: 10-year outlook. *Clin Chem, 59*(1), 60-67. doi:10.1373/clinchem.2012.184697

Stanley, M. (2003). Antibody reactivity to HPV E6 and E7 oncoproteins and early diagnosis of invasive cervical cancer. *Am J Obstet Gynecol, 188*(1), 3-4.

Stanley, M. (2006). Immune responses to human papillomavirus. *Vaccine, 24 Suppl 1*, S16-22. doi:10.1016/j.vaccine.2005.09.002

Stanley, M. (2008). Immunobiology of HPV and HPV vaccines. *Gynecol Oncol, 109*(2 Suppl), S15-21. doi:10.1016/j.ygyno.2008.02.003

Stanley, M. (2010). HPV - immune response to infection and vaccination. *Infect Agent Cancer, 5*, 19. doi:10.1186/1750-9378-5-19

Stevens, R. C. (2000). Design of high-throughput methods of protein production for structural biology. *Structure, 8*(9), R177-R185.

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Curr Opin HIV AIDS, 5*(6), 463-466. doi:10.1097/COH.0b013e32833ed177

Sud, A., Kinnersley, B., & Houlston, R. S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. *Nature Reviews Cancer, 17*, 692. doi:10.1038/nrc.2017.82
https://www.nature.com/articles/nrc.2017.82#supplementary-information

Sun, Y., Wang, M., Lin, G., Sun, S., Li, X., Qi, J., & Li, J. (2012). Serum microRNA-155 as a potential biomarker to track disease in breast cancer. *PLoS One, 7*(10), e47003. doi:10.1371/journal.pone.0047003

Swan, D. C., Tucker, R. A., Tortolero-Luna, G., Mitchell, M. F., Wideroff, L., Unger, E. R., . . . Icenogle, J. P. (1999). Human papillomavirus (HPV) DNA copy number is dependent on grade of cervical disease and HPV type. *J Clin Microbiol, 37*(4), 1030-1034.

Takulapalli, B. R., Qiu, J., Magee, D. M., Kahn, P., Brunner, A., Barker, K., . . . Wiktor, P. (2012). High density diffusion-free nanowell arrays. *J Proteome Res, 11*(8), 4382-4391. doi:10.1021/pr300467q

Tangri, S., Mothe, B. R., Eisenbraun, J., Sidney, J., Southwood, S., Briggs, K., . . . Sette, A. (2005). Rationally engineered therapeutic proteins with reduced immunogenicity. *J Immunol, 174*(6), 3187-3196.

Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M., . . . Holzhütter, H.-G. (2005). Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences, 62*(9), 1025-1037.

Thomas, K. K., Hughes, J. P., Kuypers, J. M., Kiviat, N. B., Lee, S. K., Adam, D. E., & Koutsky, L. A. (2000). Concurrent and sequential acquisition of different genital human papillomavirus types. *J Infect Dis, 182*(4), 1097-1102. doi:10.1086/315805

Thwaite, R., Pages, G., Chillon, M., & Bosch, A. (2015). AAVrh.10 immunogenicity in mice and humans. Relevance of antibody cross-reactivity in human gene therapy. *Gene Ther, 22*(2), 196-201. doi:10.1038/gt.2014.103

Tice, J. A., Miglioretti, D. L., Li, C. S., Vachon, C. M., Gard, C. C., & Kerlikowske, K. (2015). Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. *Journal of clinical oncology : official*

journal of the American Society of Clinical Oncology, 33(28), 3137-3143. doi:10.1200/JCO.2015.60.8869

Tjiong, M. Y., Zumbach, K., Schegget, J. T., van der Vange, N., Out, T. A., Pawlita, M., & Struyk, L. (2001). Antibodies against human papillomavirus type 16 and 18 E6 and E7 proteins in cervicovaginal washings and serum of patients with cervical neoplasia. *Viral Immunol, 14*(4), 415-424. doi:10.1089/08828240152716655

Transparency Market Research. (2017). Retrieved from https://www.transparencymarketresearch.com/cancer-biomarkers-market.html

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., . . . Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols, 7*(3), 562-578. doi:10.1038/nprot.2012.016

Tripathy, D., Rugo, H. S., Kaufman, P. A., Swain, S., O'Shaughnessy, J., Jahanzeb, M., . . . Hurvitz, S. (2014). The SystHERs registry: an observational cohort study of treatment patterns and outcomes in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer. *BMC Cancer, 14*, 307. doi:10.1186/1471-2407-14-307

Turtoi, A., Dumont, B., Greffe, Y., Blomme, A., Mazzucchelli, G., Delvenne, P., . . . Castronovo, V. (2011). Novel comprehensive approach for accessible biomarker identification and absolute quantification from precious human tissues. *J Proteome Res, 10*(7), 3160-3182. doi:10.1021/pr200212r

Unger, E. R., Vernon, S. D., Lee, D. R., Miller, D. L., & Reeves, W. C. (1998). Detection of human papillomavirus in archival tissues. Comparison of in situ hybridization and polymerase chain reaction. *J Histochem Cytochem, 46*(4), 535-540. doi:10.1177/002215549804600414

US Preventive Task Force. Retrieved from http://www.uspreventiveservicestaskforce.org

Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., . . . McBride, A. A. (2013). The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic acids research, 41*(D1), D571-D578.

van Duin, M., Snijders, P. J., Schrijnemakers, H. F., Voorhorst, F. J., Rozendaal, L., Nobbenhuis, M. A., . . . Meijer, C. J. (2002). Human papillomavirus 16 load in normal and abnormal cervical scrapes: an indicator of CIN II/III and viral clearance. *Int J Cancer, 98*(4), 590-595.

van Duin, M., Snijders, P. J. F., Schrijnemakers, H. F. J., Voorhorst, F. J., Rozendaal, L., Nobbenhuis, M. A. E., . . . Meijer, C. J. L. M. (2002). Human papillomavirus 16 load in normal and abnormal cervical scrapes: An indicator of CIN II/III and viral

clearance. *International Journal of Cancer, 98*(4), 590-595. doi:doi:10.1002/ijc.10232

Van Poznak, C., Somerfield, M. R., Bast, R. C., Cristofanilli, M., Goetz, M. P., Gonzalez-Angulo, A. M., . . . Harris, L. N. (2015). Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 33*(24), 2695-2704. doi:10.1200/JCO.2015.61.1459

Vester-Christensen, M. B., Halim, A., Joshi, H. J., Steentoft, C., Bennett, E. P., Levery, S. B., . . . Clausen, H. (2013). Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America, 110*(52), 21018-21023. doi:10.1073/pnas.1313446110

Vicus, D., Sutradhar, R., Lu, Y., Elit, L., Kupets, R., Paszat, L., & Investigators of the Ontario Cancer Screening Research, N. (2014). The association between cervical cancer screening and mortality from cervical cancer: a population based case-control study. *Gynecol Oncol, 133*(2), 167-171. doi:10.1016/j.ygyno.2014.02.037

Vicus, D., Sutradhar, R., Lu, Y., Kupets, R., Paszat, L., & Ontario Cancer Screening Research, N. (2015). Association between cervical screening and prevention of invasive cervical cancer in Ontario: a population-based case-control study. *Int J Gynecol Cancer, 25*(1), 106-111. doi:10.1097/IGC.0000000000000305

Villa, L. L., Ault, K. A., Giuliano, A. R., Costa, R. L., Petta, C. A., Andrade, R. P., . . . Barr, E. (2006). Immunologic responses following administration of a vaccine targeting human papillomavirus Types 6, 11, 16, and 18. *Vaccine, 24*(27-28), 5571-5583. doi:10.1016/j.vaccine.2006.04.068

Vinokurova, S., Wentzensen, N., Einenkel, J., Klaes, R., Ziegert, C., Melsheimer, P., . . . von Knebel Doeberitz, M. (2005). Clonal History of Papillomavirus-Induced Dysplasia in the Female Lower Genital Tract. *JNCI: Journal of the National Cancer Institute, 97*(24), 1816-1821. doi:10.1093/jnci/dji428

Viscidi, R. P., Kotloff, K. L., Clayman, B., Russ, K., Shapiro, S., & Shah, K. V. (1997). Prevalence of antibodies to human papillomavirus (HPV) type 16 virus-like particles in relation to cervical HPV infection among college women. *Clinical and Diagnostic Laboratory Immunology, 4*(2), 122-126.

Visser, J., Nijman, H. W., Hoogenboom, B. N., Jager, P., van Baarle, D., Schuuring, E., . . . Daemen, T. (2007). Frequencies and role of regulatory T cells in patients with (pre)malignant cervical neoplasia. *Clin Exp Immunol, 150*(2), 199-209. doi:10.1111/j.1365-2249.2007.03468.x

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., . . . Peters, B. (2015). The immune epitope database (IEDB) 3.0. *Nucleic Acids Res, 43*(Database issue), D405-412. doi:10.1093/nar/gku938

Wagner, J. A. (2002). Overview of biomarkers and surrogate endpoints in drug development. *Dis Markers, 18*(2), 41-46.

Wagner, J. A., Williams, S. A., & Webster, C. J. (2007). Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Ther, 81*(1), 104-107. doi:10.1038/sj.clpt.6100017

Wallstrom, G., Anderson, K. S., & LaBaer, J. (2013). Biomarker discovery for heterogeneous diseases. *Cancer Epidemiol Biomarkers Prev, 22*(5), 747-755. doi:10.1158/1055-9965.EPI-12-1236

Wandall, H. H., Blixt, O., Tarp, M. A., Pedersen, J. W., Bennett, E. P., Mandel, U., . . . Clausen, H. (2010). Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes. *Cancer Res, 70*(4), 1306-1313. doi:10.1158/0008-5472.CAN-09-2893

Wang, D., Mou, H., Li, S., Li, Y., Hough, S., Tran, K., . . . Xue, W. (2015). Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum Gene Ther, 26*(7), 432-442. doi:10.1089/hum.2015.087

Wang, J., Barker, K., Steel, J., Park, J., Saul, J., Festa, F., . . . Qiu, J. (2013). A versatile protein microarray platform enabling antibody profiling against denatured proteins. *Proteomics Clin Appl, 7*(5-6), 378-383. doi:10.1002/prca.201200062

Wang, J., Figueroa, J. D., Wallstrom, G., Barker, K., Park, J. G., Demirkan, G., . . . LaBaer, J. (2015). Plasma Autoantibodies Associated with Basal-like Breast Cancers. *Cancer Epidemiol Biomarkers Prev, 24*(9), 1332-1340. doi:10.1158/1055-9965.EPI-15-0047

Wang, R. F. (2006). Regulatory T cells and innate immune regulation in tumor immunity. *Springer Semin Immunopathol, 28*(1), 17-23. doi:10.1007/s00281-006-0022-7

Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., . . . Chinnaiyan, A. M. (2005). Autoantibody signatures in prostate cancer. *The New England journal of medicine, 353*(12), 1224-1235. doi:10.1056/NEJMoa051931

Wang, Y., Waters, J., Leung, M. L., Unruh, A., Roh, W., Shi, X., . . . Navin, N. E. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature, 512*(7513), 155-160. doi:10.1038/nature13600

Watanabe, N., Arase, H., Onodera, M., Ohashi, P. S., & Saito, T. (2000). The quantity of TCR signal determines positive selection and lineage commitment of T cells. *J Immunol, 165*(11), 6252-6261.

Waterboer, T., Sehr, P., Michael, K. M., Franceschi, S., Nieland, J. D., Joos, T. O., . . . Pawlita, M. (2005). Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. *Clin Chem, 51*(10), 1845-1853. doi:10.1373/clinchem.2005.052381

Weissenborn, S. J., Funke, A. M., Hellmich, M., Mallmann, P., Fuchs, P. G., Pfister, H. J., & Wieland, U. (2003). Oncogenic human papillomavirus DNA loads in human immunodeficiency virus-positive women with high-grade cervical lesions are strongly elevated. *J Clin Microbiol, 41*(6), 2763-2767.

Welters, M. J., de Jong, A., van den Eeden, S. J., van der Hulst, J. M., Kwappenberg, K. M., Hassane, S., . . . van der Burg, S. H. (2003). Frequent display of human papillomavirus type 16 E6-specific memory t-Helper cells in the healthy population as witness of previous viral encounter. *Cancer Res, 63*(3), 636-641.

Wentzensen, N., & von Knebel Doeberitz, M. (2007). Biomarkers in cervical cancer screening. *Dis Markers, 23*(4), 315-330.

Wittenberger, T., Sleigh, S., Reisel, D., Zikan, M., Wahl, B., Alunni-Fabbroni, M., . . . Widschwendter, M. (2014). DNA methylation markers for early detection of women's cancer: promise and challenges. *Epigenomics, 6*(3), 311-327. doi:10.2217/epi.14.20

Wolfl, M., Kuball, J., Ho, W. Y., Nguyen, H., Manley, T. J., Bleakley, M., & Greenberg, P. D. (2007). Activation-induced expression of CD137 permits detection, isolation, and expansion of the full repertoire of CD8+ T cells responding to antigen without requiring knowledge of epitope specificities. *Blood, 110*(1), 201-210. doi:10.1182/blood-2006-11-056168

Wong, J., Sibani, S., Lokko, N. N., LaBaer, J., & Anderson, K. S. (2009). Rapid detection of antibodies in sera using multiplexed self-assembling bead arrays. *Journal of immunological methods, 350*(1), 171-182.

Wong, J., Sibani, S., Lokko, N. N., LaBaer, J., & Anderson, K. S. (2009). Rapid detection of antibodies in sera using multiplexed self-assembling bead arrays. *J Immunol Methods, 350*(1-2), 171-182. doi:10.1016/j.jim.2009.08.013

Woodman, C. B., Collins, S. I., & Young, L. S. (2007). The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer, 7*(1), 11-22. doi:10.1038/nrc2050

Wright, T. C., Jr. (2006). HPV DNA testing for cervical cancer screening. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet, 95 Suppl 1*, S239-246. doi:10.1016/S0020-7292(06)60039-8

Wu, A. H. (2006). A selected history and future of immunoassay development and applications in clinical chemistry. *Clinica chimica acta, 369*(2), 119-124.

Wu, L., Goodwin, E. C., Naeger, L. K., Vigo, E., Galaktionov, K., Helin, K., & DiMaio, D. (2000). E2F-Rb complexes assemble and inhibit cdc25A transcription in cervical carcinoma cells following repression of human papillomavirus oncogene expression. *Mol Cell Biol, 20*(19), 7059-7067.

Wu, Q., Lu, Z., Li, H., Lu, J., Guo, L., & Ge, Q. (2011). Next-generation sequencing of microRNAs for breast cancer detection. *J Biomed Biotechnol, 2011*, 597145. doi:10.1155/2011/597145

Xue, Y., Bellanger, S., Zhang, W., Lim, D., Low, J., Lunny, D., & Thierry, F. (2010). HPV16 E2 is an immediate early marker of viral infection, preceding E7 expression in precursor structures of cervical carcinoma. *Cancer Res, 70*(13), 5316-5325. doi:10.1158/0008-5472.CAN-09-3789

Yeung, V. P., Chang, J., Miller, J., Barnett, C., Stickler, M., & Harding, F. A. (2004). Elimination of an Immunodominant CD4⁺ T Cell Epitope in Human IFN-β Does Not Result in an In Vivo Response Directed at the Subdominant Epitope. *The Journal of Immunology, 172*(11), 6658-6665. doi:10.4049/jimmunol.172.11.6658

Yewdell, J. W., & Hill, A. B. (2002). Viral interference with antigen presentation. *Nature Immunology, 3*, 1019. doi:10.1038/ni1102-1019

Yin, H., Kanasty, R. L., Eltoukhy, A. A., Vegas, A. J., Dorkin, J. R., & Anderson, D. G. (2014). Non-viral vectors for gene-based therapy. *Nat Rev Genet, 15*(8), 541-555. doi:10.1038/nrg3763

Yong, C. S. M., Dardalhon, V., Devaud, C., Taylor, N., Darcy, P. K., & Kershaw, M. H. (2016). CAR T-cell therapy of solid tumors. *Immunology And Cell Biology, 95*, 356. doi:10.1038/icb.2016.128

Yu, S., Cao, H., Shen, B., & Feng, J. (2015). Tumor-derived exosomes in cancer progression and treatment failure. *Oncotarget*. doi:10.18632/oncotarget.6022

Zaenker, P., Gray, E. S., & Ziman, M. R. (2016). Autoantibody Production in Cancer--The Humoral Immune Response toward Autologous Antigens in Cancer Patients. *Autoimmun Rev, 15*(5), 477-483. doi:10.1016/j.autrev.2016.01.017

Zawadzka, A. M., Schilling, B., Cusack, M. P., Sahu, A. K., Drake, P., Fisher, S. J., . . . Gibson, B. W. (2014). Phosphoprotein secretome of tumor cells as a source of candidates for breast cancer biomarkers in plasma. *Molecular & cellular proteomics : MCP, 13*(4), 1034-1049. doi:10.1074/mcp.M113.035485

Zeidan, B. A., Townsend, P. A., Garbis, S. D., Copson, E., & Cutress, R. I. (2015). Clinical proteomics and breast cancer. *Surgeon, 13*(5), 271-278. doi:10.1016/j.surge.2014.12.003

Zelen, M. (1993). Optimal Scheduling of Examinations for the Early Detection of Disease. *Biometrika, 80*(2), 279-293. doi:10.2307/2337199

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., . . . Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell, 163*(3), 759-771. doi:10.1016/j.cell.2015.09.038

Zhao, F. P., Liu, X., Zhong, Z. M., Lu, J., Yu, B. L., Zeng, F. Y., . . . Li, X. P. (2014). Positivity of both plasma Epstein-Barr virus DNA and serum Epstein-Barr virus capsid specific immunoglobulin A is a better prognostic biomarker for nasopharyngeal carcinoma. *BBA Clin, 2*, 88-93. doi:10.1016/j.bbacli.2014.10.003

Zheng, Y., Shen, W., Zhang, J., Yang, B., Liu, Y. N., Qi, H., . . . Yao, J. (2018). CRISPR interference-based specific and efficient gene inactivation in the brain. *Nat Neurosci, 21*(3), 447-454. doi:10.1038/s41593-018-0077-5

Zhou, J., Liu, W. J., Peng, S. W., Sun, X. Y., & Frazer, I. (1999). Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *J Virol, 73*(6), 4972-4982.

Zumbach, K., Kisseljov, F., Sacharova, O., Shaichaev, G., Semjonova, L., Pavlova, L., & Pawlita, M. (2000). Antibodies against oncoproteins E6 and E7 of human papillomavirus types 16 and 18 in cervical-carcinoma patients from Russia. *Int J Cancer, 85*(3), 313-318.

zur Hausen, H. (2002). Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer, 2*(5), 342-350. doi:10.1038/nrc798

zur Hausen, H. (2009). Papillomaviruses in the causation of human cancers - a brief historical account. *Virology, 384*(2), 260-265. doi:10.1016/j.virol.2008.11.046

APPENDIX A

REFERENCE AND CODON OPTIMIZED HPV DNA SEQUENCES USED IN THIS THESIS, THE PERCENTAGE OF NUCLEOTIDE CHANGE AND TRANSLATED AMINO ACID SEQUENCES

Gene	PaVE DNA Sequence	Optimized DNA Sequence	Protein Sequence	% Nucleotide Change
HPV6 E1	ATGGCGGACGATTCAGGTACAGAAAATGAGGGGTCTG GGTGTACAGGATGGTTTATGGTAGAAGCTATAGTGCAA CACCCAACAGGTACACAAATATCAGACGATGAGGATGA GGAGGTGGAGGACAGTGGGTATGACATGGTGGACTTT ATTGATGACAGCAATATTACACACAATTCACTGGAAGCA CAGGCATTGTTTAACAGGCAGGAGGCGGACACCCATT ATGCGACTGTGCAGGACCTAAAACGAAAGTATTTAGGT AGTCCATATGTTAGTCCTATAAACACTATAGCCGAGGC AGTGGAAAGTGAAATAAGTCCACGATTGGACGCCATTA AACTTACAAGACAGCCAAAAAAGGTAAAGCGACGGCTG TTTCAAACCAGGGAACTAACGGACAGTGGATATGGCTA TTCTGAAGTGGAAGCTGGAACGGGAACGCAGGTAGAG AAACATGGCGTACCGGAAAATGGGGGAGATGGTCAGG AAAAGGACACAGGAAGGGACATAGAGGGGGAGGAACA TACAGAGGCGGAAGCGCCCACAAACAGTGTACGGGAG CATGCAGGCACAGCAGGAATATTGGAATTGTTAAAATG TAAAGATTTACGGGCAGCATTACTTGGTAAGTTTAAAGA ATGCTTTGGGCTGTCTTTTATAGATTTAATTAGGCCATT TAAAAGTGATAAAACAACATGTTTAGATTGGGTGGTAG CAGGGTTTGGTATACATCATAGCATATCAGAGGCATTT CAAAAATTAATTGAGCCATTAAGTTTATATGCACATATA CAATGGCTAACAAATGCATGGGGAATGGTATTGTTAGT ATTATTAAGATTTAAAGTAAATAAAAGTAGAAGTACCGT TGCACGTACACTTGCAACGCTATTAAATATACCTGAAAA CCAAATGTTAATAGAGCCACCAAAAATACAAAGTGGTG TTGCAGCCCTGTATTGGTTTCGTACAGGTATATCAAATG CCAGTACAGTTATAGGGGAAGCACCAGAATGGATAACA CGCCAAACAGTTATTGAACACGGGTTGGCAGACAGTCA GTTTAAATTAACAGAAATGGTGCAGTGGGCGTATGATA ATGACATATGCGAGGAGAGTGAAATTGCATTTGAATAT GCACAAAGGGGAGATTTTGATTCTAATGCACGAGCATT TTTAAATAGCAATATGCAGGCAAAATATGTGAAAGATTG TGCAACTATGTGTAGACATTATAAACATGCAGAAATGAG GAAGATGTCTATAAAACAATGGATAAAACATAGGGGTT CTAAAATAGAAGGCACAGGAAATTGGAAACCAATTGTA CAATTCCTACGACATCAAAATATAGAATTCATTCCTTTTT TAACTAAATTTAAATTATGGCTGCACGGTACGCCAAAAA AAAACTGCATAGCCATAGTAGGCCCTCCAGATACTGGG AAATCGTACTTTTGTATGAGTTTAATAAGCTTTCTAGGA GGTACAGTTATTAGTCATGTAAATTCCAGCAGCCATTTT TGGTTGCAACCGTTAGTAGATGCTAAGGTAGCATTGTT AGATGATGCAACACAGCCATGTTGGATATATATGGATA CATATATGAGAAATTTGTTAGATGGTAATCCTATGAGTA TTGACAGAAAGCATAAAGCATTGACATTAATTAAATGTC CACCTCTGCTAGTAACGTCCAACATAGATATTACTAAAG AAGATAAATATAAGTATTTACATACTAGAGTAACAACAT TTACATTTCCAAATCCATTCCCTTTTGACAGAAATGGGA ATGCAGTGTATGAACTGTCAAATACAAACTGGAAATGTT TTTTTGAAAGACTGTCGTCAAGCCTAGACATTCAGGATT CTGAGGACGAGGAAGATGGAAGCAATAGCCAAGCGTT TAGATGCGTGCCAGGAACAGTTGTTAGAACTTTATGA	ATGGCAGACGATTCAGGGACAGAAAATGAGGGGAGCG GGTGTACTGGGTGGTTTATGGTGGAGGCTATTGTGCA GCATCCTACTGGAACCCAGATCTCCGACGATGAGGAC GAGGAAGTGGAAGATTCTGGGTACGACATGGTCGATTT TATCGACGATTCTAACATTACACACAATAGTCTGGAGG CTCAGGCACTGTTCAACAGACAGGAAGCAGACACACAT TATGCCACTGTGCAGGATCTGAAGAGGAAATACCTGGG CAGTCCCTATGTGTCACCTATCAATACCATTGCCGAGG CTGTCGAGTCTGAAATCAGTCCACGACTGGACGCCATC AAGCTGACACGGCAGCCCAAGAAAGTGAAGCGGAGAC TGTTTCAGACCCGCGAGCTGACAGATAGCGGGTACGG ATATTCCGAGGTCGAAGCCGGCACAGGGACTCAGGTG GAGAAACACGGAGTCCCAGAAAACGGAGGGGACGGAC AGGAGAAGGACACAGGACGCGATATCGAAGGCGAGGA ACACACAGAGGCAGAAGCCCCCACTAATAGCGTGCGA GAGCATGCCGGCACTGCTGGGATCCTGGAACTGCTGA AGTGCAAAGACCTGCGGGCCGCTCTGCTGGGCAAGTT CAAAGAGTGTTTTGGGCTGAGTTTCATCGATCTGATTA GACCTTTCAAGTCAGACAAAACCACATGCCTGGATTGG GTGGTCGCTGGATTTGGCATCCACCATTCCATTTCTGA GGCATTCCAGAAGCTGATCGAACCACTGTCCCTGTACG CACACATTCAGTGGCTGACTAACGCCTGGGGCATGGT GCTGCTGGTCCTGCTGAGGTTTAAGGTGAACAAGAGTA GGTCAACCGTCGCTCGCACCCTGGCAACACTGCTGAA CATCCCCGAGAATCAGATGCTGATCGAACCCCCTAAGA TTCAGAGCGGAGTGGCAGCCCTGTATTGGTTCCGCAC AGGGATCTCAAACGCTAGCACTGTGATTGGAGAGGCA CCTGAATGGATCACTCGGCAGACCGTCATTGAGCACG GCCTGGCCGACTCTCAGTTTAAGCTGACCGAAATGGTG CAGTGGGCTTACGACAACGATATCTGTGAGGAAAGCG AGATTGCCTTCGAATATGCTCAGAGAGGGGACTTTGAT TCAAATGCTAGGGCATTCCTGAACAGCAATATGCAGGC CAAGTACGTGAAAGATTGCGCTACCATGTGTCGCCACT ATAAGCATGCCGAGATGCGAAAGATGAGCATCAAACAG TGGATTAAGCATAGAGGATCCAAAATCGAAGGGACAGG AAACTGGAAGCCTATTGTGCAGTTTCTGAGGCACCAGA ATATCGAGTTCATTCCTTTTCTGACCAAGTTCAAACTGT GGCTGCATGGCACACCAAAGAAAAACTGCATCGCCATT GTGGGGCCACCCGACACAGGAAAATCTTACTTTTGTAT GTCCCTGATCTCTTTCCTGGGAGGCACTGTGATTAGTC ACGTCAATAGCTCCTCT	MADDSGTENEGSGCT GWFMVEAIVQHPTGTQI SDDEDEEVEDSGYDMV DFIDDSNITHNSLEAQAL FNRQEADTHYATVQDL KRKYLGSPYVSPINTIAE AVESEISPRLDAIKLTRQ PKKVKRRLFQTRELTDS GYGYSEVEAGTGTQVE KHGVPENGGDGQEKD TGRDIEGEEHTEAEAPT NSVREHAGTAGILELLK CKDLRAALLGKFKECFG LSFIDLIRPFKSDKTTCL DWVVAGFGIHHSISEAF QKLIEPLSLYAHIQWLTN AWGMVLLVLLRFKVNK SRSTVARTLATLLNIPEN QMLIEPPKIQSGVAALY WFRTGISNASTVIGEAP EWITRQTVIEHGLADSQ FKLTEMVQWAYDNDIC EESEIAFEYAQRGDFDS NARAFLNSNMQAKYVK DCATMCRHYKHAEMRK MSIKQWIKHRGSKIEGT GNWKPIVQFLRHQNIEF IPFLTKFKLWLHGTPKK NCIAIVGPPDTGKSYFC MSLISFLGGTVISHVNS SSHFWLQPLVDAKVALL DDATQPCWIYMDTYMR NLLDGNPMSIDRKHKAL TLIKCPPLLVTSNIDITKE DKYKYLHTRVTTFTFPN PFPFDRNGNAVYELSN TNWKCFFERLSSSLDIQ DSEDEEDGSNSQAFRC VPGTVVRTL	41.3%
HPV6 E2	ATGGAAGCAATAGCCAAGCGTTTAGATGCGTGCCAGG AACAGTTGTTAGAACTTTATGAAGAAAACAGTACTGACC TACACAAACATGTATTGCATTGGAAATGCATGAGACAT GAAAGTGTATTATTATATAAAGCAAAACAAATGGGCCTA AGCCACATAGGAATGCAAGTAGTGCCACCATTAAAGGT GTCCGAAGCAAAAGGACATAATGCCATTGAAATGCAAA TGCATTTAGAATCATTATTAAGGACTGAGTATAGTATGG AACCGTGGACATTACAAGAAACAAGTTATGAAATGTGG CAAACACCACCTAAACGCTGTTTTAAAAAACGGGGCAA AACTGTAGAAGTTAAATTTGATGGCTGTGCAAACAATAC AATGGATTATGTGGTATGGACAGATGTGTATGTGCAGG ACAATGACACCTGGGTAAAGGTGCATAGTATGGTAGAT GCTAAGGGTATATATTACACATGTGGACAATTTAAAACA TATTATGTAAACTTTGTAAAAGAGGCAGAAAAGTATGGG AGCACCAAACATTGGGAAGTATGTTATGGCAGCACAGT TATATGTTCTCCTGCATCTGTATCTAGCACTACACAAGA AGTATCCATTCCTGAATCTACTACATACACCCCCGCAC AGACCTCCACCCTTGTGTCCTCAAGCACCAAGGAAGAC GCAGTGCAAACGCCGCCTAGGAAACGAGCACGAGGAG TCCAACAGTCCCCTTGCAACGCCTTGTGTGTGGCCCAC ATTGGACCCGTGGACAGTGGAAACCACAACCTCATCAC TAACAATCACGACCAGCACCAAAGACGGAACAACAGTA ACAGTTCAGCTACGCCTATAGTGCAATTTCAAGGTGAA TCCAATTGTTTAAAGTGTTTTAGATATAGGCTAAATGAC AGACACAGACATTTATTTGATTTAATATCATCAACGTGG CACTGGGCCTCCTCAAAGGCACCACATAAACATGCCAT TGTAACTGTAACATATGATAGTGAGGAACAAAGGCAAC AGTTTTTAGATGTTGTAAAAATACCCCCTACCATTAGCC ACAAACTGGGATTTATGTCACTGCACCTATTGTAA	ATGGAGGCTATTGCTAAAAGACTGGACGCTTGTCAGGA ACAGCTGCTGGAACTGTACGAGGAAAACTCTACCGATC TGCATAAACATGTGCTGCACTGGAAGTGCATGCGCCAT GAGTCCGTGCTGCTGTACAAGGCCAAACAGATGGGCC TGTCTCACATCGGGATGCAGGTGGTCCCCCCTCTGAA GGTGAGTGAAGCTAAAGGCCACAACGCAATTGAGATG CAGATGCATCTGGAAAGCCTGCTGCGGACCGAGTACT CCATGGAACCATGGACTCTGCAGGAGACCTCCTATGAA ATGTGGCAGACCCCACCCAAGCGATGCTTCAAGAAAA GGGGCAAGACAGTGGAGGTCAAATTTGACGGGTGTGC CAACAATACCATGGACTACGTGGTCTGGACAGATGTGT ATGTCCAGGACAACGATACATGGGTGAAGGTCCACTCT ATGGTGGATGCTAAAGGGATCTACTATACATGTGGACA GTTCAAGACTTACTATGTGAATTTTGTCAAGGAGGCAG AAAAATACGGATCAACAAAACATTGGGAGGTGTGCTAT GGCAGCACTGTCATCTGTTCCCCTGCATCTGTGAGCTC CACCACACAGGAGGTCAGTATTCCAGAATCAACTACCT ATACCCCCGCCCAGACAAGCACTCTGGTGTCTAGTTCA ACTAAGGAAGACGCCGTGCAGACCCCTCCAAGGAAAC GAGCTCGAGGAGTGCAGCAGTCCCCTTGCAACGCCCT GTGTGTGGCTCACATCGGACCAGTCGACTCTGGCAAC CATAATCTGATTACTAACAATCACGATCAGCATCAGCG GAGAAACAATAGTAATAGCTCCGCCACCCCCATTGTGC AGTTCCAGGGAGAGTCAAACTGCCTGAAGTGTTTTCGG TACAGACTGAATGACAGGCACCGCCATCTGTTCGATCT GATCTCTAGTACATGGCACTGGGCATCAAGCAAGGCC CCTCACAAACATGCTATTGTGACCGTCACATATGACTC CGAGGAACAGAGACAGCAGTTCCTGGATGTGGTCAAG ATCCCCCCTACTATTTCTCACAAACTGGGGTTTATGAGT CTGCATCTGCTG	MEAIAKRLDACQEQLLE LYEENSTDLHKHVLHW KCMRHESVLLYKAKQM GLSHIGMQVVPPLKVSE AKGHNAIEMQMHLESLL RTEYSMEPWTLQETSY EMWQTPPKRCFKKRG KTVEVKFDGCANNTMD YVVWTDVYVQDNDTW VKVHSMVDAKGIYYTC GQFKTYYVNFVKEAEK YGSTKHWEVCYGSTVI CSPASVSSTTQEVSIPE STTYTPAQTSTLVSSST KEDAVQTPPRKRARGV QQSPCNALCVAHIGPV DSGNHNLITNNHDQHQ RRNNSNSSATPIVQFQ GESNCLKCFRYRLNDR HRHLFDLISSTWHWAS SKAPHKHAIVTVTYDSE EQRQQFLDVVKIPPTIS HKLGFMSLHLL	25.5%
HPV6 E4	ATGGGAGCACCAAACATTGGGAAGTATGTTATGGCAGC ACAGTTATATGTTCTCCTGCATCTGTATCTAGCACTACA CAAGAAGTATCCATTCCTGAATCTACTACATACACCCCC GCACAGACCTCCACCCTTGTGTCCTCAAGCACCAAGGA AGACGCAGTGCAAACGCCGCCTAGGAAACGAGCACGA GGAGTCCAACAGTCCCCTTGCAACGCCTTGTGTGTGG CCCACATTGGACCCGTGGACAGTGGAAACCACAACCT CATCACTAACAATCACGACCAGCACCAAAGACGGAACA ACAGTAACAGTTCAGCTACGCCTATAG	ATGGGGGCTCCTAATATCGGAAAGTATGTCATGGCCGC TCAGCTGTATGTCCTGCTGCATCTGTATCTGGCACTGC ACAAGAAGTATCCCTTCCTGAACCTGCTGCACACTCCC CCTCATAGGCCACCACCTCTGTGCCCACAGGCACCAC GAAAGACCCAGTGTAAACGGAGACTGGGCAACGAGCA CGAGGAATCTAATAGTCCTCTGGCTACACCATGCGTGT GGCCCACACTGGACCCTTGGACTGTCGAAACCACAAC TAGCTCCCTGACCATCACCACATCCACAAAGGATGGGA CTACCGTGACCGTCCAGCTGCGACTG	MGAPNIGKYVMAAQLY VLLHLYLALHKKYPFLNL LHTPPHRPPPLCPQAP RKTQCKRRLGNEHEES NSPLATPCVWPTLDPW TVETTTSSLTITTSTKDG TTVTVQLRL	23.6%

HPV6 E5A	ATGGAAGTGGTGCCTGTACAAATAGCTGCAGGAACAAC CAGCACATTCATACTGCCTGTTATAATTGCATTTGTTGT ATGTTTTGTTAGCATCATACTTATTGTATGGATATCTGA GTTTATTGTGTACACATCTGTGCTAGTACTAACACTGCT TTTATATTTACTATTGTGGCTGCTATTAACAACCCCCTT GCAATTTTTCCTACTAACTCTACTTGTGTGTTACTGTCC CGCATTGTATATACACTACTATATTGTTACCACACAGCA ATGA	ATGGAGGTCGTCCCCGTCCAGATTGCCGCCGGAACCA CTTCAACTTTCATCCTGCCCGTCATCATTGCTTTCGTCG TGTGTTTCGTGTCTATCATTCTGATCGTGTGGATTAGCG AGTTCATCGTCTACACATCCGTGCTGGTCCTGACTCTG CTGCTGTATCTGCTGCTGTGGCTGCTGCTGACCACACC CCTGCAGTTCTTTCTGCTGACCCTGCTGGTGTGCTACT GTCCTGCCCTGTATATCCACTACTATATTGTCACTACCC AGCAG	MEVVPVQIAAGTTSTFIL PVIIAFVVCFVSIILIVWIS EFIVYTSVLVLTLLLYLLL WLLLTTPLQFFLLTLLVC YCPALYIHYYIVTTQQ	27.2%
HPV6 E5B	ATGATGCTAACATGTCAATTTAATGATGGAGATACCTGG CTGGGTTTGTGGTTGTTATGTGCCTTTATTGTAGGGAT GTTGGGGTTATTATTGATGCACTATAGAGCTGTACAAG GGGATAAACACACCAAATGTAAGAAGTGTAACAAACAC AACTGTAATGATGATTATGTAACTATGCATTATACTACT GATGGTGATTATATATATATGAATTAG	ATGATGCTGACTTGTCAGTTCAACGATGGCGATACTTG GCTGGGGCTGTGGCTGCTGTGTGCTTTTATTGTCGGAA TGCTGGGGCTGCTGCTGATGCACTACCGGGCCGTGCA GGGCGACAAGCATACTAAATGCAAGAAATGTAACAAGC ACAACTGCAATGACGATTACGTCACCATGCATTATACC ACAGACGGGGATTACATCTATATGAAT	MMLTCQFNDGDTWLGL WLLCAFIVGMLGLLLMH YRAVQGDKHTKCKKCN KHNCNDDYVTMHYTTD GDYIYMN	22.4%
HPV6 E6	ATGGAAAGTGCAAATGCCTCCACGTCTGCAACGACCAT AGACCAGTTGTGCAAGACGTTTAATCTATCTATGCATAC GTTGCAAATTAATTGTGTGTTTTGCAAGAATGCACTGAC CACAGCAGAGATTTATTCATATGCATATAAACACCTAAA GGTCCTGTTTCGAGGCGGCTATCCATATGCAGCCTGC GCGTGCTGCCTAGAATTTCATGGAAAAATAAACCAATAT AGACACTTTGATTATGCTGGATATGCAACAACAGTTGAA GAAGAAACTAAACAAGACATCTTAGACGTGCTAATTCG GTGCTACCTGTGTCACAAACCGCTGTGTGAAGTAGAAA AGGTAAAACATATACTAACCAAGGCGCGGTTCATAAAG CTAAATTGTACGTGGAAGGGTCGCTGCCTACACTGCTG GACAACATGCATGGAAGACATGTTACCCTAA	ATGGAGTCCGCTAACGCTTCTACTTCCGCAACTACTAT CGACCAGCTGTGTAAGACCTTCAACCTGTCAATGCATA CCCTGCAGATTAACTGCGTGTTCTGTAAGAATGCTCTG ACCACAGCAGAAATCTACAGCTATGCCTACAAGCACCT GAAAGTCCTGTTTAGGGGCGGGTATCCCTACGCCGCT TGCGCTTGCTGTCTGGAGTTCCACGGAAAAATTAACCA GTATCGCCATTTTGACTATGCAGGCTACGCCACTACCG TGGAGGAAGAGACCAAGCAGGACATCCTGGATGTCCT GATTCGATGCTACCTGTGTCACAAACCCCTGTGTGAAG TGGAGAAGGTCAAACATATCCTGACCAAGGCCCGGTTC ATCAAGCTGAACTGCACATGGAAGGGGAGATGCCTGC ATTGTTGGACAACTTGTATGGAAGATATGCTGCCT	MESANASTSATTIDQLC KTFNLSMHTLQINCVFC KNALTTAEIYSYAYKHL KVLFRGGYPYAACACC LEFHGKINQYRHFDYAG YATTVEEETKQDILDVLI RCYLCHKPLCEVEKVK HILTKARFIKLNCTWKG RCLHCWTTCMEDMLP	22.5%
HPV6 E7	ATGCATGGAAGACATGTTACCCTAAAGGATATTGTATTA GACCTGCAACCTCCAGACCCTGTAGGGTTACATTGCTA TGAGCAATTAGTAGACAGCTCAGAAGATGAGGTGGAC GAAGTGGACGGACAAGATTCACAACCTTTAAAACAACA TTTCCAAATAGTGACCTGTTGCTGTGGATGTGACAGCA ACGTTCGACTGGTTGTGCAGTGTACAGAAACAGACATC AGAGAAGTGCAACAGCTTCTGTTGGGAACACTAAACAT AGTGTGTCCCATCTGCGCACCGAAGACCTAA	ATGCACGGAAGACACGTCACCCTGAAAGATATTGTCCT GGACCTGCAGCCTCCCGACCCTGTGGGCCTGCATTGC TATGAACAGCTGGTGGACAGCTCCGAGGACGAAGTGG ATGAGGTCGACGGCCAGGATTCTCAGCCCCTGAAGCA GCACTTCCAGATCGTGACATGCTGTTGCGGGTGTGACA GCAACGTCCGGCTGGTGGTCCAGTGCACCGAGACAGA TATTAGAGAAGTGCAGCAGCTGCTGCTGGGCACTCTGA ATATCGTCTGTCCCATTTGCGCCCCTAAAACC	MHGRHVTLKDIVLDLQP PDPVGLHCYEQLVDSS EDEVDEVDGQDSQPLK QHFQIVTCCCGCDSNV RLVVQCTETDIREVQQL LLGTLNIVCPICAPKT	22.9%
HPV6 L1	ATGTGGCGGCCTAGCGACAGCACAGTATATGTGCCTC CTCCTAACCCTGTATCCAAAGTTGTTGCCACGGATGCT TATGTTACTCGCACCAACATATTTTATCATGCCAGCAGT TCTAGACTTCTTGCAGTGGGACATCCTTATTTTTCCATA AAACGGGCTAACAAAACTGTTGTGCCAAAGGTGTCAGG ATATCAATACAGGGTATTTAAGGTGGTGTTACCAGATC CTAACAAATTTGCATTGCCTGACTCGTCTCTTTTCGATC CCACAACACAACGTTTAGTATGGGCATGCACAGGCCTA GAGGTGGGCAGGGGACAGCCATTAGGTGTGGGTGTAA GTGGACATCCTTTCCTAAATAAATATGATGATGTTGAAA ATTCAGGGAGTGGTGGTAACCCTGGACAGGATAACAG GGTTAATGTAGGTATGGATTATAAACAAACACAATTATG CATGGTTGGATGTGCCCCCCCTTTGGGCGAGCATTGG GGTAAAGGTAAACAGTGTACTAATACACCTGTACAGGC TGGTGACTGCCCGCCCTTAGAACTTATTACCAGTGTTA TACAGGATGGCGATATGGTTGACACAGGCTTTGGTGCT ATGAATTTTGCTGATTTGCAGACCAATAAATCAGATGTT CCTATTGACATATGTGGCACTACATGTAAATATCCAGAT TATTTACAAATGGCTGCAGACCCATATGGTGATAGATTA TTTTTTTTTCTACGGAAGGAACAAATGTTTGCCAGACAT TTTTTTAACAGGGCTGGCGAGGTGGGGGAACCTGTGC CTGATACACTTATAATTAAGGGTAGTGGAAATCGCACG TCTGTAGGGAGTAGTATATATGTTAACACCCCGAGCGG CTCTTTGGTGTCCTCTGAGGCACAATTGTTTAATAAGCC ATATTGGCTACAAAAAGCCCAGGGACATAACAATGGTA TTTGTTGGGGTAATCAACTGTTTGTTACTGTGGTAGATA CCACACGCAGTACCAACATGACATTATGTGCATCCGTA ACTACATCTTCCACATACACCAATTCTGATTATAAAGAG TACATGCGTCATGTGGAAGAGTATGATTTACAATTTATT TTTCAATTATGTAGCATTACATTGTCTGCTGAAGTAATG GCCTATATTCACACAATGAATCCCTCTGTTTTGGAAGAC TGGAACTTTGGGTTATCGCCTCCCCCAAATGGTACATT AGAAGATACCTATAGGTATGTGCAGTCACAGGCCATTA CCTGTCAAAAGCCCACTCCTGAAAAGGAAAAGCCAGAT CCCTATAAGAACCTTAGTTTTTGGGAGGTTAATTTAAAA GAAAAGTTTTCTAGTGAATTGGATCAGTATCCTTTGGGA CGCAAGTTTTTGTTACAAAGTGGATATAGGGGACGGTC CTCTATTCGTACAGGTGTTAAGCGCCCTGCTGTTTCCA AAGCCTCTGCTGCCCCTAAACGTAAGCGCGCCAAAACT AAAAGGTAA	ATGTGGCGGCCTTCAGATTCAACTGTCTATGTGCCCCC TCCAAACCCCGTGTCAAAAGTCGTCGCTACCGATGCTT ATGTCACCAGAACCAATATCTTTTACCACGCTAGCTCCT CTAGGCTGCTGGCAGTGGGCCATCCATATTTCTCAATT AAGCGCGCCAACAAGACAGTGGTCCCCAAGGTGTCTG GCTACCAGTATAGGGTCTTTAAGGTGGTCCTGCCTGAC CCAAACAAATTTGCTCTGCCCGACAGTTCACTGTTCGA TCCTACCACACAGCGGCTGGTGTGGGCATGCACTGGC CTGGAAGTCGGAAGAGGACAGCCACTGGGAGTGGGA GTCTCCGGACACCCCTTCCTGAATAAGTACGACGATGT GGAGAACAGCGGATCCGGAGGAAATCCAGGACAGGAC AACCGAGTGAATGTCGGCATGGATTATAAACAGACCCA GCTGTGCATGGTGGGATGTGCACCACCTCTGGGAGAA CATTGGGGCAAGGGGAAACAGTGCACTAACACCCCTG TGCAGGCTGGAGATTGTCCACCCCTGGAGCTGATCAC CTCCGTGATTCAGGACGGCGATATGGTCGACACAGGA TTTGGCGCTATGAACTTCGCAGATCTGCAGACAAATAA GAGCGACGTGCCTATCGATATTTGCGGGACTACCTGTA AATACCCTGACTATCTGCAGATGGCCGCTGACCCATAC GGAGATCGCCTGTTCTTTTTCCTGCGAAAGGAACAGAT GTTCGCCCGACACTTTTTCAATCGAGCTGGAGAAGTGG GAGAACCAGTCCCTGATACCCTGATCATCAAGGGGAGT GGAAATAGGACATCAGTGGGGAGCTCCATCTACGTCAA CACTCCTTCTGGAAGTCTGGTGTCTAGTGAGGCACAGC TGTTTAACAAGCCATATTGGCTGCAGAAAGCCCAGGGG CATAACAATGGAATTTGCTGGGGCAATCAGCTGTTCGT GACCGTGGTCGACACAACTCGAAGCACCAACATGACA CTGTGTGCCTCCGTGACCACATCAAGCACATACACTAA CTCCGACTACAAGGAGTATATGCGCCACGTGGAGGAA TATGATCTGCAGTTTATCTTCCAGCTGTGCTCCATTACT CTGTCTGCCGAAGTGATGGCTTACATCCATACCATGAA CCCATCTGTCCTGGAGGACTGGAATTTTGGACTGAGTC CTCCACCCAACGGCACTCTGGAGGATACCTACAGATAT GTGCAGAGTCAGGCAATTACATGTCAGAAGCCAACTCC CGAGAAGGAAAAACCTGACCCATATAAAAACCTGTCTT TTTGGGAAGTGAATCTGAAGGAAAAATTCTCCTCTGAG CTGGATCAGTACCCCCTGGGCCGGAAGTTCCTGCTGC AGAGCGGATATCGGGGCAGAAGTTCAATCAGAACAGG GGTGAAGAGGCCCGCAGTCTCAAAAGCCAGCGCAGCC CCTAAGAGGAAACGCGCTAAGACTAAAAGA	MWRPSDSTVYVPPPNP VSKVVATDAYVTRTNIF YHASSSRLLAVGHPYFS IKRANKTVVPKVSGYQY RVFKVVLPDPNKFALPD SSLFDPTTQRLVWACT GLEVGRGQPLGVGVSG HPFLNKYDDVENSGSG GNPGQDNRVNVGMDY KQTQLCMVGCAPPLGE HWGKGKQCTNTPVQA GDCPPLELITSVIQDGD MVDTGFGAMNFADLQT NKSDVPIDICGTTCKYP DYLQMAADPYGDRLFF FLRKEQMFARHFFNRA GEVGEPVPDTLIIKGSG NRTSVGSSIYVNTPSGS LVSSEAQLFNKPYWLQ KAQGHNNGICWGNQLF VTVVDTTRSTNMTLCAS VTTSSTYTNSDYKEYM RHVEEYDLQFIFQLCSIT LSAEVMAYIHTMNPSVL EDWNFGLSPPPNGTLE DTYRYVQSQAITCQKPT PEKEKPDPYKNLSFWE VNLKEKFSSELDQYPLG RKFLLQSGYRGRSSIRT GVKRPAVSKASAAPKR KRAKTKR	25.2%

HPV6 L2	ATGGCACATAGTAGGGCCCGACGACGCAAGCGTGCGT CAGCTACACAGCTATATCAAACATGTAAACTCACTGGA ACATGCCCCCCAGATGTAATTCCTAAGGTGGAGCACAA CACCATTGCAGATCAAATATTAAAATGGGGAAGTTTGG GGGTGTTTTTTGGAGGGTTGGGTATAGGCACGGGTTC CGGCACTGGGGGTCGTACTGGCTATGTTCCCTTACAAA CTTCTGCAAAACCTTCTATTACTAGTGGGCCTATGGCT CGTCCTCCTGTGGTGGTGGAGCCTGTGGCCCCTTCGG ATCCATCTATTGTGTCTTTAATTGAAGAATCGGCAATCA TTAACGCAGGGGCGCCTGAAATTGTGCCCCCTGCACA CGGTGGGTTTACAATTACATCCTCTGAAACAACTACCC CTGCAATATTGGATGTATCAGTTACTAGTCACACTACTA CTAGTATATTTAGAAATCCTGTCTTTACAGAACCTTCTG TAACACAACCCCAACCACCCGTGGAGGCTAATGGACAT ATATTAATTTCTGCACCCACTGTAACGTCACACCCTATA GAGGAAATTCCTTTAGATACTTTTGTGGTATCATCTAGT GATAGCGGTCCTACATCCAGTACCCCTGTTCCTGGTAC TGCACCTCGGCCTCGTGTGGGCCTATATAGTCGTGCAT TGCACCAGGTGCAGGTTACAGACCCTGCATTTCTTTCC ACTCCTCAACGCTTAATTACATATGATAACCCTGTATAT GAAGGGGAGGATGTTAGTGTACAATTTAGTCATGATTC TATACACAATGCACCTGATGAGGCTTTTATGGACATAAT TCGTTTGCACAGACCTGCCATTGCGTCCCGACGTGGC CTTGTGCGGTACAGTCGCATTGGACAACGGGGGTCTA TGCACACTCGCAGCGGAAAGCACATAGGGGCCCGCAT TCATTATTTTTATGATATTTCACCTATTGCACAGGCTGC AGAAGAAATAGAAATGCACCCTCTTGTGGCTGCACAGG ATGATACATTTGATATTTATGCTGAATCTTTTGAACCTG GCATTAACCCTACCCAACACCCTGTTACAAATATATCAG ATACATATTTAACTTCCACACCTAATACAGTTACACAAC CGTGGGGTAACACCACAGTTCCATTGTCACTTCCTAAT GACCTGTTTTTACAATCTGGCCCTGATATAACTTTTCCT ACTGCACCTATGGGAACACCCTTTAGTCCTGTAACTCC TGCTTTACCTACAGGCCCTGTTTTCATTACAGGTTCTGG ATTTTATTTGCATCCTGCATGGTATTTTGCACGTAAACG CCGTAAACGTATTCCCTTATTTTTTTCAGATGTGGCGGC CTAG	ATGGCACACTCAAGGGCACGAAGAAGAAAGAGGGCAT CCGCTACCCAGCTGTACCAGACCTGTAAACTGACCGG CACCTGTCCACCTGATGTCATCCCTAAGGTGGAGCATA ACACCATCGCCGACCAGATTCTGAAATGGGGCAGTCT GGGGGTGTTCTTTGGCGGGCTGGGCATTGGGACTGGA TCAGGAACCGGAGGACGAACAGGATACGTGCCACTGC AGACTAGCGCTAAGCCCTCTATCACCAGTGGACCTATG GCAAGACCCCCTGTGGTCGTGGAACCTGTCGCCCCAT CAGATCCCAGCATCGTGTCCCTGATTGAGGAAAGCGCT ATCATTAATGCAGGAGCTCCAGAGATCGTGCCACCAGC ACATGGGGGCTTCACCATTACCAGCTCCGAAACCACAA CTCCTGCTATCCTGGACGTCTCTGTGACCAGTCACACC ACAACTTCCATCTTCAGGAACCCTGTCTTTACTGAGCC ATCTGTGACCCAGCCCCAGCCTCCAGTCGAAGCAAAT GGACATATCCTGATTAGTGCCCCAACAGTGACTTCACA CCCTATCGAGGAAATTCCACTGGACACCTTTGTCGTGT CTAGTTCAGATTCCGGACCAACAAGCTCCACTCCAGTC CCTGGAACAGCACCACGACCACGAGTGGGACTGTACT CCCGAGCTCTGCATCAGGTCCAGGTGACCGATCCAGC ATTCCTGTCTACCCCCCAGCGCCTGATTACATACGATA ACCCCGTGTATGAGGGGGAAGATGTCAGCGTGCAGTT TTCACACGACAGCATCCATAATGCTCCAGACGAGGCAT TCATGGATATCATTAGACTGCACAGGCCCGCAATTGCC TCTCGGAGAGGCCTGGTGCGCTATAGTCGAATCGGAC AGAGGGGCTCCATGCACACACGCTCTGGGAAACATAT CGGAGCCCGCATTCACTACTTTTATGACATCAGCCCCA TTGCTCAGGCCGCTGAGGAAATTGAGATGCATCCTCTG GTGGCAGCCCAGGACGATACCTTCGATATCTACGCCG AGAGCTTTGAACCAGGCATTAACCCCACACAGCACCCT GTGACTAATATCAGCGACACCTATCTGACCTCCACACC TAACACTGTCACCCAGCCATGGGGGAATACCACAGTG CCACTGTCACTGCCCAACGATCTGTTCCTGCAGAGCG GACCTGACATCACCTTTCCTACAGCACCAATGGGCACA CCCTTCAGTCCTGTCACACCAGCCCTGCCCACTGGCC CTGTGTTCATTACTGGGTCTGGATTTTACCTGCACCCT GCCTGGTATTTCGCTCGGAAGAGGCGCAAAAGAATCC CACTGTTCTTTTCCGATGTGGCTGCA	MAHSRARRRKRASATQ LYQTCKLTGTCPPDVIP KVEHNTIADQILKWGSL GVFFGGLGIGTGSGTG GRTGYVPLQTSAKPSIT SGPMARPPVVVEPVAP SDPSIVSLIEESAIINAGA PEIVPPAHGGFTITSSET TTPAILDVSVTSHTTTSI FRNPVFTEPSVTQPQP PVEANGHILISAPTVTSH PIEEIPLDTFVVSSSDSG PTSSTPVPGTAPRPRV GLYSRALHQVQVTDPA FLSTPQRLITYDNPVYE GEDVSVQFSHDSIHNA PDEAFMDIIRLHRPAIAS RRGLVRYSRIGQRGSM HTRSGKHIGARIHYFYDI SPIAQAAEEIEMHPLVA AQDDTFDIYAESFEPGI NPTQHPVTNISDTYLTS TPNTVTQPWGNTTVPL SLPNDLFLQSGPDITFP TAPMGTPFSPVTPALPT GPVFITGSGFYLHPAWY FARKRRKRIPLFFSDVA A	26.1%
HPV11 E1	ATGGCGGACGATTCAGGTACAGAAAATGAGGGGTCGG GGTGTACAGGATGGTTTATGGTAGAAGCCATAGTAGAG CACACTACAGGTACACAAATATCAGAAGATGAGGAAGA GGAGGTGGAGGACAGTGGGTATGACATGGTGGACTTT ATTGATGACAGGCATATTACACAAAATTCTGTGGAAGC ACAGGCATTGTTTAATAGGCAGGAGGCGGATGCTCATT ATGCGACTGTGCAGGACCTAAAACGAAAGTATTTAGGC AGTCCATATGTAAGTCCTATAAGCAATGTAGCTAATGCA GTAGAAAGTGAGATAAGTCCACGGTTAGACGCCATTAA ACTTACAACACAGCCAAAAAAGGTAAAGCGACGGCTGT TTGAAACACGGGAATTAACGGACAGTGGATATGGCTAT TCTGAAGTGGAAGCTGCAACGCAGGTAGAGAAACATG GCGACCCGGAAAATGGGGGAGATGGTCAGGAAAGGG ACACAGGGAGGGACATAGAGGGTGAGGGGGTGGAAC ATAGAGAGGCGGAAGCAGTAGACGACAGCACCCGAGA GCATGCAGACACATCAGGAATATTAGAATTACTAAAATG TAAGGATATACGATCTACATTACATGGTAAGTTTAAAGA CTGCTTTGGGCTGTCATTTGTTGATTTAATTAGGCCATT TAAAAGTGATAGAACCACATGTGCCGATTGGGTGGTTG CAGGATTTGGTATACATCATAGCATAGCAGATGCATTTC AAAAGTTAATTGAGCCATTAAGTTTATATGCACATATAC AATGGCTTACAAATGCATGGGGAATGGTACTATTAGTA TTAATAAGGTTTAAAGTAAATAAGAGCAGATGTACCGTG GCACGTACATTAGGTACGTTATTAAATATACCTGAAAAT CACATGTTAATTGAGCCTCCTAAAATACAAAGTGGCGT ACGAGCCCTGTATTGGTTTAGGACAGGCATTTCAAATG CAAGTACAGTTATAGGGGAGGCGCCGGAATGGATAAC GCGCCAGACCGTTATTGAACATAGTTTGGCTGACAGTC AATTTAAATTAACTGAAATGGTGCAGTGGGCATATGATA ATGATATTTGTGAAGAAAGTGAGATAGCATTTGAATATG CACAGCGTGGAGACTTTGACTCCAATGCAAGGGCCTTT TTAAATAGTAATATGCAGGCTAAATATGTAAAAGATTGT GCAATTATGTGCAGACATTATAAACATGCAGAAATGAAA AAGATGTCTATTAAACAATGGATTAAGTATAGGGGTACT AAAGTTGACAGTGTAGGTAACTGGAAGCCAATTGTGCA GTTTCTAAGACATCAAAACATAGAATTTATTCCATTTTTA AGCAAACTAAAATTATGGCTGCACGGAACGCCCAAAAA AAATTGTATAGCCATTGTAGGGCCACCTGACACTGGGA AGTCGTGCTTTTGCATGAGTTTAATTAAGTTTTTGGGGG GAACAGTTATTAGTTATGTTAATTCCTGCAGCCATTTCT GGCTACAGCCACTAACGGATGCAAAAGTGGCATTATTG GATGATGCCACACAACCATGTTGGACATATATGGATAC ATATATGAGAAACCTATTAGATGGTAATCCTATGAGCAT AGATAGAAAACATAGAGCATTAACATTAATTAAGTGTCC ACCGCTACTGGTTACATCAAATATAGACATTAGCAAAGA GGAGAAATACAAATATTTACATAGTAGAGTTACCACATT TACATTTCCAAATCCATTCCCCTTTGACAGAAATGGGAA TGCAGTATATGAACTATCAGATGCAAACTGGAAATGTTT CTTTGAAAGACTGTCGTCCAGCCTAGACATTGAGGATT CAGAGGACGAGGAAGATGGAAGCAATAGCCAAGCGTT TAGATGCGTGCCAGGATCAGTTGTTAGAACTTTATGA	ATGGCAGATGACAGCGGGACCGAGAATGAGGGGAGC GGATGCACTGGGTGGTTTATGGTGGAGGCAATCGTGG AACACACTACTGGCACCCAGATCAGCGAGGACGAGGA AGAGGAAGTGGAAGATTCCGGATACGACATGGTCGATT TTATCGACGATCGGCACATTACTCAGAACAGCGTGGAG GCACAGGCCCTGTTCAATAGACAGGAAGCTGACGCAC ATTATGCCACCGTGCAGGATCTGAAGAGGAAATACCTG GGCAGCCCCTATGTCTCTCCTATCAGTAACGTGGCCAA TGCTGTCGAGTCAGAAATCAGCCCAAGACTGGACGCC ATTAAGCTGACCACACAGCCCAAGAAAGTGAAACGGAG ACTGTTTGAGACAAGGGAACTGACTGATAGCGGGTAC GGATATTCCGAGGTGGAAGCCGCTACCCAGGTCGAGA AGCACGGCGACCCAGAAAACGGAGGGGATGGACAGG AGCGAGACACAGGGCGGGATATCGAGGGCGAAGGGG TGGAGCACAGAGAGGCAGAAGCCGTCGACGATTCCAC TAGGGAGCATGCCGACACCTCTGGGATCCTGGAACTG CTGAAGTGCAAAGATATTCGCTCCACCCTGCATGGAAA GTTCAAAGACTGTTTTGGCCTGTCTTTCGTGGATCTGAT CCGCCCATTCAAGAGTGACCGAACTACCTGCGCCGATT GGGTGGTCGCTGGATTTGGCATCCACCATTCAATTGCT GACGCATTCCAGAAACTGATCGAGCCCCTGAGCCTGTA CGCACACATTCAGTGGCTGACAAACGCCTGGGGCATG GTGCTGCTGGTCCTGATCCGCTTTAAGGTGAACAAGTC TAGATGTACTGTCGCCAGGACCCTGGGGACACTGCTG AACATTCCTGAGAATCATATGCTGATCGAACCCCCTAA GATTCAGAGTGGAGTGCGGGCTCTGTATTGGTTCAGAA CAGGCATCTCCAACGCATCTACTGTGATTGGGGAGGC CCCAGAATGGATCACTCGGCAGACCGTCATTGAGCAC AGTCTGGCTGACTCACAGTTTAAGCTGACCGAGATGGT GCAGTGGGCATACGACAACGATATCTGCGAGGAAAGC GAGATTGCTTTCGAATATGCACAGAGGGGCGACTTTGA TAGTAATGCCCGCGCTTTCCTGAACTCAAATATGCAGG CTAAGTACGTGAAAGACTGCGCAATCATGTGTAGGCAC TATAAGCATGCCGAGATGAAGAAAATGTCCATCAAGCA GTGGATCAAGTACCGCGGGACTAAGGTGGATTCTGTC GGAAACTGGAAACCCATTGTGCAGTTTCTGCGGCACCA GAATATCGAGTTCATTCCTTTTCTGTCCAAGCTGAAACT GTGGCTGCATGGCACACCAAAGAAAAACTGCATCGCC ATTGTGGGGCCACCCGACACTGGAAAGTCTTGCTTTTG TATGAGTCTGATCAAATTCCTGGGAGGCACAGTGATTT CTTATGTCAATAGTTGCTCACACTTCTGGCTGCAGCCC CTGACTGACGCAAAGGTGGCCCTGCTGGACGATGCAA CCCAGCCTTGTTGGACCTACATGGATACATATATGAGA AACCTGCTGGACGGGAATCCCATGAGCATCGATAGGA AGCACCGCGCTCTGACCCTGATCAAGTGTCCTCCACTG CTGGTGACATCAAACATCGATATTAGCAAGGAGGAAAA GTACAAATATCTGCATAGCCGCGTGACAACTTTCACCT TTCCCAACCCTTTCCCATTTGACCGAAACGGCAATGCC GTCTACGAGCTGTCCGATGCTAATTGGAAATGCTTCTT TGAAAGGCTGAGCTCCTCTCTGGACATCGAGGATAGTG AAGACGAGGAAGATGGAAGCAATTCCCAGGCCTTCCG ATGTGTGCCTGGCTCAGTGGTCCGGACACTG	MADDSGTENEGSGCT GWFMVEAIVEHTTGTQI SEDEEEEVEDSGYDMV DFIDDRHITQNSVEAQA LFNRQEADAHYATVQD LKRKYLGSPYVSPISNV ANAVESEISPRLDAIKLT TQPKKVKRRLFETRELT DSGYGYSEVEAATQVE KHGDPENGGDGQERD TGRDIEGEGVEHREAE AVDDSTREHADTSGILE LLKCKDIRSTLHGKFKD CFGLSFVDLIRPFKSDR TTCADWVVAGFGIHHSI ADAFQKLIEPLSLYAHIQ WLTNAWGMVLLVLIRFK VNKSRCTVARTLGTLLN IPENHMLIEPPKIQSGVR ALYWFRTGISNASTVIG EAPEWITRQTVIEHSLA DSQFKLTEMVQWAYDN DICEESEIAFEYAQRGD FDSNARAFLNSNMQAK YVKDCAIMCRHYKHAE MKKMSIKQWIKYRGTK VDSVGNWKPIVQFLRH QNIEFIPFLSKLKLWLHG TPKKNCIAIVGPPDTGK SCFCMSLIKFLGGTVISY VNSCSHFWLQPLTDAK VALLDDATQPCWTYMD TYMRNLLDGNPMSIDR KHRALTLIKCPPLLVTSN IDISKEEKYKYLHSRVTT FTFPNPFPFDRNGNAV YELSDANWKCFFERLS SSLDIEDSEDEEDGSNS QAFRCVPGSVVRTL	25.7%

HPV11 E2	ATGGAAGCAATAGCCAAGCGTTTAGATGCGTGCCAGG ATCAGTTGTTAGAACTTTATGAAGAAAACAGTATTGATA TACACAAACACATTATGCATTGGAAATGCATACGATTGG AAAGTGTATTACTACACAAAGCAAAACAAATGGGCCTG AGCCACATCGGGTTACAAGTAGTACCACCATTAACTGT GTCAGAGACTAAAGGACATAATGCTATTGAAATGCAAA TGCATTTAGAATCCTTAGCAAAAACTCAGTATGGTGTG GAACCTTGGACATTACAGGACACCAGTTATGAAATGTG GCTAACACCACCCAAACGGTGCTTTAAAAAACAGGGAA ATACTGTGGAGGTAAAATTTGATGGCTGTGAAGACAAT GTAATGGAGTATGTGGTATGGACACATATATACCTGCA GGACAACGACTCATGGGTAAAAGTAACTAGTTCCGTAG ATGCCAAGGGCATATATTATACATGTGGACAATTTAAAA CATATTATGTAAATTTTAATAAAGAGGCACAAAAGTATG GTAGTACCAATCATTGGGAAGTATGTTATGGCAGCACA GTTATATGTTCTCCTGCATCTGTATCTAGCACTGTACGA GAAGTATCCATTGCTGAACCTACTACATACACCCCCGC ACAGACCACCGCCCCTACAGTGTCCGCCTGCACCACG GAAGACGGCGTGTCGGCGCCGCCTAGGAAGCGAGCA CGTGGACCGTCCACTAACAACACCCTGTGTGTGGCCA ACATCAGATCCGTGGACAGTACAATCAACAACATCGTC ACTGACAATTACAACAAGCACCAAAGAAGGAACAACTG TCACAGTGCAGCTACGCCTATAGTGCAACTGCAAGGTG ATTCCAATTGTTTAAAATGTTTTAGATATAGACTGAATGA CAAATATAAACATTTGTTTGAATTAGCATCTTCAACGTG GCATTGGGCCTCACCTGAGGCACCACATAAAAATGCAA TTGTAACATTAACATATAGCAGTGAGGAACAACGTCAG CAATTTTTAAACAGTGTAAAAATACCACCCACCATTAGG CATAAGGTGGGGTTTATGTCATTACATTTATTGTAA	ATGGAAGCAATCGCAAAGAGACTGGATGCCTGTCAGG ATCAGCTGCTGGAACTGTATGAGGAAAATAGTATCGAC ATTCACAAACATATCATGCACTGGAAGTGCATTCGGCT GGAGTCCGTGCTGCTGCACAAGGCCAAACAGATGGGC CTGTCTCATATCGGGCTGCAGGTGGTCCCCCCTCTGA CAGTGTCTGAAACTAAGGGGCACAACGCAATTGAGATG CAGATGCATCTGGAAAGTCTGGCCAAAACCCAGTACG GAGTGGAGCCTTGGACCCTGCAGGACACATCTTATGAA ATGTGGCTGACACCACCCAAGCGGTGCTTCAAGAAACA GGGGAACACTGTGGAGGTCAAATTTGACGGATGTGAG GATAATGTGATGGAATACGTGGTCTGGACCCACATCTA TCTGCAGGACAACGATAGCTGGGTGAAGGTCACAAGC TCCGTGGATGCAAAAGGAATCTACTATACTTGCGGCCA GTTCAAGACCTACTATGTCAACTTTAATAAGGAGGCTCA GAAATACGGCTCAACTAATCATTGGGAAGTGTGCTATG GGAGCACCGTCATCTGTAGTCCAGCTTCAGTGTCTAGT ACCGTGCGAGAGGTCAGCATTGCAGAACCTACCACATA CACCCCAGCACAGACTACCGCCCCCACAGTGTCTGCC TGCACAACTGAGGACGGAGTCAGTGCCCCTCCAAGGA AACGCGCTCGAGGCCCTTCCACTAACAATACCCTGTGT GTGGCCAATATCAGGAGCGTCGACTCCACAATCAACAA CATCGTGACTGATAACTACAATAAGCACCAGCGGAGAA ACAATTGTCATAGTGCCGCTACTCCCATTGTGCAGCTG CAGGGCGACTCAAACTGCCTGAAATGTTTCCGGTACAG ACTGAATGATAAGTATAAACACCTGTTTGAGCTGGCCT CAAGCACATGGCACTGGGCTTCCCCAGAAGCACCCCA TAAGAACGCTATCGTGACCCTGACATACTCCTCTGAGG AACAGAGGCAGCAGTTCCTGAATAGCGTGAAGATCCC CCCTACCATTCGCCACAAAGTCGGGTTTATGTCCCTGC ATCTGCTG	MEAIAKRLDACQDQLLE LYEENSIDIHKHIMHWK CIRLESVLLHKAKQMGL SHIGLQVVPPLTVSETK GHNAIEMQMHLESLAK TQYGVEPWTLQDTSYE MWLTPPKRCFKKQGNT VEVKFDGCEDNVMEYV VWTHIYLQDNDSWVKV TSSVDAKGIYYTCGQFK TYYVNFNKEAQKYGST NHWEVCYGSTVICSPA SVSSTVREVSIAEPTTY TPAQTTAPTVSACTTED GVSAPPRKRARGPSTN NTLCVANIRSVDSTINNI VTDNYNKHQRRNNCHS AATPIVQLQGDSNCLKC FRYRLNDKYKHLFELAS STWHWASPEAPHKNAI VTLTYSSEEQRQQFLN SVKIPPTIRHKVGFMSL HLL	24.1%
HPV11 E4	ATGGTAGTACCAATCATTGGGAAGTATGTTATGGCAGC ACAGTTATATGTTCTCCTGCATCTGTATCTAGCACTGTA CGAGAAGTATCCATTGCTGAACCTACTACATACACCCC CGCACAGACCACCGCCCCTACAGTGTCCGCCTGCACC ACGGAAGACGGCGTGTCGGCGCCGCCTAGGAAGCGA GCACGTGGACCGTCCACTAACAACACCCTGTGTGTGG CCAACATCAGATCCGTGGACAGTACAATCAACAACATC GTCACTGACAATTACAACAAGCACCAAAGAAGGAACAA CTGTCACAGTGCAGCTACGCCTTAG	ATGGTCGTCCCTATTATTGGAAAGTATGTGATGGCCGC TCAGCTGTATGTCCTGCTGCACCTGTATCTGGCTCTGT ATGAGAAGTATCCACTGCTGAACCTGCTGCACACCCCA CCTCATCGACCACCACCTCTGCAGTGCCCACCAGCAC CACGAAAGACAGCTTGTCGGAGAAGGCTGGGCAGCGA GCACGTGGACAGACCTCTGACCACACCATGCGTCTGG CCCACTTCTGATCCTTGGACCGTGCAGAGTACTACCAG CTCCCTGACAATCACAACTTCTACTAAAGAAGGGACCA CAGTGACCGTCCAGCTGCGGCTG	MVVPIIGKYVMAAQLYV LLHLYLALYEKYPLLNLL HTPPHRPPPLQCPPAP RKTACRRRLGSEHVDR PLTTPCVWPTSDPWTV QSTTSSLTITTSTKEGTT VTVQLRL	26.1%
HPV11 E5A	ATGGAGGTAGTGCCTGTACAAATTGCTGCAGCAACAAC TACAACATTGATATTGCCTGTTGTTATTGCATTTGCAGT ATGTATTCTTAGTATTGTACTTATAATATTAATATCTGAT TTTGTAGTATATACATCTGTGCTGGTACTAACACTTCTT TTATATTTGCTTTTGTGGCTTTTATTAACAACCCCTTTGC AATTCTTTTTACTAACACTGTGTGTGTGCTATTTTCCTG CCTTTTATATACACATATACATTGTGCAAACGCAACAAT AA	ATGGAGGTCGTGCCAGTCCAGATTGCTGCCGCCACCA CCACTACCCTGATCCTGCCAGTCGTCATTGCCTTCGCT GTGTGTATCCTGTCTATTGTGCTGATCATTCTGATCAGC GACTTCGTGGTCTACACTTCCGTGCTGGTCCTGACCCT GCTGCTGTATCTGCTGCTGTGGCTGCTGCTGACCACAC CCCTGCAGTTCTTTCTGCTGACACTGTGCGTGTGTTAC TTCCCTGCCTTTTACATCCACATCTACATCGTCCAGACC CAGCAG	MEVVPVQIAAATTTTLIL PVVIAFAVCILSIVLIILIS DFVVYTSVLVLTLLLYLL LWLLLTTPLQFFLLTLCV CYFPAFYIHIYIVQTQQ	30.4%
HPV11 E5B	ATGGTGATGTTAACCTGTCACTTAAATGATGGTGATACA TGGTTGTTTCTGTGGTTGTTTACTGCATTTGTTGTAGCT GTACTTGGATTGTTGTTACTACATTACAGGGCTGTACAT GGTACTGAAAAAACTAAATGTGCTAAGTGTAAATCAAAC CGCAATACTACTGTGGATTATGTGTATATGTCACATGGT GATAATGGAGATTATGTGTACATGAACTAG	ATGGTCATGCTGACTTGTCACCTGAACGATGGAGATAC CTGGCTGTTTCTGTGGCTGTTTACCGCCTTTGTGGTCG CAGTGCTGGGCCTGCTGCTGCTGCACTACCGGGCCGT GCATGGCACTGAGAAGACCAAATGCGCTAAGTGTAAAA GCAACAGAAATACCACAGTGGACTACGTCTATATGTCC CACGGCGACAACGGGGATTACGTCTATATGAAT	MVMLTCHLNDGDTWLF LWLFTAFVVAVLGLLLL HYRAVHGTEKTKCAKC KSNRNTTVDYVYMSHG DNGDYVYMN	24.9%
HPV11 E6	ATGGAAAGTAAAGATGCCTCCACGTCTGCAACATCTAT AGACCAGTTGTGCAAGACGTTTAATCTTTCTTTGCACAC TCTGCAAATTCAGTGCGTGTTTTGCAGGAATGCACTGA CCACCGCAGAGATATATGCATATGCCTATAAGAACCTA AAGGTTGTGTGGCGAGACAACTTTCCCTTTGCAGCGTG TGCCTGTTGCTTAGAACTGCAAGGGAAAATTAACCAAT ATAGACACTTTAATTATGCTGCATATGCACCTACAGTAG AAGAAGAAACCAATGAAGATATTTTAAAAGTGTTAATTC GTTGTTACCTGTGTCACAAGCCGTTGTGTGAAATAGAA AAACTAAAGCACATATTGGGAAAGGCACGCTTCATAAA ACTAAATAACCAGTGGAAGGGTCGTTGCTTACACTGCT GGACAACATGCATGGAAGACTTGTTACCCTAA	ATGGAGAGCAAGGACGCCTCAACATCCGCAACAAGCA TCGACCAGCTGTGTAAAACTTTCAACCTGTCACTGCAT ACCCTGCAGATTCAGTGCGTGTTCTGTAGGAACGCCCT GACCACAGCTGAAATCTACGCTTATGCATACAAGAACC TGAAAGTGGTCTGGCGCGACAATTTCCCATTTGCCGCT TGCGCATGCTGTCTGGAACTGCAGGGCAAGATTAACCA GTATCGACACTTCAATTATGCAGCCTACGCTCCCACAG TGGAGGAAGAGACCAATGAGGATATCCTGAAGGTCCT GATTAGATGCTACCTGTGTCACAAACCTCTGTGCGAAA TCGAGAAGCTGAAACATATTCTGGGCAAGGCCCGGTTT ATCAAACTGAACAATCAGTGGAAAGGGAGATGCCTGCA TTGTTGGACTACCTGTATGGAGGACCTGCTGCCC	MESKDASTSATSIDQLC KTFNLSLHTLQIQCVFC RNALTTAEIYAYAYKNL KVVWRDNFPFAACACC LELQGKINQYRHFNYAA YAPTVEEETNEDILKVLI RCYLCHKPLCEIEKLKHI LGKARFIKLNNQWKGR CLHCWTTCMEDLLP	24.8%
HPV11 E7	ATGCATGGAAGACTTGTTACCCTAAAGGATATAGTACTA GACCTGCAGCCTCCTGACCCTGTAGGGTTACATTGCTA TGAGCAATTAGAAGACAGCTCAGAAGATGAGGTGGACA AGGTGGACAAACAAGACGCACAACCTTTAACACAACAT TACCAAATACTGACCTGTTGCTGTGGATGTGACAGCAA CGTCCGACTGGTTGTGGAGTGCACAGACGGAGACATC AGACAACTACAAGACCTTTTGCTGGGCACACTAAATATT GTGTGTCCCATCTGCGCACCAAAACCATAA	ATGCACGGAAGACTGGTGACACTGAAAGACATCGTCCT GGATCTGCAGCCCCCTGACCCCGTCGGACTGCACTGC TATGAACAGCTGGAGGACAGCTCCGAGGACGAAGTGG ATAAGGTCGACAAACAGGATGCCCAGCCACTGACCCA GCACTACCAGATCCTGACATGCTGTTGCGGCTGTGACT CTAACGTGCGGCTGGTGGTCGAATGCACTGACGGCGA TATTAGACAGCTGCAGGATCTGCTGCTGGGGACCCTG AATATCGTCTGTCCCATTTGCGCTCCCAAGCCT	MHGRLVTLKDIVLDLQP PDPVGLHCYEQLEDSS EDEVDKVDKQDAQPLT QHYQILTCCCGCDSNV RLVVECTDGDIRQLQDL LLGTLNIVCPICAPKP	25.9%

HPV11 L1	ATGTGGCGGCCTAGCGACAGCACAGTATATGTGCCTC CTCCCAACCCTGTATCCAAGGTTGTTGCCACGGATGCG TATGTTAAACGCACCAACATATTTTATCATGCCAGCAGT TCTAGACTCCTTGCTGTGGGACATCCATATTACTCTATC AAAAAAGTTAACAAAACAGTTGTACCAAAGGTGTCTGG ATATCAATATAGAGTGTTTAAGGTAGTGTTGCCAGATCC TAACAAGTTTGCATTACCTGATTCATCCCTGTTTGACCC CACTACACAGCGTTTAGTATGGGCGTGCACAGGGTTG GAGGTAGGCAGGGGTCAACCTTTAGGCGTTGGTGTTA GTGGGCATCCATTGCTAAACAAATATGATGATGTAGAA AATAGTGGTGGGTATGGTGGTAATCCTGGTCAGGATAA TAGGGTTAATGTAGGTATGGATTATAAACAAACCCAGC TATGTATGGTGGGCTGTGCTCCACCGTTAGGTGAACAT TGGGGTAAGGGTACACAATGTTCAAATACCTCTGTACA AAATGGTGACTGCCCCCCGTTGGAACTTATTACCAGTG TTATACAGGATGGGGACATGGTTGATACAGGCTTTGGT GCTATGAATTTTGCAGACTTACAAACCAATAAATCGGAT GTTCCCCTTGATATTTGTGGAACTGTCTGCAAATATCCT GATTATTTGCAAATGGCTGCAGACCCTTATGGTGATAG GTTGTTTTTTTATTTGCGAAAGGAACAAATGTTTGCTAG ACACTTTTTTAATAGGGCCGGTACTGTGGGGGAACCTG TGCCTGATGACCTGTTGGTAAAAGGGGGTAATAACAGA TCATCTGTAGCTAGTAGTATTTATGTACATACACCTAGT GGCTCATTGGTGTCTTCAGAGGCTCAATTATTTAATAAA CCATATTGGCTTCAAAAGGCTCAGGGACATAACAATGG TATTTGCTGGGGAAACCACTTGTTTGTTACTGTGGTAG ATACCACACGCAGTACAAATATGACACTATGTGCATCT GTGTCTAAATCTGCTACATACACTAATTCAGATTATAAG GAATACATGCGCCATGTGGAGGAGTTTGATTTACAGTT TATTTTTCAATTGTGTAGCATTACATTATCTGCAGAAGT CATGGCCTATATACACACAATGAATCCTTCTGTTTTGGA GGACTGGAACTTTGGTTTATCGCCTCCACCAAATGGTA CACTGGAGGATACTTATAGATATGTACAGTCACAGGCC ATTACCTGTCAGAAACCCACACCTGAAAAAGAAAAACA GGATCCCTATAAGGATATGAGTTTTTGGGAGGTTAACT TAAAAGAAAAGTTTTCAAGTGAATTAGATCAGTTTCCCC TTGGACGTAAGTTTTTATTGCAAAGTGGATATCGAGGA CGGACGTCTGCTCGTACAGGTATAAAGCGCCCAGCTG TGTCTAAGCCCTCTACAGCCCCCAAACGAAAACGTACC AAAACCAAAAAGTAA	ATGTGGCGGCCTTCTGATTCTACAGTCTATGTGCCTCC TCCAAACCCCGTGAGCAAAGTCGTCGCTACCGATGCCT ACGTGAAAAGAACCAACATCTTCTACCACGCAAGCTCC TCTCGACTGCTGGCCGTGGGCCATCCCTACTACAGTAT TAAGAAAGTCAACAAGACAGTGGTCCCTAAAGTGTCAG GCTACCAGTATCGCGTCTTTAAGGTGGTCCTGCCTGAC CCAAACAAGTTCGCTCTGCCCGACAGTTCACTGTTTGA TCCTACCACACAGCGACTGGTGTGGGCATGCACCGGA CTGGAAGTCGGAAGAGGACAGCCACTGGGAGTGGGC GTCTCTGGACACCCACTGCTGAACAAGTACGACGATGT GGAGAATAGTGGAGGATATGGAGGAAACCCAGGACAG GACAACAGGGTGAATGTCGGAATGGATTACAAGCAGA CACAGCTGTGCATGGTGGGATGTGCACCACCTCTGGG AGAACATTGGGGGAAAGGAACTCAGTGCAGTAACACCT CAGTGCAGAATGGAGACTGTCCACCCCTGGAGCTGAT CACCAGCGTGATTCAGGACGGCGATATGGTCGACACA GGCTTCGGGGCAATGAATTTTGCCGATCTGCAGACAAA CAAGTCCGACGTGCCTCTGGATATCTGCGGGACTGTCT GTAAATACCCTGATTATCTGCAGATGGCCGCTGACCCA TACGGAGATCGCCTGTTCTTTTATCTGCGAAAGGAACA GATGTTCGCTCGACACTTCTTTAACCGAGCAGGAACTG TGGGAGAGCCAGTCCCTGACGATCTGCTGGTGAAAGG GGGAAACAATCGCAGCTCCGTGGCCTCTAGTATCTACG TCCATACACCAAGTGGCTCACTGGTGTCAAGCGAGGC ACAGCTGTTCAATAAGCCCTATTGGCTGCAGAAAGCCC AGGGCCACAACAATGGGATTTGCTGGGGAAACCATCT GTTTGTGACCGTGGTCGACACTACCAGGTCTACTAATA TGACCCTGTGTGCCAGCGTGTCCAAGTCTGCTACATAC ACTAACTCCGACTACAAAGAATATATGCGCCACGTGGA GGAATTCGATCTGCAGTTCATCTTTCAGCTGTGCTCTAT TACTCTGAGTGCTGAAGTGATGGCATATATCCATACCA TGAATCCCTCCGTCCTGGAGGACTGGAACTTTGGACTG TCTCCTCCACCCAATGGCACACTGGAGGATACTTACAG ATATGTGCAGAGCCAGGCCATTACATGTCAGAAGCCAA CTCCCGAGAAGGAAAAACAGGACCCTTACAAAGATATG TCCTTCTGGGAAGTGAATCTGAAGGAAAAATTTTCCTCT GAGCTGGATCAGTTCCCACTGGGCAGAAAGTTTCTGCT GCAGTCAGGGTATCGGGGAAGAACCAGCGCTAGAACA GGGATTAAGAGGCCCGCCGTGAGCAAACCTTCCACCG CTCCAAAGAGGAAACGCACCAAGACAAAGAAA	MWRPSDSTVYVPPPNP VSKVVATDAYVKRTNIF YHASSSRLLAVGHPYY SIKKVNKTVVPKVSGYQ YRVFKVVLPDPNKFALP DSSLFDPTTQRLVWAC TGLEVGRGQPLGVGVS GHPLLNKYDDVENSGG YGGNPGQDNRVNVGM DYKQTQLCMVGCAPPL GEHWGKGTQCSNTSV QNGDCPPLELITSVIQD GDMVDTGFGAMNFADL QTNKSDVPLDICGTVCK YPDYLQMAADPYGDRL FFYLRKEQMFARHFFN RAGTVGEPVPDDLLVK GGNNRSSVASSIYVHTP SGSLVSSEAQLFNKPY WLQKAQGHNNGICWG NHLFVTVVDTTRSTNMT LCASVSKSATYTNSDYK EYMRHVEEFDLQFIFQL CSITLSAEVMAYIHTMN PSVLEDWNFGLSPPPN GTLEDTYRYVQSQAITC QKPTPEKEKQDPYKDM SFWEVNLKEKFSSELD QFPLGRKFLLQSGYRG RTSARTGIKRPAVSKPS TAPKRKRTKTKK	25.7%
HPV11 L2	ATGAAACCTAGGGCACGCAGACGTAAACGTGCGTCAG CCACACAACTATATCAAACATGCAAGGCCACTGGTACA TGTCCCCCAGATGTAATTCCTAAAGTTGAACATACTACT ATTGCAGATCAAATATTAAAATGGGGAAGCTTAGGGGT TTTTTTTGGTGGGTTAGGTATTGGTACAGGGGCTGGTA GTGGCGGTCGTGCAGGGTATATACCCTTGGGAAGCTC TCCCAAGCCTGCTATTACTGGGGGGCCAGCAGCACGT CCGCCAGTGCTTGTGGAGCCTGTTGCCCCTTCCGATC CCTCCATTGTGTCCTTAATTGAGGAGTCTGCTATTATTA ATGCTGGTGCACCTGAGGTGGTACCCCCTACACAGGG TGGCTTTACTATAACATCATCTGAATCGACTACACCTGC TATTTTAGATGTGTCTGTTACCAATCACACTACCACTAG TGTGTTTCAAAATCCCCTGTTTACAGAACCGTCTGTAAT ACAGCCCCAACCACCTGTGGAGGCCAGTGGTCACATA CTTATATCTGCCCCAACAATAACATCCCAACATGTAGAA GACATTCCACTAGACACTTTTGTTGTATCCTCTAGTGAT AGTGGACCTACATCCAGTACTCCTCTTCCTCGTGCTTTT CCTCGGCCTCGGGTGGGTTTGTATAGTCGTGCCTTACA GCAGGTACAGGTTACGGACCCCGCGTTTTTGTCCACG CCACAGCGATTGGTAACTTATGACAACCCTGTCTATGA AGGAGAAGATGTAAGTTTACAATTTACCCATGAGTCTAT CCACAATGCACCTGATGAAGCATTTATGGATATTATTAG ACTACATAGACCAGCTATAACGTCCAGACGGGGTCTTG TGCGTTTTAGTCGCATTGGGCAACGGGGGTCCATGTAC ACACGCAGTGGACAACATATAGGTGCCCGCATACATTA TTTTCAGGACATTTCACCAGTTACACAAGCTGCAGAGG AAATAGAACTGCACCCTCTAGTGGCTGCAGAAAATGAC ACGTTTGATATTTATGCTGAACCATTTGACCCTATCCCT GACCCTGTCCAACATTCTGTTACACAGTCTTATCTTACC TCCACACCTAATACCCTTTCACAATCGTGGGGTAATAC CACAGTCCCATTGTCAATCCCTAGTGACTGGTTTGTGC AGTCTGGGCCTGACATAACTTTTCCTACTGCATCTATG GGAACACCCTTTAGTCCTGTAACTCCTGCTTTACCTACA GGCCCTGTTTTTATTACAGGTTCTGACTTCTATTTGCAT CCTACATGGTACTTTGCACGCAGACGCCGTAAACGTAT TCCCTTATTTTTTACAGATGTGGCGGCCTAG	ATGAAACCAAGAGCAAGAAGAAGAAAAAGAGCATCCGC AACTCAGCTGTATCAGACCTGTAAAGCCACTGGAACCT GTCCCCCCGACGTGATCCCCAAGGTCGAGCACACCAC AATCGCCGACCAGATTCTGAAATGGGGCTCTCTGGGG GTGTTCTTTGGCGGGCTGGGAATCGGAACCGGAGCAG GAAGCGGAGGACGAGCTGGATACATCCCACTGGGAAG CTCCCCAAAGCCTGCTATTACAGGAGGACCTGCAGCTA GACCACCTGTGCTGGTCGAACCCGTCGCACCTTCCGA CCCATCTATCGTGAGTCTGATTGAGGAAAGCGCCATCA TTAACGCAGGAGCACCAGAGGTGGTCCCACCAACCCA GGGCGGGTTTACCATCACATCTAGTGAATCCACTACCC CTGCCATTCTGGATGTCAGTGTCACCAACCATACAACT ACCTCAGTGTTCCAGAATCCTCTGTTTACAGAGCCATC CGTCATCCAGCCCCAGCCTCCAGTGGAAGCATCCGGC CACATCCTGATTTCTGCCCCAACTATCACCAGTCAGCA TGTGGAGGACATTCCCCTGGATACCTTTGTGGTCTCAA GCTCCGACAGCGGCCCTACCTCTAGTACACCACTGCC CCGAGCCTTCCCTAGACCAAGGGTGGGGCTGTATTCT CGCGCTCTGCAGCAGGTGCAGGTCACTGATCCCGCAT TCCTGAGTACCCCTCAGAGGCTGGTGACATACGACAAC CCCGTCTATGAGGGAGAAGATGTGTCCCTGCAGTTTAC CCACGAGTCTATCCATAATGCTCCAGACGAAGCATTCA TGGATATCATTCGCCTGCACCGACCCGCTATCACAAGC CGGAGAGGCCTGGTGCGGTTTTCCAGAATTGGACAGA GGGGCTCAATGTACACTCGCAGCGGGCAGCACATCGG AGCACGCATTCATTATTTCCAGGATATCAGCCCTGTGA CTCAGGCAGCCGAGGAAATTGAGCTGCACCCACTGGT GGCTGCAGAAAATGACACCTTCGATATCTACGCCGAGC CATTTGACCCCATTCCTGATCCAGTGCAGCATTCCGTC ACACAGTCTTATCTGACAAGTACTCCCAACACTCTGTCA CAGAGCTGGGGCAATACAACTGTCCCACTGTCAATCCC CAGCGACTGGTTCGTGCAGTCTGGGCCTGATATTACTT TTCCAACCGCCTCAATGGGAACACCCTTCAGCCCTGTC ACACCAGCTCTGCCCACTGGACCTGTGTTCATCACTGG CAGCGACTTTTACCTGCACCCTACCTGGTATTTCGCCA GGCGCCGACGGAAAAGGATTCCACTGTTCTTTACCGAT GTGGCCGCT	MKPRARRRKRASATQL YQTCKATGTCPPDVIPK VEHTTIADQILKWGSLG VFFGGLGIGTGAGSGG RAGYIPLGSSPKPAITG GPAARPPVLVEPVAPS DPSIVSLIEESAIINAGAP EVVPPTQGGFTITSSES TTPAILDVSVTNHTTTSV FQNPLFTEPSVIQPQPP VEASGHILISAPTITSQH VEDIPLDTFVVSSSDSG PTSSTPLPRAFPRPRVG LYSRALQQVQVTDPAFL STPQRLVTYDNPVYEG EDVSLQFTHESIHNAPD EAFMDIIRLHRPAITSRR GLVRFSRIGQRGSMYT RSGQHIGARIHYFQDIS PVTQAAEEIELHPLVAA ENDTFDIYAEPFDPIPDP VQHSVTQSYLTSTPNTL SQSWGNTTVPLSIPSD WFVQSGPDITFPTASM GTPFSPVTPALPTGPVF ITGSDFYLHPTWYFARR RRKRIPLFFTDVAA	26.6%

HPV16 E1	ATGGCTGATCCTGCAGGTACCAATGGGGAAGAGGGTA CGGGATGTAATGGATGGTTTTATGTAGAGGCTGTAGTG GAAAAAAAAACAGGGGATGCTATATCAGATGACGAGAA CGAAAATGACAGTGATACAGGTGAAGATTTGGTAGATT TTATAGTAAATGATAATGATTATTTAACACAGGCAGAAA CAGAGACAGCACATGCGTTGTTTACTGCACAGGAAGCA AAACAACATAGAGATGCAGTACAGGTTCTAAAACGAAA GTATTTGGGTAGTCCACTTAGTGATATTAGTGGATGTGT AGACAATAATATTAGTCCTAGATTAAAAGCTATATGTAT AGAAAAACAAAGTAGAGCTGCAAAAAGGAGATTATTTG AAAGCGAAGACAGCGGGTATGGCAATACTGAAGTGGA AACTCAGCAGATGTTACAGGTAGAAGGGCGCCATGAG ACTGAAACACCATGTAGTCAGTATAGTGGTGGAAGTGG GGGTGGTTGCAGTCAGTACAGTAGTGGAAGTGGGGGA GAGGGTGTTAGTGAAAGACACACTATATGCCAAACACC ACTTACAAATATTTTAAATGTACTAAAAACTAGTAATGCA AAGGCAGCAATGTTAGCAAAATTTAAAGAGTTATACGG GGTGAGTTTTTCAGAATTAGTAAGACCATTTAAAAGTAA TAAATCAACGTGTTGCGATTGGTGTATTGCTGCATTTG GACTTACACCCAGTATAGCTGACAGTATAAAAACACTAT TACAACAATATTGTTTATATTTACACATTCAAAGTTTAGC ATGTTCATGGGGAATGGTTGTGTTACTATTAGTAAGATA TAAATGTGGAAAAAATAGAGAAACAATTGAAAAATTGCT GTCTAAACTATTATGTGTGTCTCCAATGTGTATGATGAT AGAGCCTCCAAAATTGCGTAGTACAGCAGCAGCATTAT ATTGGTATAAAACAGGTATATCAAATATTAGTGAAGTGT ATGGAGACACGCCAGAATGGATACAAAGACAAACAGTA TTACAACATAGTTTTAATGATTGTACATTTGAATTATCAC AGATGGTACAATGGGCCTACGATAATGACATAGTAGAC GATAGTGAAATTGCATATAAATATGCACAATTGGCAGAC ACTAATAGTAATGCAAGTGCCTTTCTAAAAAGTAATTCA CAGGCAAAAATTGTAAAGGATTGTGCAACAATGTGTAG ACATTATAAACGAGCAGAAAAAAAACAAATGAGTATGA GTCAATGGATAAAATATAGATGTGATAGGGTAGATGAT GGAGGTGATTGGAAGCAAATTGTTATGTTTTTAAGGTAT CAAGGTGTAGAGTTTATGTCATTTTTAACTGCATTAAAA AGATTTTTGCAAGGCATACCTAAAAAAAATTGCATATTA CTATATGGTGCAGCTAACACAGGTAAATCATTATTTGGT ATGAGTTTAATGAAATTTCTGCAAGGGTCTGTAATATGT TTTGTAAATTCTAAAAGCCATTTTTGGTTACAACCATTA GCAGATGCCAAAATAGGTATGTTAGATGATGCTACAGT GCCCTGTTGGAACTACATAGATGACAATTTAAGAAATG CATTGGATGGAAATTTAGTTTCTATGGATGTAAAGCATA GACCATTGGTACAACTAAAATGCCCTCCATTATTAATTA CATCTAACATTAATGCTGGTACAGATTCTAGGTGGCCTT ATTTACATAATAGATTGGTGGTGTTTACATTTCCTAATG AGTTTCCATTTGACGAAAACGGAAATCCAGTGTATGAG CTTAATGATAAGAACTGGAAATCCTTTTTCTCAAGGACG TGGTCCAGATTAAGTTTGCACGAGGACGAGGACAAGG AAAACGATGGAGACTCTTTGCCAACGTTTAAATGTGTG TCAGGACAAAATACTAACACATTATGA	ATGGCAGACCCCGCTGGGACTAACGGAGAAGAAGGCA CTGGCTGTAATGGCTGGTTTTATGTGGAGGCTGTGGTG GAGAAGAAGACAGGCGACGCCATCTCAGACGATGAGA ACGAAAATGACAGCGATACCGGGGAGGACCTGGTGGA TTTCATTGTCAACGACAATGATTACCTGACCCAGGCAG AGACCGAAACAGCACACGCCCTGTTTACAGCACAGGA AGCCAAGCAGCATAGGGATGCCGTGCAGGTCCTGAAG CGCAAATATCTGGGGAGCCCCCTGTCCGACATCTCTG GATGCGTGGATAACAATATTAGCCCTCGACTGAAGGCC ATCTGTATTGAGAAACAGAGCCGGGCCGCTAAGCGGA GACTGTTCGAGAGTGAAGACTCAGGCTACGGGAACAC TGAGGTGGAAACCCAGCAGATGCTCCAGGTCGAGGGC AGGCACGAGACTGAAACCCCATGCAGCCAGTACTCCG GAGGGTCTGGAGGAGGGTGTTCACAGTATAGCTCCGG AAGCGGAGGAGAGGGCGTGTCCGAACGCCATACTATC TGCCAGACCCCCCTGACAAACATTCTGAATGTCCTGAA GACCAGCAACGCCAAAGCAGCCATGCTGGCTAAGTTC AAAGAGCTGTACGGGGTGTCTTTCAGTGAACTGGTCCG GCCTTTTAAGAGTAACAAGAGCACCTGCTGTGACTGGT GTATCGCTGCATTTGGCCTGACTCCAAGTATCGCTGAT TCAATTAAGACCCTGCTCCAGCAGTACTGCCTGTATCT GCACATTCAGAGCCTGGCCTGTTCCTGGGGGATGGTG GTCCTGCTGCTGGTGCGCTATAAGTGCGGAAAAAACC GAGAGACTATCGAAAAGCTGCTGTCTAAACTGCTGTGC GTGAGTCCTATGTGTATGATGATTGAGCCCCCTAAACT GCGGAGCACAGCCGCTGCACTGTACTGGTATAAGACT GGCATCAGCAATATTTCCGAGGTGTACGGGGACACCC CAGAATGGATTCAGAGACAGACAGTCCTCCAGCACTCC TTCAACGATTGTACCTTTGAGCTGTCTCAGATGGTGCA GTGGGCTTATGACAATGATATCGTGGACGATTCCGAAA TTGCATACAAATATGCTCAGCTGGCAGACACCAACTCT AATGCTAGTGCATTCCTGAAGTCAAACAGCCAGGCAAA GATCGTGAAAGATTGCGCCACAATGTGCCGGCACTACA AGCGGGCTGAGAAGAAACAGATGTCCATGTCTCAGTG GATCAAATATAGGTGCGACCGCGTGGACGATGGGGGA GATTGGAAGCAGATTGTGATGTTCCTGAGATACCAGGG AGTCGAGTTCATGTCCTTTCTGACTGCCCTGAAGCGGT TCCTCCAGGGCATCCCCAAGAAAAACTGCATTCTGCTG TATGGGGCCGCTAATACCGGAAAATCTCTGTTCGGCAT GAGTCTGATGAAGTTTCTCCAGGGGTCTGTGATCTGTT TCGTCAATAGTAAATCACACTTTTGGCTCCAGCCACTG GCCGACGCTAAGATCGGAATGCTGGACGATGCCACCG TGCCCTGCTGGAACTACATTGACGATAACCTGCGCAAT GCTCTGGACGGCAATCTGGTGAGCATGGATGTCAAAC ACCGACCCCTGGTGCAGCTGAAGTGTCCACCCCTGCT GATCACATCCAACATTAATGCCGGCACTGACTCTCGGT GGCCCTACCTGCATAACAGACTGGTGGTCTTCACATTT CCTAATGAGTTCCCATTTGACGAAAACGGCAATCCTGT GTATGAGCTGAACGATAAGAACTGGAAATCATTCTTTA GCAGAACATGGTCCAGGCTGTCTCTGCATGAGGACGA AGATAAAGAAAACGACGGAGATAGTCTGCCTACTTTTA AGTGCGTGAGCGGCCAGAACACAAATACTCTG	MADPAGTNGEEGTGC NGWFYVEAVVEKKTGD AISDDENENDSDTGEDL VDFIVNDNDYLTQAETE TAHALFTAQEAKQHRD AVQVLKRKYLGSPLSDI SGCVDNNISPRLKAICIE KQSRAAKRRLFESEDS GYGNTEVETQQMLQVE GRHETETPCSQYSGGS GGGCSQYSSGSGGEG VSERHTICQTPLTNILNV LKTSNAKAAMLAKFKEL YGVSFSELVRPFKSNKS TCCDWCIAAFGLTPSIA DSIKTLLQQYCLYLHIQS LACSWGMVVLLLVRYK CGKNRETIEKLLSKLLC VSPMCMMIEPPKLRST AAALYWYKTGISNISEV YGDTPEWIQRQTVLQH SFNDCTFELSQMVQWA YDNDIVDDSEIAYKYAQ LADTNSNASAFLKSNSQ AKIVKDCATMCRHYKR AEKKQMSMSQWIKYRC DRVDDGGDWKQIVMFL RYQGVEFMSFLTALKR FLQGIPKKNCILLYGAAN TGKSLFGMSLMKFLQG SVICFVNSKSHFWLQPL ADAKIGMLDDATVPCW NYIDDNLRNALDGNLVS MDVKHRPLVQLKCPPL LITSNINAGTDSRWPYL HNRLVVFTFPNEFPFDE NGNPVYELNDKNWKSF FSRTWSRLSLHEDEDK ENDGDSLPTFKCVSGQ NTNTL	26.6%
HPV16 E2	ATGGAGACTCTTTGCCAACGTTTAAATGTGTGTCAGGA CAAAATACTAACACATTATGAAAATGATAGTACAGACCT ACGTGACCATATAGACTATTGGAAACACATGCGCCTAG AATGTGCTATTTATTACAAGGCCAGAGAAATGGGATTTA AACATATTAACCACCAGGTGGTGCCAACACTGGCTGTA TCAAAGAATAAAGCATTACAAGCAATTGAACTGCAACTA ACGTTAGAAACAATATATAACTCACAATATAGTAATGAA AAGTGGACATTACAAGACGTTAGCCTTGAAGTGTATTTA ACTGCACCAACAGGATGTATAAAAAAACATGGATATAC AGTGGAAGTGCAGTTTGATGGAGACATATGCAATACAA TGCATTATACAAACTGGACACATATATATATTTGTGAAG AAGCATCAGTAACTGTGGTAGAGGGTCAAGTTGACTAT TATGGTTTATATTATGTTCATGAAGGAATACGAACATAT TTTGTGCAGTTTAAAGATGATGCAGAAAAATATAGTAAA AATAAAGTATGGGAAGTTCATGCGGGTGGTCAGGTAAT ATTATGTCCTACATCTGTGTTTAGCAGCAACGAAGTATC CTCTCCTGAAATTATTAGGCAGCACTTGGCCAACCACC CCGCCGCGACCCATACCAAAGCCGTCGCCTTGGGCAC CGAAGAAACACAGACGACTATCCAGCGACCAAGATCA GAGCCAGACACCGGAAACCCCTGCCACACCACTAAGT TGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTC ACTGCATTTAACAGCTCACACAAAGGACGGATTAACTG TAATAGTAACACTACACCCATAGTACATTTAAAAGGTGA TGCTAATACTTTAAAATGTTTAAGATATAGATTTAAAAAG CATTGTACATTGTATACTGCAGTGTCGTCTACATGGCAT TGGACAGGACATAATGTAAAACATAAAAGTGCAATTGTT ACACTTACATATGATAGTGAATGGCAACGTGACCAATTT TTGTCTCAAGTTAAAATACCAAAAACTATTACAGTGTCT ACTGGATTTATGTCTATATGA	ATGGAGACTCTGTGCCAGCGGCTGAACGTGTGCCAGG ATAAGATTCTGACTCACTACGAAAATGACTCAACCGAC CTGCGGGACCACATCGACTACTGGAAGCACATGCGAC TGGAGTGCGCCATCTACTATAAGGCTCGGGAAATGGG CTTCAAACACATCAATCATCAGGTGGTCCCCACCCTGG CCGTGAGCAAGAACAAGGCCCTCCAGGCAATCGAGCT GCAACTGACCCTGGAAACAATCTACAATAGTCAGTATT CAAACGAGAAGTGGACACTCCAGGACGTGAGCCTGGA AGTCTACCTGACTGCACCTACCGGATGTATTAAGAAAC ACGGCTATACCGTGGAGGTCCAGTTTGACGGCGATAT CTGCAATACAATGCATTACACAAACTGGACTCACATCTA TATTTGTGAGGAAGCTAGCGTGACTGTGGTCGAGGGG CAGGTCGATTACTATGGACTGTACTATGTGCATGAAGG GATTCGCACCTACTTCGTGCAGTTTAAGGACGATGCTG AGAAATATTCTAAGAACAAGGTCTGGGAAGTCCACGCA GGAGGACAGGTCATCCTGTGCCCTACCAGTGTGTTCA GCTCCAATGAGGTCTCTAGTCCAGAAATCATTCGACAG CACCTGGCCAACCATCCCGCCGCTACCCACACAAAGG CAGTGGCCCTGGGAACCGAGGAAACACAGACCACAAT TCAGCGGCCCAGATCCGAGCCTGACACAGGCAATCCT TGCCATACTACCAAGCTGCTGCACAGAGACAGCGTGG ATTCCGCACCAATCCTGACTGCCTTCAACTCAAGCCAT AAAGGCAGGATCAACTGTAATTCTAACACAACTCCAATT GTCCACCTGAAGGGGGATGCCAATACCCTGAAATGCC TGCGGTACAGATTCAAGAAACACTGTACTCTGTATACC GCCGTGTCCTCTACATGGCACTGGACTGGGCATAACG TGAAGCACAAATCAGCTATCGTCACTCTGACCTACGAC AGCGAGTGGCAGAGGGATCAGTTCCTGTCCCAGGTGA AGATCCCCAAAACAATTACTGTCTCTACAGGCTTCATGA GTATC	METLCQRLNVCQDKILT HYENDSTDLRDHIDYW KHMRLECAIYYKAREM GFKHINHQVVPTLAVSK NKALQAIELQLTLETIYN SQYSNEKWTLQDVSLE VYLTAPTGCIKKHGYTV EVQFDGDICNTMHYTN WTHIYICEEASVTVVEG QVDYYGLYYVHEGIRTY FVQFKDDAEKYSKNKV WEVHAGGQVILCPTSV FSSNEVSSPEIIRQHLA NHPAATHTKAVALGTEE TQTTIQRPRSEPDTGNP CHTTKLLHRDSVDSAPI LTAFNSSHKGRINCNSN TTPIVHLKGDANTLKCL RYRFKKHCTLYTAVSST WHWTGHNVKHKSAIVT LTYDSEWQRDQFLSQV KIPKTITVSTGFMSI	25.9%
HPV16 E4	TATTATGTCCTACATCTGTGTTTAGCAGCAACGAAGTAT CCTCTCCTGAAATTATTAGGCAGCACTTGGCCAACCAC CCCGCCGCGACCCATACCAAAGCCGTCGCCTTGGGCA CCGAAGAAACACAGACGACTATCCAGCGACCAAGATCA GAGCCAGACACCGGAAACCCCTGCCACACCACTAAGT TGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTC ACTGCATTTAACAGCTCACACAAAGGACGGATTAACTG TAATAGTAACACTACACCCATAG	ATGTATGTGCTGCATCTGTGCCTGGCTGCTACCAAGTA CCCCCTGCTGAAACTGCTGGGATCAACCTGGCCTACC ACCCCCCCTCGGCCCATCCCTAAGCCATCTCCCTGGG CCCCTAAGAAACACCGGCGGCTGAGCAGCGACCAGGA TCAGTCACAGACTCCTGAGACCCCAGCTACACCCCTGA GCTGCTGTACCGAAACACAGTGGACAGTGCTCCAGAG CAGCCTGCACCTGACTGCCCATACCAAAGACGGCCTG ACAGTGATTGTCACTCTGCATCCC	YYVLHLCLAATKYPLLK LLGSTWPTTPPRPIPKP SPWAPKKHRRLSSDQD QSQTPETPATPLSCCTE TQWTVLQSSLHLTAHT KDGLTVIVTLHP	29.4%
HPV16 E5	ATGACAAATCTTGATACTGCATCCACAACATTACTGGC GTGCTTTTTGCTTTGCTTTTGTGTGCTTTTGTGTGTCTG CCTATTAATACGTCCGCTGCTTTTGTCTGTGTCTACATA GCACATCATTAATAATATTGGTATTACTATTGTGGATAAC AGCAGCCTCTGCGTTTAGGTGTTTTATTGTATATATTAT ATTTGTTTATATACCATTATTTTTAATACATACACATGCA CGCTTTTTAATTACATAA	ATGACCAACCTGGATACTGCTTCTACTACCCTGCTGGC TTGTTTCCTGCTGTGTTTCTGTGTCCTGCTGTGCGTGT GCCTGCTGATTAGGCCCCTGCTGCTGAGCGTGTCCAC CTACACATCTCTGATCATTCTGGTCCTGCTGCTGTGGA TCACAGCCGCTAGCGCATTCCGGTGCTTCATCGTGTAC ATCATCTTCGTCTACATCCCTCTGTTTCTGATTCACACT CATGCCAGATTCCTGATCACC	MTNLDTASTTLLACFLL CFCVLLCVCLLIRPLLLS VSTYTSLIILVLLLWITAA SAFRCFIVYIIFVYIPLFLI HTHARFLIT	35.0%

HPV16 E6	ATGCACCAAAAGAGAACTGCAATGTTTCAGGACCCACA GGAGCGACCCAGAAAGTTACCACAGTTATGCACAGAG CTGCAAACAACTATACATGATATAATATTAGAATGTGTG TACTGCAAGCAACAGTTACTGCGACGTGAGGTATATGA CTTTGCTTTTCGGGATTTATGCATAGTATATAGAGATGG GAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTTA TTCTAAAATTAGTGAGTATAGACATTATTGTTATAGTTTG TATGGAACAACATTAGAACAGCAATACAACAAACCGTT GTGTGATTTGTTAATTAGGTGTATTAACTGTCAAAAGCC ACTGTGTCCTGAAGAAAAGCAAAGACATCTGGACAAAA AGCAAAGATTCCATAATATAAGGGGTCGGTGGACCGGT CGATGTATGTCTTGTTGCAGATCATCAAGAACACGTAG AGAAACCCAGCTGTAA	ATGCACCAGAAGAGAACCGCCATGTTTCAGGACCCTCA GGAACGACCTCGCAAACTGCCCCAGCTGTGTACCGAA CTGCAGACAACTATCCACGACATCATTCTGGAGTGCGT GTACTGTAAGCAGCAGCTGCTGCGGAGAGAAGTCTAT GACTTCGCCTTTCGCGATCTGTGCATCGTGTACCGAGA CGGAAACCCCTACGCCGTCTGCGATAAGTGTCTGAAGT TCTACTCTAAGATTAGTGAGTATCGGCATTACTGTTATA GCCTGTACGGCACCACACTGGAACAGCAGTATAACAAA CCCCTGTGCGACCTGCTGATCAGATGCATTAATTGTCA GAAGCCCCTGTGTCCTGAGGAAAAACAGAGGCACCTG GATAAGAAACAGCGCTTTCATAATATTCGAGGCCGGTG GACAGGGAGGTGCATGTCTTGCTGTAGAAGCTCCAGG ACTAGGCGCGAGACCCAGCTG	MHQKRTAMFQDPQER PRKLPQLCTELQTTIHDI ILECVYCKQQLLRREVY DFAFRDLCIVYRDGNPY AVCDKCLKFYSKISEYR HYCYSLYGTTLEQQYN KPLCDLLIRCINCQKPLC PEEKQRHLDKKQRFHNI RGRWTGRCMSCCRSS RTRRETQL	25.4%
HPV16 E7	ATGCATGGAGATACACCTACATTGCATGAATATATGTTA GATTTGCAACCAGAGACAACTGATCTCTACTGTTATGA GCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAG ATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCCA TTACAATATTGTAACCTTTTGTTGCAAGTGTGACTCTAC GCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTC GTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATT GTGTGCCCCATCTGTTCTCAGAAACCATAA	ATGCACGGCGACACTCCTACTCTGCACGAATACATGCT GGACCTGCAGCCCGAAACTACTGACCTGTACTGCTAC GAACAGCTGAATGACAGCTCCGAGGAAGAGGACGAAA TCGATGGACCTGCCGGCCAGGCTGAGCCTGACAGGGC CCACTACAACATTGTGACTTTCTGCTGTAAGTGCGATTC TACCCTGCGGCTGTGTGTGCAGAGTACCCATGTGGAC ATCAGAACCCTGGAGGACCTGCTGATGGGAACACTGG GCATCGTCTGCCCAATTTGTTCCCAGAAACCC	MHGDTPTLHEYMLDLQ PETTDLYCYEQLNDSSE EEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLR LCVQSTHVDIRTLEDLL MGTLGIVCPICSQKP	25.6%
HPV16 L1	ATGTCTCTTTGGCTGCCTAGTGAGGCCACTGTCTACTT GCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATG AATATGTTGCACGCACAAACATATATTATCATGCAGGAA CATCCAGACTACTTGCAGTTGGACATCCCTATTTTCCTA TTAAAAAACCTAACAATAACAAAATATTAGTTCCTAAAGT ATCAGGATTACAATACAGGGTATTTAGAATACATTTACC TGACCCCAATAAGTTTGGTTTTCCTGACACCTCATTTTA TAATCCAGATACACAGCGGCTGGTTTGGGCCTGTGTAG GTGTTGAGGTAGGTCGTGGTCAGCCATTAGGTGTGGG CATTAGTGGCCATCCTTTATTAAATAAATTGGATGACAC AGAAAATGCTAGTGCTTATGCAGCAAATGCAGGTGTGG ATAATAGAGAATGTATATCTATGGATTACAAACAAACAC AATTGTGTTTAATTGGTTGCAAACCACCTATAGGGGAA CACTGGGGCAAAGGATCCCCATGTACCAATGTTGCAGT AAATCCAGGTGATTGTCCACCATTAGAGTTAATAAACAC AGTTATTCAGGATGGTGATATGGTTGATACTGGCTTTG GTGCTATGGACTTTACTACATTACAGGCTAACAAAAGT GAAGTTCCACTGGATATTTGTACATCTATTTGCAAATAT CCAGATTATATTAAAATGGTGTCAGAACCATATGGCGA CAGCTTATTTTTTTATTTACGAAGGGAACAAATGTTGT TAGACATTTATTTAATAGGGCTGGTACTGTTGGTGAAAA TGTACCAGACGATTTATACATTAAAGGCTCTGGGTCTA CTGCAAATTTAGCCAGTTCAAATTATTTCCTACACCTA GTGGTTCTATGGTTACCTCTGATGCCCAAATATTCAATA AACCTTATTGGTTACAACGAGCACAGGGCCACAATAAT GGCATTTGTTGGGGTAACCAACTATTTGTTACTGTTGTT GATACTACACGCAGTACAAATATGTCATTATGTGCTGC CATATCTACTTCAGAAACTACATATAAAAATACTAACTTT AAGGAGTACCTACGACATGGGGAGGAATATGATTTACA GTTTATTTTTCAACTGTGCAAAATAACCTTAACTGCAGA CGTTATGACATACATACATTCTATGAATTCCACTATTTT GGAGGACTGGAATTTTGGTCTACAACCTCCCCCAGGA GGCACACTAGAAGATACTTATAGGTTTGTAACATCCCA GGCAATTGCTTGTCAAAAACATACACCTCCAGCACCTA AAGAAGATCCCCTTAAAAAATACACTTTTTGGGAAGTAA ATTTAAAGGAAAAGTTTTCTGCAGACCTAGATCAGTTTC CTTTAGGACGCAAATTTTTACTACAAGCAGGATTGAAG GCCAAACCAAAATTTACATTAGGAAAACGAAAAGCTAC ACCCACCACCTCATCTACCTCTACAACTGCTAAACGCA AAAAACGTAAGCTGTAA	ATGCAGGTCACTTTTATCTATATCCTGGTCATTACTTGC TACGAGAACGATGTCAACGTCTATCATATTTTCTTTCAG ATGTCCCTGTGGCTGCCATCCGAGGCAACCGTCTACCT GCCCCCTGTGCCCGTCTCTAAAGTGGTCAGTACAGATG AATATGTGGCCCGCACTAACATCTACTATCACGCCGGG ACATCTCGACTGCTGGCTGTCGGACATCCCTACTTCCC TATCAAGAAACCCAACAACAACAAAATTCTGGTCCCTAA GGTGAGTGGCCTCCAGTATAGGGTGTTCCGCATTCAC CTGCCAGATCCCAATAAGTTCGGGTTTCCTGACACCAG CTTTTACAACCCAGATACACAGCGACTGGTCTGGGCAT GCGTGGGAGTCGAAGTGGGAAGAGGACAGCCTCTGG GAGTGGGAATCAGCGGACATCCACTGCTGAACAAGCT GGACGATACCGAGAACGCTTCCGCATACGCCGCTAAT GCTGGCGTGGACAACCGGGAATGTATTTCTATGGATTA TAAGCAGACACAGCTGTGCCTGATCGGATGTAAACCAC CCATTGGAGAGCACTGGGGCAAGGGGTCCCCATGCAC TAATGTCGCCGTGAACCCCGGCGACTGTCCTCCACTG GAACTGATCAATACCGTCATTCAGGACGGAGATATGGT GCATACAGGATTCGGCGCAATGGATTTTACCACACTCC AGGCCAACAAGAGTGAGGTGCCCCTGGACATCTGCAC CTCAATTTGTAAGTACCCCGATTACATCAAGATGGTGTC CGAGCCTTACGGGGACTCTCTGTTCTTTTATCTGCGGA GAGAACAGATGTTCGTGAGACACCTGTTTAATAGGGCA GGCACTGTCGGAGAAAACGTGCCAGACGATCTGTACA TCAAGGGGTCAGGAAGCACCGCAAATCTGGCCAGCTC CAACTATTTCCCTACTCCATCCGGCTCTATGGTGACCT CTGACGCCCAGATTTTCAACAAGCCTTACTGGCTCCAG CGGGCCCAGGGACATAATAACGGCATTTGCTGGGGGA ATCAGCTGTTCGTGACAGTGGTCGATACTACCCGCTCA ACTAACATGAGCCTGTGTGCAGCCATCAGTACCTCAGA GACAACTTACAAGAACACAAACTTCAAGGAATACCTGA GACACGGAGAGGAATATGACCTCCAGTTCATCTTTCAG CTGTGCAAGATTACACTGACTGCCGATGTGATGACTTA CATCCATAGCATGAACAGCACCATTCTGGAGGACTGGA ACTTCGGACTCCAGCCACCTCCAGGCGGGACCCTGGA AGATACATATAGGTTTGTGACACAGGCCATCGCTTGTC AGAAACACACTCCCCCTGCTCCAAAGGAGGACGATCC CCTGAAGAAATACACCTTCTGGGAGGTGAACCTGAAGG AAAAGTTCAGCGCCGACCTGGACCAGTTCCCACTGGG CAGGAAGTTTCTGCTCCAGGCTGGGCTGAAGGCAAAA CCTAAGTTCACACTGGGCAAACGCAAGGCTACTCCAAC CACATCTAGTACCAGCACTACCGCAAAACGAAAGAAAC GGAAGCTG	MSLWLPSEATVYLPPV PVSKVVSTDEYVARTNI YYHAGTSRLLAVGHPY FPIKKPNNNKILVPKVSG LQYRVFRIHLPDPNKFG FPDTSFYNPDTQRLVW ACVGVEVGRGQPLGVG ISGHPLLNKLDDTENAS AYAANAGVDNRECISM DYKQTQLCLIGCKPPIG EHWGKGSPCTNVAVNP GDCPPLELINTVIQDGD MVDTGFGAMDFTTLQA NKSEVPLDICTSICKYPD YIKMVSEPYGDSLFFYL RREQMFVRHLFNRAGT VGENVPDDLYIKGSGST ANLASSNYFPTPSGSM VTSDAQIFNKPYWLQR AQGHNNGICWGNQLFV TVVDTTRSTNMSLCAAI STSETTYKNTNFKEYLR HGEEYDLQFIFQLCKITL TADVMTYIHSMNSTILE DWNFGLQPPPGGTLED TYRFVTSQAIACQKHTP PAPKEDPLKKYTFWEV NLKEKFSADLDQFPLGR KFLLQAGLKAKPKFTLG KRKATPTTSSTSTTAKR KKRKL	30.0%
HPV16 L2	ATGCGACACAAACGTTCTGCAAAACGCACAAAACGTGC ATCGGCTACCCAACTTTATAAAACATGCAAACAGGCAG GTACATGTCCACCTGACATTATACCTAAGGTTGAAGGC AAAACTATTGCTGATCAAATATTACAATATGGAAGTATG GGTGTATTTTTTGGTGGGTTAGGAATTGGAACAGGGTC GGGTACAGGCGGACGCACTGGGTATATTCCATTGGGA ACAAGGCCTCCCACAGCTACAGATACACTTGCTCCTGT AAGACCCCCTTTAACAGTAGATCCTGTGGGCCCTTCTG ATCCTTCTATAGTTTCTTTAGTGGAAGAAACTAGTTTTAT TGATGCTGGTGCACCAACATCTGTACCTTCCATTCCCC CAGATGTATCAGGATTTAGTATTACTACTTCAACTGATA CCACACCTGCTATATTAGATATTAATAATACTGTTACTA CTGTTACTACACATAATAATCCCACTTTCACTGACCCAT CTGTATTGCAGCCTCCAACACCTGCAGAAACTGGAGG GCATTTTACACTTTCATCATCCACTATTAGTACACATAAT TATGAAGAAATTCCTATGGATACATTTATTGTTAGCACA AACCCTAACACAGTAACTAGTAGCACACCCATACCAGG GTCTCGCCCAGTGGCACGCCTAGGATTATATAGTCGCA CAACACAACAGGTTAAAGTTGTAGACCCTGCTTTTGTAA CCACTCCCACTAAACTTATTACATATGATAATCCTGCAT ATGAAGGTATAGATGTGGATAATACATTATATTTTTCTA GTAATGATAATAGTATTAATATAGCTCCAGATCCTGACT TTTTGGATATAGTTGCTTTACATAGGCCAGCATTAACCT CTAGGCGTACTGGCATTAGGTACAGTAGAATTGGTAAT AAACAAACACTACGTACTCGTAGTGGAAAATCTATAGG TGCTAAGGTACATTATTATTATGATTTAAGTACTATTGAT CCTGCAGAAGAAATAGAATTACAAACTATAACACCTTCT ACATATACTACCACTTCACATGCAGCCTCACCTACTTCT ATTAATAATGGATTATATGATATTTATGCAGATGACTTTA TTACAGATACTTCTACAACCCCGGTACCATCTGTACCCT CTACATCTTTATCAGGTTATATTCCTGCAAATACAACAA TTCCTTTTGGTGGTGCATACAATATTCCTTTAGTATCAG GTCCTGATATACCCATTAATATAACTGACCAAGCTCCTT CATTAATTCCTATAGTTCCAGGGTCTCCACAATATACAA TTATTGCTGATGCAGGTGACTTTTATTTACATCCTAGTT ATTACATGTTACGAAAACGACGTAAACGTTTACCATATT TTTTTTCAGATGTCTCTTTGGCTGCCTAG	ATGAGACATAAGCGGAGTGCTAAGAGGACTAAAAGAG CCAGCGCCACCCAGCTGTATAAGACTTGTAAACAGGCC GGAACTTGCCCTCCTGACATCATTCCAAAGGTGGAGG GGAAAACCATTGCCGAACAGATCCTCCAGTATGGCAGT ATGGGGGTCTTCTTTGGCGGGCTGGGAATTGGCACAG GGTCTGGAACTGGAGGCAGGACCGGGTACATCCCACT GGGAACACGCCCCCCTACAGCAACTGATACCCTGGCA CCCGTGAGACCACCCCTGACCGTGGACCCAGTCGGAC CAAGCGATCCTTCCATTGTGTCTCTGGTCGAGGAAACC TCCTTCATCGACGCTGGCGCACCCACAAGTGTGCCTTC AATTCCTCCAGATGTCAGCGGGTTTTCCATCACCACAT CTACAGACACTACCCCAGCCATTCTGGATATCAACAAT ACTGTGACAACTGTCACCACACACAACAATCCAACATT CACTGACCCCTCCGTGCTCCAGCCACCTACACCCGCT GAGACTGGAGGACACTTTACACTGAGCAGCAGCACCA TCAGCACACATAACTATGAGGAAATTCCAATGGATACAT TCATCGTGAGCACTAACCCCAATACCGTCACAAGTTCA ACCCCAATCCCCGGCAGCCGGCCCGTGGCACGACTG GGCCTGTACTCTCGAACTACCCAGCAGGTGAAGGTGG TGGACCCCGCTTTTGTCACAACTCCAACCAAACTGATT ACCTACGACAACCCCGCATACGAAGGCATCGACGTGG ATAATACCCTGTATTTCAGCTCCAACGACAATAGCATCA ACATTGCCCCTGACCCAGATTTTCTGGATATCGTGGCT CTGCATCGCCCCGCACTGACTTCTCGGAGAACCGGCA TTAGATACAGTAGGATCGGGAATAAGCAGACTCTGCGA ACCCGGTCTGGAAAGAGTATTGGCGCAAAAGTGCACTA CTATTACGACCTGAGCACCATCGATCCTGCCGAGGAAA TTGAGCTGCAAACTATCACCCCAAGTACTTATACCACAA CTTCACATGCCGCTTCACCCACCAGCATCAACAATGGC CTGTATGACATCTACGCAGACGATTTCATCACAGATACT AGCACCACACCCGTGCCTTCCGTCCCTTCTACCAGTCT GTCAGGATATATTCCGCCAACACTACCATCCCTTTTG GCGGGGCTTACAATATCCCTCTGGTGAGCGGCCCAGA CATCCCCATTAATATCACAGATCAGGCTCCATCACTGAT TCCTATCGTCCCAGGGAGCCCCCAGTATACCATCATTG CCGACGCTGGAGATTTCTACCTGCACCCCTCCTATTAC ATGCTGCGGAAGAGGCGCAAAAGACTGCCTTACTTCTT TTCCGACGTGTCTCTGGCAGCC	MRHKRSAKRTKRASAT QLYKTCKQAGTCPPDII PKVEGKTIADQILQYGS MGVFFGGLGIGTGSGT GGRTGYIPLGTRPPTAT DTLAPVRPPLTVDPVGP SDPSIVSLVEETSFIDAG APTSVPSIPPDVSGFSIT TSTDTTPAILDINNTVTT VTTHNNPTFTDPSVLQP PTPAETGGHFTLSSSTI STHNYEEIPMDTFIVST NPNTVTSSTPIPGSRPV ARLGLYSRTTQQVKVV DPAFVTTPTKLITYDNP AYEGIDVDNTLYFSSND NSINIAPDPDFLDIVALH RPALTSRRTGIRYSRIG NKQTLRTRSGKSIGAKV HYYYDLSTIDPAEEIELQ TITPSTYTTTSHAASPTS INNGLYDIYADDFITDTS TTPVPSVPSTSLSGYIP ANTTIPFGGAYNIPLVS GPDIPINITDQAPSLPIV PGSPQYTIIADAGDFYL HPSYYMLRKRRKRLPY FFSDVSLAA	28.3%

HPV18 E1	ATGGCTGATCCAGAAGGTACAGACGGGGAGGGCACGG GTTGTAACGGCTGGTTTTATGTACAAGCTATTGTAGACA AAAAAACAGGAGATGTAATATCAGATGACGAGGACGAA AATGCAACAGACACAGGGTCGGATATGGTAGATTTTAT TGATACACAAGGAACATTTTGTGAACAGGCAGAGCTAG AGACAGCACAGGCATTGTTCCATGCGCAGGAGGTCCA CAATGATGCACAAGTGTTGCATGTTTTAAAACGAAAGTT TGCAGGAGGCAGCACAGAAAACAGTCCATTAGGGGAG CGGCTGGAGGTGGATACAGAGTTAAGTCCACGGTTAC AAGAAATATCTTTAAATAGTGGGCAGAAAAAGGCAAAA AGGCGGCTGTTTACAATATCAGATAGTGGCTATGGCTG TTCTGAAGTGGAAGCAACACAGATTCAGGTAACTACAA ATGGCGAACATGGCGGCAATGTATGTAGTGGCGGCAG TACGGAGGCTATAGACAACGGGGGCACAGAGGGCAAC AACAGCAGTGTAGACGGTACAAGTGACAATAGCAATAT AGAAAATGTAAATCCACAATGTACCATAGCACAATTAAA AGACTTGTTAAAAGTAAACAATAAACAAGGAGCTATGTT AGCAGTATTTAAAGACACATATGGGCTATCATTTACAGA TTTAGTTAGAAATTTTAAAAGTGATAAAACCACGTGTAC AGATTGGGTTACAGCTATATTTGGAGTAAACCCAACAAT AGCAGAAGGATTTAAAACACTAATACAGCCATTTATATT ATATGCCCATATTCAATGTCTAGACTGTAAATGGGGAG TATTAATATTAGCCCTGTTGCGTTACAAATGTGGTAAGA GTAGACTAACAGTTGCTAAAGGTTTAAGTACGTTGTTAC ACGTACCTGAAACTTGTATGTTAATTCAACCACCAAAAT TGCGAAGTAGTGTTGCAGCACTATATTGGTATAGAACA GGAATATCAAATATTAGTGAAGTAATGGGAGACACACC TGAGTGGATACAAAGACTTACTATTATACAACATGGAAT AGATGATAGCAATTTTGATTTGTCAGAAATGGTACAATG GGCATTTGATAATGAGCTGACAGATGAAAGCGATATGG CATTTGAATATGCCTTATTAGCAGACAGCAACAGCAAT GCAGCTGCCTTTTTAAAAAGCAATTGCCAAGCTAAATAT TTAAAAGATTGTGCCACAATGTGCAAACATTATAGGCG AGCCCAAAAACGACAAATGAATATGTCACAGTGGATAC GATTTAGATGTTCAAAAATAGATGAAGGGGGAGATTGG AGACCAATAGTGCAATTCCTGCGATACCAACAAATAGA GTTTATAACATTTTTAGGAGCCTTAAAATCATTTTTAAAA GGAACCCCCAAAAAAAATTGTTTAGTATTTTGTGGACCA GCAAATACAGGAAAATCATATTTTGGAATGAGTTTTATA CACTTTATACAAGGAGCAGTAATATCATTTGTGAATTCC ACTAGTCATTTTTGGTTGGAACCGTTAACAGATACTAAG GTGGCCATGTTAGATGATGCAACGACCACGTGTTGGAC ATACTTTGATACCTATATGAGAAATGCGTTAGATGGCAA TCCAATAAGTATTGATAGAAAGCACAAACCATTAATACA ACTAAAATGTCCTCCAATACTACTAACCACAAATATACA TCCAGCAAAGGATAATAGATGGCCATATTTAGAAAGTA GAATAACAGTATTTGAATTTCCAAATGCATTTCCATTTG ATAAAAATGGCAATCCAGTATATGAAATAAATGACAAAA ATTGGAAATGTTTTTTTGAAAGGACATGGTCCAGATTAG ATTTGCACGAGGAAGAGGAAGATGCAGACACCGAAGG AAACCCTTTCGGAACGTTTAAGTGCGTTGCAGGACAAA ATCATAGACCACTATGA	ATGGCAGACCCCGAAGGGACTGACGGCGAAGGGACT GGATGTAACGGATGGTTTTATGTGCAGGCTATTGTGGA TAAGAAGACTGGCGACGTGATCAGCGACGATGAGGAT GAAAACGCTACCGACACAGGCTCCGACATGGTCGATTT CATTGACACACAGGGGACTTTTTGCGAGCAGGCAGAG CTGGAAACCGCACAGGCCCTGTTCCACGCTCAGGAAG TGCATAACGATGCACAGGTGCTGCACGTCCTGAAGCG GAAATTTGCCGGCGGGAGTACAGAGAACAGCCCACTG GGGGAGAGACTGGAAGTGGACACTGAGCTGTCTCCCA GGCTCCAGGAAATCAGCCTGAACTCCGGACAGAAGAA AGCCAAGCGGAGACTGTTCACCATCTCAGATAGCGGG TACGGATGCAGCGAGGTGGAAGCTACACAGATTCAGG TCACCACAAACGGCGAGCACGGAGGCAATGTGTGTTC TGGGGGAAGTACAGAGGCTATTGACAATGGCGGGACT GAAGGAAACAATAGCTCCGTGGATGGCACTTCAGACAA CAGCAATATCGAAAACGTCAATCCTCAGTGCACCATTG CCCAGCTGAAGGATCTGCTGAAAGTGAACAATAAGCAG GGAGCTATGCTGGCAGTCTTCAAGGATACCTACGGCCT GAGTTTCACTGACCTGGTGAGAAACTTTAAGTCAGATA AAACTACCTGTACCGACTGGGTGACAGCAATCTTTGGG GTCAATCCCACCATTGCCGAGGGATTCAAAACACTGAT CCAGCCTTTTATTCTGTACGCCCACATCCAGTGCCTGG ACTGTAAGTGGGGCGTGCTGATTCTGGCCCTGCTGCG GTATAAGTGCGGAAAATCCAGACTGACTGTGGCTAAAG GCCTGTCTACCCTGCTGCATGTCCCCGAGACATGTATG CTGATCCAGCCCCCTAAGCTGCGCTCTAGTGTGGCCG CTCTGTACTGGTATCGAACCGGGATCTCCAACATTTCT GAGGTCATGGGAGACACCCCTGAATGGATTCAGAGGC TGACAATCATTCAGCACGGCATTGACGATAGCAACTTC GATCTGTCCGAGATGGTGCAGTGGGCTTTTGACAATGA GCTGACCGATGAATCTGACATGGCATTCGAATACGCCC TGCTGGCTGATTCCAACTCTAATGCAGCCGCTTTTCTG AAAAGTAACTGCCAGGCCAAGTACCTGAAAGACTGCGC TACAATGTGTAAACATTATAGGCGCGCCCAGAAGCGCC AGATGAATATGTCACAGTGGATCAGATTCCGGTGCAGC AAGATTGATGAGGGAGGCGACTGGAGGCCAATCGTGC AGTTTCTGCGCTACCAGCAGATCGAGTTCATCACTTTT CTGGGCGCCCTGAAGAGCTTCCTGAAGGGGACCCCTA AGAAAAACTGCCTGGTGTTCTGTGGACCAGCTAATACA GGCAAATCTTATTTTGGGATGAGTTTCATCCACTTTATT CAGGGCGCAGTGATCAGCTTCGTCAACAGTACTTCACA TTTTTGGCTGGAGCCCCTGACTGATACCAAGGTGGCAA TGCTGGACGATGCCACAACTACCTGCTGGACTTACTTC GACACCTATATGCGGAACGCCCTGGATGGGAATCCAAT CAGCATTGACAGAAAGCACAAACCCCTGATCCAGCTGA AATGTCCACCCATCCTGCTGACAACTAACATTCATCCC GCAAAGGACAATCGATGGCCTTACCTGGAGTCCCGGA TCACCGTGTTCGAATTTCCTAATGCCTTCCCATTTGATA AGAACGGCAATCCCGTCTATGAGATTAACGACAAGAAC TGGAAGTGTTTCTTTGAAAGAACATGGTCCAGGCTGGA TCTGCACGAGGAAGAGGAAGACGCAGATACTGAGGGC AACCCTTTCGGGACCTTTAAGCTGCGCGCCGGCCAGA ATCATCGACCACTG	MADPEGTDGEGTGCN GWFYVQAIVDKKTGDVI SDDEDENATDTGSDMV DFIDTQGTFCEQAELET AQALFHAQEVHNDAQV LHVLKRKFAGGSTENS PLGERLEVDTELSPRLQ EISLNSGQKKAKRRLFTI SDSGYGCSEVEATQIQ VTTNGEHGGNVCSGGS TEAIDNGGTEGNNSSV DGTSDNSNIENVNPQC TIAQLKDLLKVNNKQGA MLAVFKDTYGLSFTDLV RNFKSDKTTCTDWVTAI FGVNPTIAEGFKTLIQPF ILYAHIQCLDCKWGVLIL ALLRYKCGKSRLTVAKG LSTLLHVPETCMLIQPP KLRSSVAALYWYRTGIS NISEVMGDTPEWIQRLT IIQHGIDDSNFDLSEMV QWAFDNELTDESDMAF EYALLADSNSNAAAFLK SNCQAKYLKDCATMCK HYRRAQKRQMNMSQW IRFRCSKIDEGGDWRPI VQFLRYQQIEFITFLGAL KSFLKGTPKKNCLVFCG PANTGKSYFGMSFIHFI QGAVISFVNSTSHFWLE PLTDTKVAMLDDATTTC WTYFDTYMRNALDGNP ISIDRKHKPLIQLKCPPIL LTTNIHPAKDNRWPYLE SRITVFEFPNAFPFDKN GNPVYEINDKNWKCFF ERTWSRLDLHEEEEDA DTEGNPFGTFKCVAGQ NHRPL	25.5%
HPV18 E2	ATGCAGACACCGAAGGAAACCCTTTCGGAACGTTTAAG TGCGTTGCAGGACAAAATCATAGACCACTATGAAAATG ACAGTAAAGACATAGACAGCCAAATACAGTATTGGCAA CTAATACGTTGGGAAAATGCAATATTCTTTGCAGCAAG GGAACATGGCATACAGACATTAAACCACCAGGTGGTGC CAGCCTATAACATTTCAAAAAGTAAAGCACATAAAGCTA TTGAACTGCAAATGGCCCTACAAGGCCTTGCACAAAGT GCATACAAAACCGAGGATTGGACACTGCAAGACACATG CGAGGAACTATGGAATACAGAACCTACTCACTGCTTTA AAAAAGGTGGCCAAACAGTACAAGTATATTTTGATGGC AACAAAGACAATTGTATGACCTATGTAGCATGGGACAG TGTGTATTATATGACTGATGCAGGAACATGGGACAAAA CGGCTACCTGTGTAAGTCACAGGGGATTGTATTATGTA AAGGAAGGGTACAACACGTTTTATATAGAATTTAAAAGT GAATGTGAAAAATATGGGAACACAGGTACGTGGGAAGT ACATTTTGGGAATAATGTAATTGATTGTAATGACTCTAT GTGCAGTACCAGTGACGACACGGTATCCGCTACTCAG CTTGTTAAACAGCTACAGCACACCCCCTCACCGTATTC CAGCACCGTGTCCGTGGGCACCGCAAAGACCTACGGC CAGACGTCGGCTGCTACACGACCTGGACACTGTGGAC TCGCGGAGAAGCAGCATTGTGGACCTGTCAACCCACTT CTCGGTGCAGCTACACCTACAGGCAACAACAAAAGAC GGAAACTCTGTAGTGGTAACACTACGCCTATAATACATT TAAAAGGTGACAGAAACAGTTTAAAATGTTTACGGTACA GATTGCGAAAACATAGCGACCACTATAGAGATATATCA TCCACCTGGCATTGGACAGGTGCAGGCAATGAAAAAAC AGGAATACTGACTGTAACATACCATAGTGAAACACAAA GAACAAAATTTTTAAATACTGTTGCAATTCCAGATAGTG TACAAATATTGGTGGGATACATGACAATGTAA	ATGCAGACCCCTAAGGAAACACTGAGCGAACGACTGT CATGCGTCCAGGATAAAATCATCGACCATTACGAAAAC GACTCCAAAGATATCGACAGCCAGATTCAGTACTGGCA GCTGATCCGGTGGGAGAACGCAATTTTCTTTGCCGCTA GAGAACACGGAATCCAGACCCTGAACCATCAGGTGGT CCCCGCCTACAATATCTCAAAGAGCAAAGCCCACAAGG CTATTGAGCTGCAAATGGCACTCCAGGGCCTGGCCCA GTCCCGATATAAAACAGAGGACTGGACTCTCCAGGATA CCTGCGAGGAACTGTGGAATACAGAACCTACTCATTGT TTCAAGAAAGGCGGGCAGACCGTGCAGGTCTACTTTG ACGGGAACAAGGATAATTGCATGACCTACGTGGCCTG GGATTCCGTCTACTATATGACAGACGCTGGAACTTGGG ATAAGACTGCAACCTGTGTGTCTCACAGGGGCCTGTAC TATGTGAAAGAGGGGTACAACACCTTCTATATCGAGTT CAAGTCTGAGTGCGAAAAATACGGGAATACAGGAACTT GGGAGGTGCACTTCGGGAACAATGTCATTGACTGCAA CGATAGCATGTGTTCCACCTCTGACGATACAGTGTCCG CCACTCAGCTGGTCAAGCAGCTCCAGCATACACCCAGT CCTTACAGCTCCACTGTGTCAGTCGGAACCGCCAAAAC CTACGGCCAGACCAGTGCAGCCACACGGCCAGGCCAC TGCGGACTGGCTGAAAAGCAGCATTGTGGCCCAGTGA ATCCCCTGCTGGGGGCTGCAACCCCTACAGGAAACAA TAAGCGGAGAAAACTGTGCAGCGGAAACACCACACCA ATCATTCACCTGAAGGGCGACCGGAACAGCCTGAAGT GTCTGCGGTACAGACTGCGAAAGCACAGTGACCATTAT CGCGATATCTCTAGTACTTGGCACTGGACCGGAGCTG GCAACGAGAAGACCGGCATTCTGACTGTGACCTACCAT TCAGAAACTCAGCGGACCAAATTTCTGAATACTGTGGC CATCCCCGATAGCGTGCAGATTCTGGTCGGGTATATGA CAATG	MQTPKETLSERLSALQ DKIIDHYENDSKDIDSQI QYWQLIRWENAIFFAAR EHGIQTLNHQVVPAYNI SKSKAHKAIELQMALQG LAQSAYKTEDWTLQDT CEELWNTEPTHCFKKG GQTVQVYFDGNKDNC MTYVAWDSVYYMTDA GTWDKTATCVSHRGLY YVKEGYNTFYIEFKSEC EKYGNTGTWEVHFGNN VIDCNDSMCSTSDDTV SATQLVKQLQHTPSPY SSTVSVGTAKTYGQTS AATRPGHCGLAEKQHC GPVNPLLGAATPTGNN KRRKLCSGNTTPIIHLK GDRNSLKCLRYRLRKH SDHYRDISSTWHWTGA GNEKTGILTVTYHSETQ RTKFLNTVAIPDSVQILV GYMTM	25.2%
HPV18 E4	ATGACTCTATGTGCAGTACCAGTGACGACACGGTATCC GCTACTCAGCTTGTTAAACAGCTACAGCACACCCCCTC ACCGTATTCCAGCACCGTGTCCGTGGGCACCGCAAAG ACCTACGGCCAGACGTCGGCTGCTACACGACCTGGAC ACTGTGGACTCGCGGAGAAGCAGCATTGTGGACCTGT CAACCCACTTCTCGGTGCAGCTACACCTACAGGCAACA ACAAAAGACGGAAACTCTGTAGTGGTAACACTACGCCT ATAA	ATGACCCTGTGTGCTGTCCCTGTGACTACAAGATACCC CCTGCTGTCCCTGCTGAACTCCTACTCCACCCCTCCCC ATAGAATCCCCGCACCATGCCCTTGGGCTCCACAGAG ACCAACTGCACGGAGAAGGCTGCTGCACGACCTGGAT ACCGTGGACAGCCGGCGGAGCAGCATCGTGGATCTGT CTACACACTTCAGTGTCCAGCTGCACCTCCAGGCCACC ACAAAGGACGGCAACTCTGTGGTCGTGACCCTGCGGC TG	MTLCAVPVTTRYPLLSL LNSYSTPPHRIPAPCPW APQRPTARRRLLHDLD TVDSRRSSIVDLSTHFS VQLHLQATTKDGNSVV VTLRL	27.6%
HPV18 E5	ATGTTATCACTTATTTTTTTATTTTGCTTTTGTGTATGCA TGTATGTGTGCTGCCATGTCCCGCTTTTGCCATCTGTC TGTATGTGTGCGTATGCATGGGTATTGGTATTTGTGTAT ATTGTGGTAATAACGTCCCCTGCCACAGCATTCACAGT ATATGTATTTTGTTTTTTATTGCCCATGTTACTATTGCAT ATACATGCTATATTGTCTTTACAGTAA	ATGCTGAGTCTGATTTTCCTGTTTTGTTTCTGCGTGTGT ATGTATGTCTGCTGTCATGTCCCACTGCTGCCTTCTGT CTGTATGTGTGCCTACGCTTGGGTGCTGGTCTTCGTGT ATATCGTGGTCATTACCAGCCCCGCAACCGCCTTTACA GTCTACGTGTTCTGCTTTCTGCTGCCTATGCTGCTGCT GCACATCCATGCTATTCTGAGCCTCCAG	MLSLIFLFCFCVCMYVC CHVPLLPSVCMCAYAW VLVFVYIVVITSPATAFT VYVFCFLLPMLLLHIHAI LSLQ	29.2%

HPV18 E6	ATGGCGCGCTTTGAGGATCCAACACGGCGACCCTACA AGCTACCTGATCTGTGCACGGAACTGAACACTTCACTG CAAGACATAGAAATAACCTGTGTATATTGCAAGACAGTA TTGGAACTTACAGAGGTATTTGAATTTGCATTTAAAGAT TTATTTGTGGTGTATAGAGACAGTATACCGCATGCTGC ATGCCATAAATGTATAGATTTTTATTCTAGAATTAGAGA ATTAAGACATTATTCAGACTCTGTGTATGGAGACACATT GGAAAAACTAACTAACACTGGGTTATACAATTTATTAAT AAGGTGCCTGCGGTGCCAGAAACCGTTGAATCCAGCA GAAAAACTTAGACACCTTAATGAAAAACGACGATTTCAC AACATAGCTGGGCACTATAGAGGCCAGTGCCATTCGTG CTGCAACCGAGCACGACAGGAACGACTCCAACGACGC AGAGAAACACAAGTATAA	ATGGCTAGATTTGAGGACCCAACAAGACGCCCCTATAA ACTGCCCGACCTGTGTACCGAACTGAATACTTCTCTGC AGGACATTGAGATCACATGCGTGTACTGTAAAACTGTC CTGGAGCTGACCGAAGTGTTCGAGTTTGCTTTCAAGGA CCTGTTTGTGGTCTACAGAGATTCTATTCCCCACGCCG CTTGCCATAAATGTATCGACTTCTATAGTCGGATTAGAG AACTGAGGCACTACAGCGACTCCGTCTATGGAGATACT CTGGAGAAGCTGACCAACACAGGCCTGTACAATCTGCT GATCCGATGCCTGAGGTGTCAGAAGCCCCTGAACCCT GCCGAAAAACTGCGGCACCTGAACGAGAAGCGGAGAT TTCACAATATTGCAGGCCATTATCGCGGGCAGTGCCAT TCCTGCTGTAATAGGGCCCGCCAGGAACGACTCCAGA GGCGCCGAGAGACCCAGGTG	MARFEDPTRRPYKLPD LCTELNTSLQDIEITCVY CKTVLELTEVFEFAFKD LFVVYRDSIPHAACHKC IDFYSRIRELRHYSDSV YGDTLEKLTNTGLYNLLI RCLRCQKPLNPAEKLR HLNEKRRFHNIAGHYR GQCHSCCNRARQERL QRRRETQV	27.1%
HPV18 E7	ATGCATGGACCTAAGGCAACATTGCAAGACATTGTATT GCATTTAGAGCCCCAAAATGAAATTCCGGTTGACCTTC TATGTCACGAGCAATTAAGCGACTCAGAGGAAGAAAAC GATGAAATAGATGGAGTTAATCATCAACATTTACCAGCC CGACGAGCCGAACCACAACGTCACACAATGTTGTGTAT GTGTTGTAAGTGTGAAGCCAGAATTGAGCTAGTAGTAG AAAGCTCAGCAGACGACCTTCGAGCATTCCAGCAGCT GTTTCTGAACACCCTGTCCTTTGTGTGTCCGTGGTGTG CATCCCAGCAGTAA	ATGCATGGACCCAAGGCTACTCTGCAGGATATTGTGCT GCACCTGGAACCTCAGAATGAAATCCCCGTGGACCTG CTGTGTCACGAACAGCTGTCTGACAGTGAGGAAGAGA ACGACGAGATCGATGGCGTGAATCACCAGCATCTGCC CGCCCGGAGAGCTGAACCTCAGAGGCACACCATGCTG TGCATGTGCTGTAAGTGTGAAGCTCGCATTGAGCTGGT GGTCGAAAGCTCCGCAGACGATCTGCGAGCCTTCCAG CAGCTGTTTCTGAACACACTGAGCTTCGTCTGCCCCTG GTGTGCCTCTCAGCAG	MHGPKATLQDIVLHLEP QNEIPVDLLCHEQLSDS EEENDEIDGVNHQHLP ARRAEPQRHTMLCMCC KCEARIELVVESSADDL RAFQQLFLNTLSFVCP WCASQQ	23.8%
HPV18 L1	ATGGCTTTGTGGCGGCCTAGTGACAATACCGTATATCT TCCACCTCCTTCTGTGGCAAGAGTTGTAAATACCGATG ATTATGTGACTCGCACAAGCATATTTTATCATGCTGGCA GCTCTAGATTATTAACTGTTGGTAATCCATATTTTAGGG TTCCTGCAGGTGGTGGCAATAAGCAGGATATTCCTAAG GTTTCTGCATACCAATATAGAGTATTTAGGGTGCAGTTA CCTGACCCAAATAAATTTGGTTTACCTGATACTAGTATT TATAATCCTGAAACACAACGTTTAGTGTGGGCCTGTGC TGGAGTGGAAATTGGCCGTGGTCAGCCTTTAGGTGTTG GCCTTAGTGGGCATCCATTTTATAATAAATTAGATGACA CTGAAAGTTCCCATGCCGCCACGTCTAATGTTTCTGAG GACGTTAGGGACAATGTGTCTGTAGATTATAAGCAGAC ACAGTTATGTATTTTGGGCTGTGCCCCTGCTATTGGGG AACACTGGGCTAAAGGCACTGCTTGTAAATCGCGTCCT TTATCACAGGGCGATTGCCCCCCTTTAGAACTTAAAAA CACAGTTTTGGAAGATGGTGATATGGTAGATACTGGAT ATGGTGCCATGGACTTTAGTACATTGCAAGATACTAAAT GTGAGGTACCATTGGATATTTGTCAGTCTATTTGTAAAT ATCCTGATTATTTACAAATGTCTGCAGATCCTTATGGGG ATTCCATGTTTTTTTTGCTTACGGCGTGAGCAGCTTTTTG CTAGGCATTTTTTGGAATAGAGCAGGTACTATGGGTGAC ACTGTGCCTCAATCCTTATATATTAAAGGCACAGGTATG CGTGCTTCACCTGGCAGCTGTGTGTATTCTCCCTCTCC AAGTGGCTCTATTGTTACCTCTGACTCCCAGTTGTTTAA TAAACCATATTGGTTACATAAGGCACAGGGTCATAACA ATGGTGTTTGCTGGCATAATCAATTATTTGTTACTGTGG TAGATACCACTCGCAGTACCAATTTAACAATATGTGCTT CTACACAGTCTCCTGTACCTGGGCAATATGATGCTACC AAATTTAAGCAGTATAGCAGACATGTTGAGGAATATGAT TTGCAGTTTATTTTTCAGTTGTGTACTATTACTTTAACTG CAGATGTTATGTCCTATATTCATAGTATGAATAGCAGTA TTTTAGAGGATTGGAACTTTGGTGTTCCCCCCCCGCCA ACTACTAGTTTGGTGGATACATATCGTTTTGTACAATCT GTTGCTATTACCTGTCAAAAGGATGCTGCACCGGCTGA AAATAAGGATCCCTATGATAAGTTAAAGTTTTGGAATGT GGATTTAAAGGAAAAGTTTTCTTTAGACTTAGATCAATA TCCCCTTGGACGTAAATTTTTGGTTCAGGCTGGATTGC GTCGCAAGCCCACCATAGGCCCTCGCAAACGTTCTGC TCCATCTGCCACTACGTCTTCTAAACCTGCCAAGCGTG TGCGTGTACGTGCCAGGAAGTAA	ATGTGCCTGTATACCCGCGTGCTGATTCTGCATTACCA TCTGCTGCCCCTGTACGGACCCCTGTACCATCCAAGAC CTCTGCCTCTGCACTCCATTCTGGTGTACATGGTCCAC ATCATTATCTGCGGGCATTATATTATCCTGTTCCTGCGC AACGTGAATGTCTTCCCTATCTTCCTCCAGATGGCTCT GTGGCGACCAAGTGACAACACCGTGTACCTGCCCCCT CCATCAGTCGCACGGGTGGTCAATACAGACGATTACGT GACCCCTACAAGCATTTTCTATCATGCAGGCAGCTCCC GACTGCTGACCGTGGGAAACCCTTATTTTCGGGTCCCA GCAGGCGGGGGAAATAAGCAGGATATCCCAAAAGTGT CTGCCTACCAGTATCGGGTGTTCAGAGTCCAGCTGCC CGACCCTAACAAGTTTGGCCTGCCAGATACTAGCATCT ACAATCCCGAGACCCAGAGACTGGTGTGGGCTTGTGC AGGAGTCGAAATCGGAAGGGGACAGCCACTGGGAGTG GGACTGTCTGGACACCCTTTCTACAACAAGCTGGACGA TACAGAGTCTAGTCATGCCGCTACTTCAAACGTGAGCG AAGACGTCAGAGATAATGTGAGCGTGGACTATAAACAG ACTCAGCTGTGCATTCTGGGATGTGCACCAGCTATCGG AGAGCACTGGGCTAAGGGAACCGCATGCAAATCTAGG CCTCTGAGTCAGGGCGACTGTCCCCCTCTGGAGCTGA AGAATACCGTGCTGGAAGACGGGGATATGGTCGATAC AGGGTACGGAGCTATGGACTTTTCTACACTCCAGGATA CTAAGTGCGAGGTGCCTCTGGACATTTGCCAGAGTATC TGTAAATACCCAGATTACCTCCAGATGTCCGCTGACCC CTATGGCGATTCTATGTTCTTTTGTCTGCGGAGAGAAC AGCTGTTCGCCAGGCACTTTTGGAACCGCGCTGGCAC AATGGGCGACACAGTGCCCCAGAGCCTGTACATTAAG GGCACAGGAATGCCAGCATCCCCTGGATCTTGCGTGT ATAGTCCATCACCCAGCGGCTCCATCGTGACCTCTGAC AGTCAGCTGTTCAATAAGCCATACTGGCTGCACAAAGC TCAGGGGCATAACAATGGAGTGTGCTGGCATAACCAG CTGTTTGTCACAGTGGTCGATACCACACCCAGCACTAA TCTGACCATCTGTGCCAGTACACAGTCACCTGTGCCAG GACAGTACGACGCTACTAAGTTCAAACAGTACTCCAGA CACGTGGAGGAATATGACCTCCAGTTCATCTTTCAGCT GTGCACTATCACCCTGACAGCCGACGTGATGTCATACA TTCATAGCATGAACTCAAGCATCCTGGAGGATTGGAAT TTCGGCGTGCCACCCCCTCCAACTACCAGCCTGGTGG ACACTTATCGCTTTGTGCAGTCCGTCGCTATTACCTGT CAGAAGGATGCAGCCCCCGCAGAGAACAAAGACCCTT ACGATAAGCTGAAATTCTGGAATGTGGACCTGAAGGAA AAATTTTCCCTGGACCTGGATCAGTATCCCCTGGGACG GAAGTTTCTGGTGCAGGCAGGACTGAGGCGAAAGCCA ACCATCGGACCACGAAAACGGAGCGCACCTTCCGCCA CAACTTCCTCTAAGCCAGCAAAAAGAGTGAGGGTCCGC GCCCGAAAA	MALWRPSDNTVYLPPP SVARVVNTDDYVTRTSI FYHAGSSRLLTVGNPYF RVPAGGGNKQDIPKVS AYQYRVFRVQLPDPNK FGLPDTSIYNPETQRLV WACAGVEIGRGQPLGV GLSGHPFYNKLDDTES SHAATSNVSEDVRDNV SVDYKQTQLCILGCAPA IGEHWAKGTACKSRPL SQGDCPPLELKNTVLE DGDMVDTGYGAMDFS TLQDTKCEVPLDICQSI CKYPDYLQMSADPYGD SMFFCLRREQLFARHF WNRAGTMGDTVPQSL YIKGTGMRASPGSCVY SPSPSGSIVTSDSQLFN KPYWLHKAQGHNNGV CWHNQLFVTVVDTTRS TNLTICASTQSPVPGQY DATKFKQYSRHVEEYD LQFIFQLCTITLTADVMS YIHSMNSSILEDWNFGV PPPPTTSLVDTYRFVQS VAITCQKDAAPAENKDP YDKLKFWNVDLKEKFSL DLDQYPLGRKFLVQAG LRRKPTIGPRKRSAPSA TTSSKPAKRVRVRARK	34.1%
HPV18 L2	ATGGTATCCCACCGTGCCGCACGACGCAAACGGGCTT CGGTAACTGACTTATATAAAACATGTAAACAATCTGGTA CATGTCCACCTGATGTTGTTCCTAAGGTGGAGGGCACC ACGTTAGCAGATAAAATATTGCAATGGTCAAGCCTTGG TATATTTTTGGGTGGACTTGGCATAGGTACTGGCAGTG GTACAGGGGGTCGTACAGGGTACATTCCATTGGGTGG GCGTTCCAATACAGTGGTGGATGTTGGTCCTACACGTC CCCCAGTGGTTATTGAACCTGTGGGCCCCACAGACCC ATCTATTGTTACATTAATAGAGGACTCCAGTGTGGTTAC ATCAGGTGCACCTAGGCCTACGTTTACTGGCACGTCTG GGTTTGATATAACATCTGCGGGTACAACTACACCTGCG GTTTTGGATATCACACCTTCGTCTACCTCTGTGTCTATT TCCACAACCAATTTTACCAATCCTGCATTTTCTGATCCG TCCATTATTGAAGTTCCACAAACTGGGGAGGTGGCAGG TAATGTATTTGTTGGTACCCCTACATCTGGAACACATGG GTATGAGGAAATACCTTTACAAACATTTGCTTCTTCTGG TACGGGGGAGGAACCCATTAGTAGTACCCCATTGCCTA CTGTGCGGCGTGTAGCAGGTCCCCGCCTTTACAGTAG GGCCTACCAACAAGTGTCAGTGGCTAACCCTGAGTTTC TTACACGTCCATCCTCTTTAATTACATATGACAACCCGG CCTTTGAGCCTGTGGACACTACATTAACATTTGATCCTC GTAGTGATGTTCCTGATTCAGATTTTATGGATATTATCC GTCTACATAGGCCTGCTTTAACATCCAGGCGTGGGACT GTTCGCTTTAGTAGATTAGGTCAACGGGCAACTATGTT TACCCGCAGCGGTACACAAATAGGTGCTAGGGTTCACT TTTATCATGATATAAGTCCTATTGCACCTTCCCCAGAAT ATATTGAACTGCAGCCTTTAGTATCTGCCACGGAGGAC AATGACTTGTTTGATATATATGCAGATGACATGGACCCT GCAGTGCCTGTACCATCGCGTTCTACTACCTCCTTTGC ATTTTTTAAATATTCGCCCACTATATCTTCTGCCTCTTCC TATAGTAATGTAACGGTCCCTTTAACCTCCTCTTGGGAT GTGCCTGTATACACGGGTCCTGATATTACATTACCATCT ACTACCTCTGTATGGCCCATTGTATCACCCACGGCCCC TGCCTCTACACAGTATATTGGTATACATGGTACACATTA TTATTTGTGGCCATTATATTATTTTATTCCTAAGAAACGT	ATGGTGTCTCATCGCGCAGCACGACGGAAAAGAGCCA GTGTGACCGACCTGTATAAAACCTGTAAGCAGAGCGGA ACTTGCCCTCCTGACGTGGTCCCCAAGGTGGAGGGAA CCACACTGGCTGATAAGATCCTCCAGTGGAGCAGCCT GGGAATCTTCCTGGGAGGACTGGGAATTGGGACTGGA AGCGGCACCGGAGGCCGAACAGGCTACATCCCTCTGG GGGGACGCAGCAACACCGTGGTGGACGTGGGACCAA CAAGGCCCCCTGTGGTCATTGAGCCTGTGGGGCCAAC TGACCCCTCCATCGTCACCCTGATTGAAGATTCTAGTG TGGTCACATCTGGGGCCCCACGACCAACATTCACTGG CACCTCCGGGTTTGACATCACCTCTGCTGGAACTACCA CACCCGCCGTGCTGGACATCACTCCATCAAGCACCAG TGTGTCAATTAGCACTACCAACTTCACAAATCCAGCCTT TAGTGATCCCTCAATCATTGAGGTGCCCCAGACTGGCG AAGTCGCTGGGAATGTGTTCGTCGGCACACCCACTAG CGGAACCCACGGCTACGAGGAAATCCCTCTCCAGACA TTTGCATCCTCTGGGACTGGAGAGGAACCAATTAGTTC AACACCTCTGCCAACTGTGCGGAGAGTCGCAGGACCA CGACTGTACAGCAGAGCATATCAGCAGGTGTCCGTCG CCAACCCCGAGTTCCTGACTCGGCCTAGCTCCCTGATC ACCTATGACAATCCCGCTTTCGAACCTGTGGATACAAC TCTGACCTTTGACCCTCGGAGCGATGTGCCAGACAGT GATTTTATGGACATCATTAGACTGCATAGGCCAGCACT GACTAGCAGGCGCGGGACCGTGCGCTTCAGCCGACTG GGACAGAGGGCCACCATGTTTACACGCTCCGGAACAC AGATTGGCGCTAGGGTGCACTTCTACCATGATATCTCA CCAATTGCACCCAGCCCTGAGTATATCGAGCTGCAACC CCTGGTGTCTGCCACCGAGGACAACGATCTGTTCGAC ATCTACGCAGACGATATGGACCCCGCCGTGCCCGTGC CCAGCCGGAGCACCACATCTTTTGCTTTCTTTAAGTAC AGTCCCACCATCTCTAGTGCATCAAGCTATAGCAATGT GACCGTCCCTCTGACATCCTCTTGGGACGTGCCCGTCT ATACAGGCCCTGATATCACTCTGCCAAGCACTACCTCC GTGTGGCCTATTGTCAGTCCTACCGCACCAGCCTCAAC ACAGTACATCGGGATTCACGGAACCCATTACTATCTGT	MVSHRAARRKRASVTD LYKTCKQSGTCPPDVV PKVEGTTLADKILQWSS LGIFLGGLGIGTGSGTG GRTGYIPLGGRSNTVV DVGPTRPPVVIEPVGPT DPSIVTLIEDSSVVTSGA PRPTFTGTSGFDITSAG TTTPAVLDITPSSTSVSI STTNFTNPAFSDPSIIEV PQTGEVAGNVFVGTPT SGTHGYEEIPLQTFASS GTGEEPISSTPLPTVRR VAGPRLYSRAYQQVSV ANPEFLTRPSSLITYDN PAFEPVDTTLTFDPRSD VPDSDFMDIIRLHRPAL TSRRGTVRFSRLGQRA TMFTRSGTQIGARVHFY HDISPIAPSPEYIELQPL VSATEDNDLFDIYADDM DPAVPVPSRSTTSFAFF KYSPTISSASSYSNVTV PLTSSWDVPVYTGPDIT LPSTTSVWPIVSPTAPA STQYIGIHGTHYYLWPL YYFIPKKRKRVPYFFAD GFVAA	27.2%

	AAACGTGTTCCCTATTTTTTTGCAGATGGCTTTGTGGCG GCCTAG	GGCCACTGTACTATTTCATCCCCAAGAAACGAAAACGG GTGCCATATTTCTTTGCCGACGGCTTTGTGGCCGCT		
HPV31 E1	ATGGCTGATCCAGCAGGTACAGATGGGGAGGGGACGG GATGCAATGGTTGGTTTTATGTAGAAGCAGTAATTGAC AGACAGACAGGGGACAACATTTCAGAGGACGAAAATG AAGACAGTAGTGATACTGGGGAGGATATGGTTGACTTT ATTGACAATTGTAATGTATACAACAATCAGGCAGAAGCA GAGACAGCACAGGCATTGTTTCATGCACAGGAAGCGG AGGAACATGCAGAGGCTGTGCAGGTTCTAAAACGAAA GTATGTAGGTAGTCCTTTAAGTGATATTAGTAGTTGTGT GGATTATAATATTAGTCCACGGTTAAAAGCTATATGCAT AGAAAATAACAGTAAAACAGCAAAACGAAGACTCTTTG AACTTCCAGACAGCGGGTATGGCAATACTGAAGTGGAA ACGCAGCAGATGGTACAGGTAGAGGAGCAACAAACAA CATTAAGTTGTAATGGTAGTGACGGGACACATAGTGAA CGAGAGAATGAAACTCCAACACGTAATATATTGCAAGT GTTAAAAACTAGCAATGGTAAAGCTGCTATGTTAGGTAA ATTTAAAGAATTATATGGTGTAAGTTTTATGGAACTAATT AGGCCATTTCAAAGCAATAAAAGCACATGTACTGATTG GTGTGTAGCTGCGTTTGGAGTTACAGGTACAGTTGCAG AAGGATTTAAAACCCTATTGCAACCATATTGTTTGTATT GCCATTTACAAAGTTTAGCATGTTCCTGGGGCATGGTT ATGTTAATGCTTGTGAGATTTAAATGTGCAAAAAATAGA ATAACAATTGAAAAATTATTAGAAAAATTATTGTGTATAT CTACAAATTGTATGTTAATTCAGCCACCCAAATTACGTA GCACAGCTGCAGCATTATATTGGTACAGAACAGGAATG TCAAACATTAGCGATGTATATGGTGAAACACCAGAATG GATAGAAAGACAAACAGTATTACAGCATAGTTTTAATGA CACAACATTTGATTTGTCCCAAATGGTACAATGGGCATA TGACAATGATGTTATGGATGATAGTGAAATTGCCTATAA ATATGCACAATTAGCTGACAGTGATAGTAATGCATGTG CATTTTTAAAAAGTAATTCGCAGGCAAAAATAGTTAAAG ATTGTGGAACAATGTGTAGACATTATAAACGAGCAGAA AAACGACAAATGTCCATGGGACAGTGGATTAAAAGTAG ATGTGACAAAGTTAGTGACGAAGGTGACTGGAGGGAC ATAGTAAAGTTTTTAAGATATCAACAAATAGAATTTGTGT CATTTTTATCTGCATTAAAGCTGTTTTTAAAAGGAGTGC CAAAGAAAAACTGTATTTTAATACATGGTGCACCTAATA CAGGTAAATCATATTTTGGAATGAGCCTTATTAGCTTTT TACAAGGATGTATAATATCATATGCAAATTCAAAAAGTC ATTTTTGGTTACAACCACTGGCTGATGCTAAAATAGGCA TGTTAGATGATGCTACAACGCCATGTTGGCATTATATAG ACAATTACCTACGAAATGCACTAGATGGCAACCCTGTA TCTATAGATGTAAAGCATAAAGCTTTAATGCAGTTAAAA TGTCCTCCTTTATTGATTACATCTAATATAAATGCAGGT AAGGATGACAGATGGCCATACCTACATAGCAGACTGGT GGTTTTTACATTTCCAAATCCATTTCCATTTGACAAAAA CGGAAATCCAGTATATGAATTAAGTGATAAAAACTGGAA ATCCTTTTTCTCAAGGACGTGGTGCAGATTAAATTTGCA CGAGGAAGAGGACAAAGAAAACGATGGAGACTCTTTCT CAACGTTTAAATGTGTGTCAGGACAAAATATTAGAACAT TATGA	ATGGCAGACCCAGCAGGAACAGACGGAGAGGGGACA GGATGTAATGGATGGTTTTATGTGGAGGCAGTGATTGA CAGGCAGACCGGCGACAACATCAGCGAAGATGAGAAT GAAGACAGCTCCGATACCGGCGAGGACATGGTGGATT TCATTGACAACTGCAATGTCTACAACAATCAGGCTGAG GCAGAAACAGCCCAGGCTCTGTTTCACGCACAGGAGG CCGAGGAACATGCAGAAGCCGTGCAGGTCCTGAAGAG AAAATACGTGGGGTCTCCCCTGAGTGACATCTCTAGTT GCGTCGATTATAACATTTCTCCTAGGCTGAAGGCCATC TGTATTGAGAACAATAGTAAGACCGCTAAACGGAGACT GTTCGAACTGCCAGACTCTGGCTACGGGAACACCGAG GTGGAAACACAGCAGATGGTGCAGGTCGAGGAACAGC AGACCACACTGTCATGCAATGGAAGCGATGGCACACA CAGCGAGAGGGAAAACGAGACCCCCACACGCAATATC CTGCAGGTGCTGAAGACTAGCAACGGCAAAGCCGCTA TGCTGGGGAAGTTTAAAGAGCTGTATGGCGTGTCCTTC ATGGAACTGATTCGCCCCTTTCAGTCAAATAAGAGCAC TTGCACCGACTGGTGTGTGGCAGCCTTCGGGGTGACA GGAACTGTCGCCGAGGGCTTCAAGACACTGCTGCAGC CTTACTGCCTGTATTGTCACCTGCAGTCCCTGGCCTGC TCTTGGGGCATGGTCATGCTGATGCTGGTCCGATTCAA GTGTGCTAAAAACCGGATCACTATTGAGAAGCTGCTGG AAAAACTGCTGTGCATCTCAACCAATTGTATGCTGATTC AGCCCCCTAAGCTGCGCAGCACAGCTGCAGCCCTGTA CTGGTATCGGACTGGAATGTCCAATATCTCTGATGTGT ACGGCGAAACTCCTGAGTGGATTGAACGCCAGACCGT CCTGCAGCACAGCTTCAACGACACTACCTTTGATCTGT CCCAGATGGTGCAGTGGGCATATGACAATGATGTCATG GACGATTCTGAGATCGCCTACAAGTATGCTCAGCTGGC AGACTCAGATAGCAACGCTTGCGCATTTCTGAAATCCA ATTCTCAGGCAAAGATCGTGAAAGACTGCGGCACAATG TGTAGGCATTACAAGCGCGCCGAGAAACGACAGATGA GCATGGGCCAGTGGATTAAGAGTCGGTGTGATAAAGT GTCAGATGAGGGGGACTGGAGAGATATCGTCAAGTTC CTGAGGTATCAGCAGATTGAATTCGTGAGTTTTCTGTC AGCTCTGAAGCTGTTTCTGAAAGGCGTGCCTAAGAAAA ACTGCATCCTGATTCACGGCGCCCCAAATACTGGGAAG AGTTACTTCGGAATGAGCCTGATCTCCTTTCTGCAGGG GTGTATCATTAGCTATGCCAACAGTAAGTCACATTTCTG GCTGCAGCCACTGGCCGACGCTAAAATCGGAATGCTG GACGATGCCACAACTCCCTGCTGGCACTACATTGATAA CTATCTGAGAAATGCTCTGGACGGCAATCCCGTGTCCA TCGATGTCAAGCATAAAGCACTGATGCAGCTGAAGTGT CCACCCCTGCTGATCACTTCCAACATTAATGCCGGGAA AGACGATCGCTGGCCTTACCTGCACTCTCGACTGGTG GTCTTCACCTTTCCTAACCCATTCCCCTTTGACAAGAAC GGAAATCCAGTGTATGAGCTGAGCGATAAGAATTGGAA ATCTTTCTTTAGTCGGACCTGGTGCAGACTGAACCTGC ATGAGGAAGAGGACAAGGAGAATGACGGGGATAGCTT CTCCACCTTTAAATGTGTGTCCGGACAGAACATCAGGA CACTG	METLSQRLNVCQDKILE HYENDSKRLCDHIDYW KHIRLECVLMYKAREM GIHSINHQVVPALSVSK AKALQAIELQMMLETLN NTEYKNEDWTMQQTSL ELYLTAPTGCLKKHGYT VEVQFDGDVHNTMHYT NWKFIYLCIDGQCTVVE GQVNCKGIYYVHEGHIT YFVNFTEEAKKYGTGK KWEVHAGGQVIVFPES VFSSDEISFAGIVTKLPT ANNTTTSNSKTCALGTS EGVRRATTSTKRPRTE PEHRNTHHPNKLLRGD SVDSVNCGVISAAACTN QTRAVSCPATTPIIHLKG DANILKCLRYRLSKYKQ LYEQVSSTWHWTCTDG KHKNAIVTLTYISTSQRD DFLNTVKIPNTVSVSTG YMTI	24.7%
HPV31 E2	ATGGAGACTCTTTCTCAACGTTTAAATGTGTGTCAGGA CAAAATATTAGAACATTATGAAAATGATAGTAAACGACT TTGTGATCATATAGACTATTGGAAACATATTCGACTTGA ATGTGTATTAATGTATAAAGCAAGAGAAATGGGAATACA CAGTATTAACCACCAGGTGGTGCCAGCGTTGTCAGTAT CAAAGGCCAAAGCCTTACAAGCTATTGAACTACAAATG ATGTTGGAAACATTAAATAACACTGAATACAAAAATGAG GACTGGACAATGCAGCAAACAAGTCTTGAACTGTATTT AACTGCACCTACAGGGTGTTTAAAAAAACATGGATATA CTGTAGAGGTGCAATTTGATGGTGATGTACACAACACC ATGCATTATACTAACTGGAAATTTATATACCTATGTATA GATGGCCAATGTACTGTTGTGGAAGGGCAAGTTAATTG TAAGGGCATTTATTATGTACATGAAGGACATATAACATA TTTTGTAAATTTTACAGAAGAGGCAAAAAAATATGGGAC TGGTAAAAAATGGGAAGTGCATGCGGGTGGTCAGGTA ATTGTTTTTCCTGAATCTGTATTTAGCAGTGACGAAATA TCCTTTGCTGGGATTGTTACAAAGCTACCAACAGCCAA CAACACCACCACATCGAATTCCAAAACCTGCGCCTTGG GCACCAGTGAAGGTGTGCGGCGGGCGACGACGTCTAC TAAGCGACCAAGAACAGAGCCAGAGCACAGAAACACC CACCACCCCAACAAGTTGTTGCGAGGCGACTCCGTGG ACAGTGTCAACTGTGGGGTTATCAGTGCAGCTGCATGC ACAAACCAAACAAGGGCTGTCAGTTGTCCTGCAACTAC ACCTATAATACACTTAAAAGGTGATGCAAATATATTAAA ATGTTTAAGATATAGGCTGTCAAAATATAAACAATTGTA TGAACAAGTGTCATCTACATGGCATTGGACATGTACAG ATGGAAAACATAAAAATGCTATTGTAACCTTAACATATA TAAGTACATCACAAAGAGACGATTTTTTAAATACTGTAA AAATACCTAACACAGTATCAGTGTCAACAGGATATATGA CTATTTAG	ATGGAAACTCTGTCCCAGCGACTGAACGTCTGCCAGG ATAAGATTCTGGAACACTACGAAAATGATTCTAAAAGAC TGTGCGACCACATCGACTACTGGAAGCACATTAGACTG GAGTGCGTGCTGATGTATAAAGCCAGGGAAATGGGCA TCCACAGCATTAACCATCAGGTGGTCCCCGCACTGTCA GTGAGCAAGGCCAAAGCTCTGCAGGCCATCGAGCTGC AGATGATGCTGGAAACCCTGAACAATACAGAGTACAAG AATGAAGACTGGACTATGCAGCAGACCTCCCTGGAGCT GTACCTGACTGCCCCTACCGGCTGCCTGAAGAAACAC GGGTATACAGTGGAAGTCCAGTTCGACGGCGATGTGC ACAACACAATGCATTACACTAACTGGAAGTTTATCTATC TGTGCATTGATGGGCAGTGTACCGTGGTCGAGGGACA GGTGAACTGTAAAGGCATCTACTATGTCCACGAAGGAC ATATCACTTACTTCGTGAACTTCACCGAGGAAGCTAAG AAATATGGAACCGGCAAGAAATGGGAGGTCCATGCAG GCGGGCAGGTCATCGTCTTCCCTGAGTCAGTGTTTAGC TCCGATGAAATCAGCTTCGCTGGCATTGTCACCAAGCT GCCAACAGCAAACAATACCACAACTTCCAACTCTAAAA CATGCGCACTGGGAACTTCCGAGGGAGTGCGGAGAGC TACCACATCTACCAAGAGGCCCCGCACAGAGCCTGAA CACCGCAACACCCACCATCCAAACAAGCTGCTGCGAG GGGACTCTGTGGATAGTGTCAACTGCGGAGTGATCAG TGCCGCTGCATGTACAAATCAGACTAGGGCAGTCAGCT GCCCAGCCACTACCCCCATCATTCATCTGAAGGGCGA CGCTAACATTCTGAAATGTCTGCGATACCGGCTGTCTA AGTACAAACAGCTGTATGAGCAGGTGTCTAGTACATGG CACTGGACATGTACTGATGGGAAGCATAAAAATGCCAT CGTGACCCTGACATACATTAGTACCTCACAGCGGGACG ATTTTCTGAACACAGTGAAGATCCCCAATACTGTGAGC GTCTCCACTGGCTATATGACCATT	METLSQRLNVCQDKILE HYENDSKRLCDHIDYW KHIRLECVLMYKAREM GIHSINHQVVPALSVSK AKALQAIELQMMLETLN NTEYKNEDWTMQQTSL ELYLTAPTGCLKKHGYT VEVQFDGDVHNTMHYT NWKFIYLCIDGQCTVVE GQVNCKGIYYVHEGHIT YFVNFTEEAKKYGTGK KWEVHAGGQVIVFPES VFSSDEISFAGIVTKLPT ANNTTTSNSKTCALGTS EGVRRATTSTKRPRTE PEHRNTHHPNKLLRGD SVDSVNCGVISAAACTN QTRAVSCPATTPIIHLKG DANILKCLRYRLSKYKQ LYEQVSSTWHWTCTDG KHKNAIVTLTYISTSQRD DFLNTVKIPNTVSVSTG YMTI	24.7%
HPV31 E4	TTGTTTTTCCTGAATCTGTATTTAGCAGTGACGAAATAT CCTTTGCTGGGATTGTTACAAAGCTACCAACAGCCAAC AACACCACCACATCGAATTCCAAAACCTGCGCCTTGGG CACCAGTGAAGGTGTGCGGCGGGCGACGACGTCTACT AAGCGACCAAGAACAGAGCCAGAGCACAGAAACACCC ACCACCCCAACAAGTTGTTGCGAGGCGACTCCGTGGA CAGTGTCAACTGTGGGGTTATCAGTGCAGCTGCATGCA CAAACCAAACAAGGGCTGTCAGTTGTCCTGCAACTACA CCTATAA	ATGTTTTTCCTGAACCTGTATCTGGCCGTGACAAAGTAT CCTCTGCTGGGCCTGCTGCAGTCTTATCAGCAGCCTAC CACCCCCCCTCACCGAATCCCAAAGCCTGCACCATGG GCTCCAGTGAAAGTCTGCGGAGGACGAAGAAGGCTGC TGTCAGACCAGGAGCAGAGCCAGTCCACTGAAACCCC CACCACACCTACAAGCTGCTGTGAGGCAACACCCTGG ACTGTGTCTACCGTCGGACTGAGTGTGCAGCTGCACG CCCAGACCAAGCAGGGCCTGTCTGTGGTCCTGCAGCT GCATCTG	LFFLNLYLAVTKYPLLGL LQSYQQPTTPPHRIPKP APWAPVKVCGGRRRLL SDQEQSQSTETPTTPT SCCEATPWTVSTVGLS VQLHAQTKQGLSVVLQ LHL	26.8%
HPV31 E5	ATGATTGAACTAAATATTTCTACAGTAAGCATTGTGCTA TGCTTTTTGCTTTGCTTTTGTGTGCTACTATTTGTGTGT CTTGTCATACGTCCACTTGTGCTGTCTGTGTCGGTATAT GCAACACTACTATTATTAATTGTGATTTTATGGGTTATT GCAACCTCTCCATTACGTTGTTTTTGTATATATGTTGTG TTTATATATATTCCATTATTTGTAATTCATACACATGCAT CTTTTTTAAGTCAACAGTAA	ATGATTGAACTGAATATCTCTACCGTGTCCATTGTCCTG TGTTTTCTGCTGTGTTTCTGCGTCCTGCTGTTTGTCTGC CTGGTCATCCGGCCCCTGGTGCTGAGCGTGTCCGTCT ACGCCACCCTGCTGCTGCTGATCGTGATTCTGTGGGTC ATCGCTACATCCCCCCTGAGATGCTTCTGTATCTACGT GGTCTTTATCTATATTCCTCTGTTCGTGATCCACACCCA TGCCTCTTTTCTGAGTCAGCAG	MIELNISTVSIVLCFLLCF CVLLFVCLVIRPLVLSVS VYATLLLLIVILWVIATSP LRCFCIYVVFIYIPLFVIH THASFLSQQ	28.6%

HPV31 E6	ATGTTCAAAAATCCTGCAGAAAGACCTCGGAAATTGCA TGAACTAAGCTCGGCATTGGAAATACCCTACGATGAAC TAAGATTGAATTGTGTCTACTGCAAAGGTCAGTTAACAG AAACAGAGGTATTAGATTTTGCATTTACAGATTTAACAA TAGTATATAGGGACGACACACCACACGGAGTGTGTACA AAATGTTTAAGATTTTATTCAAAAGTAAGTGAATTTAGAT GGTATAGATATAGTGTGTATGGAACAACATTAGAAAAAT TGACAAACAAAGGTATATGTGATTTGTTAATTAGGTGTA TAACGTGTCAAAGACCGTTGTGTCCAGAAGAAAAACAA AGACATTTGGATAAAAAGAAACGATTCCACAACATAGG AGGAAGGTGGACAGGACGTTGCATAGCATGTTGGAGA AGACCTCGTACTGAAACCCAAGTGTAA	ATGTTTAAGAACCCCGCCGAGAGACCAAGAAAGCTGCA CGAACTGTCATCCGCCCTGGAAATCCCCTACGACGAAC TGAGGCTGAACTGCGTGTACTGTAAAGGGCAGCTGAC TGAGACCGAAGTCCTGGACTTCGCCTTTACCGATCTGA CAATCGTGTATAGGGACGATACTCCACACGGAGTCTGC ACCAAATGTCTGCGGTTCTACAGCAAGGTGTCCGAGTT TAGGTGGTACCGCTATTCTGTCTATGGAACCACACTGG AAAAACTGACAAACAAGGGCATTTGCGACCTGCTGATC AGATGCATTACTTGTCAGAGGCCCCTGTGTCCTGAGGA AAAGCAGCGCCACCTGGATAAGAAAAAGCGATTCCATA ATATCGGAGGACGATGGACCGGACGATGCATTGCTTG TTGGCGGAGACCCCGGACAGAGACTCAGGTG	MFKNPAERPRKLHELS SALEIPYDELRLNCVYC KGQLTETEVLDFAFTDL TIVYRDDTPHGVCTKCL RFYSKVSEFRWYRYSV YGTTLEKLTNKGICDLLI RCITCQRPLCPEEKQR HLDKKKRFHNIGGRWT GRCIACWRRPRTETQV	27.5%
HPV31 E7	ATGCGTGGAGAAACACCTACGTTGCAAGACTATGTGTT AGATTTGCAACCTGAGGCAACTGACCTCCACTGTTATG AGCAATTACCCGACAGCTCAGATGAGGAGGATGTCATA GACAGTCCAGCTGGACAAGCAGAACCGGACACATCCA ATTACAATATCGTTACCTTTTGTTGTCAGTGTAAGTCTA CACTTCGTTTGTGTGTACAGAGCACACAAGTAGATATT CGCATATTGCAAGAGCTGTTAATGGGCTCATTTGGAAT CGTGTGCCCCAACTGTTCTACTAGACTGTAA	ATGAGAGGAGAAACCCCAACACTGCAGGACTATGTGCT GGACCTGCAGCCCGAAGCCACTGATCTGCATTGCTAC GAACAGCTGCCAGACAGCTCCGATGAGGAAGACGTGA TCGATTCCCCAGCAGGACAGGCTGAGCCTGACACCAG TAACTACAATATTGTCACATTCTGCTGTCAGTGCAAGTC TACTCTGCGGCTGTGTGTGCAGAGTACCCAGGTCGATA TCAGAATTCTGCAGGAACTGCTGATGGGCTCATTTGGG ATCGTGTGCCCCAACTGTAGCACAAGGCTG	MRGETPTLQDYVLDLQ PEATDLHCYEQLPDSS DEEDVIDSPAGQAEPDT SNYNIVTFCCQCKSTLR LCVQSTQVDIRILQELL MGSFGIVCPNCSTRL	25.9%
HPV31 L1	ATGTCTCTGTGGCGGCCTAGCGAGGCTACTGTCTACTT ACCACCTGTCCCAGTGTCTAAAGTTGTAAGCACGGATG AATATGTAACACGAACCAACATATATTATCACGCAGGCA GTGCTAGGCTGCTTACAGTAGGCCATCCATATTATTCC ATACCTAAATCTGACAATCCTAAAAAAATAGTTGTACCA AAGGTGTCAGGATTACAATATAGGGTATTTAGGGTTCG TTTACCAGATCCAAACAAATTTGGATTTCCTGATACATC TTTTTATAATCCTGAAACTCAACGCTTAGTTTGGGCCTG TGTTGGTTTAGAGGTAGGTCGCGGGCAGCCATTAGGT GTAGGTATTAGTGGTCATCCATTATTAAATAAATTTGAT GACACTGAAAACTCTAATAGATATGCCGGTGGTCCTGG CACTGATAATAGGGAATGTATATCAATGGATTATAAACA AACACAACTGTGTTTACTTGGTTGCAAACCACCTATTGG AGAGCATTGGGGTAAAGGTAGTCCTTGTAGTAACAATG CTATTACCCCTGGTGATTGTCCTCCATTAGAATTAAAAA ATTCAGTTATACAAGATGGGGATATGGTTGATACAGGC TTTGGAGCTATGGATTTTACTGCTTTACAAGACACTAAA AGTAATGTTCCTTTGGACATTTGTAATTCTATTTGTAAAT ATCCAGATTATCTTAAAATGGTTGCTGAGCCATATGGC GATACATTATTTTTTTATTTACGTAGGGAACAAATGTTTG TAAGGCATTTTTTTTAATAGATCAGGCACGGTTGGTGAAT CGGTCCCTACTGACTTATATATTAAAGGCTCCGGTTCA ACAGCTACTTTAGCTAACAGTACATACTTTCCTACACCT AGCGGCTCCATGGTTACTTCAGATGCACAAATTTTTAAT AAACCATATTGGATGCAACGTGCTCAGGGACACAATAA TGGTATTTGTTGGGGCAATCAGTTATTTGTTACTGTGGT AGATACCACACGTAGTACCAATATGTCTGTTTGTGCTG CAATTGCAAACAGTGATACTACATTTAAAAGTAGTAATT TTAAAGAGTATTTAAGACATGGTGAGGAATTTGATTTAC AATTTATATTTCAGTTATGCAAAATAACATTATCTGCAGA CATAATGACATATATTCACAGTATGAATCCTGCTATTTT GGAAGATTGGAATTTTGGATTGACCACACCTCCCTCAG GTTCTTTGGAGGATACCTATAGGTTTGTCACCTCACAG GCCATTACATGTCAAAAAACTGCCCCCCAAAAGCCCAA GGAAGATCCATTTAAAGATTATGTATTTTGGGAGGTTAA TTTAAAAGAAAAGTTTTCTGCAGATTTAGATCAGTTTCC ACTGGGTCGCAAATTTTTATTACAGGCAGGATATAGGG CACGTCCTAAATTTAAAGCAGGTAAACGTAGTGCACCC TCAGCATCTACCACTACACCAGCAAAACGTAAAAAAAC TAAAAAGTAA	ATGTCTCTGTGGCGACCTAGTGAAGCAACTGTCTACCT GCCTCCTGTCCCTGTGTCCAAAGTGGTGTCTACCGACG AGTATGTGACTCGGACTAATATCTACTATCACGCAGGA TCCGCAAGACTGCTGACCGTGGGGCATCCCTACTATTC TATCCCTAAGAGTGACAACCCAAAGAAAATTGTGGTCC CTAAAGTGTCCGGACTGCAGTACAGGGTGTTCAGGGT CCGCCTGCCAGACCCTAATAAGTTCGGCTTTCCCGATA CATCTTTTTATAACCCTGAGACTCAGAGGCTGGTGTGG GCATGCGTCGGACTGGAAGTGGGACGAGGACAGCCAC TGGGAGTGGGAATTTCAGGACACCCTCTGCTGAATAAG TTCGACGATACCGAGAACAGCAATCGATACGCTGGAG GACCAGGAACAGACAACCGAGAATGTATCTCTATGGAT TATAAACAGACCCAGCTGTGCCTGCTGGGCTGTAAGCC CCCTATCGGCGAGCATTGGGGCAAAGGGAGCCCTTGC TCCAACAATGCCATTACACCAGGCGACTGTCCACCCCT GGAACTGAAGAATTCCGTCATCCAGGACGGGGATATG GTGGATACTGGATTCGGCGCTATGGACTTTACCGCACT GCAGGATACAAAAAGTAACGTCCCCCTGGACATCTGCA ATTCAATCTGTAAGTACCCAGATTATCTGAAGATGGTG GCCGAGCCCTACGGGGACACACTGTTCTTTTATCTGCG GAGAGAACAGATGTTCGTGAGACACTTCTTTAATAGGT CCGGAACCGTCGGAGAGTCTGTGCCAACAGACCTGTA CATTAAGGGGTCTGGAAGTACCGCTACACTGGCAAACT CTACTTATTTCCCAACCCCCTCAGGCAGCATGGTGACC AGTGATGCACAGATTTTTAATAAGCCTTACTGGATGCA GCGGGCCCAGGGACATAACAATGGCATCTGCTGGGGG AACCAGCTGTTCGTCACAGTGGTCGACACCACAAGATC AACTAACATGAGCGTGTGTGCCGCTATCGCCAATAGCG ATACTACCTTCAAGAGCTCCAACTTTAAAGAGTACCTGA GACACGGCGAGGAATTTGACCTGCAGTTCATCTTTCAG CTGTGCAAGATTACTCTGAGCGCTGATATCATGACCTA TATTCATTCCATGAACCCAGCAATTCTGGAGGACTGGA ATTTCGGGCTGACAACTCCTCCATCCGGATCTCTGGAA GATACTTACAGGTTTGTGACCAGCCAGGCCATCACATG TCAGAAGACTGCTCCTCAGAAGCCAAAAGAAGACCCCT TCAAAGATTATGTCTTTTGGGAGGTGAACCTGAAGGAA AAATTCAGTGCCGACCTGGATCAGTTCCCTCTGGGACG CAAGTTTCTGCTGCAGGCAGGATACCGAGCACGACCA AAGTTTAAAGCCGGCAAGCGATCCGCCCCAAGTGCTTC AACCACAACTCCCGCTAAACGCAAGAAAACCAAGAAA	MSLWRPSEATVYLPPV PVSKVVSTDEYVTRTNI YYHAGSARLLTVGHPY YSIPKSDNPKKIVVPKVS GLQYRVFRVRLPDPNK FGFPDTSFYNPETQRLV WACVGLEVGRGQPLG VGISGHPLLNKFDDTEN SNRYAGGPGTDNRECI SMDYKQTQLCLLGCKP PIGEHWGKGSPCSNNA ITPGDCPPLELKNSVIQ DGDMVDTGFGAMDFTA LQDTKSNVPLDICNSIC KYPDYLKMVAEPYGDT LFFYLRREQMFVRHFF NRSGTVGESVPTDLYIK GSGSTATLANSTYFPTP SGSMVTSDAQIFNKPY WMQRAQGHNNGICWG NQLFVTVVDTTRSTNM SVCAAIANSDTTFKSSN FKEYLRHGEEFDLQFIF QLCKITLSADIMTYIHSM NPAILEDWNFGLTTPPS GSLEDTYRFVTSQAITC QKTAPQKPKEDPFKDY VFWEVNLKEKFSADLD QFPLGRKFLLQAGYRA RPKFKAGKRSAPSAST TTPAKRKKTKK	27.0%
HPV31 L2	ATGCGGTCCAAACGCTCTACAAAACGCACTAAACGTGC GTCTGCTACACAATTATATCAAACATGTAAAGCAGCAG GTACTTGTCCATCAGACGTTATACCTAAAATAGAACATA CTACCATTGCAGACCAAATATTAAGGTATGGTAGTATG GGTGTTTTTTTTGGTGGGTTGGGTATTGGGTCCGGCTC TGGTACTGGGGGTCGCACTGGATATGTCCCTCTTAGTA CACGTCCTTCTACAGTATCTGAGGCAAGTATACCTATTA GACCACCAGTTAGCATTGACCCTGTAGGTCCCTTGGAC CCCTCTATAGTAAGTCTTGTTGAAGAATCTGGAATTGTT GATGTTGGTGCCCCTGCTCCTATACCACACCCTCCTAC AACATCTGGGTTTGACATTGCTACAACTGCAGACACAA CACCTGCAATTTTAGATGTAACAAGTGTTAGCACACATG AAAATCCTACTTTTACTGATCCATCTGTATTCAGCCTC CTACACCTGCAGAAACATCAGGTCATTTACTACTTTCAT CATCATCTATTAGCACACATAATTATGAGGAAATACCTA TGGATACATTTATTGTTTCTACTAATAATGAAAACATAAC AAGTAGCACACCCATTCCAGGGGTGCGCCGTCCTGCA CGTTTAGGGTTATATAGTAAGGCTACACAACAAGTAAAA GTTATTGATCCAACGTTTCTTAGTGCTCCAAAACAGCTA ATTACATATGAAAACCCTGCCTATGAAACTGTAAATGCT GAAGAATCTTTATACTTTTCCAATACATCGCATAATATA GCCCCTGATCCCGACTTTCTAGATATTATAGCATTACAT AGGCCTGCCCTTACCTCACGTAGGAACACTGTTAGATA TAGTAGACTAGGTAATAAACAAACTTTGCGCACTCGTA GTGGTGCTACTATTGGTGCAAGGGTGCATTATTATTAT GATATTAGTAGTATTAATCCTGCAGGTGAAAGTATTGAA ATGCAACCTTTAGGGGCGTCTGCAACTACTACTTCTAC TTTAAATGATGGCTTATATGACATTTATGCAGACACTGA TTTTACTGTGGATACACCTGCCACACATAATGTTTCCCC TTCTACTGCTGTACAGTCCACATCTGCTGTGTCTGCCT ATGTACCTACAAATACCACTGTGCCACTAAGTACAGGT TTTGACATTCCCATATTTTCTGGGCCTGATGTACCTATA GAGCATGCACCTACACAGGTTTTCCCATTTCCTTTGGC CCCTACAACGCCACAAGTGTCTATTTTTGTTGATGGGG GTGATTTTTATTTGCACCCTAGTTATTATATGTTAAAACG TCGACGTAAACGTGTATCATATTTTTTTACAGATGTCTC TGTGGCGGCCTAG	ATGAGAAGCAAAAGAAGCACTAAAAGAACTAAAAGAGC CTCCGCAACCCAGCTGTATCAGACCTGTAAAGCCGCA GGAACTTGTCCATCTGACGTGATCCCTAAGATTGAGCA CACCACAATCGCCGATCAGATTCTGCGCTATGGGAGCA TGGGAGTGTTCTTTGGCGGGCTGGGCATTGGGAGTGG ATCAGGCACAGGAGGCAGAACTGGCTACGTGCCTCTG AGTACCAGGCCATCCACAGTCTCTGAAGCCAGTATCCC AATTAGACCCCCTGTGAGCATCGACCCCGTCGGACCT CTGGATCCATCAATCGTGAGCCTGGTCGAGGAAAGCG GAATTGTGGACGTCGGAGCACCAGCACCTATCCCACA CCCACCCACTACCTCCGGCTTCGACATTGCCACAACTG CTGATACCACACCCGCTATCCTGGACGTGACTAGCGTC TCCACCCATGAGAACCCCACCTTTACAGATCCTTCCGT GCTGCAGCCTCCAACACCCGCAGAAACTTCTGGGCAC CTGCTGCTGAGCTCCTCTAGTATCAGTACCCATAACTA TGAGGAAATCCCTATGGACACCTTCATTGTGTCTACAA ACAATGAGAATATCACTTCAAGCACCCCCATTCCTGGG GTCCGGAGACCAGCTAGGCTGGGACTGTACTCCAAGG CAACACAGCAGGTGAAAGTCATTGATCCAACCTTTCTG TCTGCCCCCAAGCAGCTGATCACCTATGAGAACCCCG CATACGAAACAGTGAATGCCGAGGAAAGCCTGTATTTC TCCAACACCTCTCACAATATCGCCCCGAGACCCCGATTT TCTGGATATCATTGCCCTGCATCGCCCTGCTCTGACTT CTAGGCGCAACACCGTGCGATACAGTCGGCTGGGCAA CAAGCAGACACTGAGGACTCGCAGCGGCGCTACAATT GGGGCACGAGTGCACTACTATTACGACATCTCCTCTAT TAACCCAGCCGGAGAGTCCATCGAAATGCAGCCCCTG GGCGCTAGTGCAACTACCACATCAACCCTGAATGACG GCCTGTATGATATCTACGCTGACACCGATTTCACAGTG GATACTCCCGCCACCCATAACGTGTCTCCTAGTACAGC TGTCCAGTCAACTAGCGCAGTGAGCGCCTACGTCCCTA CTAATACTACCGTGCCACTGTCAACCGGCTTCGACATC CCTATTTTTAGCGGGCCCGATGTGCCTATTGAGCACGC ACCAACACAGGTCTTCCCTTTTCCACTGGCCCCAACAA CTCCCCAGGTGTCAATCTTCGTCGACGGGGGAGATTTT TATCTGCATCCCAGCTATTACATGCTGAAGCGACGGAG AAAACGGGTGAGCTACTTCTTTACCGACGTGTCCGTCG CCGCT	MRSKRSTKRTKRASAT QLYQTCKAAGTCPSDVI PKIEHTTIADQILRYGSM GVFFGGLGIGSGSGTG GRTGYVPLSTRPSTVS EASIPIRPPVSIDPVGPL DPSIVSLVEESGIVDVG APAPIPHPPTTSGFDIAT TADTTPAILDVTSVSTH ENPTFTDPSVLQPPTPA ETSGHLLLSSSSISTHN YEEIPMDTFIVSTNNENI TSSTPIPGVRRPARLGL YSKATQQVKVIDPTFLS APKQLITYENPAYETVN AEESLYFSNTSHNIAPD PDFLDIIALHRPALTSRR NTVRYSRLGNKQTLRT RSGATIGARVHYYYDIS SINPAGESIEMQPLGAS ATTTSTLNDGLYDIYAD TDFTVDTPATHNVSPST AVQSTSAVSAYVPTNTT VPLSTGFDIPIFSGPDVP IEHAPTQVFPFPLAPTT PQVSIFVDGGDFYLHPS YYMLKRRRKRVSYFFT DVSVAA	28.3%

HPV33 E1	ATGGCCGATCCTGAAGGTACAAATGGGGCTGGGATGG GGTGTACTGGTTGGTTTGAGGTAGAAGCAGTCATAGAG AGAAGAACAGGAGATAATATTTCAGAAGATGAGGATGA AACAGCAGATGACAGTGGCACGGATTTACTAGAGTTTA TAGATGATTCTATGGAAAATAGTATACAGGCAGACACA GAGGCAGCCCGGGCATTGTTTAATATACAGGAAGGGG AGGATGATTTAAATGCTGTGTGTGCACTAAAACGAAAG TTTGCCGCATGTTCACAAAGTGCTGCGGAGGACGTTGT TGATCGTGCTGCAAACCCGTGTAGAACGTCTATTAATA AAAATAAAGAATGCACATACAGAAAACGAAAAATAGATG AGCTAGAAGACAGCGGATATGGCAATACTGAAGTGGAA ACTCAGCAGATGGTACAACAGGTAGAAAGTCAAAATGG CGACACAAACTTAAATGACTTAGAATCTAGTGGGGTGG GGGATGATTCAGAAGTAAGCTGTGAGACAAATGTAGAT AGCTGTGAAAATGTTACGTTGCAGGAAATTAGTAATGTT CTACATAGTAGTAATACAAAAGCAAATATATTATATAAAT TTAAAGAGGCCTATGGAATAAGTTTTATGGAATTAGTAA GACCATTTAAAAGTGATAAAACAAGCTGTACAGATTGGT GTATAACAGGATATGGAATTAGTCCATCAGTAGCAGAA AGTTTAAAAGTATTAATTAAACAGCATAGTTTGTATACTC ATTTACAATGTTTAACTTGCGATAGAGGAATAATAATAT TATTGTTAATTAGATTTAGGTGTAGCAAAAACAGGTTAA CAGTAGCAAAACTAATGAGTAATTTATTATCAATACCTG AAACATGTATGGTTATAGAGCCACCAAAATTACGGAGC CAAACATGTGCATTGTATTGGTTTAGAACAGCAATGTCA AACATTAGTGATGTACAAGGTACAACACCTGAATGGAT AGATAGACTAACTGTTTTACAACATAGCTTTAATGATAA TATATTTGATTTAAGTGAAATGGTACAGTGGGCATATGA TAACGAGTTAACGGACGATAGTGACATTGCATATTATTA TGCACAACTTGCAGATTCAAATAGTAATGCTGCTGCATT TTTAAAAAGTAACTCACAAGCAAAAATAGTAAAGGACTG TGGAATAATGTGTAGACATTATAAAAAAGCAGAAAAACG TAAAATGTCAATAGGACAATGGATACAAAGTAGATGTG AAAAAACAAATGATGGAGGAAATTGGAGACCAATAGTA CAGTTGTTAAGATATCAAAACATTGAATTTACAGCATTT TTAGGTGCATTTAAAAAGTTTTTAAAAGGTATACCAAAA AAAAGCTGTATGCTAATTTGTGGACCAGCAAATACAGG AAAGTCATATTTTGGAATGAGTTTAATACAGTTTTTAAAA GGGTGTGTTATATCATGTGTAAATTCTAAAAGTCACTTT TGGTTGCAGCCATTATCAGATGCAAAAATAGGAATGAT AGATGATGTAACGCCAATAAGTTGGAGACATATAGATG ATTACATGAGAAATGCGTTAGATGGAAATGAAATTTCAA TAGATGTGAAACATAGGGCATTAGTGCAATTAAAATGTC CACCACTGCTTCTTACCTCAAATACAAATGCAGGCACA GACTCTAGATGGCCATATTTACATAGTAGATTAACAGTA TTTGAATTTAAAAATCCATTCCCATTTGATGAAAATGGT AACCCAGTGTATGCAATAAATGATGAAAATTGGAAATCC TTTTTCTCAAGGACGTGGTGCAAATTAGATTTAATAGAG GAAGAGGACAAGGAAAACCATGGAGGAAATATCAGCA CGTTTAAATGCAGTGCAGGAGAAAATACTAGATCTTTAC GAAGCTGA	ATGGCTGACCCTGAAGGCACTAACGGGGCTGGGATGG GCTGTACTGGCTGGTTTGAGGTGGAGGCTGTGATTGA AAGACGGACTGGGGACAACATCAGCGAAGACGAGGAT GAAACAGCCGACGATTCCGGGACTGATCTGCTGGAGT TCATTGACGATTCTATGGAAAATAGTATCCAGGCAGAC ACCGAGGCAGCTCGAGCACTGTTCAACATCCAGGAGG GAGAAGACGATCTGAATGCAGTGTGCGCCCTGAAGAG AAAATTTGCAGCCTGTTCACAGAGCGCTGCAGAGGAC GTGGTCGATAGAGCCGCTAACCCTTGCAGGACATCCAT TAACAAGAACAAGGAATGTACTTATCGGAAGAGAAAAA TCGACGAGCTGGAAGATAGTGGGTACGGAAACACCGA GGTGGAAACACAGCAGATGGTGCAGCAGGTCGAGAGC CAGAATGGCGACACAAACCTGAATGATCTGGAAAGCTC CGGCGTGGGGGACGATTCAGAGGTCAGCTGCGAAACC AACGTGGATTCTTGTGAGAATGTCACACTGCAGGAAAT CAGTAATGTGCTGCACTCTAGTAACACCAAGGCCAACA TCCTGTACAAGTTTAAAGAGGCTTACCGGCATCAGCTTC ATGGAACTGGTGCGGCCTTTTAAGTCAGACAAAACTAG CTGCACCGATTGGTGTATTACAGGATATGGCATCTCCC CATCTGTGGCCGAGTCCCTGAAGGTCCTGATCAAGCA GCACTCTCTGTACACCCATCTGCAGTGCCTGACATGTG ACCGCGGGATCATTATCCTGCTGCTGATCAGGTTCCGC TGCAGCAAGAACCGACTGACCGTGGCCAAACTGATGT CCAATCTGCTGTCTATTCCAGAGACATGCATGGTCATC GAACCCCCTAAGCTGCGATCCCAGACTTGTGCTCTGTA TTGGTTTCGGACCGCAATGTCCAACATTTCTGACGTGC AGGGCACCACACCCGAGTGGATCGATAGGCTGACAGT CCTGCAGCACAGTTTCAACGACAATATTTTTGATCTGTC AGAGATGGTGCAGTGGGCATACGACAACGAACTGACT GACGATTCTGATATCGCCTACTATTACGCTCAGCTGGC AGATAGTAACTCAAATGCAGCCGCTTTCCTGAAAAGCA ATTCCCAGGCCAAGATTGTGAAAGACTGCGGCATCATG TGTAGGCATTATAAGAAAGCTGAGAAGCGCAAAATGTC TATTGGGCAGTGGATCCAGAGTCGCTGCGAAAAGACTA ACGACGGCGGGAATTGGCGCCCCATTGTGCAGCTGCT GCGATACCAGAACATCGAGTTCACCGCCTTTCTGGGG GCTTTCAAGAAATTTCTGAAAGGAATTCCCAAGAAAAG CTGCATGCTGATCTGTGGGCCTGCTAACACCGGAAAG AGTTACTTCGGCATGTCACTGATTCAGTTTCTGAAAGG ATGCGTGATCTCATGTGTCAATTCTAAGAGTCACTTTTG GCTGCAGCCACTGTCCGATGCCAAGATTGGCATGATC GACGATGTGACTCCCATTTCTTGGACCTATATCGACGA TTACATGAGAAACGCTCTGGACGGGAATGAGATTTCCA TCGATGTGAAGCACAGGGCACTGGTCCAGCTGAAATG TCCACCCCTGCTGCTGACATCTAACACTAATGCAGGAA CAGACTCACGGTGGCCCTATCTGCATAGCAGACTGACT GTGTTCGAGTTTAAGAACCCTTTCCCATTTGACGAAAAC GGCAATCCTGTCTACGCCATCAACGATGAGAATTGGAA GAGTTTCTTTTCAAGAACCTGGTGCAAACTGGACCTGA TTGAGGAAGAGGATAAGGAGAACCATGGAGGCAATAT CAGCACTTTCAAGTGTTCCGCCGGCGAAAATACCCGAA GCCTGCGGTCC	MADPEGTNGAGMGCT GWFEVEAVIERRTGDNI SEDEDETADDSGTDLL EFIDDSMENSIQADTEA ARALFNIQEGEDDLNAV CALKRKFAACSQSAAE DVVDRAANPCRTSINKN KECTYRKRKIDELEDSG YGNTEVETQQMVQQVE SQNGDTNLNDLESSGV GDDSEVSCETNVDSCE NVTLQEISNVLHSSNTK ANILYKFKEAYGISFMEL VRPFKSDKTSCTDWCIT GYGISPSVAESLKVLIKQ HSLYTHLQCLTCDRGIII LLLIRFRCSKNRLTVAKL MSNLLSIPETCMVIEPP KLRSQTCALYWFRTAM SNISDVQGTTPEWIDRL TVLQHSFNDNIFDLSEM VQWAYDNELTDDSDIA YYYAQLADSNSNAAAFL KSNSQAKIVKDCGIMCR HYKKAEKRKMSIGQWI QSRCEKTNDGGNWRPI VQLLRYQNIEFTAFLGA FKKFLKGIPKKSCMLIC GPANTGKSYFGMSLIQF LKGCVISCVNSKSHFWL QPLSDAKIGMIDDVTPIS WTYIDDYMRNALDGNEI SIDVKHRALVQLKCPPL LLTSNTNAGTDSRWPY LHSRLTVFEFKNPFPFD ENGNPVYAINDENWKS FFSRTWCKLDLIEEEDK ENHGGNISTFKCSAGE NTRSLRS	27.0%
HPV33 E2	ATGGAGGAAATATCAGCACGTTTAAATGCAGTGCAGGA GAAAATACTAGATCTTTACGAAGCTGATAAAACTGATTT ACCATCACAAATTGAACATTGGAAACTGATACGCATGG AGTGTGCTTTATTGTATACAGCCAAACAAATGGGATTTT CACATTTATGCCACCAGGTGGTGCCTTCTTTGTTAGCA TCAAAGACCAAAGCATTTCAAGTAATTGAACTACAAATG GCATTAGAGACATTAAGTAAATCACAGTATAGTACAAGC CAATGGACATTGCAACAAACAAGCTTAGAGGTGTGGCT TTGTGAACCACCAAAATGTTTTAAAAAACAAGGAGAAAC AGTAACTGTGCAATATGACAATGACAAAAAAAATACAAT GGATTATACAAACTGGGGTGAAATATATATTATAGAGGA AGATACATGTACTATGGTTACAGGGAAAGTAGATTATAT AGGTATGTATTATATACATAACTGTGAAAAGGTATATTT TAAATATTTTAAAGAGGATGCTGCAAAGTATTCTAAAAC ACAAATGTGGGAAGTACATGTGGGTGGTCAGGTAATTG TTTGTCCTACGTCTATATCTAGCAACCAAATATCCACTA CTGAAACTGCTGACATACAGACAGACAACGATAACCGA CCACCACAAGCAGCGGCCAAACGACGACGACCTGCAG ACACCACAGACACCGCCCAGCCCCTTACAAAGCTGTTC TGTGCAGACCCCGCCTTGGACAATAGAACAGCACGTA CTGCAACTAACTGCACAAACAAGCAGCGGACTGTGTGT AGTTCTAACGTTGCACCTATAGTGCATTTAAAAGGTGAA TCAAATAGTTTAAAATGTTTAAGATACAGATTAAAACCTT ATAAAGAGTTGTATAGTTCTATGTCATCCACCTGGCATT GGACCAGTGACAACAAAAATAGTAAAAATGGAATTGTA ACTGTAACATTTGTAACTGAACAGCAACAACAAATGTTT TTAGGTACCGTAAAAATACCACCTACTGTGCAAATAAGT ACTGGATTTATGACATTATAA	ATGGAGGAGATTAGCGCAAGACTGAACGCCGTGCAGG AAAAGATTCTGGACCTGTATGAAGCCGACAAGACCGAC CTGCCAAGCCAGATCGAGCACTGGAAGCTGATTCGAAT GGAATGCGCCCTGCTGTACACCGCTAAACAGATGGGG TTCAGCCACCTGTGTCATCAGGTGGTCCCATCACTGCT GGCAAGCAAGACTAAAGCCTTTCAGGTCATCGAGCTGC AGATGGCCCTGGAAACCCTGAGTAAGTCACAGTACAG CACCTCCCAGTGGACACTGCAGCAGACTTCCCTGGAA GTGTGGCTGTGCGAACCCCCTAAGTGTTTCAAGAAACA GGGCGAGACAGTGACTGTCCAGTATGACAACGATAAG AAAAATACCATGGACTACACAAACTGGGGGGAAATCTA TATCATTGAGGAAGACACCTGCACAATGGTGACCGGAA AGGTCGATTACATCGGCATGTACTATATTCACAACTGT GAGAAGGTGTACTTCAAGTACTTCAAGGAAGACGCCGC TAAGTATTCTAAAACACAGATGTGGGAGGTGCATGTCG GCGGGCAGGTCATCGTCTGCCCCACTAGTATCAGCTC CAACCAGATTAGCACCACAGAAACCGCTGATATTCAGA CAGACAACGATAATAGGCCACCACAGGCAGCAGCTAA GCGAAGAAGGCCAGCTGACACTACCGATACTGCACAG CCTCTGACCAAACTGTTCTGCGCAGACCCAGCCCTGG ATAATCGGACAGCTAGAACTGCAACCAACTGCACAAAT AAGCAGCGCACTGTGTGTTCTAGTAACGTGGCCCCCAT CGTCCACCTGAAGGGAGAGTCTAATAGTCTGAAATGTC TGCGCTATCGACTGAAGCCTTACAAAGAACTGTATTCA AGCATGTCCTCTACTTGGCATTGGACCTCCGATAACAA GAATTCTAAAAACGGCATTGTGACAGTCACTTTCGTGA CCGAGCAGCAGCAGCAGATGTTTCTGGGGACCGTGAA GATCCCTCCAACAGTCCAGATTAGCACAGGCTTCATGA CTCTG	MEEISARLNAVQEKILDL YEADKTDLPSQIEHWKL IRMECALLYTAKQMGFS HLCHQVVPSLLASKTKA FQVIELQMALETLSKSQ YSTSQWTLQQTSLEVW LCEPPKCFKKQGETVT VQYDNDKKNTMDYTN WGEIYIIEEDTCTMVTG KVDYIGMYYIHNCEKVY FKYFKEDAAKYSKTQM WEVHVGGQVIVCPTSIS SNQISTTETADIQTDND NRPPQAAAKRRRPADT TDTAQPLTKLFCADPAL DNRTARTATNCTNKQR TVCSSNVAPIVHLKGES NSLKCLRYRLKPYKELY SSMSSTWHWTSDNKN SKNGIVTVTFVTEQQQQ MFLGTVKIPPTVQISTGF MTL	26.8%
HPV33 E4	TTGTTTGTCCTACGTCTATATCTAGCAACCAAATATCCA CTACTGAAACTGCTGACATACAGACAGACAACGATAAC CGACCACCACAAGCAGCGGCCAAACGACGACGACCTG CAGACACCACAGACACCGCCCAGCCCCTTACAAAGCT GTTCTGTGCAGACCCCGCCTTGGACAATAGAACAGCAC GTACTGCAACTAACTGCACAAACAAGCAGCGGACTGTG TGTAGTTCTAACGTTGCACCTATAG	ATGTTCGTGCTGAGGCTGTACCTGGCAACTAAGTATCC CCTGCTGAAACTGCTGACCTACCGACAGACCACCATCA CTGACCATCACAAGCAGCGGCCCAACGACGATGACCT GCAGACCCCTCAGACACCCCCTTCCCCACTGCAGTCTT GCAGTGTGCAGACACCACCCTGGACTATCGAGCAGCA CGTCCTGCAGCTGACTGCCCAGACCAGCTCCGGCCTG TGTGTGGTCCTGACCCTGCATCTG	LFVLRLYLATKYPLLKLL TYRQTTITDHHKQRPND DDLQTPQTPPSPLQSC SVQTPPWTIEQHVLQLT AQTSSGLCVVLTLHL	27.2%
HPV33 E5	ATGATATTTGTTTTTGTATTATGTTTTATATTGTTTTTATG CTTATCCTTATTATTACGTCCTTTAATACTTTCCATTTCT ACCTATGCTTGGTTGCTGGTGTTGGTATTGCTGCTTTG GGTGTTTGTGGGATCTCCTTTAAAAATTTTTTTTTGCTAT TTGTTGTTTTTATATTTACCAATGATGTGTATTAATTTTC ATGCACAGCATATGACACAACAAGAGTAA	ATGATTTTTGTGTTTGTCCTGTGTTTTATCCTGTTTCTGT GCCTGAGCCTGCTGCTGAGACCACTGATTCTGTCCATT TCTACTTATGCCTGGCTGCTGGTGCTGGTCCTGCTGCT GTGGGTGTTCGTCGGCAGCCCCCTGAAGATCTTCTTTT GCTACCTGCTGTTCCTGTATCTGCCTATGATGTGTATTA ACTTTCACGCTCAGCATATGACCCAGCAGGAG	MIFVFVLCFILFLCLSLLL RPLILSISTYAWLLVLVL LLWVFVGSPLKIFFCYLL FLYLPMMCINFHAQHM TQQE	26.1%

HPV33 E6	ATGTTTCAAGACACTGAGGAAAAACCACGAACATTGCA TGATTTGTGCCAAGCATTGGAGACAACTATACACAACA TTGAACTACAGTGCGTGGAATGCAAAAAACCTTTGCAA CGATCTGAGGTATATGATTTTGCATTTGCAGATTTAACA GTTGTATATAGAGAGGGAAATCCATTTGGAATATGTAAA CTGTGTTTGCGGTTCTTATCTAAAATTAGTGAATATAGA CATTATAATTATTCTGTATATGGAAATACATTAGAACAAA CAGTTAAAAAACCTTTAAATGAAATATTAATTAGGTGTAT TATATGTCAAAGACCTTTGTGTCCTCAAGAAAAAAAACG ACATGTGGATTTAAACAAACGATTTCATAATATTTCGGG TCGTTGGGCAGGGCGCTGTGCGGCGTGTTGGAGGTCC CGACGTAGAGAAACTGCACTGTGA	ATGTTTCAGGACACCGAGGAGAAGCCAAGAACTCTGCA TGATCTGTGCCAGGCTCTGGAGACCACCATTCACAATA TCGAACTGCAGTGCGTGGAGTGTAAGAAACCACTGCA GCGCAGCGAAGTCTACGACTTCGCATTTGCCGATCTGA CTGTGGTCTATCGGGAGGGCAACCCCTTCGGGATCTG CAAGCTGTGTCTGCGATTTCTGAGCAAAATTTCCGAAT ACAGGCACTACAACTATTCTGTGTATGGGAATACCCTG GAGCAGACAGTCAAGAAACCCCTGAATGAAATCCTGAT TCGGTGCATCATTTGTCAGAGACCCCTGTGCCCTCAGG AGAAGAAAAGGCACGTGGACCTGAACAAGCGCTTCCA TAATATCTCTGGACGATGGGCTGGACGATGCGCAGCTT GTTGGAGAAGTCGGAGAAGGGAAACCGCCCTG	MFQDTEEKPRTLHDLC QALETTIHNIELQCVECK KPLQRSEVYDFAFADLT VVYREGNPFGICKLCLR FLSKISEYRHYNYSVYG NTLEQTVKKPLNEILIRC IICQRPLCPQEKKRHVD LNKRFHNISGRWAGRC AACWRSRRRETAL	26.9%
HPV33 E7	ATGAGAGGACACAAGCCAACGTTAAAGGAATATGTTTT AGATTTATATCCTGAACCAACTGACCTATACTGCTATGA GCAATTAAGTGACAGCTCAGATGAGGATGAAGGCTTGG ACCGGCCAGATGGACAAGCACAACCAGCCACAGCTGA TTACTACATTGTAACCTGTTGTCACACTTGTAACACCAC AGTTCGTTTATGTGTCAACAGTACAGCAAGTGACCTAC GAACCATACAGCAACTACTTATGGGCACAGTGAATATT GTGTGCCCTACCTGTGCACAACAATAA	ATGCGAGGCCACAAGCCCACCCTGAAAGAGTATGTCC TGGACCTGTATCCCGAGCCCACCGACCTGTATTGTTAT GAGCAGCTGTCAGACAGCTCCGACGAGGATGAAGGAC TGGACAGGCCAGATGGACAGGCTCAGCCTGCAACCGC TGATTACTATATCGTGACTTGCTGTCACACCTGCAACAC CACAGTGCGGCTGTGTGTCAATTCTACAGCAAGCGACC TGAGAACTATCCAGCAGCTGCTGATGGGCACCGTGAA CATTGTCTGCCCCACATGTGCCCAGCAG	MRGHKPTLKEYVLDLY PEPTDLYCYEQLSDSS DEDEGLDRPDGQAQPA TADYYIVTCCHTCNTTV RLCVNSTASDLRTIQQL LMGTVNIVCPTCAQQ	24.0%
HPV33 L1	ATGTCCGTGTGGCGGCCTAGTGAGGCCACAGTGTACC TGCCTCCTGTACCTGTATCTAAAGTTGTCAGCACTGAT GAATATGTGTCTCGCACAAGCATTTATTATTATGCTGGT AGTTCCAGACTTCTTGCTGTTGGCCATCCATATTTTTCT ATTAAAAATCCTACTAACGCTAAAAAATTATTGGTACCC AAAGTATCAGGCTTGCAATATAGGGTTTTTAGGGTCCG TTTACCAGATCCTAATAAATTTGGATTTCCTGACACCTC CTTTTATAACCCTGATACACAACGATTAGTATGGGCATG TGTAGGCCTTGAAATAGGTAGAGGGCAGCCATTAGGC GTTGGCATAAGTGGTCATCCTTTATTAAACAAATTTGAT GACACTGAAACCGGTAACAAGTATCCTGGACAACCGG GTGCTGATAATAGGGAATGTTTATCCATGGATTATAAAC AAACACAGTTATGTTTACTTGGATGTAAGCCTCCAACAG GGGAACATTGGGGTAAAGGTGTTGCTTGTACTAATGCA GCACCTGCCAATGATTGTCCACCTTTAGAACTTATAAAT ACTATTATTGAGGATGGTGATATGGTGGACACAGGATT TGGTTGCATGGATTTTAAAACATTGCAGGCTAATAAAAG TGATGTTCCTATTGATATTTGTGGCAGTACATGCAAATA TCCAGATTATTTAAAAATGACTAGTGAGCCTTATGGTGA TAGTTTATTTTTCTTTCTTCGACGTGAACAAATGTTTGTA AGACACTTTTTTAATAGGGCTGGTACATTAGGAGAGGC TGTTCCCGATGACCTGTACATTAAAGGTTCAGGAACTA CTGCCTCTATTCAAAGCAGTGCTTTTTTTCCCACTCCTA GTGGATCAATGGTTACTTCCGAATCTCAGTTATTTAATA AGCCATATTGGCTACAACGTGCACAAGGTCATAATAAT GGTATTTGTTGGGGCAATCAGGTATTTGTTACTGTGGT AGATACCACTCGCAGTACTAATATGACTTTATGCACACA AGTAACTAGTGACAGTACATATAAAAATGAAAATTTTAA AGAATATATAAGACATGTTGAAGAATATGATCTACAGTT TGTTTTTCAACTATGCAAAGTTACCTTAACTGCAGAAGT TATGACATATATTCATGCTATGAATCCAGATATTTTAGA AGATTGGCAATTTGGTTTAACACCTCCTCCATCTGCTAG TTTACAGGATACCTATAGGTTTGTTACCTCTCAGGCTAT TACGTGTCAAAAAACAGTACCTCCAAAGGAAAAGGAAG ACCCCTTAGGTAAATATACATTTTGGGAAGTGGATTTAA AGGAAAAATTTTCAGCAGATTTAGATCAGTTTCCTTTGG GACGCAAGTTTTTATTACAGGCAGGTCTTAAAGCAAAA CCTAAACTTAAACGTGCAGCCCCCACATCCACCCGCAC ATCGTCTGCAAAACGCAAAAAGGTTAAAAAATAA	ATGTCAGTGTGGAGACCCAGCGAGGCTACCGTGTATC TGCCCCCAGTCCCCGTGAGCAAAGTGGTGTCAACCGA TGAGTATGTGAGCCGCACCTCCATCTACTATTACGCTG GAAGCTCCCGACTGCTGGCAGTGGGACACCCCTATTTT AGCATTAAGAACCCTACAAATGCCAAGAAACTGCTGGT GCCTAAAGTCTCCGGGCTGCAGTATAGGGTGTTTAGG GTCCGCCTGCCCGACCCTAACAAGTTTGGATTCCCAGA CACATCTTTCTACAATCCCGATACTCAGCGACTGGTGT GGGCATGCGTCGGACTGGAGATCGGAAGAGGACAGC CACTGGGAGTGGGCATTAGTGGACACCCTCTGCTGAA CAAGTTCGACGATACAGAGACTGGCAACAAGTATCCTG GGCAGCCAGGAGCTGACAACCGCGAATGTCTGAGCAT GGATTACAAGCAGACCCAGCTGTGCCTGCTGGGCTGT AAGCCCCCTACAGGCGAGCATTGGGGGAAAGGAGTGG CCTGCACTAACGCCGCTCCAGCTAATGACTGTCCACCC CTGGAGCTGATCAACACCATCATTGAAGACGGCGATAT GGTCGACACTGGCTTTGGGTGCATGGATTTCAAGACCC TGCAGGCCAACAAGAGTGACGTGCCCATCGATATTTGC GGCTCAACCTGTAAGTATCCAGACTACCTGAAAATGAC TTCCGAGCCCTATGGGGATTCTCTGTTCTTTTTCCTGC GGAGAGAACAGATGTTTGTCCGACACTTTTTCAACCGA GCAGGAACCCTGGGAGAGGCTGTGCCCGACGATCTGT ACATCAAGGGATCAGGCACCACAGCAAGCATTCAGTCT AGTGCCTTTTTCCCAACCCCCTCCGGCTCTATGGTGAC AAGTGAATCACAGCTGTTTAATAAGCCTTACTGGCTGC AGCGAGCCCAGGGACATAACAATGGCATCTGCTGGGG GAACCAGGTGTTCGTCACTGTGGTCGACACTACCCGCT CTACTAATATGACCCTGTGTACACAGGTCACTAGCGAT TCCACATACAAGAACGAGAACTTCAAGGAATACATTCG GCACGTGGAGGAATACGACCTGCAGTTTGTGTTCCAG CTGTGCAAGGTCACCCTGACAGCAGAAGTGATGACCTA CATCCATGCCATGAATCCCGACATTCTGGAAGATTGGC AGTTTGGACTGACACCTCCACCCTCTGCTAGTCTGCAG GATACTTATAGATTCGTCACCAGCCAGGCAATCACCTG TCAGAAGACAGTGCCTCCAAAGGAGAAAGAAGACCCT CTGGGCAAATACACCTTTTGGGAGGTGGATCTGAAGGA AAAATTCAGCGCCGACCTGGATCAGTTTCCACTGGGCA GGAAGTTCCTGCTGCAGGCTGGGCTGAAGGCAAAACC TAAGCTGAAACGCGCAGCCCCAACTTCCACCAGAACAT CAAGCGCTAAAAGGAAGAAAGTGAAGAAA	MSVWRPSEATVYLPPV PVSKVVSTDEYVSRTSI YYYAGSSRLLAVGHPY FSIKNPTNAKKLLVPKV SGLQYRVFRVRLPDPN KFGFPDTSFYNPDTQR LVWACVGLEIGRGQPL GVGISGHPLLNKFDDTE TGNKYPGQPGADNREC LSMDYKQTQLCLLGCK PPTGEHWGKGVACTNA APANDCPPLELINTIIED GDMVDTGFGCMDFKTL QANKSDVPIDICGSTCK YPDYLKMTSEPYGDSL FFFLRREQMFVRHFFN RAGTLGEAVPDDLYIKG SGTTASIQSSAFFPTPS GSMVTSESQLFNKPYW LQRAQGHNNGICWGN QVFVTVVDTTRSTNMTL CTQVTSDSTYKNENFK EYIRHVEEYDLQFVFQL CKVTLTAEVMTYIHAMN PDILEDWQFGLTPPPSA SLQDTYRFVTSQAITCQ KTVPPKEKEDPLGKYTF WEVDLKEKFSADLDQF PLGRKFLLQAGLKAKPK LKRAAPTSTRTSSAKRK KVKK	26.4%
HPV33 L2	ATGAGACACAAACGATCTACAAGGCGCAAGCGTGCATC TGCAACACAACTATACCAAACATGCAAGGCCACAGGCA CCTGCCCACCCGATGTTATTCCTAAAGTGGAAGGAAGT ACCATAGCAGATCAAATTCTTAAATATGGCAGTTTAGGG GTTTTTTTTGGTGGTTTAGGTATTGGCACAGGCTCTGG TTCAGGTGGAAGGACTGGCTATGTACCTATTGGTACTG ACCCACCTACAGCTGCAATCCCCTTGCAGCCTATACGT CCTCCGGTTACTGTAGACACTGTTGGACCTTTAGACTC GTCTATAGTGTCATTAATAGAAGAAACAAGTTTTATAGA GGCAGGTGCACCAGCCCCATCTATTCCTACACCATCAG GTTTTGATGTTACTACATCTGCAGATACTACACCTGCAA TTATTAATGTTTCATCTGTTGGGGAGTCATCTATTCAAA CTATTTCTACACATTTAAATCCCACATTTACTGAACCAT CTGTACTACACCCTCCAGCGCCTGCAGAAGCCTCTGG ACATTTTATATTTTCTTCCCCTACTGTTAGCACACAAAG TTATGAAAACATACCAATGGATACCTTTGTTGTTTCCAC AGACAGTAGTAATGTAACATCAAGCACGCCCATTCCAG GGTCTCGCCCTGTGGCACGCCTTGGTTTATATAGTCGC AATACCCAACAGGTTAAGGTTGTTGACCCTGCTTTTTTA ACATCGCCTCATAAACTTATAACATATGATAATCCTGCA TTTGAAAGCTTTGACCCTGAAGACACATTACAATTTCAA CATAGTGATATATCACCTGCTCCTGATCCTGACTTTCTA GATATTATTGCATTACATAGGCCTGCTATTACATCTCGT AGACATACTGTGCGTTTTAGTAGAGTAGGTCAAAAAGC CACACTTAAAACTCGCAGTGGTAAACAAATTGGAGCTA GAATACATTATTATCAGGATTTAAGTCCTATTGTGCCTT TAGACCACACCGTGCCAAATGAACAATATGAATTACAG CCTTTACATGATACTTCTACATCGTCTTATAGTATTAATG ATGGTTTGTATGATGTTTATGCTGACGATGTGGATAATG TACACACCCCAATGCAACACTCATACAGTACGTTTGCA ACAACACGTACCAGCAATGTGTCTATACCTTTAAATACA GGATTTGATACTCCTGTTATGTCTGGCCCTGATATACCT TCCCCTTTATTTCCCACATCTAGCCCATTTGTTCCTATT TCGCCTTTTTTTCCTTTTGACACCATTGTTGTAGACGGT GCTGACTTTGTTTTACATCCTAGTTATTTTATTTTACGTC GCAGGCGTAAACGTTTTCCATATTTTTTTACAGATGTCC GTGTGGCGGCCTAG	ATGCGACACAAGAGAAGCACCAGAAGGAAGAGAGCAA GCGCCACCCAGCTGTATCAGACCTGTAAAGCCACCGG GACCTGCCCTCCAGACGTGATCCCCAAGGTCGAGGGC AGTACCATCGCCGATCAGATTCTGAAATACGGATCACT GGGCGTGTTCTTTGGAGGACTGGGAATCGGAACTGGA TCCGGATCTGGAGGACGAACCGGATATGTGCCAATTG GGACTGACCCACCTACCGCAGCTATCCCACTGCAGCC CATTCGCCCACCCGTGACAGTCGACACTGTGGGCCCC CTGGATAGCTCCATCGTCAGCCTGATTGAGGAAACATC CTTCATCGAGGCTGGAGCACCAGCACCTTCCATTCCAA CTCCCTCTGGGTTTGACGTGACCACATCCGCTGATACT ACCCCTGCAATCATTAACGTGTCTAGTGTCGGAGAATC AAGCATCCAGACCATTAGCACACATCTGAATCCTACCT TCACAGAGCCATCCGTGCTGCACCCTCCAGCTCCAGC AGAAGCCTCTGGCCATTTCATCTTTTCCTCTCCAACTGT GAGCACCCAGTCCTACGAGAACATTCCCATGGACACCT TTGTGGTCAGCACAGATAGTTCAAATGTGACAAGCTCC ACTCCTATCCCAGGATCCCGACCAGTCGCACGACTGG GACTGTACTCTAGAAACACTCAGCAGGTGAAGGTGGTC GACCCAGCTTTCCTGACCAGTCCCCATAAACTGATCAC ATATGATAATCCTGCATTCGAGTCTTTTGACCCAGAAGA TACACTGCAGTTCCAGCACTCTGACATCAGTCCCGCTC CTGACCCAGATTTTCTGGATATCATTGCCCTGCACAGG CCTGCTATTACTTCACGGAGACATACCGTGAGATTCAG CAGGGTCGGGCAGAAGGCAACCCTGAAAACACGGTCC GGGAAGCAGATCGGAGCCAGAATTCACTACTATCAGG ACCTGAGCCCTATCGTGCCACTGGATCATACCGTCCCC AACGAGCAGTACGAACTGCAGCCTCTGCACGACACTA GTACCTCTAGTTATTCAATTAACGACGGCCTGTACGAT GTGTATGCCGACGATGTGGATAATGTCCACACCCCCAT GCAGCATAGTTATTCAACATTCGCTACAACTCGGACTT CAAACGTGAGCATCCCTCTGAATACAGGGTTTGACACT CCCGTGATGTCTGGACCTGATATTCCCAGTCCTCTGTT CCCCACCTCAAGCCCCTTTGTGCCTATCTCCCCATTCT TTCCCTTCGACACAATTGTGGTCGACGGCGCCGATTTC GTGCTGCACCCTAGCTACTTTATCCTGAGGCGCCGAC GGAAAAGGTTTCCATATTTCTTTACCGATGTGCGCGTC GCAGCC	MRHKRSTRRKRASATQ LYQTCKATGTCPPDVIP KVEGSTIADQILKYGSL GVFFGGLGIGTGSGSG GRTGYVPIGTDPPTAAI PLQPIRPPVTVDTVGPL DSSIVSLIEETSFIEAGA PAPSIPTPSGFDVTTSA DTTPAIINVSSVGESSIQ TISTHLNPTFTEPSVLHP PAPAEASGHFIFSSPTV STQSYENIPMDTFVVST DSSNVTSSTPIPGSRPV ARLGLYSRNTQQVKVV DPAFLTSPHKLITYDNP AFESFDPEDTLQFQHS DISPAPDPDFLDIIALHR PAITSRRHTVRFSRVGQ KATLKTRSGKQIGARIH YYQDLSPIVPLDHTVPN EQYELQPLHDTSTSSYS INDGLYDVYADDVDNV HTPMQHSYSTFATTRT SNVSIPLNTGFDTPVMS GPDIPSPLFPTSSPFVPI SPFFPFDTIVVDGADFV LHPSYFILRRRRKRFPY FFTDVRVAA	28.4%

HPV35 E1	ATGGCTGATCCTGCAGGTACAGATGAAGGGGAGGGGA CGGGATGTAATGGATGGTTTTTTGTAGAAGCAGTAGTT AGTAGACGTACGGGGGATCCAGTGTCAGAGGACGAAA ATGAAGATGACTGTGACAGGGGGGAGGATATGGTGGA CTTTATAAATGATACAGATATATTAAACATACAGGCAGA AACAGAGACAGCACAAGCATTATTTCATGCACAGGAGG AGCAAACACACAAAGAGGCTGTACAGGTCCTAAAACGA AAGTATGCTAGTAGTCCACTTAGCAGCGTGAGCTTATG TGTTAATAATAACATAAGTCCACGTTTAAAAGCTATTTG CATTGAAAATAAAAATACAGCAGCAAAGCGACGATTATT TGAACTACCAGACAGCGGTTATGGCAATTCTGAAGTGG AAATACAGCAGATACAACAGGTAGAGGGGCATGATACA GTTGAACAATGTAGTATGGGCAGTGGGGATAGTATAAC CTCTAGTAGCGATGAAAGACATGATGAGACTCCAACGC GAGACATAATACAAATACTAAAATGTAGTAATGCAAACG CAGCTATGTTGGCTAAATTTAAAGAACTATTTGGTATTA GTTTTACAGAACTTATTAGACCATTTAAGAGTGATAAAT CCACATGTACAGATTGGTGTGTGGCCGCATTTGGAATA GCCCCAAGTGTGGCGGAAAGTTTAAAAACATTAATTAA ACCATATTGTTTATATATACATATACAATGTTTATCGTGT TCATGGGGTATGGTAATTCTAGCATTATTACGATTTAAA TGTGCAAAAAAACAGAACAACAATTGAAAAACTATTATCA AAATTGCTATGTATTTCAGCTGCAAGTATGCTAATACAA CCACCAAAATTACGTAGTACCCCAGCTGCGTTATATTG GTTTAAAACAGCAATGTCAAATATTAGTGAGGTTGATGG AGAAACACCAGAATGGATTCAAAGACAAACAGTATTAC AGCATAGTTTTAATGATGCAATATTTGACCTATCTGAAA TGGTACAATGGGCATATGACAATGATTTTATAGATGATA GTGATATAGCATATAAATATGCACAATTGGCAGAAACTA ATAGTAATGCATGTGCTTTTTTAAAAAGTAATTCGCAAG CTAAAATTGTAAAAGATTGTGCAACAATGTGTAGACATT ATAAACGAGCTGAAAAAAGAGAAATGACAATGTCACAG TGGATTAAAAGGCGATGTGAAAAGGTGGACGATGACG GTGACTGGAGGGACATAGTACGATTTTTAAGATATCAA CAAGTAGATTTTGTGGCATTTTTATCTGCACTAAAAAAT TTTTTACATGGTGTGCCTAAAAAAAATTGCATACTTATAT ATGGAGCACCAAACACAGGTAAATCATTATTTGGAATG AGTCTAATGCATTTCTTACAAGGAGCTATTATATCCTAT GTAAATTCTAAAAGCCATTTTTGGTTGCAGCCATTATAT GATGCCAAAATAGCTATGTTAGATGATGCTACATCGCC ATGTTGGGCATATATAGACCAATATTTAAGAAATGCACT AGATGGAAATCCTATTTCATTAGATGTAAAGCATAAAGC ATTAGTGCAATTAAAATGCCCACCTTTACTTATTACATC AAATATAAATGCAGGCAAAGATGACAGGTGGCCATACT TACATAGCAGGGTAGTGGTCTTTACATTTCACAATGAAT TCCCATTTGATAAAAATGGAAACCCAGTGTATGGGCTT AATGATAAAAACTGGAAATCCTTTTTCTCAAGGACGTGG TGCAGATTAAATTTGCACGAGGAAGAGGACAAAGAAAA TGATGGAGACGCTTTCCCAGCGTTTAAGTGTGTGTCAG GACAAAATACTAGAACATTACGAGACTGA	ATGGCTGACCCCGCAGGGACCGATGAGGGAGAAGGC ACAGGATGTAATGGATGGTTTTTCGTGGAGGCAGTCGT GAGCAGGAGAACAGGAAGCTCCGTGGAAGATGAGAAC GAAGACGATTGCGACCGGGGCGAGGATATGGTGGACT TTATCAATGATACTGACATCCTGAACATTCAGGCCGAG ACCGAAACAGCTCAGGCACTGTTCCACGCCCAGGAGG AACAGACCCATAAGGAGGCTGTCCAGGTGCTGAAGAG GAAATATGCATCTAGTCCACTGTCAAGCGTCAGCCTGT GCGTGAACAATAACATCTCCCCCCGCCTGAAGGCCATC TGTATTGAGAATAAGAACACTGCCGCTAAACGGAGACT GTTCGAACTGCCTGATTCCGGCTACGGGAATTCTGAGG TGGAAATCCACGAGATTCAGCAGGTCGAAGGGCATGA TACTGTGGAGCAGTGCAGTATGGGATCAGGCGACAGC ATTACCTCCTCTAGTGATGAGCGGCACGACGAAACTCC AACCAGAGACATCATTCAGATCCTGAAGTGTTCCAATG CAAACGCAGCCATGCTGGCCAAGTTCAAAGAGCTGTTT GGCATCTCTTTCACAGAACTGATTAGGCCCTTCAAGTC CGATAAATCTACATGCACTGACTGGTGTGTCGCTGCAT TTGGGATTGCACCTAGCGTGGCCAACTTCAAGCACATC ACCTACGTCTATATCTACAATGTCTATCGCGTGCATGG AGCCATGGTCATCCTGGCTCTGCTGCGCTTTAAGGTCG AGAAACGAGAACAGCAGCTGAAGACAATCGATGCCAAA CTGCTGTGTATTAGTGCCGCTTCAATGCTGATCCAGCC ACCTAAGCTGCGATCCACCCCAGCAGCCCTGTATTGGT TCAAAACAGCCATGAGCAACATTTCCGAGGTGGACGG CGAGACACCCGAATGGATCCAGAGACAGACTGTCCTG CAGCACTCTTTTAACGATGCCATCTTCGACCTGAGTGA GATGGTGCAGTGGGCTTACGATAATGACTTTATCGACG ATTCAGATATTGCCTATAAGTACGCACAGCTGGCCGAA ACCAATAGCAACGCCTGCGCTTTCCTGAAATCAAATAG CCAGGCTAAGATCGTGAAAGACTGCGCAACAATGTGTC GCCACTACAAGCGAGCAGAGAAACGGGAAATGACTAT GTCTCAGTGGATTAAGAGGCGATGTGCACAGGTGGAC GATGACGGCGATTGGAGGGACATCGTCCGATTTCTGC GGTATCAGCAGGTCGATTTCGTGGCTTTTCTGAGCGCA CTGAAGAATTTCCTGCATGGGGTGCCCAAGAAAAACTG CATCCTGATCTACGGGGCTCCTAATACCGGAAAAAGTC TGTTTGGCATGTCACTGATGCACTTCCTGCAGGGAGCC ATCATTAGTTATGTGAACTCCAAGTCTCATTTTTGGCTG CAGCCCCTGTACGACGCTAAAATTGCAATGCTGGATGA CGCCACCAGCCCATGCGGCATCTATCGACCCATTTTTA AGAAATGTACACGGTGGAAGTCTTACATCAGTTTCAGA TGTAAAGCTCTGAGCATCGTGCACATTATGCCTACCTT CACATACTATATCAATATTAACGCCGGCAAGGATGACA GGTGGCCATATCTGCACTCCCGCGTGGTCGTGTTCAC ATTCCATAACGAGTTCCCTTTCGATAAGAATGGGAACC CAGAATACGGACTGAATGACAAGAACTGGAAATCATTC TTTAGCAGAACTTGGTGCAGGCTGAACCTGCATGAGGA AGAGGTGAAGGAGAATGATGGCGACGCCTTCCCTGCT TTTAAATGTGTGTCTGGGCAGAATACTAGAACCCTGAG GGAC	MADPAGTDEGEGTGC NGWFFVEAVVSRRTGD PVSEDENEDDCDRGED MVDFINDTDILNIQAETE TAQALFHAQEEQTHKE AVQVLKRKYASSPLSSV SLCVNNNISPRLKAICIE NKNTAAKRRLFELPDS GYGNSEVEIQQIQQVE GHDTVEQCSMGSGDSI TSSSDERHDETPTRDII QILKCSNANAAMLAKFK ELFGISFTELIRPFKSDK STCTDWCVAAFGIAPSV AESLKTLIKPYCLYIHIQ CLSCSWGMVILALLRFK CAKNRTTIEKLLSKLLCI SAASMLIQPPKLRSTPA ALYWFKTAMSNISEVD GETPEWIQRQTVLQHS FNDAIFDLSEMVQWAY DNDFIDDSDIAYKYAQL AETNSNACAFLKSNSQ AKIVKDCATMCRHYKR AEKREMTMSQWIKRRC EKVDDDGDWRDIVRFL RYQQVDFVAFLSALKNF LHGVPKKNCILIYGAPN TGKSLFGMSLMHFLQG AIISYVNSKSHFWLQPL YDAKIAMLDDATSPCW AYIDQYLRNALDGNPIS LDVKHKALVQLKCPPLLI TSNINAGKDDRWPYLH SRVVVFTFHNEFPFDKN GNPVYGLNDKNWKSFF SRTWCRLNLHEEEDKE NDGDAFPAFKCVSGQN TRTLRD	27.6%
HPV35 E2	ATGATGGAGACGCTTTCCCAGCGTTTAAGTGTGTGTCA GGACAAAATACTAGAACATTACGAGACTGATAGCACAT GTTTGTCTGATCACATACAGTATTGGAAACTGATTCGTC TTGAATGTGCAGTATTTTATAAAGCAAGAGAAATGGGAA TTAAAACTCTTAACCACCAAGTGGTTCCAACGCAGGCC ATTTCAAAAGCCAAAGCAATGCAAGCAATTGAACTGCA ATTAATGTTAGAGACATTAAATACAACTGAGTATAGCAC AGAAACATGGACACTGCAAGAAACAAGTATTGAATTATA TACAACAGTTCCACAAGGATGTTTTAAAAAACATGGGG TTACAGTGGAAGTACAATTTGATGGTGATAAACAAAATA CTATGCATTATACTAATTGGACACATATATATATATTAGA GGACAGTATATGTACTGTTGTAAAGGGACTGGTAAATT ATAAAGGTATTTATTATGTGCATCAGGGTGTAGAAACAT ATTATGTTACTTTTAGGGAAGAGGCTAAAAAGTATGGAA AAAAAAATATATGGGAAGTGCATGTGGGTGGTCAGGTA ATTGTTTGTCCTGAATCTGTATTTAGCAGCACAGAACTA TCCACTGCTGAAATTGCTACACAGCTACACGCCTACAA CACCACCGAGACCCATACCAAAGCCTGCTCCGTGGGC ACCACAGAAACCCAGAAGACAAATCACAAACGACTTCG AGGGGGTACCGAGCTCCCCTACAACCCCACCAAGCGA GTGCGACTCAGTGCCGTGGACAGTGTTGACAGAGGGG TCTACTCTACATCTGACTGCACAAACAAAGACCGGTGT GGTAGTTGTAGTACAACTACACCTATAGTACATTTAAAA GGTGATGCAAATACATTAAAGTGTTTAAGATATAGATTG GGTAAATATAAAGCATTGTATCAAGATGCTTCATCTACA TGGAGATGGACATGTACAAACGATAAAAAACAAATAGC AATTGTAACATTAACTTACACAACAGAATATCAAAGGGA TAAATTTTTAACTACAGTAAAAATACCTAACACAGTTACA GTGTCTAAAGGATATATGTCTATATGA	ATGATGGAAACTCTGTCACAGCGACTGAGCGTCTGCCA GGATAAGATTCTGGAGCATTACGAAACCGATAGCACTT GTCTGAGCGACCACATCCAGTACTGGAAGCTGATTAGA CTGGAGTGCGCTGTGTTCTATAAGGCAAGGGAAATGG GGATCAAAACTCTGAACCATCAGGTGGTCCCAACCCAG GCAATCAGCAAGGCCAAAGCTATGCAGGCCATTGAGC TGCAGCTGATGCTGGAAACCCTGAATACCACAGAGTAC TCAACCGAAGACTGGACACTGCAGGAGACTAGCATTGA ACTGTACACTACCGTGCCCACACGGTGCCTGAAGAAA GATGTGTATACTGTCGAGGCTCAGTTTGACGGCGATAA GCAGAACACCATGCACTACACTAATTGGACCCATATCT ATATTCTGGAAGACAGTATCTGTACAGTGGTCAAGGGG CTGGTGAACTACAAAGGAATCTACTATGTGCACCAGGG CGTCGAGACCTACTATGTCACATTCAGAGAGGAAGCCA AGAAATATGGGAAGAAAAATATCTGGGAGGTGCATGTC GGCGGGCAGGTCATCGTCTGCCCCGAATCTGTGTTTA GCTCCACTGAGCTGAGTACCGCAGAAATCGCCACCCA GCTGCACGCTTACAACACAACTGAGACCCATACAAAGG CATGTTCCGTGGGAACCACAGAAACACAGAAGACTAAC CACAAACGGCTGAGAGGAGGCACCGAGCTGCCCTACA ATCCTACAAAGAGGGTGCGCCTGAGTGCCGTGGACTC AGTCGATCGCGGCGTCTATTCAACAAGCGACTGTACTA ACAAAGATCGATGCGGCTCCTGTTCTACTACCACACCT ATCGTGCATCTGAAGGGGGACGCTAATACCCTGAAATG CTCTCGATACCGGCTGGGAAAGTACAAAGCCCTGTATC AGGACGCTTCTAGTACATGGAGGTGGACTTGTACCAAC GATAAGAAACAGATCGCAATTGTGACACTGACTTACAC TACCGAGTATCAGCGCGATAAGTTCCTGACAACTGTGA AAATCCCAAATACCGTGACAGTCAGCAAGGGCTATATG TCCATT	MMETLSQRLSVCQDKIL EHYETDSTCLSDHIQY WKLIRLECAVFYKAREM GIKTLNHQVVPTQAISK AKAMQAIELQLMLETLN TTEYSTETWTLQETSIE LYTTVPQGCFKKHGVT VEVQFDGDKQNTMHYT NWTHIYILEDSICTVVKG LVNYKGIYYVHQGVETY YVTFREEAKKYGKKNIW EVHVGGQVIVCPESVFS STELSTAEIATQLHAYN TTETHTKACSVGTTETQ KTNHKRLRGGTELPYN PTKRVRLSAVDSVDRG VYSTSDCTNKDRCGSC STTTPIVHLKGDANTLK CLRYRLGKYKALYQDA SSTWRWTCTNDKKQIAI VTLTYTTEYQRDKFLTT VKIPNTVTVSKGYMSI	25.7%
HPV35 E4	TTGTTTGTCCTGAATCTGTATTTAGCAGCACAGAACTAT CCACTGCTGAAATTGCTACACAGCTACACGCCTACAAC ACCACCGAGACCCATACCAAAGCCTGCTCCGTGGGCA CCACAGAAACCCAGAAGACAAATCACAAACGACTTCGA GGGGGTACCGAGCTCCCCTACAACCCCACCAAGCGAG TGCGACTCAGTGCCGTGGACAGTGTTGACAGAGGGGT CTACTCTACATCTGACTGCACAAACAAAGACCGGTGTG GTAGTTGTAGTACAACTACACCTATAG	ATGTTTGTCCTGAACCTGTACCTGGCTGCCCAGAACTA TCCCCTGCTGAAGCTGCTGCATTCCTATACCCCTACCA CTCCTCCACGGCCAATCCCCAAGCCTGCCCCATGGGC TCCCCAGAAACCTCGGAGACAGATTACCAACGACTTCG AGGGAGTGCCAAGCTCCCCAACCACACCACCTTCTGA GTGCGATAGTGTCCCTTGGACAGTGCTGACTGAAGGC TCCACCCTGCACCTGACAGCCCAGACTAAGACCGGGG TGGTCGTGGTCGTGCAGCTGCATCTG	LFVLNLYLAAQNYPLLK LLHSYTPTTPPRPIPKPA PWAPQKPRRQITNDFE GVPSSPTTPPSECDSV PWTVLTEGSTLHLTAQT KTGVVVVVQLHL	25.6%
HPV35 E5	ATGATAGACCTTACAGCTTCCAGTACTGTGTTGCTGTG CTTTTTGTTGTGCTTTTGTGTGCTTTTGTGCTTGTGTCT GCTTGTACGTTCGCTATTGCTATCTGTGTCATTATACTC AGCATTAATATTACTGGTTTTAATACTGTGGGTTACTGT AGCAACACCACTACGTTGCTTTTGTTGTTTCTTTGCTT TTTGTATATACCTATGGGAATGATTAACGCTCATGCACA ATATTTGGCAGTACAGTAA	ATGATTGACCTGACTGCTTCCTCCACTGTGCTGCTGTG TTTTCTGCTGTGTTTCTGCGTCCTGCTGTGCCTGTGTCT GCTGGTGCGGTCTCTGCTGCTGAGCGTGTCCCTGTAC AGTGCTCTGATCCTGCTGGTCCTGATTCTGTGGGTGAC CGTCGCAACACCCCTGCTGGCCTTCGTGGTCTCCTGC TTTTGTATCTACCTGTGGATGATTAACGCCCACGCTCA GTATCTGGCCGTGCAG	MIDLTASSTVLLCFLLCF CVLLCLCLLVRSLLLSVS LYSALILLVLILWVTVAT PLRCFCCFLCFLYIPMG MINAHAQYLAVQ	28.1%

HPV35 E6	ATGTTTCAGGACCCAGCTGAACGACCTTACAAACTGCA TGATTTGTGCAACGAGGTAGAAGAAAGCATCCATGAAA TTTGTTTGAATTGTGTATACTGCAAACAAGAATTACAGC GGAGTGAGGTATATGACTTTGCATGCTATGATTTGTGT ATAGTATATAGAGAAGGCCAGCCATATGGAGTATGCAT GAAATGTTTAAAATTTTATTCAAAAATAAGTGAATATAGA TGGTATAGATATAGTGTGTATGGAGAAACGTTAGAAAA ACAATGCAACAAACAGTTATGTCATTTATTAATTAGGTG TATTACATGTCAAAAACCGCTGTGTCCAGTTGAAAAGC AAAGACATTTAGAAGAAAAAAAACGATTCCATAACATCG GTGGACGGTGGACAGGTCGGTGTATGTCCTGTTGGAA ACCAACACGTAGAGAAACCGAGGTGTAA	ATGTTCCAGGACCCCGCCGAAAGGCCCTATAAACTGCA CGATCTGTGTAACGAAGTCGAAGAGAGCATCCACGAAA TCTGTCTGAATTGCGTGTACTGTAAGCAGGAGCTGCAG CGCTCTGAAGTCTACGACTTCGCCTGCTATGATCTGTG TATCGTGTACCGAGAGGGACAGCCATATGGCGTCTGC ATGAAGTGTCTGAAGTTCTACAGCAAGATCTCCGAATA CAGGTGGTACCGCTATAGTGTGTATGGGGAGACTCTG GAAAAGCAGTGCAACAAACAGCTGTGTCACCTGCTGAT CAGGTGCATTACCTGTCAGAAGCCCCTGTGCCCTGTC GAGAAACAGAGACACCTGGAGGAAAAGAAAAGGTTCC ATAATATCGGAGGACGATGGACAGGACGATGCATGTC CTGTTGGAAGCCCACCCGGAGAGAGACAGAAGTG	MFQDPAERPYKLHDLC NEVEESIHEICLNCVYC KQELQRSEVYDFACYD LCIVYREGQPYGVCMK CLKFYSKISEYRWYRYS VYGETLEKQCNKQLCH LLIRCITCQKPLCPVEKQ RHLEEKKRFHNIGGRW TGRCMSCWKPTRRETE V	24.1%
HPV35 E7	ATGCATGGAGAAATAACTACATTGCAAGACTATGTTTTA GATTTGGAACCCGAGGCAACTGACCTATACTGTTATGA GCAATTGTGTGACAGCTCAGAGGAGGAGGAAGATACT ATTGACGGTCCAGCTGGACAAGCAAAACCAGACACCTC CAATTATAATATTGTAACGTCCTGTTGTAAATGTGAGGC GACACTACGTCTGTGTGTACAGAGCACACACATTGACA TACGTAAATTGGAAGATTTATTAATGGGCACATTTGGAA TAGTGTGCCCCGGCTGTTCACAGAGAGCATAA	ATGCACGGCGAAATTACCACTCTGCAGGATTATGTCCT GGATCTGGAGCCCGAAGCCACTGACCTGTATTGTTATG AGCAGCTGTGTGACAGCTCCGAGGAAGAGGAAGACAC AATCGATGGACCAGCAGGACAGGCTAAGCCTGATACC TCTAACTACAATATTGTGACAAGTTGCTGTAAATGCGAG GCAACTCTGCGGCTGTGTGTCCAGAGCACCCACATCG ACATTAGAAAGCTGGAAGATCTGCTGATGGGAACCTTC GGCATCGTGTGCCCCGGGTGTTCTCAGAGGGCC	MHGEITTLQDYVLDLEP EATDLYCYEQLCDSSE EEEDTIDGPAGQAKPDT SNYNIVTSCCKCEATLR LCVQSTHIDIRKLEDLLM GTFGIVCPGCSQRA	23.2%
HPV35 L1	ATGTCTCTGTGGCGGTCTAACGAAGCCACTGTCTACCT GCCTCCAGTGTCAGTGTCTAAGGTTGTTAGCACTGATG AATATGTAACACGCACAAACATCTACTATCATGCAGGC AGTTCTAGGCTATTAGCTGTGGGTCACCCATACTATGC TATTAAAAAACAAGATTCTAATAAAATAGCAGTACCCAA GGTATCTGGTTTGCAATACAGAGTATTTAGAGTAAAATT ACCAGATCCTAATAAGTTTGGATTTCCAGACACATCATT TTATGATCCTGCCTCCCAGCGTTTGGTTTGGGCCTGTA CAGGAGTTGAAGTAGGTCGTGGTCAGCCATTGGGTGT AGGTATTAGTGGTCATCCTTTATTAAATAAATTGGATGA TACTGAAAATTCTAATAAATATGTTGGTAACTCTGGTAC AGATAACAGGGAATGCATTTCTATGGATTATAAACAAAC ACAATTGTGTTTAATAGGTTGTAGGCCTCCTATAGGTGA ACATTGGGGAAAAGGCACACCTTGTAATGCTAACCAGG TAAAAGCAGGAGAATGTCCTCCTTTGGAGTTACTAAAC ACTGTACTACAAGACGGGGACATGGTAGACACAGGATT TGGTGCAATGGATTTTACTACATTACAAGCTAATAAAAG TGATGTTCCCCTAGATATATGCAGTTCCATTTGCAAATA TCCTGATTATCTAAAAATGGTTTCTGAGCCATATGGAGA TATGTTATTTTTTTATTTACGTAGGGAGCAAATGTTTGTT AGACATTTATTTAATAGGGCTGGAACTGTAGGTGAAAC AGTACCTGCAGACCTATATATTAAGGGTACCACTGGCA CATTGCCTAGTACTAGTTATTTTCCTACTCCTAGTGGCT CTATGGTAACCTCCGATGCACAAATATTTAATAAACCAT ATTGGTTGCAACGTGCACAAGGCCATAATAATGGTATT TGTTGGAGTAACCAATTGTTTGTTACTGTAGTTGATACA ACCCGTAGTACAAATATGTCTGTGTGTTCTGCTGTGTCT TCTAGTGACAGTACATATAAAAATGACAATTTTAAGGAA TATTTAAGGCATGGTGAAGAATATGATTTACAGTTTATT TTTCAGTTATGTAAAATAACACTAACAGCAGATGTTATG ACATATATTCATAGTATGAACCCGTCATTTTAGAGGAT TGGAATTTTGGCCTTACACCACCGCCTTCTGGTACCTT AGAGGACACATATCGCTATGTAACATCACAGGCTGTAA CTTGTCAAAAACCCAGTGCACCAAAACCTAAAGATGAT CCATTAAAAAATTATACTTTTTGGGAGGTTGATTTAAAG GAAAAGTTTTCTGCAGACTTAGATCAATTTCCGTTGGG CCGTAAATTTTTGTTACAAGCAGGACTAAAGGCCAGGC CTAATTTTAGATTAGGCAAGCGTGCAGCTCCAGCATCT ACATCTAAAAAATCTTCTACTAAACGTAGAAAAGTAAAA AGTTAA	ATGAGCCTGTGGAGGAGCAATGAAGCAACCGTCTATCT GCCCCCTGTGAGCGTGAGCAAAGTCGTGAGCACTGAT GAATACGTGACAAGGACCAACATCTACTATCACGCAGG AAGCTCCCGACTGCTGGCTGTGGGGCATCCTTACTATG CAATCAAGAAACAGGACTCCAACAAGATTGCCGTGCCA AAAGTCTCTGGACTGCAGTACAGAGTGTTCAGGGTCAA GCTGCCTGATCCAAACAAGTTCGGCTTTCCTGACACTT CCTTTTATGATCCATGCCTGCAGCCGACTGGTGTGGGCC TGTACCGGAGTGGAGGTCGGACGAGGACAGCCACTGG GAGTCGGCATCTCTGGACACCCTCTGCTGAACAAGCT GGACGATACCGAGAACCTGAACAAGTACGTGGGAAAC AGCGGCAATTCCGGGACCGACAATCGGGAATGCATTA GCATGGATTATAAGCAGACACAGCTGTGCCTGATCGGA TGTAGACCCCCTATTGGCGAACATTGGGGGAAGGGAA CACCCTGCAACGCTAATCAGGTGAAAGCAGGCGAGTG TCCACCCCTGGAACTGCTGAACACAGTGCTGCAGGAC GGGGATATGGTCGACACTGGCTTCGGGGCCATGGATT TTACCACACTGCAGGCTAATAAGTCTGACGTGCCTCTG GATATCTGCTCTAGTATTTGTAAGTACCCAGACTATCTG AAAATGGTCAGTGAGCCCTACGGGGATATGCTGTTCTT TTATCTGCGGAGAGAACAGATGTTCGTGCGGCACCTGT TTAACAGAGCTGGAACTGTGGGCGAGACCGTCCCAGC AGACCTGTACATCAAGGGGACTACCGGAACACTGCCC TCAACTAGCTATTTCCCCACCCCTTCCGGCTCTATGGT GACATCCGATGCCCAGATCTTCAACAAGCCTTACTGGC TGCAGAGGGCTCAGGGCCATAACAATGGGATTTGCTG GAGCAACCAGCTGTTCGTGACTGTGGTCGACACAACTC GCTCCACCAATATGTCTGTGTGTAGTGCTGTCTCAAGC TCCGACTCTACCTACAAGAACGATAACTTCAAGGAGTA CCTGAGACACGGCGAGGAATATGACCTGCAGTTCATCT TTCAGCTGTGCAAGATTACCCTGACAGCCGATGTGATG ACATATATCCATTCAATGAACCCAAGCATTCTGGAGGA CTGGAATTTCGGGCTGACTCCTCCACCCAGCGGAACC CTGGAAGATACATACAGATATGTGACTAGTCAGGCAGT CACCTGTCAGAAGCCTTCAGCCCCAAAGCCCAAAGAC GATCCACTGAAAAACTACACATTCTGGGAGGTGGACCT GAAGGAAAAGTTCAGCGCAGACCTGGATCAGTTCCCC CTGGGACGGAAGTTTCTGCTGCAGGCAGGCCTGAAAG CACGACCAAATTTCCGACTGGGAAGGCGAGCAGCTCC TGCAAGTACATCAAAGAAATCTAGTACTAAGCGACGGA AGGTGAAAAGC	MSLWRSNEATVYLPPV SVSKVVSTDEYVTRTNI YYHAGSSRLLAVGHPY YAIKKQDSNKIAVPKVS GLQYRVFRVKLPDPNK FGFPDTSFYDPASQRL VWACTGVEVGRGQPL GVGISGHPLLNKLDDTE NSNKYVGNSGTDNREC ISMDYKQTQLCLIGCRP PIGEHWGKGTPCNANQ VKAGECPPLELLNTVLQ DGDMVDTGFGAMDFTT LQANKSDVPLDICSSIC KYPDYLKMVSEPYGDM LFFYLRREQMFVRHLFN RAGTVGETVPADLYIKG TTGTLPSTSYFPTPSGS MVTSDAQIFNKPYWLQ RAQGHNNGICWSNQLF VTVVDTTRSTNMSVCS AVSSSDSTYKNDNFKE YLRHGEEYDLQFIFQLC KITLTADVMTYIHSMNP SILEDWNFGLTPPPSGT LEDTYRYVTSQAVTCQ KPSAPKPKDDPLKNYTF WEVDLKEKFSADLDQF PLGRKFLLQAGLKARPN FRLGKRAAPASTSKKSS TKRRKVKS	27.7%
HPV35 L2	ATGCGACACAAAAGGTCTACAAAACGTGTTAAACGTGC ATCTGCAACACAACTATATCGTACTTGCAAAGCTGCAG GAACTTGTCCACCAGATGTTATACCTAAGGTTGAGGGT AATACTGTTGCTGATCAAATTTTAAAATATGGCAGCATG GCTGTGTTTTTTTGGGGGGTTAGGAATTGGTTCTGGATC TGGCACAGGTGGAAGATCTGGATATGTTCCACTGGGTA CAACACCTCCAACGGCTGCCACAAACATTCCTATACGA CCCCCTGTAACTGTGGAAAGTATACCATTAGACACAAT TGGCCCTTTAGATTCTTCTATAGTGTCATTAGTAGAGGA AACTAGTTTTATTGAGTCTGGTGCCCCTGTTGTTACACC AAGGGTCCCACCTACAACAGGTTTTACAATAACCACAT CTACAGATACCACACCTGCTATTTTAGATGTGACATCCA TAAGTACACATGATAATCCTACTTTCACTGATCCTTCTG TTTTACACCCACCCACGCCTGCAGAAACTTCAGGTCAT TTTGTACTTTCATCATCTTCTATTAGTACACATAATTATG AAGAAATCCCTATGGATACTTTTATTGTTTCCACAGACA GCAATAATATAACTAATAGCACGCCTATTCCAGGGTCT CGCCCTACGACACGCCTAGGATTATATAGTAAAGGTAC CCAGCAGGTTAAGGTTGTTGACCCTGCCTTTATGACTT CTCCTGCAAAACTTATTACATATGATAATCCTGCATATG AAGGCCTTAACCCTGATACAACCTTACAATTTGAGCAT GAGGATATTAGCTTAGCTCCGGATCCTGACTTTATGGA CATTATAGCTTTACATAGGCCTGCACTAACATCTAGGAA AGGCACTATTAGATATAGTAGAGTAGGTAATAAACGTA CTATGCATACACGAAGTGGAAAAGCTATAGGGGCACG GGTACATTATTATCAGGATTTAAGTAGTATTACTGAAGA TATAGAATTACAACCCTTACAACATGTACCATCCTCTTT ACCACATACCACTGTTTCAACATCATTAAATGATGGTAT GTTTGATATTTATGCTCCTATAGATACTGAGGAAGATAT TATATTTTCAGCATCTTCTAACAATACTTTATATACTACA TCTAACACTGCATATGTTCCTAGCAATACTACTATACCA TTAAGTAGTGGCTATGATATTCCTATAACAGCAGGGCC AGACATTGTATTTAACTCTAATACTATTACTAACACTGTA CTACCGGTACCCACAGGTCCTATATATTCTATTATTGCA GATGGGGGGTGACTTTTATTTACACCCTAGTTATTATTTA TTAAAACGACGTCGTAAACGTATCCCATATTTTTTTGCA GATGTCTCTGTGGCGGTCTAA	ATGAGACATAAAAGAAGCACAAAGAGAGTCAAGAGAGC AAGCGCAACACAGCTGTACCGAACCTGCAAAGCCGCC GGAACATGCCCTCCAGACGTCATCCCCAAGGTGGAGG GAAACACCGTCGCTGATCAGATTCTGAAATACGGCTCC ATGGCAGTGTTCTTTGGAGGACTGGGAATCGGATCAG GAAGCGGAACAGGAGGACGATCTGGCTATGTGCCACT GGGAACCACACCACCTACAGCAGCTACTAATATCCCCA TTCGGCCACCCGTGACCGTCGAGTCTATCCCCCTGGA CACAATTGGCCCTCTGGATAGCTCCATCGTCAGTCTGG TGGAGGAAACTTCTTTCATTGAAAGTGGGGCCCCTGTG GTCACCCCAAGAGTGCCTCCAACTACCGGCTTCACCAT CACAACTAGCACCGACACCACACCCGCCATCCTGGAT GTGACATCCATTTCTACTCACGACAACCCAACCTTCAC AGATCCATCTGTCCTGCACCCACCTACCCCAGCAGAGA CAAGTGGCCATTTTGTGCTGTCTAGTTCAAGCATCTCA ACCCATAACTACGAGGAAATCCCTATGGACACATTCAT TGTGAGCACTGATTCCAACAATATCACCAATTCAACACC AATTCCCGGGAGCCGGCCTACTACCAGACTGGGACTG TATAGCAAGGGCACCCAGCAGGTGAAAGTGGTCGACC CAGCCTTCATGACTAGCCCCGCCAAGCTGATCACCTAC GATAACCCCGCATATGAAGGCCTGAATCCTGACACAAC TCTGCAGTTCGAGCACGAAGATATTAGCCTGGCCCCTG ACCCAGATTTTATGGACATCATTGCTCTGCATCGACCA GCACTGACCAGCAGGAAAGGGACAATCCGCTACTCCC GAGTCGGAAACAAGAGGACTATGCACACCCGCAGCGG GAAAGCAATTGGAGCCAGGGTGCATTACTATCAGGACC TGTCCTCTATCACCGAGGATATTGAACTGCAGCCACTG CAGCACGTCCCCAGTTCACTGCCTCATACCACAGTGAG TACATCACTGAATGACGGCATGTTCGATATCTACGCCC CCATTGACACTGAGGAAGATATCATCTTCAGCGCTAGC TCCAACAATACACTGTACACTACCAGTAACACTGCTTAT GTGCCTTCAAATACAACTATCCCACTGTCTAGTGGCTAT GACATCCCTATTACCGCAGGGCCAGATATCGTGTTCAA CTCCAATACTATTACCAATTCTGTCCTGCCCGTGCCTAC AGGCCCTATCTACAGCATCATTGCCGACGGGGGAGAT TTTTATCTGCACCCTTCCTACTATCTGCTGAAGCGGAG AAGGAAAGCTATTCCATACTTCTTTGCCGACGTGTCTG TCGCTGTG	MRHKRSTKRVKRASAT QLYRTCKAAGTCPPDVI PKVEGNTVADQILKYGS MAVFFGGLGIGSGSGT GGRSGYVPLGTTPPTA ATNIPIRPPVTVESIPLD TIGPLDSSIVSLVEETSFI ESGAPVVTPRVPPTTG FTITTSTDTTPAILDVTSI STHDNPTFTDPSVLHPP TPAETSGHFVLSSSSIS THNYEEIPMDTFIVSTD SNNITNSTPIPGSRPTT RLGLYSKGTQQVKVVD PAFMTSPAKLITYDNPA YEGLNPDTTLQFEHEDI SLAPDPDFMDIIALHRP ALTSRKGTIRYSRVGNK RTMHTRSGKAIGARVH YYQDLSSITEDIELQPLQ HVPSSLPHTTVSTSLND GMFDIYAPIDTEEDIIFS ASSNNTLYTTSNTAYVP SNTTIPLSSGYDIPITAG PDIVFNSNTITNTVLPVP TGPIYSIIADGGDFYILHP SYYLLKRRRKRIPYFFA DVSVAV	28.9%

HPV39 E1	ATGGCCAATCGTGAAGGTACAGACGGGGATGGGTCGG GATGTAACGGATGGTTTCTAGTACAGGCAATAGTAGAT AAACAAACAGGCGACACAGTGTCGGAGGATGAGGATG AAAATGCAACAGATACAGGTTCAGACCTGGCAGACTTT ATTGATGATTCCACAGATATTTGTGTACAGGCAGAGCG TGAGACAGCACAGGTACTTTTACATATGCAAGAGGCCC AAAGGGATGCACAAGCAGTGCGTGCCTTAAAACGAAA GTATACAGACAGCAGTGGCGACACTAGACCGTATGGA AAAAAAGTAGGCAGGAATACCAGGGGAACACTACAGG AAATTTCATTAAATGTAAGCAGTACGCAGGCAACACAAA CGGTGTATTCCGTGCCAGACAGCGGATATGGCAATATG GAAGTGGAAACAGCTGAAGTGGAGGAGGTAACTGTAG CAACTAATACAAATGGGGATGCTGAAGGGGAACATGG CGGCAGTGTACGGGAGGAGTGCAGTAGTGTGGATAGT GCTATAGATAGTGAAAACCAGGATCCCAAATCTCCAAC TGCACAAATTAAATTATTGTTACAATCCAATAACAAAAA GGCTGCAATGCTAACACAATTTAAAGAAACATATGGAC TATCCTTTACTGACCTGGTACGTACGTTTAAAAGTGATA AAACAACATGTACAGACTGGGTGGCAGCCATATTTGGA GTACATCCAACTATTGCAGAAGGATTTAAAACATTAATC AACAAATATGCCTTATATACACATATACAAAGCTTAGAC ACAAAACAAGGAGTACTAATTTTAATGCTAATAAGATAT ACATGTGGAAAAAATAGGGTTACTGTAGGAAAGGGATT AAGTACATTGTTACATGTTCCAGAAAGTTGTATGCTTCT GGAGCCTCCTAAACTGCGCAGCCCTGTAGCAGCACTA TATTGGTATCGCACAGGTATATCCAATATTAGTGTGGTA ACAGGGGATACGCCAGAATGGATACAACGATTAACTGT TATACAACATGGAATAGATGATAGTGTATTTGACCTATC GGACATGGTACAATGGGCATTTGACAATGAATATACTG ATGAAAGTGACATAGCATTTAATTATGCAATGTTAGCAG ATTGTAACAGTAATGCTGCAGCCTTTTTAAAAAGTAACT GCCAGGCAAAATATGTAAAAGATTGTGCAACAATGTGT AAACATTACAAGCGAGCACAAAAAAGGCAAATGTCCAT GTCTCAATGGATAAAATTTAGGTGTAGTAAATGTGATGA AGGCGGGGACTGGAGACCCATAGTACAATTCTTAAGAT ATCAAGGAATAGAATTTATATCCTTTTTATGTGCATTAAA GGAATTTTTAAAGGGTACTCCCAAAAAAAACTGTATAGT TATATATGGACCTGCGAATACAGGAAAGTCACATTTTTG TATGAGCCTTATGCATTTTTTACAGGGCACAGTTATTTC ATATGTAAACTCCACCAGCCACTTTTGGCTAGAACCAC TTGCAGATGCAAAACTAGCAATGTTAGATGATGCAACC GGTACCTGCTGGTCATATTTCGATAATTATATGAGAAAT GCATTAGATGGGTATGCAATAAGTTTAGATAGGAAATAT AAAAGTTTACTACAAATGAAATGTCCACCATTATTAATA ACCTCCAATACCAATCCTGTGGAAGACGATAGGTGGCC ATATTTACGTAGTAGGCTAACAGTGTTTAAATTTCCTAA TGCATTTCCATTTGACCAAAACAGGAATCCAGTGTACA CAATCAATGATAAAAACTGGAAATGTTTTTTTGAAAAGA CTTGGTGCAGATTAGACTTGCAGCAGGACGAGGATGA AGGAGACAATGATGAAAACACTTTCACAACGTTTAAATG TGTTACAGGACAAAATACTAGAATACTATGA	ATGGCTAATCGGGAGGGGACTGATGGGGATGGGAGC GGCTGTAATGGCTGGTTCCTGGTGCAGGCAATCGTGG ATAAGCAGACTGGCGACACCGTGAGCGAGGACGAAGA TGAGAACGCTACAGATACTGGCTCCGACCTGGCAGATT TCATCGACGATTCTACAGACATTTGCGTGCAGGCCGAA AGAGAGACTGCTCAGGTCCTGCTGCACATGCAGGAGG CACAGAGGGATGCACAGGCTGTGCGAGCTCTGAAGCG GAAATACACCGACAGCTCCGGAGATACAAGGCCATATG GAAAGAAAGTGGGCAGGAACACCCGCGGCACACTGCA GGAAATCTCCCTGAATGTCTCTAGTACCCAGGCTACCC AGACAGTGTACTCTGTCCCCGACAGTGGGTATGGAAAC ATGGAAGTGGAGACTGCCGAGGTCGAGGAAGTGACCG TCGCCACTAACACCAATGGGGATGCTGAAGGAGAGCA TGGCGGGAGCGTGCGGGAGGAATGCTCAAGCGTCGA CTCAGCTATCGATAGCGAAAATCAGGACCCTAAGAGCC CAACAGCCCAGATCAAGCTGCTGCTGCAGTCCAACAAT AAGAAAGCCGCTATGCTGACTCAGTTTAAGGAGACCTA CGGGCTGAGTTTCACAGATCTGGTGAGAACTTTTAAGT CAGACAAAACCACATGTACCGATTGGGTGGCAGCCATC TTCGGAGTCCACCCCACAATTGCAGAGGGCTTTAAGAC TCTGATCAACAAATACGCCCTGTATACCCATATTCAGTC TCTGGACACAAAGCAGGGCGTGCTGATCCTGATGCTG ATTCGCTACACTTGCGGGAAGAATCGAGTGACCGTCG GCAAAGGGCTGTCTACACTGCTGCACGTGCCCGAAAG TTGTATGCTGCTGGAGCCCCCTAAGCTGAGGAGCCCT GTCGCTGCACTGTACTGGTATCGCACAGGGATCAGCA ACATTTCCGTGGTCACTGGAGATACCCCTGAGTGGATC CAGCGGCTGACTGTGATCCAGCATGGCATTGACGATTC CGTGTTCGACCTGTCTGATATGGTCCAGTGGGCTTTTG ACAACGAATACACAGACGAGTCCGATATTGCATTCAAT TATGCTATGCTGGCAGATTGCAACTCAAATGCCGCTGC ATTTCTGAAGAGCAATTGTCAGGCAAAGTACGTGAAAG ACTGCGCCACCATGTGTAAGCACTATAAAAGAGCCCAG AAAAGGCAGATGTCCATGTCTCAGTGGATCAAGTTCAG ATGCTCTAAATGTGACGAGGGAGGCGATTGGCGGCCT ATTGTGCAGTTTCTGAGATACCAGGGCATCGAATTCAT TAGCTTTCTGTGCGCCCTGAAGGAGTTCCTGAAAGGGA CTCCTAAGAAAAACTGTATCGTGATCTACGGCCCAGCC AATACCGGGAAGAGTCACTTCTGCATGTCACTGATGCA TTTTCTGCAGGGCACTGTGATCAGCTATGTCAACAGTA CCTCACATTTCTGGCTGGAACCACTGGCAGACGCCAA GCTGGCAATGCTGGACGATGCCACAGGAACTTGCTGG TCCTACTTTGATAACTATATGCGAAATGCTCTGGACGG CTACGCAATCAGCCTGGATCGGAAGTATAAATCCCTGC TGCAGATGAAGTGTCCACCCCTGCTGATTACATCTAAC ACTAATCCAGTGGAGGACGATAGGTGGCCCTACCTGC GGAGTAGACTGACCGTCTTCAAATTTCCCAACGCCTTC CCTTTTGACCAGAACCGCAATCCCGTGTATACCATCAA CGATAAGAACTGGAAGTGCTTCTTTGAAAAGACATGGT GTCGACTGGACCTGCAGCAGGACGAAGATGAGGGGGA CAACGATGAGAATACCTTCACTACCTTTAAATGTGTGAC CGGACAGAATACACGCATCCTG	MANREGTDGDGSGCN GWFLVQAIVDKQTGDT VSEDEDENATDTGSDL ADFIDDSTDICVQAERE TAQVLLHMQEAQRDAQ AVRALKRKYTDSSGDT RPYGKKVGRNTRGTLQ EISLNVSSTQATQTVYS VPDSGYGNMEVETAEV EEVTVATNTNGDAEGE HGGSVREECSSVDSAI DSENQDPKSPTAQIKLL LQSNNKKAAMLTQFKE TYGLSFTDLVRTFKSDK TTCTDWVAAIFGVHPTI AEGFKTLINKYALYTHIQ SLDTKQGVLILMLIRYTC GKNRVTVGKGLSTLLH VPESCMLLEPPKLRSPV AALYWYRTGISNISVVT GDTPEWIQRLTVIQHGI DDSVFDLSDMVQWAFD NEYTDESDIAFNYAMLA DCNSNAAAFLKSNCQA KYVKDCATMCKHYKRA QKRQMSMSQWIKFRCS KCDEGGDWRPIVQFLR YQGIEFISFLCALKEFLK GTPKKNCIVIYGPANTG KSHFCMSLMHFLQGTVI SYVNSTSHFWLEPLAD AKLAMLDDATGTCWSY FDNYMRNALDGYAISLD RKYKSLLQMKCPPLLIT SNTNPVEDDRWPYLRS RLTVFKFPNAFPFDQNR NPVYTINDKNWKCFFEK TWCRLDLQQDEDEGD NDENTFTTFKCVTGQN TRIL	25.8%
HPV39 E2	ATGAAGGAGACAATGATGAAAACACTTTCACAACGTTTA AATGTGTTACAGGACAAAATACTAGAATACTATGAACAA GACAGTAAATCAATATATGATCAAATTAATTATTGGAAA TGTGTGCGAATGGAAAATGCAATATTTTATGCAGCACG AGAACGTGGCATGCATACTATTGACCACCAGGTGGTGC CAACCATAAACATTTCAAAATGTAAAGCATATCAAGCTA TTGAACTGCAGATGGCACTAGAAAGTGTTGCACAAACT GAATACAATACAGAGGAGTGGACATTAAAAGACACTAG TAATGAACTGTGGCATACACAGCCAAAACAATGTTTTAA AAAACAAGGAACTACAGTGGAGGTGTGGTATGATGGG GACAAATGTAATGCTATGAACTATGTATTATGGGGTGCT ATATATTATAAAAATAATATAGACATATGGTGTAAAACA GAAGGGTGTGTGGACTATTGGGGTATATATTATATGAA CGAGCACCTAAAAGTATACTATGAAGTGTTTATTCAAGA TGCGGAAAGGTATGGGACTAGTGGCAAATGGGAAGTG CATTATAATGGCAACATAATTCATTGTCCTGACTCTATG TGCAGTACCAGTGACGGATCGGTACCCACTACTGAACT TACTACCGAATTATCAAACACCACCGCGACCCATTCCA CCGCAACAACCCCATGCACCCAAAAAACAATCCCGCC GCCGTCTCGAAAGCGACCTCGACAGTGTGCAGTCACA GAGCCCACTGAGCCCGACGGAGTGTCCCTGGACCATC TTAACAACCCACTCCACAGTAACAGTACAGGCCACAAC ACAAGACGGTACCTCAGTTGTGGTAACACTACGCCTAT AATACATTTAAAAGGTGACAAAAATGGTTTAAAATGTTT AAGATATAGACTACAAAAATATGACACATTGTTTGAAAA TATTTCATGTACCTGGCATTGGATACGGGGTAAGGGAA CCAAAAACGCTGGCATATTAACTGTTACATATGCCACA GAGTCACAACGCCAAAAATTTTTGGACACTGTTAAAATA CCTTCTAGTGTACATGTTTCATTGGGTTACATGACATTG TAA	ATGAAGGAAACTATGATGAAAACACTGAGCCAGAGACT GAACGTGCTGCAGGATAAGATTCTGGAGTATTACGAAC AGGATTCTAAAAGTATCTACGACCAGATTAACTATTGGA AGTGCGTGCGGATGGAGAATGCCATCTTCTACGCCGC TAGGGAACGCGGGATGCACACTATTGATCATCAGGTG GTCCCAACTTCAAGAAACAGGGAACCACAGTGGAAGTC TGGTACGACGGCGATAAGTGTAACGCAATGAATTATGT GCTGTGGGGCGCCATCTACTATAAGAACAATATCGACA TTTGGTGCAAAACCGAGGGCTGTGTCGATTATTGGGG GATCTACTATATGAACGAACATCTGAAGGTGTACTATGA GGTCTTTATTCAGGATGCCGAACGGTACGGAACTAGCG GCAAATGGGAGGTGCATTATAACGGCAATATCATTCAC TGCCCAGACAGTATGTGTTCTACAAGTGATGGCTCAGT GCCCACTACCGAGCTGACAACTGAACTGTCCAATACCA CAGCCACTCACTCTACCGCTACTACCCCTTGCACACAG AAGACTATCCCCCCTCCATCACGAAAACGGCCAAGACA GTGTGCTGTGACCGAGCCCACAGAACCTGACGGCGTC AGCCTGGATCACCTGAACAATCCCCTGCATTCAAACAG CACCGGGCACAATACACGGAGATACCTGTCCTGCGGA AACACAACTCCTATCATTCATCTGAAGGGGGACAAAAA TGGACTGAAGTGCCTGAGGTACCGCCTGCAGAAATAT GATACACTGTTCGAGAACATCTCTTGTACTTGGCACTG GATTCGAGGCAAGGGGACCAAAAATGCTGGAATCCTG ACCGTGACATACGCAACCGAATCTCAGAGGCAGAAGTT TCTGGACACAGTCAAAATTCCTAGCTCCGTGCACGTCA GTCTGGGCTATATGACACTG	MKETMMKTLSQRLNVL QDKILEYYEQDSKSIYD QINYWKCVRMENAIFYA ARERGMHTIDHQVVPTI NISKCKAYQAIELQMAL ESVAQTEYNTEEWTLK DTSNELWHTQPKQCFK KQGTTVEVWYDGDKC NAMNYVLWGAIYYKNNI DIWCKTEGCVDYWGIY YMNEHLKVYYEVFIQDA ERYGTSGKWEVHYNG NIIHCPDSMCSTSDGSV PTTELTTELSNTTATHS TATTPCTQKTIPPPSRK RPRQCAVTEPTEPDGV SLDHLNNPLHSNSTGH NTRRYLSCGNTTPIIHLK GDKNGLKCLRYRLQKY DTLFENISCTWHWIRGK GTKNAGILTVTYATESQ RQKFLDTVKIPSSVHVS LGYMTL	35.4%
HPV39 E4	TTCATTGTCCTGACTCTATGTGCAGTACCAGTGACGGA TCGGTACCCACTACTGAACTTACTACCGAATTATCAAAC ACCACCGCGACCCATTCCACCGCAACAACCCCATGCA CCCAAAAAACAATCCCGCCGCCGTCTCGAAAGCGACC TCGACAGTGTGCAGTCACAGAGCCCACTGAGCCCGAC GGAGTGTCCCTGGACCATCTTAACAACCCACTCCACAG TAACAGTACAGGCCACAACACAAGACGGTACCTCAGTT GTGGTAACACTACGCCTATAA	ATGATTGTGCTGACACTGTGCGCCGTGCCTGTGACCG ACCGCTACCCCCTGCTGAACCTGCTGCCCAACTACCA GACCCCTCCCCGCCCAATCCCACCTCAGCAGCCACAC GCACCCAAGAAACAGAGTCGGAGAAGGCTGGAGTCTG ACCTGGATAGTGTCCAGAGCCAGTCCCCTCTGTCACCA ACCGAATGCCCTTGGACAATTCTGACCACACATAGCAC AGTCACTGTGCAGGCTACTACCCAGGACGGCACTTCC GTGGTCGTGACACTGCGGCTG	FIVLTLCAVPVTDRYPLL NLLPNYQTPPRPIPPQQ PHAPKKQSRRRLESDL DSVQSQSPLSPTECPW TILTTHSTVTVQATTQD GTSVVVTLRL	34.8%
HPV39 E5	ATGATATTATTGGTATTTTTGGTGTGGTTTGGTGTGTGT ATATATATATGTTGCAATGTCCCGCTTTTGCCGTCTGTG CATGTGTGTGCGTATGTGTGGATAATTGTGTTTGTGTTT ATTCTTATACGTACCACACCATTGGAGGTGTTTTTTGTA TATTTACTATTTTTTTGTATTGCCCATGTGGTTGTTGCATA GATGGCAATGGATATGATATAG	ATGATTCTGCTGGTCTTTCTGGTCTGGTTCGGCGTGTG TATCTATATCTGTTGTAATGTCCCTCTGCTGCCATCCGT CCATGTGTGTGCCTACGTGTGGATCATTGTGTTCGTCT TTATCCTGATTCGGACCACACCCCTGGAGGTGTTCTTT GTCTATCTGCTGTTCTTTGTCCTGCCTATGTGGCTGCT GCACAGACTGGCTATGGACATGATC	MILLVFLVWFGVCIYICC NVPLLPSVHVCAYVWII VFVFILIRTTPLEVFFVYL LFFVLPMWLLHRLAMD MI	21.9%

HPV39 E6	ATGGCGCGATTTCACAATCCTGCAGAACGGCCATACAA ATTGCCAGACCTGTGCACAACGCTGGACACCACCTTGC AGGACATTACAATAGCCTGTGTCTATTGCAGACGACCA CTACAGCAAACCGAGGTATATGAATTTGCATTTAGTGAT TTATATGTAGTATATAGGGACGGGGAACCACTAGCTGC ATGCCAATCATGTATAAAATTTTATGCTAAAATACGGGA GCTACGATATTACTCGGACTCGGTGTATGCAACTACAT TAGAAAATATAACTAATACAAAGTTATATAATTTATTAAT AAGGTGCATGTGTTGTCTGAAACCGCTGTGTCCAGCAG AAAAATTAAGACACCTAAATAGCAAACGAAGATTTCATA AAATAGCAGGAAGCTATACAGGACAGTGTCGACGGTG CTGGACCACAAAACGGGAGGACCGCAGACTAACACGA AGAGAAACCCAAGTATAA	ATGGCTAGATTTCATAACCCCGCCGAACGACCTTACAA ACTGCCCGACCTGTGCACTACACTGGATACAACTCTGC AGGACATCACCATCGCCTGCGTGTACTGTCGGAGACC ACTGCAGCAGACTGAGGTCTATGAATTCGCTTTTAGCG ACCTGTACGTGGTCTATAGGGATGGAGAGCCACTGGC AGCTTGCCAGTCTTGTATCAAGTTCTACGCAAAAATTAG GGAGCTGCGCTACTATAGCGACTCCGTGTACGCCACC ACACTGGAAAACATCACAAATACTAAGCTGTATAACCTG CTGATTCGGTGCATGTGCTGTCTGAAGCCCCTGTGTCC TGCTGAAAAACTGAGACACCTGAATTCTAAGAGGCGCT TTCATAAAATTGCAGGCAGTTATACAGGGCAGTGCCGA CGGTGTTGGACTACCAAACGGGAGGATAGAAGGCTGA CCCGCCGAGAAACACAGGTC	MARFHNPAERPYKLPD LCTTLDTTLQDITIACVY CRRPLQQTEVYEFAFS DLYVVYRDGEPLAACQ SCIKFYAKIRELRYYSDS VYATTLENITNTKLYNLL IRCMCCLKPLCPAEKLR HLNSKRRFHKIAGSYTG QCRRCWTTKREDRRLT RRETQV	24.8%
HPV39 E7	ATGCGTGGACCAAAGCCCACCTTGCAGGAAATTGTATT AGATTTATGTCCTTACAATGAAATACAGCCGGTTGACCT TGTATGTCACGAGCAATTAGGAGAGTCAGAGGATGAAA TAGATGAACCCGACCATGCAGTTAATCACCAACATCAA CTACTAGCCAGACGGGATGAACCACAGCGTCACACAAT ACAGTGTTCGTGTTGTAAGTGTAACAACACACTGCAGC TGGTAGTAGAAGCCTCACGGGATACTCTGCGACAACTA CAGCAGCTGTTTATGGACTCACTAGGATTTGTGTGTCC GTGTGTGCAACTGCAAACCAGTAA	ATGCGAGGCCCAAAGCCAACTCTGCAGGAAATTGTGCT GGACCTGTGCCCCTATAATGAGATTCAGCCCGTGGAC CTGGTGTGTCACGAACAGCTGGGCGAGTCTGAAGACG AGATCGATGAGCCCGACCACGCAGTGAACCACCAGCA TCAGCTGCTGGCCCGGAGAGATGAACCTCAGCGACAT ACCATTCAGTGCAGCTGCTGTAAGTGTAACAATACACT GCAGCTGGTGGTCGAGGCTAGCAGGGATACTCTGCGC CAGCTGCAGCAGCTGTTCATGGACTCCCTGGGGTTTGT CTGCCCCTGGTGTGCCACCGCTAATCAG	MRGPKPTLQEIVLDLCP YNEIQPVDLVCHEQLGE SEDEIDEPDHAVNHQH QLLARRDEPQRHTIQCS CCKCNNTLQLVVEASR DTLRQLQQLFMDSLGF VCPWCATANQ	23.5%
HPV39 L1	ATGGCTATGTGGCGGTCTAGTGACAGCATGGTGTATTT GCCTCCACCTTCTGTGGCGAAGGTTGTCAATACTGATG ATTATGTTACACGCACAGGCATATATTATTATGCTGGCA GCTCTAGATTATTAACAGTAGGACATCCATATTTTAAAG TGGGTATGAATGGTGGTCGCAAGCAGGACATTCCAAA GGTGTCTGCATATCAATATAGGGTATTTCGCGTGACAT TGCCCGATCCTAATAAATTCAGTATTCCAGATGCATCCT TATATAATCCAGAAACACAACGTTTAGTATGGGCTTGTG TAGGGGTGGAGGTGGGCAGGGGCCAGCCATTGGGTG TTGGTATTAGTGGACACCCATTATATAATAGACAGGATG ATACTGAAAACTCACCATTTTCATCAACCACCAATAAGG ACAGTAGGGATAATGTGTCTGTGGATTATAAACAGACA CAGTTGTGCATTATAGGCTGTGTTCCCGCCATTGGGGA GCACTGGGGTAAGGGAAAGGCATGCAAGCCCAATAAT GTATCTACGGGGGACTGTCCTCCTTTGGAACTAGTAAA CACCCCTATTGAGGATGGTGATATGATTGATACTGGCT ATGGAGCTATGGACTTTGGTGCATTGCAGGAAACCAAA AGTGAGGTGCCTTTAGATATTTGTCAATCCATTTGTAAA TATCCTGATTATTTGCAAATGTCTGCAGATGTGTATGGG GACAGTATGTTCTTCTGTTTACGTAGGGAACAACTGTTT GCAAGACATTTTTGGAATCGTGGTGGTATGGTGGGTGA CGCCATTCCTGCCCAATTGTATATTAAGGGCACAGATA TACGTGCAAACCCCGGTAGTTCTGTATACTGCCCCTCT CCCAGCGGTTCCATGGTAACCTCTGATTCCCAGTTATT TAATAAGCCTTATTGGCTACATAAGGCCCAGGGCCACA ACAATGGTATATGTTGGCATAATCAATTATTTCTTACTG TTGTGGACACTACCCGTAGTACCAACTTTACATTATCTA CCTCTATAGAGTCTTCCATACCTTCTACATATGATCCTT CTAAGTTTAAGGAATATACCAGGCACGTGGAGGAGTAT GATTTACAATTTATATTTCAACTGTGTACTGTCACATTAA CAACTGATGTTATGTCTTATATTCACACTATGAATTCCT CTATATTGGACAATTGGAATTTTGCTGTAGCTCCTCCAC CATCTGCCAGTTTGGTAGACACTTACAGATACCTACAG TCTGCAGCCATTACATGTCAAAAGGATGCTCCAGCACC TGAAAAGAAAGATCCATATGACGGTCTAAAGTTTTGGA ATGTTGACTTAAGGGAAAAGTTTAGTTTGGAACTTGATC AATTCCCTTTGGGACGTAAATTTTTGTTGCAGGCCAGG GTCCGCAGGCGCCCTACTATAGGTCCCCGAAAGCGGC CTGCTGCATCCACTTCCTCGTCCTCAGCTACTAAACAC AAACGTAAACGTGTGTCTAAATAA	ATGGCTATGTGGCGAAGCTCCGATTCTATGGTCTATCT GCCCCCCCCCTCCGTGGCTAAAGTCGTGAACACCGAT GATTATGTGACACGGACAGGAATCTACTATTACGCAGG CAGCTCCAGACTGCTGACTGTCGGCCACCCATATTTCA AAGTGGGGATGAATGGCGGGAGAAAACAGGACATCCC CAAGGTGAGCGCTTATCAGTACCGAGTCTTCCGGGTG ACCCTGCCAGACCCCAACAAGTTTAGTATTCCTGATGC ATCACTGTACAATCCAGAGACACAGAGGCTGGTCTGG GCATGCGTGGGAGTCGAAGTGGGACGAGGACAGCCA CTGGGAGTGGGAATCAGTGGACACCCTCTGTATAACC GACAGGACGATACAGAGAATTCTCCCTTCTCTAGTACC ACAAACAAAGACAGCCGGGATAATGTCTCCGTGGACTA CAAGCAGACTCAGCTGTGCATCATTGGGTGTGTGCCTG CCATTGGAGAACATTGGGGAAAGGGCAAAGCTTGCAA GCCAAACAATGTCAGCACAGGCGATTGTCCCCCTCTG GAGCTGGTGAACACACCTATCGAAGACGGGGATATGA TTGACACTGGGTATGGAGCTATGGATTTTGGAGCACTG CAGGAGACCAAATCCGAAGTCCCACTGGACATCTGCC AGTCTATTTGTAAGTATCCCGATTACCTGCAGATGTCAG CTGACGTGTACGGGGATAGCATGTTCTTTTGTCTGCGG AGAGAGCAGCTGTTCGCAAGACACTTTTGGAACAGGG GAGGAATGGTGGGCGACGCAATCCCAGCACAGCTGTA TATCAAGGGAACTGATATTAGGGCCAATCCTGGCTCAA GCGTCTACTGCCCTTCACCAAGCGGCTCCATGGTGAC CTCTGACAGTCAGCTGTTTAACAAACCCTACTGGCTGC ACAAGGCCCAGGGCCATAACAATGGGATTTGTTGGCAT AACCAGCTGTTCCTGACAGTGGTCGATACTACCAGAAG CACTAATTTTACCCTGTCAACAAGCATCGAGTCCTCTAT TCCATCTACCTATGACCCCAGTAAGTTCAAAGAATATAC AAGGCACGTGGAGGAATACGATCTGCAGTTCATCTTTC AGCTGTGCACCGTCACACTGACAACTGACGTGATGTCC TACATCCATACCATGAACAGTTCAATCCTGGATAACTG GAACTTCGCTGTCGCACCACCCCCTTCCGCCTCTCTGG TGGACACTTATCGCTACCTGCAGTCCGCCGCTATTACC TGTCAGAAAGATGCCCCCGCTCCTGAGAAGAAAGACC CTTACGATGGCCTGAAATTCTGGAACGTGGACCTGCG GGAGAAGTTTTCTCTGGAACTGGATCAGTTCCCACTGG GACGCAAATTTCTGCTGCAGGCACGAGTGAGGCGACG ACCAACCATCGGACCACGAAAGAGACCTGCAGCAAGC ACTAGCTCCTCTAGTGCTACCAAGCATAAAAGGAAGCG CGTGAGCAAG	MAMWRSSDSMVYLPP PSVAKVVNTDDYVTRT GIYYYAGSSRLLTVGHP YFKVGMNGGRKQDIPK VSAYQYRVFRVTLPDP NKFSIPDASLYNPETQR LVWACVGVEVGRGQPL GVGISGHPLYNRQDDT ENSPFSSTTNKDSRDN VSVDYKQTQLCIIGCVP AIGEHWGKGKACKPNN VSTGDCPPLELVNTPIE DGDMIDTGYGAMDFGA LQETKSEVPLDICQSICK YPDYLQMSADVYGDSM FFCLRREQLFARHFWN RGGMVGDAIPAQLYIKG TDIRANPGSSVYCPSPS GSMVTSDSQLFNKPYW LHKAQGHNNGICWHNQ LFLTVVDTTRSTNFTLS TSIESSIPSTYDPSKFKE YTRHVEEYDLQFIFQLC TVTLTTDVMSYIHTMNS SILDNWNFAVAPPPSAS LVDTYRYLQSAAITCQK DAPAPEKKDPYDGLKF WNVDLREKFSLELDQF PLGRKFLLQARVRRRP TIGPRKRPAASTSSSSA TKHKRKRVSK	26.6%
HPV39 L2	ATGGTTTCCCACCGTGCTGCCAGGCGTAAGCGTGCAT CTGCAACTGACCTATATAGAACCTGTAAACAATCGGGT ACCTGTCCACCAGACGTTGTTGATAAAGTTGAGGGTAC TACACTTGCTGACAAAATTTTACAGTGGACTAGTTTAGG TATATTTTTGGGTGGGTTAGGCATAGGCACAGGTACTG GTACTGGGGGACGCACAGGATATATACCCCTGGGGGG TAGGCCTAATACTGTTGTAGATGTGTCTCCTGCACGTC CACCTGTAGTTATTGAACCTGTTGGTCCTTCTGAGCCA TCTATTGTGCAATTGGTGGAGGACTCAAGTGTTATAAC CTCTGGAACACCAGTACCAACATTTACAGGCACCTCTG GATTTGAAATTACTTCTTCTTCTACTACTACGCCTGCGG TATTGGATATTACACCCTCCTCTGGGTCTGTACAAATAA CCTCTACTAGTTATACTAACCCTGCCTTTACGGATCCTT CCTTAATTGAGGTTCCCCAAACAGGTGAAACCTCGGGT AATATATTTGTCAGTACCCCTACATCAGGTACACATGGC TATGAGGAAATACCTATGGAAGTGTTGCCACACATGG CACAGGTACCGAACCTATTAGCAGCACACCTACACCTG GAATCAGTCGTGTGGCAGGACCACGTTTATATAGTAGA GCACATCAGCAGGTTCGTGTTAGTAATTTTGATTTTGTA ACTCACCCTTCATCATTTGTAACATTTGATAATCCTGCT TTTGAGCCTGTTGATACTACATTAACATATGAAGCTGCT GACATAGCTCCAGATCCGGATTTTCTGGACATTGTTCG TTTACATAGGCCTGCCTTAACCTCGCGTAAAGGAACAG TAAGGTTTAGTAGGCTTGGCAAAAAGGCTACCATGGTT ACCCGGCGTGGCACACAAATTGGAGCGCAAGTACATT ATTACCATGACATTAGTAGTATTGCTCCTGCTGAAAGCA TTGAATTACAGCCCCTAGTTCACGCTGAGCCCTCTGAT GCTTCAGATGCATTATTTGATATATATGCTGATGTGGAC AATAACACATATTTAGATACTGCATTTAATAATACAAGG GATTCGGGCACTACATATAACACAGGCTCACTACCTTC TGTGGCTTCTTCAGCATCTACTAAATATGCCAATACAAC TATTCCTTTTAGTACCTCATGGAATATGCCTGTAAATAC TGGTCCTGATATTGCTTTACCAAGTACTACTCCACAGTT GCCATTGGTGCCTTCTGGACCAATAGACACAACATATG CAATAACCATTCAGGGTTCCAATTATTATTTGTTGCCAT TATTGTATTTTTTCCTAAAAAAACGTAAACGTATTCCCTA TTTTTTTTCAGATGGCTATGTGGCGGTCTAG	ATGGTCTCCCACCGGGCAGCAAGAAGGAAACGGGCAT CAGCAACCGACCTGTATCGCACCTGTAAGCAGAGCGG AACCTGTCCCCCCGACGTGGTCGATAAGGTGGAGGGG ACCACACTGGCTGACAAAATTCTGCAGTGGACCTCTCT GGGAATCTTCCTGGGAGGACTGGGAATTGGAACCGGA ACAGGCACTGGAGGCCGGACTGGCTATATCCCACTGG GGGGAAGACCCAACACCGTGGTCGACGTGTCCCCAGC AAGACCACCTGTGGTCATCGAGCCAGTCGGACCATCA GAACCTAGCATCGTGCAGCTGGTCGAGGATAGCTCCG TGATCACTTCAGGGACCCCAGTCCCCACCTTCACAGGC ACTAGCGGGTTTGAAATCACATCTAGTTCAACTACCAC ACCAGCTGTGCTGGACATCACCCCCAGCTCCGGCAGC GTCCAGATTACCAGCACATCCTACACAAACCCTGCATT CACTGATCCATCCCTGATTGAGGTGCCTCAGACAGGG GAAACTTCTGGAAATATCTTTGTCAGCACTCCTACCTCC GGAACACACGGCTATGAGGAAATCCCAATGGAGGTGT TCGCCACCCATGGGACAGGAACTGAACCAATCTCTAGT ACCCCTACACCAGGAATTTCAAGGGTGGCAGGACCAC GACTGTACAGCCGAGCACACCAGCAGGTGCGCGTCTC CAACTTCGATTTTGTGACTCATCCCTCAAGCTTCGTCAC CTTTGACAATCCCGCCTTTGAGCCTGTGGATACTACCC TGACATACGAAGCCGCTGACATCGCACCCGACCCTGA TTTCCTGGATATTGTGCGACTGCACCGACCTGCACTGA CCTCTCGAAAGGGCACCGTGCGGTTCAGCAGGCTGGG GAAGAAAGCCACCATGGTGACACGGAGAGGCACCCAG ATTGGGGCTCAGGTCCACTACTATCATGACATCTCCTC TATTGCACCCGCCGAGAGCATCGAACTGCAGCCACTG GTGCATGCAGAGCCTTCCGACGCTTCTGATGCACTGTT CGATATCTACGCTGACGTGGATAACAATACTTATCTGG ACACCGCCTTCAACAATACCCGCGATAGTGGGACAACT TACAACACAGGAAGTCTGCCTTCAGTGGCCAGTTCAGC TAGCACCAAGTATGCTAATACCACAATTCCATTTTCTAC AAGTTGGAACATGCCCGTGAATACTGGCCCTGACATCG CACTGCCATCTACTACCCCACAGCTGCCCCTGGTGCCT AGTGGACCAATTGATACAACTTACGCCATCACCATTCA GGGCAGCAATTACTATCTGCTGCCCCTGCTGTATTTCT TTCTGAAGAAAAGGAAACGCATCCCTTACTTCTTTTCCG ACGGCTATGTGGCCGTC	MVSHRAARRKRASATD LYRTCKQSGTCPPDVV DKVEGTTLADKILQWTS LGIFLGGLGIGTGTGTG GRTGYIPLGGRPNTVV DVSPARPPVVIEPVGPS EPSIVQLVEDSSVITSGT PVPTFTGTSGFEITSSS TTTPAVLDITPSSGSVQI TSTSYTNPAFTDPSLIEV PQTGETSGNIFVSTPTS GTHGYEEIPMEVFATH GTGTEPISSTPTPGISR VAGPRLYSRAHQQVRV SNFDFVTHPSSFVTFDN PAFEPVDTTLTYEAADI APDPDFLDIVRLHRPAL TSRKGTVRFSRLGKKA TMVTRRGTQIGAQVHY YHDISSIAPAESIELQPL VHAEPSDASDALFDIYA DVDNNTYLDTAFNNTR DSGTTYNTGSLPSVAS SASTKYANTTIPFSTSW NMPVNTGPDIALPSTTP QLPLVPSGPIDTTYAITI QGSNYYLLPLLYFFLKK RKRIPYFFSDGYVAV	27.6%

HPV45 E1	ATGGCGGATCCAGAAGGTACCGACGGGGAGGGAACG GGGTGTAATGGCTGGTTCTTTGTAGAAACAATTGTAGA GAAAAAAACAGGGGATGTAATATCAGATGATGAGGATG AAACTGCAACAGATACAGGGTCGGATATGGTAGATTTT ATTGACACACAATTATCCATTTGTGAACAGGCAGAGCA AGAGACAGCACAGGCATTGTTCCATGCGCAGGAAGTT CAGAATGATGCACAGGTGTTGCATCTTTTAAAACGAAA GTTTGCAGGAGGCAGCAAGGAAAACAGTCCATTAGGG GAGCAGCTAAGTGTGGATACGGATCTAAGTCCACGGTT ACAAGAAATTTCATTAAATAGTGGGCACAAAAAAGCAAA ACGACGGTTGTTTACAATATCAGATAGTGGCTATGGCT GTTCTGAAGTGGAAGCTGCAGAGACTCAGGTAACTGTA AACACTAATGCGGAAAATGGCGGCAGTGTACATAGTAC ACAAAGTAGTGGTGGGGATAGTAGTGACAATGCAGAAA ATGTAGATCCGCATTGCAGTATTACAGAACTAAAGGAG CTATTACAAGCAAGTAACAAAAAGGCTGCAATGCTGGC AGTATTTAAAGACATATATGGGCTGTCATTTACGGATTT GGTTAGAAATTTTAAAAGTGATAAAACAACATGTACAGA TTGGGTAATGGCTATATTTGGAGTTAATCCAACGGTAG CAGAAGGCTTTAAAACATTAATTAAACCAGCAACGTTAT ACGCCCATATCCAATGTTTAGATTGTAAATGGGGAGTA TTAATATTAGCTTTATTAAGATATAAATGTGGCAAAAATA GACTAACTGTTGCAAAAGGCTTAAGCACATTGTTGCAC GTACCTGAAACATGTATGTTAATTGAACCACCAAAATTG CGAAGTAGTGTTGCAGCATTATACTGGTATAGAACAGG TATATCCAATATTAGTGAAGTAAGTGGAGACACACCTG AGTGGATACAAAGACTGACAATTATTCAACATGGTATTG ACGATAGTAATTTTGATTTGTCAGACATGGTGCAATGG GCATTTGATAATGACCTTACAGATGAAAGTGATATGGC ATTTCAATATGCCCAATTAGCAGACTGCAACAGTAATGC AGCTGCATTTTTAAAAAGTAACTGCCAAGCCAAATATTT AAAAGATTGTGCTGTAATGTGTAGACATTATAAAAGAGC ACAAAAACGCCAAATGAATATGTCTCAATGGATTAAATA TAGATGTTCCAAAATAGATGAAGGTGGGGATTGGAGAC CCATAGTACAATTCCTAAGATATCAGGGAGTAGAATTTA TTAGCTTTTTAAGGGCACTAAAGGAATTTCTTAAAGGAA CACCAAAAAAAAAATTGTATACTGTTATATGGACCTGCAA ATACAGGAAAATCGTATTTTGGAATGAGTTTTATACATT TCCTACAAGGTGCAATAATATCATTTGTAAATTCAAACA GCCATTTTTGGTTAGAACCGTTAGCAGATACTAAGGTA GCCATGTTGGATGATGCCACACACACGTGTTGGACATA TTTTGATAATTATATGAGAAATGCATTAGATGGTAATCC TATAAGTATAGACAGAAAGCATAAACCATTATTACAGCT AAAATGTCCTCCAATCCTATTAACATCCAATATTGATCC AGCAAAAGATAATAAATGGCCATATTTAGAAAGTAGGG TGACGGTATTTACATTTCCACATGCATTTCCATTTGATA AAAATGGTAATCCAGTATATGAAATAAATGATAAAAATT GGAAATGTTTTTTTGAAAGGACATGGTCCAGATTAGATT TGCACGAGGACGATGAAGATGCAGACACCGAAGGAAT CCCTTTCGGAACGTTTAAGTGCGTTACAGGACAAAATA CTAGACCACTATGA	ATGGCAGATCCTGAGGGGACTGACGGGGAGGGGACT GGGTGTAATGGGTGGTTTTTCGTGGAGACTATCGTGGA GAAGAAGACTGGGGATGTGATCTCCGACGATGAGGAT GAAACTGCCACCGACACAGGCTCCGACATGGTCGATTT CATCGACACTCAGCTGTCTATTTGCGAGCAGGCTGAGC AGGAAACCGCCCAGGCTCTGTTCCACGCCCAGGAAGT GCAGAACGACGCTCAGGTCCTGCATCTGCTGAAGCGC AAATTTGCCGGCGGGTCTAAGGAGAATAGTCCACTGG GAGAACAGCTGAGTGTGGACACAGATCTGTCACCCCG GCTGCAGGAGATCAGCCTGAACTCCGGCCACAAGAAA GCCAAACGGAGACTGTTCACCATTTCCGATTCTGGATA CGGCTGCAGCGAGGTGGAAGCCGCTGAGACACAAGTG ACTGTCAACACCAATGCAGAAAATGGAGGCTCTGTGCA CAGTACACAGAGCTCCGGGGGAGATTCTAGTGACAAC GCCGAAAATGTCGACCCTCATTGTAGTATCACTGAGCT GAAGGAACTGCTGCAGGCTTCAAACAAGAAAGCAGCC ATGCTGGCAGTGTTTAAAGATATCTACGGCCTGTCATT CACCGACCTGGTCCGGAACTTTAAGAGCGATAAAACCA CATGTACAGACTGGGTCATGGCCATCTTCGGCGTGAAT CCCACTGTCGCTGAGGGGTTTAAGACTCTGATCAAACC TGCTACCCTGTACGCACACATTCAGTGCCTGGATTGTA AGTGGGGCGTGCTGATCCTGGCTCTGCTGAGGTATAA GTGCGGAAAAAATCGCCTGACAGTGGCAAAGGGCCTG TCCACTCTGCTGCATGTCCCCGAGACCTGTATGCTGAT CGAACCCCCTAAACTGAGATCAAGCGTGGCTGCACTGT ACTGGTATAGGACAGGGATCTCAAACATTAGCGAGGTC TCCGGAGACACCCCTGAATGGATCCAGAGACTGACAAT CATTCAGCACGGCATTGACGATTCAAACTTCGATCTGA GCGACATGGTGCAGTGGGCATTTGACAATGATCTGACC GATGAGTCTGACATGGCCTTCCAGTACGCACAGCTGG CCGATTGCAACAGCAATGCCGCTGCATTTCTGAAGTCC AACTGTCAGGCAAAGTACCTGAAAGACTGCGCCGTGAT GTGTAGGCATTATAAGCGCGCCCAGAAACGACAGATG AATATGTCTCAGTGGATCAAGTACAGATGCAGTAAAATT GATGAGGGCGGGGACTGGCGACCAATCGTGCAGTTCC TGCGGTATCAGGGCGTCGAGTTCATTTCTTTTCTGAGG GCCCTGAAGGAATTTCTGAAAGGGACCCCTAAGAAAAA CTGTATCCTGCTGTACGGCCCAGCCAATACCGGGAAG TCCTATTTCGGAATGTCTTTCATTCACTTTCTGCAGGGG GCTATCATCAGTTTCGTGAACAGTAACTCACATTTCTGG CTGGAGCCCCTGGCCGATACTAAAGTCGCTATGCTGG ACGATGCAACTCACACCTGCTGGACCTACTTTGACAAC TATATGCGCAATGCCCTGGATGGAAATCCAATCAGCAT TGACCGAAAGCATAAACCCCTGCTGCAGCTGAAGTGTC CACCCATCCTGCTGACCAGCAACATTGATCCCGCTAAG GACAACAAGTGGCCTTACCTGGAGTCCCGCGTGACAG TCTTCACTTTTCCTCACGCATTCCCATTTGATAAGAACG GCAATCCCGTGTATGAGATCAACGACAAGAATTGGAAA TGCTTCTTTGAACGGACCTGGAGCAGACTGGACCTGCA TGAGGACGATGAAGACGCCGATACAGAAGGGATTCCT TTCGGAACTTTTAAGTGTGTGACCGGGCAGAACACAAG GCCACTG	MADPEGTDGEGTGCN GWFFVETIVEKKTGDVI SDDEDETATDTGSDMV DFIDTQLSICEQAEQET AQALFHAQEVQNDAQV LHLLKRKFAGGSKENSP LGEQLSVDTDLSPRLQE ISLNSGHKKAKRRLFTIS DSGYGCSEVEAAETQV TVNTNAENGGSVHSTQ SSGGDSSDNAENVDPH CSITELKELLQASNKKA AMLAVFKDIYGLSFTDL VRNFKSDKTTCTDWVM AIFGVNPTVAEGFKTLIK PATLYAHIQCLDCKWG VLILALLRYKCGKNRLTV AKGLSTLLHVPETCMLI EPPKLRSSVAALYWYR TGISNISEVSGDTPEWI QRLTIIQHGIDDSNFDLS DMVQWAFDNDLTDESD MAFQYAQLADCNSNAA AFLKSNCQAKYLKDCA VMCRHYKRAQKRQMN MSQWIKYRCSKIDEGG DWRPIVQFLRYQGVEFI SFLRALKEFLKGTPKKN CILLYGPANTGKSYFGM SFIHFLQGAIISFVNSNS HFWLEPLADTKVAMLD DATHTCWTYFDNYMRN ALDGNPISIDRKHKPLL QLKCPPILLTSNIDPAKD NKWPYLESRVTVFTFP HAFPFDKNGNPVYEIND KNWKCFFERTWSRLDL HEDDEDADTEGIPFGTF KCVTGQNTRPL	26.4%
HPV45 E2	ATGAAGATGCAGACACCGAAGGAATCCCTTTCGGAACG TTTAAGTGCGTTACAGGACAAAATACTAGACCACTATGA AAATGACAGTAAAGACATAAACAGCCAAATAAGTTATTG GCAACTTATACGTTTGGAAAATGCAATACTATTTACAGC AAGGGAACATGGTATTACCAAACTAAACCACCAGGTGG TGCCTCCTATTAACATTTCAAAAGCAAAGCACATAAAG CTATTGAACTGCAAATGGCCTTAAAGGGCCTTGCACAA AGCAAGTATAACAATGAGGAATGGACACTGCAAGATAC ATGCGAGGAACTATGGAATACAGAACCGTCGCAGTGTT TTAAAAAAGGCGGTAAAACCGTGCACGTATACTTTGAT GGCAACAAGGACAACTGTATGAACTATGTAGTATGGGA CAGTATATATTATATAACTGAGACAGGGATATGGGACA AAACAGCAGCATGTGTTAGCTATTGGGGTGTATATTATA TAAAAGATGGAGATACCACATATTATGTACAATTTAAAA GCGAATGTGAGAAATATGGAAATAGTAATACGTGGGAA GTACAATATGGGGGCAATGTAATTGATTGTAATGACTCT ATGTGCAGTACCAGTGACGACACGGTATCCGCTACTCA GATTGTTAGACAGCTACAACACGCCTCCACGTCGACCC CCAAAACCGCATCCGTGGGCACCCCAAAACCCCACAT CCAGACGCCGGCTACTAAGCGACCTAGACAGTGTGGA CTCACAGAGCAGCACCACCACGGACGTGTCAACACCCACG TGCACAACCCGCTCCTGTGTTCAAGTACAAGTAACAAC AAAAGAAGGAAAGTGTGTAGTGGTAACACTACGCCTAT AATACACTTAAAAGGTGACAAAAACAGTTTGAAATGTTT AAGATATAGGCTACGCAAATATGCAGACCATTACTCAG AAATATCCTCCACCTGGCATTGGACAGGTTGTAATAAA AACACTGGTATATTAACTGTAACATATAATAGTGAGGTA CAAAGAAATACCTTTTTGGATGTAGTTACTATTCCTAAC AGTGTACAAATCTCGGTGGGATACATGACTATATGA	ATGAAGATGCAGACTCCCAAGGAAAGCCTGAGCGAGA GACTGAGCGCACTGCAGGATAAGATTCTGGACCATTAC GAAAACGACTCCAAGGACATCAATAGCCAGATTTCCTA CTGGCAGCTGATCAGGCTGGAGAACGCTATTCTGTTCA CTGCACGCGAACACGGAATCACCAAGCTGAATCATCAG GTGGTCCCCCCTATCAACATTTCAAAGAGCAAAGCACA CAAAGCCATTGAGCTGCAGATGGCCCTGAAGGGCCTG GCTCAGAGCAAATATAACAATGAGGAATGGACCCTGCA GGATACATGCGAGGAACTGTGGAATACCGAACCATCC CAGTGTTTCAAGAAAGGCGGGAAGACAGTGCATGTCTA CTTTGACGGCAACAAAGATAATTGCATGAACTATGTGG TCTGGGACTCTATCTACTATATTACAGAGACTGGGATCT GGGATAAGACAGCCGCTTGTGTGAGTTACTGGGGCGT CTACTATATCAAGGACGGGGATACCACATACTATGTGC AGTTTAAGTCAGAGTGCGAAAAATACGGGAACAGCAAT ACCTGGGAAGTGCAGTATGGAGGCAATGTCATCGACT GCAACGATAGTATGTGTTCCACTTCTGACGATACCGTG TCAGCTACACAGATTGTGCGACAGCTGCAGCACGCAA GTACCTCAACACCCAAGACAGCCTCTGTGGGCACTCCA AAACCCCATATCCAGACACCTGCCACTAAGAGGCCACG CCAGTGTGGACTGACAGAGCAGCACCATGGCAGGGTG AATACTCACGTCCATAACCCCCTGCTGTGCAGCTCCAC CTCCAACAATAAGCGGAGAAAAGTGTGTTCTGGGAATA CTACCCCTATCATTCACCTGAAGGGAGACAAAAACAGC CTGAAGTGTCTGCGATACCGGCTGAGAAAATACGCCG ATCACTATTCCGAGATCTCTAGTACTTGGCATTGGACC GGGTGCAACAAGAATACTGGAATTCTGACTGTGACCTA CAATAGCGAAGTGCAGCGGAACACCTTCCTGGACGTG GTCACAATCCCTAACTCTGTGCAGATTAGTGTCGGCTA TATGACCATC	MKMQTPKESLSERLSA LQDKILDHYENDSKDIN SQISYWQLIRLENAILFT AREHGITKLNHQVVPPI NISKSKAHKAIELQMAL KGLAQSKYNNEEWTLQ DTCEELWNTEPSQCFK KGGKTVHVYFDGNKDN CMNYVVWDSIYYITETG IWDKTAACVSYWGVYYI KDGDTTYYVQFKSECE KYGNSNTWEVQYGGN VIDCNDSMCSTSDDTV SATQIVRQLQHASTSTP KTASVGTPKPHIQTPAT KRPRQCGLTEQHHGRV NTHVHNPLLCSSTSNN KRRKVCSGNTTPIIHLK GDKNSLKCLRYRLRKY ADHYSEISSTWHWTGC NKNTGILTVTYNSEVQR NTFLDVVTIPNSVQISVG YMTI	25.4%
HPV45 E4	ATGACTCTATGTGCAGTACCAGTGACGACACGGTATCC GCTACTCAGATTGTTAGACAGCTACAACACGCCTCCAC GTCGACCCCCAAAACCGCATCCGTGGGCACCCCAAAA CCCCACATCCAGACGCCGGCTACTAAGCGACCTAGAC AGTGTGGACTCACAGAGCAGCACCACGGACGTGTCAA CACCCACGTGCACAACCCGCTCCTGTGTTCAAGTACAA GTAACAACAAAAGAAGGAAAGTGTGTAGTGGTAACACT ACGCCTATAA	ATGACCCTGTGTGCCGTCCCCGTGACCACCCGCTACC CACTGCTGCGACTGCTGGATAGTTACAATACCCCACCC CGAAGACCCCCTAAGCCACACCCTTGGGCACCACAGA ACCCAACTTCACGGAGAAGGCTGCTGAGCGACCTGGA TTCTGTGGACAGTCAGAGCTCCACCACAGATGTGTCCA CCCCCACATGCACTACCCGCAGTTGTGTCCAGGTGCA GGTCACAACTAAGGAGGGCAAATGCGTGGTCGTGACA CTGCGACTG	MTLCAVPVTTRYPLLRL LDSYNTPPRRPPKPHP WAPQNPTSRRRLLSDL DSVDSQSSTTDVSTPT CTTRSCVQVQVTTKEG KCVVVTLRL	28.5%
HPV45 E5	ATGCTATCTTTAGTGTTTTTATTGTGCTTTTCTGTGTGC CTTTATGTGTGCTGCAATGTCCCGCTTGTGCAGTCTGT CTATGTGTGTGCTTTTGCTTGGTTGTTGGTGTTTCTTTT TATAGTTGTTATTACATCCCCATTAACAGCATTTGCTGT ATACATTTGTTGCTATTTACTACCTATGTTTGTATTACAT ATGCATGCTTTACACACCATACAATAA	ATGCTGAGTCTGGTCTTTCTGCTGTGTTTTTCCGTCTGT CTGTATGTGTGTTGTAATGTCCCCCTGGTGCAGTCCGT CTATGTCTGTGCCTTCGCTTGGCTGCTGGTGTTCCTGT TTATCGTGGTCATTACCAGCCCCCTGACAGCATTCGCC GTGTACATCTGCTGTTATCTGCTGCCTATGTTTGTCCTG CACATGCATGCTCTGCACACCATTCAG	MLSLVFLLCFSVCLYVC CNVPLVQSVYVCAFAW LLVFLFIVVITSPLTAFAV YICCYLLPMFVLHMHAL HTIQ	24.2%

HPV45 E6	ATGGCGCGCTTTGACGATCCAAAGCAACGACCCTACAA GCTACCAGATTTGTGCACAGAATTGAATACATCACTACA AGACGTATCTATTGCCTGTGTATATTGCAAAGCAACATT GGAACGCACAGAGGTATATCAATTTGCTTTTAAAGATTT ATGTATAGTGTATAGAGACTGTATAGCATATGCTGCATG CCATAAATGTATAGACTTTTATTCCAGAATTAGAGAATT AAGATATTATTCAAACTCTGTATATGGAGAGACACTGGA AAAAATAACTAATACAGAGTTGTATAATTTGTTAATAAG GTGCCTGCGGTGCCAGAAACCATTGAACCCAGCAGAA AAACGTAGACACCTTAAGGACAAACGAAGATTTCACAG CATAGCTGGACAGTACCGAGGGCAGTGTAATACATGTT GTGACCAGGCACGGCAAGAAAGACTTCGCAGACGTAG GGAAACACAAGTATAG	ATGGCCCGATTTGATGACCCTAAACAGCGCCCCTATAA ACTGCCCGATCTGTGTACCGAACTGAATACTTCTCTGC AGGATGTGAGCATCGCATGCGTGTACTGTAAGGCCACT CTGGAGCGAACCGAAGTCTATCAGTTCGCTTTTAAAGA CCTGTGCATCGTGTACCGCGATTGTATTGCATATGCCG CTTGCCACAAGTGTATCGACTTCTACTCTCGCATTCGA GAGCTGAGATACTATAGCAACTCCGTCTACGGCGAGAC CCTGGAAAAAATCACCAACACAGAGCTGTATAATCTGC TGATTCGGTGCCTGAGATGTCAGAAGCCCCTGAACCCT GCCGAAAAACGGAGACACCTGAAGGACAAAAGGCGCT TTCATAGCATTGCCGGCCAGTATCGGGGGCAGTGCAA TACATGCTGTGATCAGGCTAGGCAGGAGCGCCTGCGA CGGAGAAGGGAAACTCAGGTG	MARFDDPKQRPYKLPD LCTELNTSLQDVSIACV YCKATLERTEVYQFAFK DLCIVYRDCIAYAACHK CIDFYSRIRELRYYSNS VYGETLEKITNTELYNLL IRCLRCQKPLNPAEKRR HLKDKRRFHSIAGQYR GQCNTCCDQARQERL RRRRETQV	25.7%
HPV45 E7	ATGCATGGACCCCGGGAAACACTGCAAGAAATTGTATT GCATTTGGAACCTCAGAATGAATTAGATCCTGTTGACC TGTTGTGTTACGAGCAATTAAGCGAGTCAGAGGAGGAA AACGATGAAGCAGATGGAGTTAGTCATGCACAACTACC AGCCCGACGAGCCGAACCACAGCGTCACAAAATTTTGT GTGTATGTTGTAAGTGTGACGGCAGAATTGAGCTTACA GTAGAGAGCTCGGCAGAGGACCTTAGAACACTACAGC AGCTGTTTTTGAGCACCTTGTCCTTTGTGTGTCCGTGG TGTGCACTAACCAATAA	ATGCATGGACCAAGAGAAACCCTGCAGGAAATCGTGCT GCATCTGGAGCCTCAGAATGAACTGGACCCTGTGGAC CTGCTGTGCTATGAACAGCTGTCTGAGAGTGAGGAAGA GAACGACGAGGCAGATGGCGTGTCTCACGCACAGCTG CCAGCACGAAGAGCTGAACCTCAGAGGCATAAGATCC TGTGCGTGTGCTGTAAATGTGACGGGCGCATTGAACTG ACTGTCGAGAGCTCCGCTGAAGATCTGCGAACCCTGC AGCAGCTGTTCCTGTCAACACTGAGCTTTGTCTGCCCC TGGTGTGCCACCAATCAG	MHGPRETLQEIVLHLEP QNELDPVDLLCYEQLSE SEEENDEADGVSHAQL PARRAEPQRHKILCVCC KCDGRIELTVESSAEDL RTLQQLFLSTLSFVCPW CATNQ	27.3%
HPV45 L1	ATGGCTTTGTGGCGGCCTAGTGACAGTACGGTATATCT TCCACCACCTTCTGTGGCCAGAGTTGTCAGCACTGATG ATTATGTGTCTCGCACAAGCATATTTTATCATGCAGGCA GTTCCCGATTATTAACTGTAGGCAATCCATATTTTAGGG TTGTACCTAATGGTGCAGGTAATAAACAGGCTGTTCCT AAGGTATCCGCATATCAGTATAGGGTGTTTAGAGTAGC TTTACCCGATCCTAATAAATTTGGATTACCTGATTCTAC TATATATAATCCTGAAACACAACGTTTGGTTTGGGCATG TGTAGGTATGGAAATTGGTCGTGGGCAGCCTTTAGGTA TTGGCCTAAGTGGCCATCCATTTTATAATAAATTGGATG ATACAGAAAGTGCTCATGCAGCTACAGCTGTTATTACG CAGGATGTTAGGGATAATGTGTCAGTTGATTATAAGCA AACACAGCTGTGTATTTTAGGTTGTGTACCTGCTATTGG TGAGCACTGGGCCAAGGGCACACTTTGTAAACCTGCA CAATTGCAACCTGGTGACTGTCCTCCTTTGGAACTTAA AAACACCATTATTGAGGATGGTGATATGGTGGATACAG GTTATGGGGCAATGGATTTTAGTACATTGCAGGATACA AAGTGCGAGGTTCCATTAGACATTTGTCAATCCATCTGT AAATATCCAGATTATTTGCAAATGTCTGCTGATCCCTAT GGGGATTCTATGTTTTTTTGCCTACGCCGTGAACAACT GTTTGCAAGACATTTTTGGAATAGGGCAGGTGTTATGG GTGACACAGTACCTACGGACCTATATTAAAGGCACT AGCGCTAATATGCGTGAAACCCCTGGCAGTTGTGTGTA TTCCCCTTCTCCCAGTGGCTCTATTATTACTTCTGATTC TCAATTATTTAATAAGCCATATTGGTTACATAAGGCCCA GGGCCATAACAATGGTATTTGTTGGCATAATCAGTTGTT TGTTACTGTAGTGGACACTACCCGCAGTACTAATTTAAC ATTATGTGCCTCTACACAAAATCCTGTGCCAAGTACATA TGACCCTACTAAGTTTAAGCAGTATAGTAGACATGTGG AGGAATATGATTTACAGTTTATTTTTCAGTTGTGCACTA TTACTTTAACTGCAGAGGTTATGTCATATATCCATAGTA TGAATAGTAGTATATTAGAAAATTGGAATTTTGGTGTCC CTCCACCACCTACTACAAGTTTGGTGGATACATATCGTT TTGTGCAATCAGTTGCTGTTACCTGTCAAAAGGATACTA CACCTCCAGAAAAGCAGGATCCATATGATAAATTAAAG TTTTGGACTGTTGACCTAAAGGAAAAATTTTCCTCCGAT TTGGATCAATATCCCCTTGGTCGAAAGTTTTTAGTTCAG GCTGGGTTACGTCGTAGGCCTACCATAGGACCTCGTAA GCGTCCTGCTGCTTCCACGTCTACTGCATCTACTGCAT CTAGGCCTGCCAAACGTGTACGTATACGTAGTAAGAAA TAA	ATGGCTCACAATATTATCTACGGGCACGGGATTATCAT CTTCCTGAAAAATGTCAATGTCTTCCCTATCTTCCTGCA GATGGCACTGTGGAGGCCCTCCGACTCTACAGTGTAC CTGCCCCCTCCATCTGTCGCTCGAGTGGTCAGTACTGA CGATTACGTGTCCCGGACCTCTATCTTCTATCACGCAG GCAGCTCCAGACTGCTGACCGTGGGGAACCCTTATTTT AGGGTGGTCCCAAACGGCGCCGGGAATAAGCAGGCTG TGCCTAAAGTCAGCGCATACCAGTATCGGGTGTTCAGA GTCGCCCTGCCAGACCCCAACAAGTTTGGCCTGCCCG ATTCCACCATCTACAATCCTGAGACACAGCGACTGGTG TGGGCATGCGTCGGAATGGAAATCGGACGAGGCCAGC CACTGGGGATTGGACTGTCTGGCCACCCCTTCTACAAC AAACTGGACGATACCGAGAGTGCACATGCCGCTACCG CCGTGATCACACAGGACGTCCGCGATAATGTGTCCGT CGATTATAAGCAGACTCAGCTGTGCATCCTGGGCTGTG TGCCAGCTATTGGGGAACATTGGGCAAAGGGAACCCT GTGCAAACCAGCACAGCTGCAGCCAGGGGACTGTCCA CCTCTGGAGCTGAAGAATACCATCATTGAAGACGGCGA TATGGTGGATACAGGCTACGGGGCCATGGACTTTTCTA CTCTGCAGGATACCAAGTGCGAGGTGCCTCTGGACAT CTGCCAGAGCATTTGTAAATACCCTGATTATCTGCAGAT GTCAGCCGACCCATATGGCGATAGCATGTTCTTTTGTC TGCGGAGAGAGCAGCTGTTCGCCAGGCACTTTTGGAA CCGCGCTGGAGTGATGGGCGACACTGTCCCCACCGAT CTGTACATCAAGGGAACAAGTGCCAATATGCGGGAAAC TCCTGGCTCATGTGTGTATAGTCCTTCACCAAGCGGGT CCATCATTACATCTGACAGTCAGCTGTTCAACAAGCCTT ACTGGCTGCACAAAGCTCAGGGGCATAACAATGGAATT TGCTGGCATAATCAGCTGTTTGTGACAGTGGTCGATAC CACACGGTCTACAAACCTGACTCTGTGTGCATCCACTC AGAATCCCGTGCCTTCTACATACGACCCAACTAAGTTC AAACAGTACAGCAGACACGTCGAGGAATATGATCTGCA GTTCATCTTTCAGCTGTGCACCATTACACTGACTGCCG AAGTGATGAGCTACATCCATTCCATGAACTCTAGTATTC TGGAAAACTGGAATTTCGGCGTGCCACCCCCTCCAACT ACCAGCCTGGTGGACACTTATAGATTTGTCCAGTCCGT GGCTGTCACCTGTCAGAAGGATACAACTCCCCCTGAGA AACAGGACCCATACGATAAGCTGAAATTCTGGACCGTG GACCTGAAGGAAAAATTTTCAAGCGACCTGGATCAGTA TCCCCTGGGAAGGAAGTTTCTGGTGCAGGCAGGACTG AGGCGACGACCAACCATCGGGCCCCGGAAAAGACCTG CAGCCTCAACCAGCACAGCTAGTACAGCATCACGCCC CGCTAAGAGGGTGCGCATTCGAAGCAAGAAA	MALWRPSDSTVYLPPP SVARVVSTDDYVSRTSI FYHAGSSRLLTVGNPYF RVVPNGAGNKQAVPKV SAYQYRVFRVALPDPN KFGLPDSTIYNPETQRL VWACVGMEIGRGQPLG IGLSGHPFYNKLDDTES AHAATAVITQDVRDNVS VDYKQTQLCILGCVPAI GEHWAKGTLCKPAQLQ PGDCPPLELKNTIIEDG DMVDTGYGAMDFSTLQ DTKCEVPLDICQSICKY PDYLQMSADPYGDSMF FCLRREQLFARHFWNR AGVMGDTVPTDLYIKGT SANMRETPGSCVYSPS PSGSIITSDSQLFNKPY WLHKAQGHNNGICWH NQLFVTVVDTTRSTNLT LCASTQNPVPSTYDPTK FKQYSRHVEEYDLQFIF QLCTITLTAEVMSYIHS MNSSILENWNFGVPPP PTTSLVDTYRFVQSVAV TCQKDTTPPEKQDPYD KLKFWTVDLKEKFSSDL DQYPLGRKFLVQAGLR RRPTIGPRKRPAASTST ASTASRPAKRVRIRSKK	29.6%
HPV45 L2	ATGGTATCCCACCGTGCAGCACGTCGCAAGCGGGCCT CTGCAACTGACTTATATAGAACATGTAAGCAATCCGGT ACGTGCCCCCCTGATGTTATTAACAAAGTGGAAGGCAC AACCTTAGCTGATAAAATTTTACAGTGGTCTAGCCTTGG GATATTTTTGGGTGGCCTTGGCATTGGTACCGGCAGTG GTTCTGGAGGCCGTACGGGCTATGTACCCTTAGGGGG CAGGTCTAATACTGTTGTGGATGTTGGCCCCACTAGGC CACCTGTGGTTATTGAACCTGTAGGGCCTACTGATCCA TCTATTGTTACGTTGGTAGAGGATTCCAGTGTTGTTGC CTCTGGTGCTCCGGTTCCCACATTTACCGGAACCTCTG GGTTTGAAATTACGTCTTCTGGTACTACCACACCAGCT GTGTTGGACATCACACCTACCGTGGACTCTGTTTCTAT TTCGTCAACTAGTTTTACAAATCCTGCATTTTCTGATCC CTCTATTATTGAGGTGCCCCAAACAGGGGAGGTATCAG GTAATATATTTGTTGGTACACCAACATCGGGCAGCCAT GGATATGAGGAAATACCTTTACAAACATTTGCATCTTCT GGGTCAGGTACGGAACCCATTAGTAGTACCCCCCTCC CTACTGTGCGGCGGGTACGGGGTCCCCGCCTGTATAG TAGGGCTAATCAACAGGTCCGTGTGTCCACCTCACAGT TTTTAACACATCCCTCATCGTTGGTTACATTTGATAATC CAGCTTATGAGCCCCTGGACACCACACTATCCTTTGAG CCTACCAGTAATGTTCCTGATTCCGATTTTATGGATATT ATTCGTTTGCATAGGCCAGCATTATCCTCTAGACGTGG CACTGTTAGATTTAGTAGATTGGGTCAAAGGGCAACCA TGTTTACACGTAGTGGTAAACAAATAGGGGGTAGGGTA CATTTTTACCATGATATAAGCCCCATTGCTGCTACAGAG GAAATTGAATTGCAGCCTTTAATTAGTGCTACAAATGAT AGTGACCTGTTTGATGTATATGCAGACTTCCCACCTCC TGCGTCCACTACACCTAGCACTATACACAAATCATTTAC ATATCCAAAGTATTCCTTGACCATGCCTTCTACTGCTGC ATCCTCTTACAGTAATGTTACAGTACCATTAACATCTGC ATGGGATGTACCTATATATACTGGCCCGGACATTATATT GCCATCCCATACTCCTATGTGGCCTAGTACATCTCCTA CCAATGCTTCCACCACCACCTATATAGGTATTCATGGC ACACAATATTATTTATGGCCATGGTATTATTATTTTCCTA AAAAACGTAAACGTATTCCCTATTTTTTTGCAGATGGCT TTGTGGCGGCCTAG	ATGGTGTCACATAGAGCAGCCAGAAGAAAAAGGGCAT CAGCAACCGACCTGTATAGAACCTGTAAGCAGAGCGG AACTTGTCCTCCCGACGTGATCAACAAGGTCGAGGGAA CCACACTGGCCGATAAAATTCTGCAGTGGAGCTCCCTG GGAATCTTCCTGGGAGGACTGGGAATTGGAACCGGAA GTGGATCAGGAGGACGAACAGGATATGTCCCACTGGG AGGAAGAAGCAATACAGTGGTCGACGTGGGCCCTACT AGGCCACCTGTGGTCATCGAGCCAGTCGGACCTACCG ACCCATCCATTGTGACACTGGTCGAAGATTCTAGTGTG GTCGCATCCGGAGCTCCAGTGCCAACCTTCACAGGAA CTTCTGGCTTTGAGATCACCTCAAGCGGAACTACCACA CCTGCCGTGCTGGACATCACCCCAACAGTGGATAGCG TCTCCATTTCCTCTACAAGCTTCACTAACCCTGCTTTTA GTGATCCATCAATCATTGAGGTGCCTCAGACCGGAGAA GTCAGCGGCAATATCTTCGTGGGCACTCCAACCTCTGG GAGTCACGGATACGAGGAAATTCCCCTGCAGACTTTTG CCAGTTCAGGCTCCGGGACCGAACCAATCAGCTCCAC ACCTCTGCCAACTGTGCGGAGAGTCAGAGGGCCCAGG CTGTATTCTAGGGCAAACCAGCAGGTGCGCGTCTCAAC AAGCCAGTTCCTGACTCACCCCTCTAGTCTGGTGACAT TTGACAACCCAGCCTACGAGCCCCTGGATACTACCCTG TCTTTCGAACCCACCAGTAATGTGCCTGACTCAGATTTT ATGGACATCATTCGCCTGCATCGACCTGCTCTGTCAAG CAGGCGAGGAACCGTGCGGTTCAGTAGACTGGGACAG CGAGCAACCATGTTTACACGAAGCGGCAAGCAGATCG GAGGACGAGTGCACTTCTATCATGATATCAGCCCTATT GCCGCTACAGAGGAAATCGAGCTGCAGCCACTGATTA GCGCCACCAATGACTCCGATCTGTTCGACGTGTACGCA GATTTTCCACCCCCTGCCAGCACAACTCCATCCACTAT TCATAAGAGCTTTACCTATCCCAAATACTCTCTGACCAT GCCTAGTACAGCAGCCTCCTCTTATTCAAACGTGACTG TCCCTCTGACCAGCGCTTGGGACGTGCCAATCTACACA GGCCCCGATATCATTCTGCCTTCCCACACTCCCATGTG GCCTTCCACTTCTCCAACCAATGCATCTACCACAACTTA TATCGGCATTCATGGGACCCAGTACTATCTGTGGCCAT GGTACTATTACTTCCCCAAGAAACGAAAACGGATCCCC TACTTCTTTGCTGACGGCTTTGTGGCTGCA	MVSHRAARRKRASATD LYRTCKQSGTCPPDVIN KVEGTTLADKILQWSSL GIFLGGLGIGTGSGSGG RTGYVPLGGRSNTVVD VGPTRPPVVIEPVGPTD PSIVTLVEDSSVVASGA PVPTFTGTSGFEITSSG TTTPAVLDITPTVDSVSI SSTSFTNPAFSDPSIIEV PQTGEVSGNIFVGTPTS GSHGYEEIPLQTFASSG SGTEPISSTPLPTVRRV RGPRLYSRANQQVRVS TSQFLTHPSSLVTFDNP AYEPLDTTLSFEPTSNV PDSDFMDIIRLHRPALS SRRGTVRFSRLGQRAT MFTRSGKQIGGRVHFY HDISPIAATEEIELQPLIS ATNDSDLFDVYADFPPP ASTTPSTIHKSFTYPKY SLTMPSTAASSYSNVTV PLTSAWDVPIYTGPDIIL PSHTPMWPSTSPTNAS TTTYIGIHGTQYYLWPW YYYFPKKRKRIPYFFAD GFVAA	28.0%

HPV51 E1	ATGGACTGTGAAGGTACAGAGGATGAGGGGGCGGGGT GTAATGGGTGGTTTTTTGTTGAAGCAATAGTAGAAAAAA AAACAGGAGATAATGTTTCGGATGATGAGGATGAAAAT GCAGATGATACAGGATCTGATTTAATAAACTTTATAGAT AGTGAAACTAGTATTTGCAGTCAGGCGGAACAGGAGAC AGCACGGGCGTTGTTTCAGGCCCAAGAATTACAGGCA AACAAAGAGGCTGTGCATCAGTTAAAACGAAAGTTTCT AGTCAGCCCGCGAAGCAGCCCATTAGGAGACATTACA AATCAAAACAACACACACAGCCATAGTCAGGCAAACGA GTCACAAGTTAAAAGGAGATTACTGGACAGTTATCCGG ACAGCGGATATGGCAATACACAAGTGGAAACTGTGGAA GCAACGTTGCAGGTAGATGGGCAACATGGCGGTTCAC AGAACAGTGTGTGTAGTAGCGGGGGGGGCAGTGTTAT GGATGTGGAAACAACAGAAAGCTGTGCAAATGTAGAAC TAAACAGTATATGTGAAGTATTAAAAAGCAGTAATGCAA AAGCAACGTTAATGGCAAAATTTAAAGAGTTGTATGGTA TTAGTTATAATGAGTTGGTACGGGTGTTTAAAAGTGATA AAACATGTTGTATAGATTGGGTTTGTGCATTGTTTGGCG TTTCCCCAATGGTAGCAGAAAATTTAAAAACACTAATTA AGCCATTTTGCATGTACTACCATATACAATGTTTATCAT GTGATTGGGGCACCATTGTATTAATGCTAATTAGGTTTT CATGTGCAAAAAACAGAACAACAATTGCTAAGTGTTTAA GTACATTAGTAAATATCCCACAATCACAAATGTTTATAG AACCACCAAAATTACGTAGTACACCTGTGGCATTATATT TTTATAGAACAGGCATATCAAACATTAGCAATACATATG GAGAGACACCTGAATGGATTACACGACAAACGCAACTA CAACATAGTTTTGAGGATAGTACCTTTGAATTATCACAA ATGGTGCAATGGGCATTTGACCATGAAGTATTAGATGA TAGTGAAATAGCATTTCATTATGCACAATTAGCAGATAT AGATAGTAATGCTGCAGCGTTTTTAAAGAGTAATTGCCA AGCAAAATATGTAAAAGATTGTGGGACCATGGCACGGC ATTACAAACGAGCACAAAGAAAATCATTATCTATGTCAG CCTGGATAAGGTATAGATGTGATAGAGCAAAGGATGGA GGCAACTGGAGAGAAATTGCTAAATTTTTAAGATATCAA GGTGTAAACTTTATGTCCTTTATTCAAATGTTTAAACAG TTTTTAAAAGGAACACCAAAACACAATTGCATAGTCATA TATGGCCCACCAAACACAGGCAAGTCATTATTTGCAAT GAGCCTAATGAAGTTTATGCAAGGGTCCATTATTTCATA TGTAAACTCTGGTAGTCATTTTTGGTTACAGCCACTAGA GGATGCTAAAATAGCATTGTTAGATGATGCTACGTATG GGTGTTGGACATATATTGATCAGTATTTAAGAAACTTTT TAGATGGTAATCCATGTAGTATAGATAGAAAACATAGGA GTTTAATACAATTAGTATGTCCACCATTACTAATAACGT CAAACATAAATCCACAAGAGGATGCAAACCTAATGTATT TACATACAAGGGTAACAGTATTAAAGTTTTTAAATACAT TTCCATTTGATAACAATGGGAATGCTGTGTATACATTGA ATGATGAAAATTGGAAAAATTTTTTTTCCACCACATGGT CCAGATTAGATTTGGAGGAGGAAGAGGACAAAGAAAAT GGAGACCCTATGCCACCGTTTAAATGTGTGCCAGGAGA AAATACTAGACTGTTATGA	ATGGACTGCGAGGGAACTGAGGATGAGGGGGCTGGG TGTAACGGCTGGTTTTTCGTGGAGGCTATTGTGGAGAA GAAGACTGGGGATAATGTGAGCGACGATGAGGACGAA AACGCCGACGATACCGGCTCCGACCTGATCAATTTCAT TGATAGTGAGACCTCAATCTGCAGCCAGGCAGAGCAG GAAACAGCCCGGGCTCTGTTTCAGGCTCAGGAGCTGC AGGCAAACAAGGAAGCCGTGCACCAGCTGAAGCGCAA ATTCCTGGTCTCTCCACGAAGCTCCCCCCTGGGAGATA TTACCAATCAGAACAATACACACAGCCATTCTCAGGCA AACGAGAGCCAGGTGAAGCGGAGACTGCTGGACTCTT ACCCTGATAGTGGGTATGGAAATACACAGGTGGAGACT GTCGAAGCAACCCTGCAGGTGGACGGACAGCATGGAG GGTCCCAGAACTCTGTCTGTTCTAGTGGAGGCGGGAG CGTGATGGATGTCGAGACCACAGAATCTTGCGCTAATG TGGAGCTGAACAGTATCTGTGAAGTCCTGAAATCAAGC AATGCAAAGGCCACTCTGATGGCAAAGTTCAAAGAGCT GTACGGCATCAGTTATAACGAACTGGTGAGGGTCTTTA AGTCAGACAAAACCTGCTGTATTGATTGGGTGTGCGCA CTGTTCGGGGTGTCCCCTATGGTGGCCGAGAACCTGA AGACCCTGATCAAACCTTTCTGTATGTACTACCACATCC AGTGCCTGTCTTGTGACTGGGGGACAATCGTGCTGAT GCTGATTCGGTTCAGTTGCGCCAAAAATAGAACTACCA TCGCTAAGTGTCTGTCAACCCTGGTGAACATCCCACAG AGCCAGATGTTTATTGAACCCCCTAAGCTGCGATCCAC ACCCGTCGCCCTGTACTTCTATCGGACTGGAATCAGTA ACATTTCAAATACCTACGGCGAGACTCCAGAATGGATC ACCCGCCAGACACAGCTGCAGCACTCATTCGAGGACA GCACATTTGAACTGAGCCAGATGGTGCAGTGGGCATTC GATCACGAGGTCCTGGACGATTCCGAAATCGCCTTTCA TTACGCTCAGCTGGCAGACATTGATTCCAATGCCGCTG CATTTCTGAAGTCTAACTGCCAGGCCAAGTACGTGAAA GACTGTGGCACTATGGCTAGACATTATAAACGAGCACA GCGGAAGAGTCTGTCAATGAGCGCTTGGATCAGATAC CGGTGCGACAGGGCAAAAGATGGAGGCAATTGGAGGG AGATCGCCAAGTTCCTGCGCTATCAGGGGGTGAACTTC ATGTCTTTTATTCAGATGTTCAAGCAGTTTCTGAAAGGA ACACCCAAGCACAATTGTATCGTCATCTACGGCCCACC CAACACTGGGAAAAGTCTGTTCGCTATGTCACTGATGA AGTTTATGCAGGGGAGCATCATTAGCTACGTGAACAGC GGATCTCATTTTTGGCTGCAGCCCCTGGAGGACGCCA AAATCGCTCTGCTGGACGATGCCACTTACGGATGCTGG ACCTACATTGATCAGTATCTGCGGAATTTCCTGGACGG CAACCCTTGCAGCATCGATAGAAAGCACAGGTCCCTGA TTCAGCTGGTGTGTCCTCCACTGCTGATCACAAGCAAC ATTAATCCTCAGGAGGACGCCAACCTGATGTACCTGCA TACTAGAGTGACCGTCCTGAAATTTCTGAATACTTTCCC ATTTGACAACAATGGCAACGCTGTGTATACCCTGAACG ATGAAAATTGGAAGAACTTCTTTTCCACAACTTGGTCTC GCCTGGATCTGGAGGAAGAGGAAGACAAAGAGAATGG CGATCCTATGCCCCCTTTCAAGTGCGTGCCAGGGGAA AACACCCGACTGCTG	MDCEGTEDEGAGCNG WFFVEAIVEKKTGDNVS DDEDENADDTGSDLINF IDSETSICSQAEQETAR ALFQAQELQANKEAVH QLKRKFLVSPRSSPLGD ITNQNNTHSHSQANES QVKRRLLDSYPDSGYG NTQVETVEATLQVDGQ HGGSQNSVCSSGGGS VMDVETTESCANVELN SICEVLKSSNAKATLMA KFKELYGISYNELVRVF KSDKTCCIDWVCALFG VSPMVAENLKTLIKPFC MYYHIQCLSCDWGTIVL MLIRFSCAKNRTTIAKCL STLVNIPQSQMFIEPPKL RSTPVALYFYRTGISNIS NTYGETPEWITRQTQL QHSFEDSTFELSQMVQ WAFDHEVLDDSEIAFHY AQLADIDSNAAAFLKSN CQAKYVKDCGTMARHY KRAQRKSLSMSAWIRY RCDRAKDGGNWREIAK FLRYQGVNFMSFIQMF KQFLKGTPKHNCIVIYG PPNTGKSLFAMSLMKF MQGSIISYVNSGSHFWL QPLEDAKIALLDDATYG CWTYIDQYLRNFLDGN PCSIDRKHRSLIQLVCP PLLITSNINPQEDANLMY LHTRVTVLKFLNTFPFD NNGNAVYTLNDENWKN FFSTTWSRLDLEEEEDK ENGDPMPPFKCVPGEN TRLL	26.9%
HPV51 E2	ATGGAGACCCTATGCCACCGTTTAAATGTGTGCCAGGA GAAAATACTAGACTGTTATGAACTGGACAGTGATAAATT AGTAGATCAAATTAACTATTGGACATTGTTACGATATGA AGCTGCTATGTTTTATGCAGCACGGGAAAGAAACTTAC GAACAATCAATCACCAGGTAGTACCAGCAACAACAGTA TCAAAACAAAAGGCCTGTCAAGCAATTGAAATGCACAT GGCCTTACAATCGCTTAACAAATCAGACTATAACATGG AACCATGGACAATGCGGGAGACATGTTATGAACTATGG TGTGTGGCTCCCAAGCAATGTTTCAAAAAGGGGGGCAT AACTGTAACAGTTATATTTGATGGAAATAAGGACAATGC AATGGACTATACAAGCTGGAAATTTATATATATATATGA TAATGATAAGTGGGTAAAGACAAATGGAAATGTGGACT ATACGGGTATATATTACACTGTAAATTCAAAAAAAGAAT ATTATGTACAGTTTAAAGATGAAGCCAAAATATATGGGG CACAACAGTGGGAGGTCTATATGTATGGTACTGTAATA ACATGTCCTGAATATGTATCTAGTACCTGCAGCGACGC GTTATCCACTACTACAACTGTTGAACAACTATCAAACAC CCCAACGACCAATCCCCTTACCACCTGCGTGGGCGCC AAAGAAGCCCAGACACAACAGCGAAAACGACAGCGAC TTACTGAGCCCGACTCCTCCACAATCTCCCCACTGTCC GTGGACAATACAAACAACCAAATACACTGTGGAAGTGG AAGCACTAACACTGGAGGGCACCAAAGTGCAACTCAG ACTGCGTTTATAGTGCATTTAAAAGGTGATACAAATTGT TTAAAATGTTTTAGATACAGATTTACAAAACACAAAGGG TTATATAAAAACGTATCCTCAACCTGGCATTGGACCAGT AATACTAAAACAGGCATTGTTACCATTGTGTTTGACAGT GCACATCAACGGGAAACATTTATAAAAACCATTAAAGTA CCCCCAAGTGTAACACTGTCATTGGGAATTATGACACT GTAA	ATGGAAACTCTGTGTCATAGACTGAACGTGTGTCAGGA AAAGATTCTGGACTGCTACGAACTGGACTCCGATAAAC TGGTGGATCAGATCAACTACTGGACACTGCTGAGATAT GAGGCCGCTATGTTCTACGCAGCCCGGGAAAGAAACC TGCGCACCATCAATCATCAGGTGGTCCCAGCCACCACA GTGTCTAAGCAGAAGGCCTGCCAGGCAATTGAGATGC ACATGGCTCTGCAGAGCCTGAACAAGTCCGATTATAAT ATGGAACCCTGGACTATGCGAGAGACCTGCTACGAACT GTGGTGTGTGGCACCTAAGCAGTGTTTCAAGAAAGGC GGGATCACAGTGACTGTCATTTTTGACGGCAACAAGGA TAATGCTATGGACTATACAAGTTGGAAATTCATCTACAT CTACGACAACGATAAGTGGGTGAAAACTAACGGCAATG TCGATTACACTGGGATCTACTATACCGTGAATAGCAAG AAAGAGTACTATGTCCAGTTTAAGGACGAAGCTAAAAT CTATGGAGCACAGCAGTGGGAGGTGTACATGTATGGC ACCGTCATTACATGCCCCGAATACGTGAGCTCCACCTG TTCTGACGCCCTGAGTACTACCACAACTGTCGAGCAGC TGAGCAACACACCAACCACAAATCCCCTGACTACCTGC GTGGGAGCAAAGGAGGCTCAGACCCAGCAGAGGAAAC GCCAGCGACTGACAGAACCTGATTCTAGTACTATCAGT CCACTGTCAGTGGACAACACCAACAATCAGATTCACTG TGGGTCCGGATCTACCAATACAGGAGGCCATCAGTCC GCAACTCAGACCGCCTTTATCGTGCACCTGAAGGGGG ATACTAATTGCCTGAAATGTTTCCGGTACAGATTCACCA AGCATAAAGGACTGTACAAGAACGTGTCAAGCACATGG CACTGGACTTCCAATACAAAAACTGGCATCGTGACCAT TGTCTTCGACTCTGCCCATCAGAGGGAGACCTTTATCA AGACCATCAAGGTGCCCCCTTCAGTCACCCTGAGCCT GGGGATTATGACACTG	METLCHRLNVCQEKILD CYELDSDKLVDQINYWT LLRYEAAMFYAARERNL RTINHQVVPATTVSKQK ACQAIEMHMALQSLNK SDYNMEPWTMRETCY ELWCVAPKQCFKKGGI TVTVIFDGNKDNAMDYT SWKFIYIYDNDKWVKTN GNVDYTGIYYTVNSKKE YYVQFKDEAKIYGAQQ WEVYMYGTVITCPEYV SSTCSDALSTTTTVEQL SNTPTTNPLTTCVGAKE AQTQQRKRQRLTEPDS STISPLSVDNTNNQIHC GSGSTNTGGHQSATQT AFIVHLKGDTNCLKCFR YRFTKHKGLYKNVSST WHWTSNTKTGIVTIVFD SAHQRETFIKTIKVPPSV TLSLGIMTL	25.1%
HPV51 E4	ATGTATCTAGTACCTGCAGCGACGCGTTATCCACTACT ACAACTGTTGAACAACTATCAAACACCCCAACGACCAA TCCCCTTACCACCTGCGTGGGCGCCAAAGAAGCCCAG ACACAACAGCGAAAACGACAGCGACTTACTGAGCCCG ACTCCTCCACAATCTCCCCACTGTCCGTGGACAATACA AACAACCAAATACACTGTGGAAGTGGAAGCACTAACAC TGGAGGGCACCAAAGTGCAACTCAGACTGCGTTTATAG	ATGTACCTGGTCCCTGCCGCTACCCGCTATCCCCTGCT GCAGCTGCTGAACAACTATCAGACTCCACAGAGGCCC ATTCCCCTGCCACCCGCCTGGGCTCCAAAGAAACCCA GGCACAACAGCGAGAATGACTCCGATCTGCTGTCTCCT ACCCCCCCCTCAGAGCCCTCATTGCCCCTGGACAATCCA GACCACAAAGTACACTGTGGAGGTCGAAGCCCTGACT CTGGAAGGCACCAAAGTGCAGCTGCGGCTGAGACTG	MYLVPAATRYPLLQLLN NYQTPQRPIPLPPAWA PKKPRHNSENDSDLLS PTPPQSPHCPWTIQTTK YTVEVEALTLEGTKVQL RLRL	27.6%
HPV51 E5	ATGTATAGACATATTGTAACCATTGCAGTGTTTATTATTT TGCTATTTGTGCTTTGCTTGTGTGTGTGTCTTGTGTTGT GTTGTTTGTTGCCGCTACTGCTGTCCCAATACGTGTTT GCAGCTGCCTTATTATTAATTTTATGTTTTTGGTTTGTTG TTGCAACATCCCAATTAACTACATTTTTTGTATATTTGAT TTTTTTTTACTTACCTTGTTTACTTTTACATCTATATACAT TTTTACTTTTGCAATAA	ATGTATCGCCATATTGTGACTATCGCCGTGTTCATCATT CTGCTGTTCGTGCTGTGCCTGTGTGTGTGTCTGGTGCT GTGCTGCCTGCTGCCCCTGCTGCTGAGCCAGTACGTG TTCGCCGCTGCACTGCTGCTGATCCTGTGCTTCTGGTT TGTGGTCGCCACTTCCCAGCTGACCACATTCTTTGTCT ACCTGATTTTCTTTTATCTGCCTTGTCTGCTGCTGCACC TGTATACCTTTCTGCTGCTGCAG	MYRHIVTIAVFIILLFVLC LCVCLVLCCLLPLLLSQ YVFAAALLLILCFWFVVA TSQLTTFFVYLIFFYLPC LLLHLYTFLLLQ*X	27.5%

HPV51 E6	ATGTTCGAAGACAAGAGGGAAAGACCACGAACGCTGC ATGAATTATGTGAAGCTTTGAACGTTTCTATGCACAATA TACAGGTAGTGTGTGTGTATTGTAAAAAGGAATTATGTA GAGCAGATGTATATAATGTAGCATTTACTGAAATTAAGA TTGTATATAGGGATAATAATCCATATGCAGTATGCAAAC AATGTTTACTGTTTTATTCAAAAATTAGAGAGTATAGAC GTTATAGCAGGTCTGTGTATGGTACTACATTAGAGGCA ATTACTAAAAAAAGCTTATATGATTTATCGATAAGGTGT CATAGATGTCAAAGACCACTTGGGCCTGAAGAAAAGCA AAAATTGGTGGACGAAAAAAAAAGGTTCCATGAAATAG CGGGACGTTGGACGGGGCAATGCGCTAATTGCTGGCA ACGTACACGACAACGTAACGAAACCCAAGTGTAA	ATGTTTGAAGACAAGCGGGAACGCCCAAGGACTCTGC ACGAACTGTGTGAAGCACTGAATGTGAGCATGCATAAT ATCCAGGTGGTGTGCGTGTACTGTAAGAAAGAGCTGTG CCGGGCCGACGTGTATAACGTCGCTTTCACCGAAATCA AGATTGTGTACAGAGATAACAATCCCTATGCTGTCTGC AAGCAGTGTCTGCTGTTTTACTCCAAAATCCGCGAGTA CCGGAGATATAGCCGATCCGTGTACGGAACCACACTG GAAGCCATCACAAAGAAATCTCTGTATGACCTGAGTAT TCGGTGCCACAGATGTCAGAGGCCCCTGGGCCCTGAG GAAAAGCAGAAACTGGTGGATGAGAAGAAAAGGTTCCA TGAAATTGCAGGACGATGGACTGGACAGTGCGCAAAC TGTTGGCAGAGGACTCGCCAGCGAAATGAGACCCAGG TC	MFEDKRERPRTLHELC EALNVSMHNIQVVCVY CKKELCRADVYNVAFT EIKIVYRDNNPYAVCKQ CLLFYSKIREYRRYSRS VYGTTLEAITKKSLYDLS IRCHRCQRPLGPEEKQ KLVDEKKRFHEIAGRWT GQCANCWQRTRQRNE TQV	26.3%
HPV51 E7	ATGCGTGGTAATGTACCACAATTAAAAGATGTAGTATTG CATTTAACACCACAGACTGAAATTGACTTGCAATGCTAC GAGCAATTTGACAGCTCAGAGGAGGAGGATGAAGTAG ATAATATGCGTGACCAGCTACCAGAAAGACGGGCTGG ACAGGCTACGTGTTACAGAATTGAAGCTCCGTGTTGCA GGTGTTCAAGTGTAGTACAACTGGCAGTGGAAAGCAGT GGAGACACCCTTCGCGTTGTACAGCAGATGTTAATGGG CGAACTAAGCCTGGTTTGCCCGTGTTGTGCGAACAACT AG	ATGAGAGGCAACGTCCCCCAGCTGAAAGATGTCGTCC TGCACCTGACTCCACAGACCGAGATTGACCTGCAGTGT TATGAACAGTTTGACAGCTCCGAGGAAGAGGACGAAGT GGATAACATGCGAGATCAGCTGCCAGAGCGAAGAGCA GGACAGGCTACCTGCTACAGGATCGAAGCACCTTGCT GTCGCTGTTCTAGTGTGGTCCAGCTGGCCGTGGAGTC AAGCGGCGACACACTGAGAGTGGTCCAGCAGATGCTG ATGGGGGAACTGTCTCTGGTCTGCCCCTGCTGTGCTAA CAAT	MRGNVPQLKDVVLHLT PQTEIDLQCYEQFDSSE EEDEVDNMRDQLPERR AGQATCYRIEAPCCRC SSVVQLAVESSGDTLR VVQQMLMGELSLVCPC CANN*X	25.9%
HPV51 L1	ATGGCATTGTGGCGCACTAATGACAGCAAGGTGTATTT GCCACCTGCACCTGTGTCTCGAATTGTGAATACAGAAG AATATATCACACGCACCGGCATATATTACTATGCAGGC AGTTCCAGACTAATAACATTAGGACATCCCTATTTTCCA ATACCTAAAACCTCAACGCGTGCTGCTATTCCTAAAGTA TCTGCATTTCAATACAGGGTATTTAGGGTACAGTTACCA GATCCTAACAAGTTTGGACTCCCGGATCCAAATTTATAT AATCCAGACACAGATAGGTTGGTGTGGGGTTGTGTGG GCGTTGAGGTGGGCAGAGGACAGCCCCTTGGTGTTGG CCTTAGTGGTCATCCCTTATTTAATAAATATGATGACAC AGAAAATTCACGCATAGCAAATGGCAATGCACAACAAG ATGTTAGAGATAACACATCTGTTGACAACAAACAGACTC AGTTATGTATAATAGGCTGTGCTCCACCTATTGGGGAA CACTGGGGTATTGGCACTACATGCAAAAACACACCTGT ACCTCCAGGAGACTGCCCCCCCCTGGAACTTGTATCCT CTGTCATTCAGGATGGCGATATGATTGATACAGGGTTT GGAGCTATGGATTTCGCTGCCCTACAGGCCACCAAATC AGACGTCCCTTTGGATATTTCACAGTCTGTTTGTAAATA TCCTGATTATTTAAAAATGTCTGCAGACACATATGGTAA TTCCATGTTTTTTCATTTACGCAGGGAGCAAATCTTTGC TAGGCACTATTATAATAAACTTGTAGGTGTTGGGGAAG ACATTCCTAACGATTATTATATTAAGGGTAGTGGTAATG GCCGTGACCCTATAGAAAGTTATATATACTCTGCTACTC CCAGTGGGTCTATGATAACATCTGATTCTCAAATTTTTA ATAAGCCTTATTGGCTCCACCGTGCGCAGGGTCACAAT AATGGCATTTGCTGGAACAATCAGCTTTTTATTACCTGT GTTGATACTACCAGAAGTACAAATTTAACTATTAGCACT GCCACTGCTGCGGTTTCCCCAACATTTACTCCAAGTAA CTTTAAGCAATATATTAGGCATGGGGAAGAGTATGAATT GCAATTTATTTTTCAATTATGTAAAATTACTTTAACTACA GAGGTAATGGCTTATTTACACACAATGGATCCTACCATT CTTGAACAGTGGAATTTTGGATTAACATTACCTCCGTCT GCTAGTTTGGAGGATGCATATAGGTTTGTTAGAAATGC AGCTACTAGCTGTCAAAAGGACACCCCTCCACAGGCTA AGCCAGATCCTTTGGCCAAATATAAATTTTGGGATGTTG ATTTAAAGGAACGATTTTCTTTAGATTTAGACCAATTTG CATTGGGTCGCAAGTTTTTGTTGCAGGTTGGCGTACAA CGCAAGCCCAGACCAGGCCTTAAACGCCCGGCCTCAT CGGCATCCTCTTCCTCTTCCTCTTCAGCCAAACGTAAA CGTGTTAAAAAGTAA	ATGGCTCTGTGGCGAACTAATGACTCTAAGGTCTATCT GCCCCCTGCCCCCGTCTCCCGAATCGTGAATACTGAA GAATACATCACCCGGACAGGAATCTACTATTACGCTGG CAGCTCCAGACTGATCACTCTGGGGCACCCTTATTTCC CTATTCCAAAGACCTCAACACGGGCCGCTATCCCAAAA GTGTCTGCATTCCAGTACCGAGTCTTTCGGGTGCAGCT GCCCGACCCTAACAAGTTTGGACTGCCAGATCCCAACC TGTATAATCCAGACACAGATCGACTGGTCTGGGGATGC GTGGGAGTCGAAGTGGGACGAGGGCAGCCTCTGGGA GTGGGCCTGAGTGGACACCCACTGTTCAACAAGTACG ACGATACTGAAAATTCTAGGATCGCTAACGGCAATGCA CAGCAGGACGTCCGCGATAACACCTCCGTGGACAATA AGCAGACACAGCTGTGCATCATTGGGTGTGCCCCCCC TATCGGAGAGCATTGGGGGATTGGAACCACATGCAAAA ATACCCCCGTGCCACCAGGCGATTGTCCTCCACTGGA ACTGGTCTCTAGTGTGATCCAGGACGGGGATATGATTG ACACCGGCTTCGGGGCTATGGATTTTGCAGCCCCTGCA GGCAACAAAGTCCGACGTCCCACTGGATATTAGTCAGT CAGTGTGTAAGTATCCCGACTACCTGAAAATGTCTGCC GATACCTACGGCAACAGCATGTTCTTTCACCTGCGGAG AGAGCAGATTTTCGCTCGGCATTATTACAATAAGCTGG TCGGAGTGGGCGAAGACATCCCCAACGATTATTACATC AAGGGGAGCGGAAATGGCAGAGACCCCATCGAGTCCT ATATCTACTCTGCAACTCCTAGCGGCTCCATGATCACC TCTGATAGTCAGATTTTCAACAAGCCTTACTGGCTGCA CCGAGCACAGGGACATAACAATGGGATCTGCTGGAAC AATCAGCTGTTTATTACCTGTGTGGACACTACCCGAAG TACCAACCTGACAATTTCAACTGCCACCGCTGCAGTGA GCCCAACATTCACTCCCTCCAATTTTAAGCAGTATATCA GGCACGGCGAGGAATACGAGCTGCAGTTCATCTTTCA GCTGTGCAAAATTACTCTGACAACTGAAGTGATGGCCT ACCTGCATACTATGGACCCTACCATCCTGGAACAGTGG AACTTCGGACTGACCCTGCCACCTTCAGCCAGCCTGG AGGATGCTTATAGATTTGTGAGGAATGCCGCTACATCC TGTCAGAAGGACACTCCACCCCAGGCAAAACCTGATCC ACTGGCCAAGTACAAATTCTGGGACGTGGATCTGAAGG AACGGTTCAGCCTGGACCTGGACCAGTTCGCCCTGGG CAGGAAATTTCTGCTGCAGGTCGGAGTGCAGCGAAAG CCACGACCTGGACTGAAACGCCCCGCATCAAGCGCCT CCTCTAGTTCAAGCTCCTCTGCTAAGAGGAAACGCGTG AAGAAA	MALWRTNDSKVYLPPA PVSRIVNTEEYITRTGIY YYAGSSRLITLGHPYFPI PKTSTRAAIPKVSAFQY RVFRVQLPDPNKFGLP DPNLYNPDTDRLVWGC VGVEVGRGQPLGVGLS GHPLFNKYDDTENSRIA NGNAQQDVRDNTSVD NKQTQLCIIGCAPPIGEH WGIGTTCKNTPVPPGD CPPLELVSSVIQDGDMI DTGFGAMDFAALQATK SDVPLDISQSVCKYPDY LKMSADTYGNSMFFHL RREQIFARHYYNKLVGV GEDIPNDYYIKGSGNGR DPIESYIYSATPSGSMIT SDSQIFNKPYWLHRAQ GHNNGICWNNQLFITCV DTTRSTNLTISTATAAVS PTFTPSNFKQYIRHGEE YELQFIFQLCKITLTTEV MAYLHTMDPTILEQWN FGLTLPPSASLEDAYRF VRNAATSCQKDTPPQA KPDPLAKYKFWDVDLK ERFSLDLDQFALGRKFL LQVGVQRKPRPGLKRP ASSASSSSSSSAKRKR VKK	24.8%
HPV51 L2	ATGGTGGCTACACGTGCACGGCGTCGGAAGCGAGCAT CTGTAACACAATTATATTCTACATGCAAAGCTGCTGGTA CATGTCCTCCTGATGTTGTGAATAAGGTTGAAGGTACT ACATTGGCCGATAAAATATTACAGTGGAGTGGGTTGGG TATATTTTTGGGTGGCCTAGGTATTGGTACTGGGTCTG GATCTGGGGGGCGTACTGGATATATCCCTTTAGGTGGT GGGGGTCGCCCAGGCGTGGTGGATATTGCTCCTGCAA GGCCACCTATTATAATTGACCTATGGCACCATACTGAA CCTTCTATAGTAAATTTGGTTGAGGACTCTAGTATTATT CAGTCTGGGTCTCCTATACCTACCTTTACTGGTACCGA TGGCTTTGAAATTACTTCATCTTCCACAACAACCCCTGC TGTGTTGGACATCACCCCATCTGCTGGTACTGTACATG TTTCTAGTACTAACATTGAAAATCCTTTATATATTGAACC TCCATCCATTGAGGCTCCACAATCTGGAGAAGTGTCAG ATATATATTTACTAGTACACTACTCTGGTACTCATGGGT ATGAAGAAATACCTATGGAAGTGTTTGCATCCAATGTCA GTACTGGTACTGAACCTATTAGCAGCACACCTACTCCA GGGGTTAGTCGCATAGCTGCTCCCCGCTTGTATAGTAA GTCCTACACACAGGTTAAAGTTACAAATCCTGATTTTAT TAGTAAGCCATCCACATTTGTTACATTTAATAATCCTGC TTTTGAGCCTATTGACACATCCATAACTTTTGAGGAACC TGATGCTGTTGCACCTGATCCTGATTTTCTGGATATTAT TACACTGCACCGCCCTGCCCTTACATCTCGTAGAGGCA CAGTACGCTTTAGTAGGTTAGGTCAAAAGGCCACCATG CGCACTCGTAGTGGCAAACAAATTGGTGCTCGTGTACA TTATTATCATGATATTAGTAGAATTGCACCAGCTGATGA ACTTGAAATGCAGCCTTTACTTTCACCTTCTAATAATTAT AGTTATGACATTTATGCTGATTTAGATGAAGCTGAAACA GGTTTTATACAGCCCACACACACCACACCTATGTCACA CTCCTCTTTGTCTAGGCAGTTGCCCTCCTTATCTTCATC TATGTCTTCATCTTATGCAAATGTTACTATTCCATTTTCA ACTACATATTCTGTTCCTATTCATACAGGGCCTGATGTG GTATTGCCCACATCTCCTACAGTATGGCCTTATGTTCC CCACACTTCCATTGACACCAAGCATTCTATTGTTATACT AGGTGGGGATTACTATTTGTGGCCCTATACACATTTACT ACGCAAACGCCGTAAACGTATACCCTATTTTTTTACAGA TGGCATTGTGGCGCACTAA	ATGGTCGCAACTCGGGCAAGAAGAAGGAAAAGAGCCT CAGTCACTCAGCTGTATAGCACCTGTAAAGCCGCCGGA ACCTGTCCCCCAGACGTGGTCAACAAGGTGGAGGGCA CCACACTGGCCGATAAAATTCTGCAGTGGAGCGGGCT GGGAATTTTCCTGGGAGGACTGGGAATCGGGACAGGA TCTGGCAGTGGAGGCCGGACTGGATACATTCCACTGG GAGGAGGCGGGCGACCTGGAGTGGTGGACATCGCAC CAGCCCGCCCCCCTATCATTATCGATCTGTGGCACCAT ACCGAGCCCAGCATCGTGAATCTGGTCGAAGACAGCT CCATTATCCAGTCAGGCAGCCCTATTCCAACCTTCACA GGGACTGACGGATTTGAGATCACCTCTAGTTCAACTAC CACACCCGCTGTGCTGGATATTACTCCTTCCGCAGGGA CCGTGCACGTCAGCTCCACAAACATCGAGAATCCACTG TACATTGAGCCACCCTCTATCGAAGCACCCCAGTCAGG AGAAGTGAGCGATATCTACCTGCTGGTCCACTATAGTG GCACCCACGGCTATGAGGAAATCCCTATGGAGGTGTT CGCCAGCAACGTCTCCACCGGAACAGAACCTATTTCTA GTACTCCAACCCCTGGCGTGTCTCGCATCGCAGCTCCT CGACTGTACTCCAAGTCTTATACACAGGTGAAAGTCAC TAATCCTGACTTTATCTCTAAGCCAAGTACATTCGTGAC TTTTAACAATCCAGCTTTCGAGCCCATTGACACCTCCAT CACATTTGAGGAACCTGATGCTGTGGCACCAGACCCC GATTTCCTGGATATTATCACCCTGCACAGACCAGCACT GACCAGTCGGAGAGGGACAGTGAGATTTTCAAGGCTG GGACAGAAGGCCACTATGCGAACCCGGAGCGGAAAAC AGATCGGCGCTAGGGTGCACTACTATCATGACATTTCC CGCATCGCCCCCCGCTGATGAGCTGGAAATGCAGCCTC TGCTGAGTCCATCAAACAATTACAGCTATGACATCTAC GCAGACCTGGATGAGGCCGAAACTGGCTTCATCCAGC CCACCCACACTACCCCTATGAGTCATTCAAGCCTGTCA AGGCAGCTGCCAAGCCTGTCCTCTAGTATGTCAAGCTC CTACGCCAACGTGACCATTCCCTTTTCCACAACTTATTC TGTCCCAATCCATACAGGGCCCGACGTGGTCCTGCCT ACATCCCCAACTGTGTGGCCCTATGTCCCTCACACCTC CATCGACACAAAACATTCTATTGTGATCCTGGGAGGCG ATTACTATCTGTGGCCTTACACTCACCTGCTGCGGAAG AGGCGCAAAAGAATCCCCTATTTCTTTACAGATGGCAT CGTGGCTCAT	MVATRARRRKRASVTQ LYSTCKAAGTCPPDVV NKVEGTTLADKILQWSG LGIFLGGLGIGTGSGSG GRTGYIPLGGGGRPGV VDIAPARPPIIIDLWHHT EPSIVNLVEDSSIIQSGS PIPTFTGTDGFEITSSST TTPAVLDITPSAGTVHV SSTNIENPLYIEPPSIEA PQSGEVSDIYLLVHYSG THGYEEIPMEVFASNVS TGTEPISSTPTPGVSRIA APRLYSKSYTQVKVTNP DFISKPSTFVTFNNPAF EPIDTSITFEEPDAVAPD PDFLDIITLHRPALTSRR GTVRFSRLGQKATMRT RSGKQIGARVHYYHDIS RIAPADELEMQPLLSPS NNYSYDIYADLDEAETG FIQPTHTTPMSHSSLSR QLPSLSSSMSSSYANV TIPFSTTYSVPIHTGPDV VLPTSPTVWPYVPHTSI DTKHSIVILGGDYYLWP YTHLLRKRRKRIPYFFT DGIVAH	28.1%

HPV52 E1	ATGGAGGACCCTGAAGGTACAGAGGGCGAAAGGGAG GGATGTACAGGCTGGTTTGAAGTAGAGGCAATAATAGA AAAACAAACAGGAGATAACATTTCAGAGGACGAGGATG AAAATGCATATGATAGTGGAACAGATCTAATAGATTTTA TAGATGATTCAAATATAAATAATGAACAGGCAGAACATG AGGCAGCCCGGGCATTGTTTAATGCACAGGAAGGGGA GGATGATTTACATGCTGTGTCTGCAGTAAAACGAAAGT TTACAAGCAGTCCGGAAAGTGCTGGGCAAGATGGTGT AGAAAAACATGGTAGTCCGCGTGCAAAACACATTTGTG TAAATACAGAGTGTGTTTTACCAAAACGCAAACCATGTC ACGTAGAAGACAGCGGCTATGGCAATAGTGAAGTGGA AGCGCAGCAGATGGCAGACCAGGTAGACGGGCAAAAT GGCGACTGGCAAAGTAACAGTAGTCAATCAAGTGGGG TGGGGGCTAGTAATTCAGATGTAAGTTGTACTAGTATA GAGGACAATGAGGAAAATAGTAATAGAACGCTAAAAAG CATACAAAATATTATGTGCGAAAATAGCATAAAAACAAC TGTATTATTTAAATTTAAAGAAACATATGGTGTTAGCTTT ATGGAATTAGTAAGACCATTTAAAAGTAATAGAAGTAGT TGTACAGATTGGTGTATTATAGGAATGGGAGTAACACC ATCAGTTGCAGAAGGATTAAAAGTATTAATACAGCCCTA TAGCATATATGCCCATTTGCAATGTTTAACATGTGACAG AGGCGTGCTTATACTGCTGCTAATTAGGTTTAAATGTG GAAAAAACAGATTAACAGTGTCCAAACTAATGTCACAG CTGTTAAATATACCAGAAACACATATGGTAATAGAACCA CCAAAATTACGAAGTGCTACCTGTGCATTATATTGGTAT AGAACAGGTTTGTCTAATATTAGTGAGGTATATGGTACC ACCCCAGAATGGATAGAACAACAAACAGTATTACAGCA TAGCTTTGACAATAGCATATTCGATTTTGGAGAAATGGT GCAATGGGCATATGATCATGATATAACAGATGATAGTG ACATAGCATATAAATATGCACAGTTAGCAGATGTAAATA GCAATGCTGCAGCATTCCTAAAAAGCAATTCGCAAGCA AAAATAGTAAAGGACTGTGCAACCATGTGTAGACATTAT AAACGGGCAGAAAGAAAACATATGAATATTGGACAATG GATACAGTATAGATGTGATAGAATAGATGATGGTGGAG ATTGGAGGCCTATAGTAAGATTTTTAAGATATCAAGACA TAGAATTTACAGCCTTTTTAGACGCATTTAAAAAATTTTT AAAAGGTATACCTAAAAAAAATTGTTTAGTATTATATGG ACCTGCAAACACAGGAAAATCATATTTTGGAATGAGTTT AATTAGGTTCTTAAGTGGATGTGTAATATCCTATGTAAA CTCAAAAAGCCATTTTTGGCTACAACCATTAACAGATGC AAAAGTGGGTATGATAGATGATGTAACACCTATATGTTG GACATATATAGATGATTATATGAGAAATGCACTGGATG GAAATGATATATCAGTAGATGTAAAGCATAGAGCCTTA GTACAAATAAAATGCCCACCATTAATTTTAACAACAAAT ACAAATGCAGGAACAGATCCTAGGTGGCCATATTTACA TAGTAGATTGGTTGTGTTTCATTTCAAAAACCCATTTCC ATTTGATGAAAATGGCAATCCTATATATGAAATTAACAA CGAAAATTGGAAATCCTTTTTCTCAAGGACGTGGTGCA AATTAGATTTAATACAGGAAGAGGACAAGGAAAACGAT GGAGTCGATACCGGCACGTTTAAATGCAGTGCAGGAA AAAATACTAGATCTATACGAAGCTGA	ATGGAAGACCCCGAGGGAACTGAAGGAGAAAGAGAAG GATGCACAGGCTGGTTTGAGGTGGAGGCTATCATTGAA AAGCAGACCGGAGACAACATTTCAGAGGACGAAGATG AGAATGCTTACGATAGCGGCACCGACCTGATCGATTTC ATTGACGATAGCAACATCAACAATGAGCAGGCAGAACA CGAGGCAGCTCGGGCACTGTTCAATGCCCAGGAAGGA GAGGACGATCTGCATGCAGTGTCCGCCGTCAAGAGAA AGTTCACCAGCAGCCCAGAAAGTGCAGGACAGGACGG AGTGGAGAAGCACGGCTCACCCAGAGCTAAACATATCT GCGTGAACACCGAATGTGTCCTGCCAAAGAGGAAACC CTGCCACGTGGAGGACTCCGGATACGGCAATTCTGAA GTGGAGGCTCAGCAGATGGCAGACCAGGTCGATGGGC AGAACGGAGATTGGCAGAGCAATTCTAGTCAGTCAAGC GGGGTGGGAGCCAGTAACTCAGACGTCTCATGCACAA GCATTGAAGATAATGAGGAAAACTCTAATCGGACTCTG AAAAGTATCCAGAACATTATGTGTGAGAACAGCATCAA GACCACAGTGCTGTTCAAGTTTAAAGAAACCTACGGCG TGAGCTTCATGGAGCTGGTCCGCCCTTTTAAGTCTAAC CGATCCTCTTGCACTGACTGGTGTATCATTGGAATGGG AGTGACCCCAAGCGTCGCAGAGGGACTGAAGGTGCTG ATCCAGCCTTACTCCATCTACGCCCACCTGCAGTGCCT GACTTGTGATAGAGGGGTGCTGATCCTGCTGCTGATTC GGTTTAAGTGCGGAAAAAACAGACTGACCGTGTCTAAA CTGATGAGTCAGCTGCTGAATATCCCCGAAACACACAT GGTCATCGAGCCCCCTAAGCTGCGATCTGCTACCTGT GCACTGTACTGGTATCGGACAGGACTGTCCAACATTTC TGAAGTGTACGGCACTACCCCTGAATGGATCGAGCAG CAGACAGTCCTGCAGCACTCATTCGACAATAGCATCTT CGATTTTGGCGAGATGGTGCAGTGGGCTTATGACCATG ATATCACTGACGATTCTGACATTGCATACAAATATGCCC AGCTGGCTGATGTGAACAGTAATGCAGCCGCTTTTCTG AAGAGCAACTCCCAGGCAAAGATCGTCAAAGACTGCG CCACCATGTGTAGGCACTACAAGCGGGCCGAGAGAAA ACACATGAATATCGGCCAGTGGATTCAGTATAGGTGCG ACCGAATCGACGATGGAGGGGATTGGCGACCAATTGT GCGATTCCTGAGATACCAGGACATCGAGTTCACCGCCT TTCTGGATGCTTTCAAGAAATTTCTGAAAGGCATCCCCA AGAAGAACTGCCTGGTGCTGTACGGACCAGCTAATACA GGCAAGAGTTATTTCGGGATGTCACTGATCAGGTTTCT GAGCGGCTGTGTGATTTCCTATGTCAACTCTAAAAGTC ACTTTTGGCTGCAGCCTCTGACAGACGCCAAAGTGGG GATGATTGACGATGTCACACCAATCTGCTGGACTTACA TTGACGATTATATGCGCAACGCTCTGGACGGAAATGAT ATCTCTGTGGATGTCAAGCATCGAGCACTGGTGCAGAT CAAATGTCCACCCCTGATTCTGACAACTAACACTAATG CAGGCACCGACCCCAGGTGGCCTTACCTGCACAGCCG CCTGGTGGTCTTCCATTTTAAGAACCCTTTCCCATTTGA TGAAAACGGGAACCCCATCTATGAAATTAACAACGAGA ACTGGAAGAGTTTCTTTTCACGCACTTGGTGCAAACTG GACCTGATTCAGGAGGAAGATAAGGAGAACGACGGCG TGGATACCGGGACATTCAAGTGTAGCGCCGGCAAAAAT ACCAGAAGCATCAGGTCC	MEDPEGTEGEREGCTG WFEVEAIIEKQTGDNISE DEDENAYDSGTDLIDFI DDSNINNEQAEHEAAR ALFNAQEGEDDLHAVS AVKRKFTSSPESAGQD GVEKHGSPRAKHICVN TECVLPKRKPCHVEDS GYGNSEVEAQQMADQ VDGQNGDWQSNSSQS SGVGASNSDVSCTSIED NEENSNRTLKSIQNIMC ENSIKTTVLFKFKETYG VSFMELVRPFKSNRSS CTDWCIIGMGVTPSVAE GLKVLIQPYSIYAHLQCL TCDRGVLILLLLIRFKCGK NRLTVSKLMSQLLNIPE THMVIEPPKLRSATCAL YWYRTGLSNISEVYGTT PEWIEQQTVLQHSFDN SIFDFGEMVQWAYDHDI TDDSDIAYKYAQLADVN SNAAAFLKSNSQAKIVK DCATMCRHYKRAERKH MNIGQWIQYRCDRIDD GGDWRPIVRFLRYQDIE FTAFLDAFKKFLKGIPKK NCLVLYGPANTGKSYF GMSLIRFLSGCVISYVN SKSHFWLQPLTDAKVG MIDDVTPICWTYIDDYM RNALDGNDISVDVKHR ALVQIKCPPLILTTNTNA GTDPRWPYLHSRLVVF HFKNPFPFDENGNPIYE INNENWKSFFSRTWCK LDLIQEEDKENDGVDTG TFKCSAGKNTRSIRS	25.5%
HPV52 E2	ATGGAGTCGATACCGGCACGTTTAAATGCAGTGCAGGA AAAAATACTAGATCTATACGAAGCTGATAGTAATGACCT AAACGCACAAATTGAACATTGGAAATTGACTCGAATGG AATGTGTTTTGTTTTACAAAGCAAAGGAACTGGGAATAA CTCATATAGGCCACCAGGTGGTGCCACCAATGGCAGT GTCTAAGGCAAAGGCCTGCCAAGCTATTGAACTACAAT TGGCATTGGAGGCATTAAACAAAACACAATATAGCACA GATGGATGGACATTACAACAAACAAGTCTAGAAATGTG GCGTGCAGAACCACAAAAATACTTTAAAAAACATGGGT ATACAATAACAGTGCAATACGATAATGATAAAAACAATA CTATGGATTATACAAACTGGAAGGAAATTTATTTACTTG GTGAGTGTGAATGTACAATTGTAGAAGGACAAGTAGAT TACTATGGGTTATATTATTGGTGTGATGGAGAAAAAATA TATTTTGTAAAATTTAGTAACGATGCAAAGCAATATTGT GTAACAGGAGTATGGGAAGTACATGTGGGTGGTCAGG TAATTGTTTGTCCTGCATCTGTATCTAGTAACGAAGTAT CCACTACTGAAACTGCTGTCCACCTATGCACCGAAACC TCCAAGACCTCCGCAGTGTCCGTGGGTGCCAAAGACA CACACCTACAACCACCACAGAAACGACGACGACCAGA CGTCACAGACTCCAGAAACACCAAGTACCCCAACAACC TTTTGCGGGGACAACAATCCGTGGACAGTACTACACGG GGACTCGTCACTGCAACTGAGTGCACAAACAAAGGAC GGGTTGCACATACAACTTGTACTGCACCTATAATACAC CTAAAAGGTGATCCTAATAGTTTAAAATGTTTAAGATAT AGGGTAAAAACACATAAAAGTTTGTATGTTCAAATTTCA TCTACCTGGCATTGGACCAGTAATGAATGTACAAATAAT AAACTAGGTATTGTAACAATAACGTACAGTGATGAAACA CAACGTCAACAATTTTTAAAAACTGTTAAAATACCAAAT ACTGTGCAAGTTATACAAGGTGTCATGTCATTGTGA	ATGGAAAGTATCCCCGCAAGACTGAACGCCGTGCAGG AAAAAATTCTGGACCTGTATGAAGCCGACTCAAATGAT CTGAACGCCCAGATCGAGCACTGGAAGCTGACTCGCA TGGAATGCGTGCTGTTCTATAAGGCCAAAGAGCTGGGA ATCACACACATTGGCCATCAGGTGGTCCCCCCTATGGC AGTGAGCAAGGCCAAGGCCTGCCAGGCCATTGAGCTG CAGCTGGCACTGGAAGCCCTGAACAAAACTCAGTACTC CACCGACGGCTGGACACTGCAGCAGACTTCTCTGGAG ATGTGGCGAGCTGAACCACAGAAGTACTTTAAGAAACA CGGGTATACCATCACAGTGCAGTACGACAACGATAAGA ACAACACAATGGACTACACAAACTGGAAGGAAATCTAC CTGCTGGGGGAGTGCGAATGTACCATTGTGGAAGGGC AGGTGGACTACTATGGACTGTACTATTGGTGCGATGGG GAGAAAATCTACTTCGTGAAGTTCAGCAACGACGCCAA GCAGTACTGCGTGACCGGAGTCTGGGAAGTGCATGTC GGCGGGCAGGTCATCGTCTGTCCCGCTTCAGTGAGCT CCAATGAGGTCAGCACCACAGAAACAGCAGTGCACCT GTGTACTGAGACCTCAAAGACTAGCGCTGTGTCCGTCG GCGCAAAAGATACCCATCTGCAGCCACCCCAGAAGCG GAGAAGGCCCGACGTGACAGATAGCCGGAATACTAAA TATCCTAACAATCTGCTGAGAGGCCAGCAGTCTGTGGA CAGTACTACCAGGGGGCTGGTCACAGCCACTGAGTGC ACAAACAAGGGAAGGGTGGCCCACACAACTTGTACTG CTCCTATCATTCATCTGAAGGGCGATCCAAATAGTCTG AAATGCCTGCGCTATCGAGTGAAGACCCACAAATCACT GTACGTCCAGATCAGCAGCACCTGGCATTGGACCAGC AACGAGTGTACCAACAATAAGCTGGGAATCGTGACCAT TACATACTCCGACGAAACACAGCGGCAGCAGTTCCTGA AGACCGTGAAAATCCCTAATACAGTGCAGGTCATTCAG GGCGTCATGTCTCTG	MESIPARLNAVQEKILDL YEADSNDLNAQIEHWK LTRMECVLFYKAKELGI THIGHQVVPPMAVSKA KACQAIELQLALEALNK TQYSTDGWTLQQTSLE MWRAEPQKYFKKHGYT ITVQYDNDKNNTMDYT NWKEIYLLGECECTIVE GQVDYYGLYYWCDGE KIYFVKFSNDAKQYCVT GVWEVHVGGQVIVCPA SVSSNEVSTTETAVHLC TETSKTSAVSVGAKDTH LQPPQKRRRPDVTDSR NTKYPNNLLRGQQSVD STTRGLVTATECTNKG RVAHTTCTAPIIHLKGDP NSLKCLRYRVKTHKSLY VQISSTWHWTSNECTN NKLGIVTITYSDETQRQ QFLKTVKIPNTVQVIQG VMSL	25.1%
HPV52 E4	TTGTTTGTCCTGCATCTGTATCTAGTAACGAAGTATCCA CTACTGAAACTGCTGTCCACCTATGCACCGAAACCTCC AAGACCTCCGCAGTGTCCGTGGGTGCCAAAGACACAC ACCTACAACCACCACAGAAACGACGACGACCAGACGT CACAGACTCCAGAAACACCAAGTACCCCAACAACCTTT TGCGGGGACAACAATCCGTGGACAGTACTACACGGGG ACTCGTCACTGCAACTGAGTGCACAAACAAAGGACGG GTTGCACATACAACTTGTACTGCACCTATAA	ATGTTCGTCCTGCACCTGTACCTGGTCACTAAATACCC ACTGCTGAAACTGCTGTCAACTTACGCACCTAAACCTC CTCGCCCACCCCAGTGCCCCTGGGTGCCTAAGACTCA CACCTACAACCACCATCGGAATGACGATGACCAGACAT CTCAGACTCCAGAGACCCCCAGTACACCTACCACATTC TGTGGCGATAACAATCCCTGGACTGTGCTGCATGGGG ACAGCTCCCTGCAGCTGTCCGCACAGACAAAAGATGG CCTGCACATCCAGCTGGTCCTGCATCTG	LFVLHLYLVTKYPLLKLL STYAPKPPRPPQCPWV PKTHTYNHHRNDDDQT SQTPETPSTPTTFCGD NNPWTVLHGDSSLQLS AQTKDGLHIQLVLHL	23.1%
HPV52 E5	ATGTTAGGATTATTTGTATTTTGTTTTATTTTGCTTATGG TGTTTTGTGCAGTGCTTAGGCCGCTCTTGCTATCTATAT CGGTGTATGCGCAGGTGTTGGTGCTGGTGCTTTTGCTA TGGGTATCTATTGGGTCACCATTTAAAGTGTTTTTTTTG TACCTACTGTTTTTATATTTTCCAATGTTTTGTATTCACT GTCATGCACAGTATTTGGCACAACTGCAATAA	ATGCTGGGACTGTTTGTGTTCTGCTTTATTCTGCTGATG GTGTTTTGTGCCGTGCTGAGACCCCTGCTGCTGAGTAT TAGCGTGTATGCCCAGGTGCTGGTCCTGGTGCTGCTG CTGTGGGTCAGCATCGGCTCCCCCTTTAAGGTGTTCTT TCTGTACCTGCTGTTCCTGTATTTTCCTATGTTCTGCAT TCACTGTCATGCCCAGTACCTGGCTCAGCTGCAG	MLGLFVFCFILLMVFCA VLRPLLLSISVYAQVLVL VLLLWVSIGSPFKVFFL YLLFLYFPMFCIHCHAQ YLAQLQ	23.0%

HPV52 E6	ATGTTTGAGGATCCAGCAACACGACCCCGGACCCTGC ACGAATTGTGTGAGGTGCTGGAAGAATCGGTGCATGAA ATAAGGCTGCAGTGTGTGCAGTGCAAAAAAGAGCTACA ACGAAGAGAGGTATACAAGTTTCTATTTACAGATTTACG AATAGTATATAGAGACAATAATCCATATGGCGTGTGTAT TATGTGCCTACGCTTTTTATCTAAGATAAGTGAATATAG GCATTATCAATATTCACTGTATGGGAAAACATTAGAAGA GAGGGTAAAAAAACCATTAAGTGAAATAACTATTAGATG TATAATTTGTCAAACGCCATTATGTCCTGAAGAAAAAGA AAGACATGTTAATGCAAACAAGCGATTTCATAATATTAT GGGTCGTTGGACAGGGCGCTGTTCAGAGTGTTGGAGA CCCCGACCTGTGACCCAAGTGTAA	ATGTTTGAAGACCCCGCTACAAGACCAAGAACCCTGCA TGAACTGTGCGAAGTGCTGGAGGAATCCGTCCACGAA ATCAGACTGCAGTGCGTGCAGTGTAAGAAAGAGCTGC AGCGGAGAGAAGTCTACAAGTTCCTGTTTACAGACCTG CGAATCGTGTACCGGGATAACAATCCTTATGGAGTCTG CATCATGTGTCTGAGGTTCCTGAGCAAGATTTCCGAGT ACCGCCACTACCAGTATTCTCTGTATGGCAAAACCCTG GAGGAACGGGTGAAGAAACCCCTGAGTGAGATCACCA TTAGATGCATCATTTGTCAGACACCACTGTGCCCCGAG GAAAAGGAACGCCACGTGAACGCCAACAAGCGATTTC ATAACATTATGGGCAGATGGACTGGGAGGTGCTCCGA ATGTTGGAGGCCCCGCCCTGT	MFEDPATRPRTLHELC EVLEESVHEIRLQCVQC KKELQRREVYKFLFTDL RIVYRDNNPYGVCIMCL RFLSKISEYRHYQYSLY GKTLEERVKKPLSEITIR CIICQTPLCPEEKERHV NANKRFHNIMGRWTGR CSECWRPRPVTQV	27.5%
HPV52 E7	ATGCGTGGAGACAAAGCAACTATAAAAGATTATATATTA GATCTGCAACCTGAAACAACTGACCTACACTGCTATGA GCAATTAGGTGACAGCTCAGATGAGGAGGATACAGAT GGTGTGGACCGGCCAGATGGACAAGCAGAACAAGCCA CAAGCAATTACTACATTGTGACATATTGTCACAGTTGTG ATAGCACACTACGGCTATGCATTCATAGCACTGCGACG GACCTTCGTACTCTACAGCAAATGCTGTTGGGCACATT ACAAGTTGTGTGCCCCGGCTGTGCACGGCTATAA	ATGAGAGGAGACAAAGCCACCATCAAGGATTACATTCT GGACCTGCAGCCTGAGACAACTGACCTGCATTGCTATG AACAGCTGGGGGACAGCTCCGATGAGGAAGACACCGA TGGAGTGGACAGGCCAGATGGACAGGCAGAGCAGGCT ACTAGCAACTACTATATCGTCACCTACTGCCACTCTTGT GACAGTACACTGCGGCTGTGCATTCATTCTACCGCAAC AGATCTGAGAACACTGCAGCAGATGCTGCTGGGAACT CTGCAGGTGGTCTGCCCTGGCTGTGCCCGGCTG	MRGDKATIKDYILDLQP ETTDLHCYEQLGDSSD EEDTDGVDRPDGQAEQ ATSNYYIVTYCHSCDST LRLCIHSTATDLRTLQQ MLLGTLQVVCPGCARL	23.7%
HPV52 L1	ATGTCCGTGTGGCGGCCTAGTGAGGCCACTGTGTACC TGCCTCCTGTACCTGTCTCTAAGGTTGTAAGCACTGAT GAGTATGTGTCTCGCACAAGCATCTATTATTATGCAGG CAGTTCTCGATTACTAACAGTAGGACATCCCTATTTTTC TATTAAAAACACCAGTAGTGGTAATGGTAAAAAAGTTTT AGTTCCCAAGGTGTCTGGCCTGCAATACAGGGTATTTA GAATTAAATTGCCGGACCCTAATAAATTTGGTTTTCCAG ATACATCTTTTTATAACCCAGAAACCCAAAGGTTGGTGT GGGCCTGTACAGGCTTGGAAATTGGTAGGGGACAGCC TTTAGGTGTGGGTATTAGTGGGCATCCTTTATTAAACAA GTTTGATGATACTGAAACCAGTAACAAATATGCTGGTAA ACCTGGTATAGATAATAGGGAATGTTTATCTATGGATTA TAAGCAGACTCAGTTATGCATTTTAGGATGCAAACCTC CTATAGGTGAACATTGGGGTAAGGGAACCCCTTGTAAT AATAATTCAGGAAATCCTGGGGATTGTCCTCCCCTACA GCTCATTAACAGTGTAATACAGGATGGGGACATGGTAG ATACAGGATTTGGTTGCATGGATTTTAATACCTTGCAAG CTAGTAAAAGTGATGTGCCCATTGATATATGTAGCAGT GTATGTAAGTATCCAGATTATTTGCAAATGGCTAGCGA GCCATATGGTGACAGTTTGTTCTTTTTTCTTAGACGTGA GCAAATGTTTGTTAGACACTTTTTTAATAGGGCCGGTAC CTTAGGTGACCCTGTGCCAGGTGATTTATATATACAAG GGTCTAACTCTGGCAATACTGCCACTGTACAAAGCAGT GCTTTTTTTCCTACTCCTAGTGGTTCTATGGTAACCTCA GAATCCCAATTATTTAATAAACCGTACTGGTTACAACGT GCGCAGGGCCACAATAATGGCATATGTTGGGGCAATC AGTTGTTTGTCACAGTTGTGGATACCACTCGTAGCACT AACATGACTTTATGTGCTGAGGTTAAAAAGGAAAGCAC ATATAAAAATGAAAATTTTAAGGAATACCTTCGTCATGG CGAGGAATTTGATTTACAATTTATTTTTCAATTGTGCAAA ATTACATTAACAGCTGATGTTATGACATACATTCATAAG ATGGATGCCACTATTTTAGAGGACTGGCAATTTGGCCT TACCCCACCACCGTCTGCATCTTTGGAGGACACATACA GATTTGTCACTTCTACTGCTATAACTTGTCAAAAAAACA CACCACCTAAAGGAAAGGAAGATCCTTTAAAGGACTAT ATGTTTTGGGAGGTGGATTTAAAAGAAAAGTTTTCTGCA GATTTAGATCAGTTTCCTTTAGGTAGGAAGTTTTTGTTA CAGGCAGGGCTACAGGCTAGGCCCAAACTAAAACGCC CTGCATCATCGGCCCCACGTACCTCCACAAAGAAGAAA AAGGTTAAAAGGTAA	ATGGTCCAGATCCTGTTTTATATCCTGGTCATTTTCTAT TATGTCGCCGGGGTCAACGTCTTTCACATTTTCCTGCA GATGAGCGTCTGGAGGCCTAGCGAGGCCACCGTGTAT CTGCCCCCTGTGCCAGTCTCTAAGGTGGTCAGTACAGA CGAATACGTGAGCCGGACTTCCATCTACTATTACGCTG GAAGCTCCAGACTGCTGACAGTGGGCCACCCCTACTTT TCTATTAAGAATACTTCTAGTGGCAACGGGAAGAAAGT GCTGGTCCCTAAAGTGAGTGGACTGCAGTATAGGGTCT TTCGCATCAAGCTGCCAGACCCCAACAAGTTTGGCTTC CCAGATACCAGCTTCTACAACCCCGAGACACAGAGGCT GGTGTGGGCTTGCACCGGCCTGGAAATCGGACGAGGA CAGCCACTGGGGGTCGGAATTAGTGGGCACCCTCTGC TGAATAAGTTCGACGATACTGAGACCTCAAACAAGTAT GCCGGGAAACCTGGAATTGACAATCGCGAATGTCTGA GCATGGACTACAAACAGACCCAGCTGTGCATCCTGGG CTGTAAGCCACCCATTGGGGAGCATTGGGGCAAAGGG ACACCTTGCAACAATAACAGCGGCAATCCAGGGGACT GTCCTCCACTGCAGCTGATCAACTCCGTGATTCAGGAC GGCGATATGGTGGACACCGGATTTGGCTGCATGGATTT CAACACACTGCAGGCTAGCAAGTCCGACGTGCCCATC GATATTTGCTCAAGCGTCTGTAAATATCCAGACTACCTG CAGATGGCATCAGAGCCCTATGGCGATAGCCTGTTCTT TTTCCTGCGGAGAGAACAGATGTTCGTGCGACACTTTT TCAATCGAGCAGGAACCCTGGGCGACCCTGTCCCAGG GGATCTGTACATCCAGGGGTCTAATAGTGGAAACACAG CTACTGTGCAGTCCTCTGCATTTTTCCCCACTCCTTCAG GAAGCATGGTCACCTCCGAGTCTCAGCTGTTTAACAAG CCCTATTGGCTGCAGCGAGCACAGGGCCATAATAACG GGATTTGCTGGGGAAATCAGCTGTTCGTGACTGTGGTC GATACCACACGCTCCACCAACATGACACTGTGTGCCGA GGTGAAGAAAGAATCTACATACAAGAACGAGAACTTCA AGGAATACCTGAGGCACGGCGAGGAGTTCGACCTGCA GTTTATCTTCCAGCTGTGCAAGATTACCCTGACAGCCG ATGTGATGACATACATCCATAAAATGGACGCTACTATTC TGGAGGATTGGCAGTTTGGCCTGACTCCCCCTCCAAGT GCATCACTGGAAGACACCTATCGGTTCGTGACTTCTAC CGCCATCACTTGTCAGAAGAATACCCCCCCTAAGGGGA AAGAGGACCCACTGAAAGATTACATGTTTTGGGAGGTG GATCTGAAGGAAAAATTCAGCGCCGACCTGGATCAGTT TCCCCTGGGGAGAAAGTTCCTGCTGCAGGCAGGACTG CAGGCCAGACCAAAGCTGAAAAGGCCCGCCAGTTCAG CTCCTCGCACAAGCACTAAGAAAAAGAAAGTGAAGCGA	MSVWRPSEATVYLPPV PVSKVVSTDEYVSRTSI YYYAGSSRLLTVGHPYF SIKNTSSGNGKKVLVPK VSGLQYRVFRIKLPDPN KFGFPDTSFYNPETQRL VWACTGLEIGRGQPLG VGISGHPLLNKFDDTET SNKYAGKPGIDNRECLS MDYKQTQLCILGCKPPI GEHWGKGTPCNNNSG NPGDCPPLQLINSVIQD GDMVDTGFGCMDFNTL QASKSDVPIDICSSVCK YPDYLQMASEPYGDSL FFFLRREQMFVRHFFN RAGTLGDPVPGDLYIQ GSNSGNTATVQSSAFF PTPSGSMVTSESQLFN KPYWLQRAQGHNNGIC WGNQLFVTVVDTTRST NMTLCAEVKKESTYKN ENFKEYLRHGEEFDLQ FIFQLCKITLTADVMTYI HKMDATILEDWQFGLT PPPSASLEDTYRFVTST AITCQKNTPPKGKEDPL KDYMFWEVDLKEKFSA DLDQFPLGRKFLLQAGL QARPKLKRPASSAPRT STKKKKVKR	31.1%
HPV52 L2	ATGAGATACAGACGGTCTACACGGCACAAACGTGCTTC TGCAACACAGCTATATCAAACATGCAAAGCCTCTGGCA CCTGCCCCCCCCGATGTTATTCCTAAAGTGGAAGGCACA ACTATTGCAGATCAACTTTTAAAATATGGCAGCCTAGG GGTGTTTTTTGGAGGTTTGGGTATAGGTACAGGTGCAG GCTCTGGTGGTAGGGCAGGCTATGTGCCATTGTCCAC TCGTCCTCCCACTAGTAGTATTACCACGTCCACCATTC GTCCCCCTGTAACTGTAGAACCCATTGGTCCCTTAGAA CCATCTATAGTTTCTATGATAGAAGAAACAACATTTATT GAGTCTGGCGCACCTGCTCCATCTATTCCATCAGCAAC AGGGTTTGATGTTACAACATCTGCAAATAATACTCCTGC AATAATTAATGTAACATCTATAGGTGAATCATCTGTACA ATCAGTTTCTACACATTTAAATCCTACATTCACTGAACC ATCTATAATACAGCCCCCGGCACCTGCAGAAGCATCTG GTCATGTATTGTTTTCTAGTCCAACTATTAGTACACACA CCTATGAAGAAATCCCTATGGATACATTTGTTACCTCTA CTGACAGCAGCAGTGTAACAAGTAGTACACCTATTCCA GGGTCTCGCCCTACGACACGCCTTGGTTTATATAGCCG TGCCACACAACAGGTTAAGGTAGTCGACCCTGCTTTTA TGTCATCACCACAGAAATTAGTAACATATAACAATCCTG TTTTTGAGGGCGTTGATACAGATGAAACTATAATTTTTG ATCGTTCACAACTTTTACCTGCACCGGATCCTGATTTTT TAGACATTATAGCTTTGCATAGGCCTGCATTAACCTCTC GAAGAGGTACTGTTAGGTTTAGCAGGCTTGGTAATAAG GCCACCCTACGTACACGTAGTGGAAAACAAATTGGGG CACGGGTACATTATTATCATGATATTAGTCCTATCCAGC CTGCTGAAGTTCAGGAAGACATAGAATTGCAACCTTTA TTACCACAGTCTGTGTCCCCTTACACTATTAATGATGGT TTGTATGATGTGTATGCAGATTCTTTGCAGCAACCCAC GTTTCACTTACCTTCCACACTTTCTACCCATAATAATAC TTTCACTGTACCTATTAATAGTGGTATTGACTTTGTATAT CAACCCACTATGTCCATTGAGTCAGGTCCTGACATTCC ATTACCTTCGTTACCCACACATACTCCTTTTGTTCCTAT AGCCCCTACAGCTCCATCTACATCTATTATTGTTGATGG TACAGATTTTATTTTACATCCTAGTTATTTTTTACTACGT CGCAGGCGTAAACGTTTTCCATATTTTTTTTACAGATGTC CGTGTGGCGGCCTAG	ATGAGGTATCGGCGAAGCACTCGGCATAAAAGGGCAT CCGCAACCCAGCTGTATCAGACTTGTAAGGCATCCGG CACTTGTCCCCCAGATGTGATCCCCAAGGTCGAAGGC ACCACAATTGCTGACCAGCTGCTGAAATACGGAAGCCT GGGCGTGTTCTTTGGAGGACTGGGAATCGGAACCGGA GCTGGGTCAGGAGGCAGGGCAGGATATGTGCCTCTGA GCACACGCCCCCCTACTAGCTCCATCACTACCTCTACA ATTAGGCCACCCGTGACTGTCGAACCCATCGGCCCTCT GGAGCCATCAATCGTGAGCATGATTGAGGAAACAACTT TCATCGAAAGCGGAGCACCAGCACCTTCCATTCCATCT GCCACCGGATTTGATGTGACCACATCCGCCAACAATAC CCCTGCTATCATTAACGTCACATCTATCGGGGAGTCTA GTGTGCAGAGTGTCTCAACCCACCTGAATCCAACATTC ACTGAACCCAGTATCATTCAGCCTCCAGCTCCCGCAGA GGCCTCAGGACACGTGCTGTTCTCAAGCCCAACTATCA GCACCCATACATACGAGGAAATTCCTATGGACACCTTT GTGACTAGCACCGACTCCTCTAGTGTCACTTCAAGCAC CCCAATCCCAGGCTCCCGGCCAACTACCAGACTGGGG CTGTACTCTAGGGCCACCCAGCAGGTGAAGGTGGTGG ACCCCGCTTTTATGTCCTCTCCTCAGAAACTGGTGACA TATAACAATCCCGTGTTCGAAGGCGTGGACACAGATGA GACTATCATTTTTGATCGGTCCCAGCTGCTGCCTGCAC CAGACCCCGATTTCCTGGACATCATTGCACTGCATAGA CCCGCCCTGACTTCTCGGAGAGGCACCGTGCGGTTCA GCCGGCTGGGAAACAAGGCAACCCTGAGAACAAGGAG TGGGAAACAGATCGGAGCCCGCGTGCACTACTATCAT GATATCAGCCCCATTCAGCCTGCTGAGGTGCAGGAAG ACATCGAGCTGCAGCCACTGCTGCCACAGAGCGTGTC CCCTTACACAATTAACGACGGCCTGTACGATGTCTATG CAGACTCTCTGCAGCAGCCTACTTTCCACCTGCCAAGT ACTCTGTCAACCCATAACAATACATTCACTGTGCCAATC AATAGCGGCATTGATTTTGTCTATCAGCCAACCATGAG CATCGAGTCCGGGCCCGACATTCCTCTGCCATCCCTG CCTACCCACACACCTTTCGTGCCAATCGCTCCCACCGC ACCTTCTACAAGTATCATTGTGGACGGGACCGATTTCA TTCTGCATCCTAGCTACTTTCTGCTGAGGCGCCGACGG AAGCGATTTCCATATTTCTTTACAGATGTGCGCGTCGC CGCT	MRYRRSTRHKRASATQ LYQTCKASGTCPPDVIP KVEGTTIADQLLKYGSL GVFFGGLGIGTGAGSG GRAGYVPLSTRPPTSSI TTSTIRPPVTVEPIGPLE PSIVSMIEETTFIESGAP APSIPSATGFDVTTSAN NTPAIINVTSIGESSVQS VSTHLNPTFTEPSIIQPP APAEASGHVLFSSPTIS THTYEEIPMDTFVTSTD SSSVTSSTPIPGSRPTT RLGLYSRATQQVKVVD PAFMSSPQKLVTYNNP VFEGVDTDETIIFDRSQL LPAPDPDFLDIIALHRPA LTSRRGTVRFSRLGNK ATLRTRSGKQIGARVHY YHDISPIQPAEVQEDIEL QPLLPQSVSPYTINDGL YDVYADSLQQPTFHLP STLSTHNNTFTVPINSGI DFVYQPTMSIESGPDIP LPSLPTHTPFVPIAPTAP STSIIVDGTDFILHPSYFL LRRRRKRFPYFFTDVR VAA	27.8%

HPV58 E1	ATGGATGACCCTGAAGGTACAAACGGGGTAGGGGCGG GCTGTACTGGCTGGTTTGAGGTAGAAGCGGTAATAGAA CGAAGAACAGGAGATAATATTTCAGATGATGAGGACGA AACAGCAGACGATAGTGGTACAGATTTAATAGAGTTTAT AGATGATTCAGTACAAAGTACTACACAGGCAGAAGCAG AGGCAGCCCGAGCGTTGTTTAATGTACAGGAAGGGGT GGACGATATAAATGCTGTGTGTGCACTAAAACGAAAGT TTGCAGCATGCTCAGAAAGTGCTGTAGAGGACTGTGTG GACCGGGCTGCAAATGTGTGTGTATCGTGGAAATATAA AAATAAAGAATGCACACACAGAAAACGAAAAATTATTGA GCTAGAAGACAGCGGATATGGCAATACTGAAGTGGAAA CTGAGCAGATGGCACACCAGGTAGAAAGCCAAAATGG CGACGCAGACTTAAATGACTCGGAGTCTAGTGGGGTG GGGGCTAGTTCAGATGTAAGCAGTGAAACGGATGTAG ACAGTTGTAATACTGTTCCATTACAAAATATTAGTAATAT TCTACATAACAGTAATACTAAAGCAACGCTATTATATAA ATTCAAAGAAGCTTATGGAGTAAGTTTTATGGAATTAGT TAGACCATTTAAAAGTGATAAAACAAGCTGTACAGATTG GTGTATAACAGGGTATGGAATAAGTCCCTCCGTAGCAG AAAGTTTAAAAGTACTAATTAAACAGCACAGTATATATA CACACCTACAATGTTTAACGTGTGACAGAGGAATTATAT TATTATTGTTAATTAGATTTAAATGTAGCAAAAATAGATT AACTGTGGCAAAATTAATGAGTAATTTACTATCAATTCC TGAAACATGTATGATTATCGAGCCACCAAAATTACGAA GTCAAGCATGTGCCTTATATTGGTTTAGAACAGCAATGT CAAATATAAGTGATGTGCAAGGGACAACACCAGAATGG ATAGATAGATTAACAGTGTTACAGCATAGCTTTAATGAT GATATATTTGATTTAAGTGAAATGATACAATGGGCATAT GATAATGACATTACAGATGATAGTGACATTGCATATAAA TATGCACAGTTAGCAGATGTTAATAGTAATGCAGCAGC ATTTTTAAGAAGCAATGCACAAGCAAAAATAGTAAAAGA CTGTGGCGTTATGTGCAGACATTATAAAAGAGCAGAAA AGCGTGGTATGACAATGGGACAATGGATACAAAGTAGG TGTGAAAAAACAAATGATGGAGGTAATTGGAGACCAAT AGTACAATTTTTTAAGATATCAAAATATTGAATTTACAGCA TTTTTAGTTGCATTTAAACAGTTTTTACAAGGTGTACCA AAAAAAAGTTGTATGTTACTGTGTGGCCCAGCAAATAC AGGGAAATCATATTTTGGAATGAGTTTAATACATTTTTTA AAAGGATGCATTATTTCATATGTAAATTCCAAAAGTCAT TTTTGGTTGCAGCCATTATCAGATGCTAAACTAGGTATG ATAGATGATGTAACAGCCATAAGCTGGACATATATAGAT GATTATATGAGAAATGCATTAGATGGTAACGACATTTCA ATAGATGTAAAACATAGGGCATTAGTACAATTAAAATGT CCACCATTAATAATTACCTCAAATACAAATGCAGGCAAA GATTCACGATGGCCATATTTGCACAGTAGACTAACAGT ATTTGAATTTAACAATCCATTTCCATTTGATGCAAATGG TAATCCAGTGTATAAAATAAATGATGAAAATTGGAAATC CTTTTTCTCAAGGACGTGGTGCAAATTAGGCTTAATAGA GGAAGAGGACAAGGAAAACGATGGAGGAAATATCAGC ACGTTTAAGTGCAGTGCAGGACAAAATCCTAGACATAT ACGAAGCTGA	ATGGACGATCCTGAGGGAACTAACGGGGTGGGGGCTG GCTGCACTGGCTGGTTTGAGGTGGAGGCTGTCATTGA AAGAAGAACTGGCGACAACATCTCCGACGATGAAGATG AGACTGCTGACGATTCTGGGACCGACCTGATCGAGTTC ATTGACGATTCTGTGCAGAGTACCACACAGGCAGAAGC TGAGGCAGCTCGCGCCCTGTTCAACGTGCAGGAAGGA GTGGACGATATCAATGCCGTGTGTGCTCTGAAGAGGAA ATTTGCAGCCTGCTCCGAGTCTGCTGTGGAAGACTGTG TCGATCGCGCTGCAAACGTGTGCGTCTCCTGGAAGTA CAAAAATAAGGAGTGCACCCACCGGAAAAGAAAGATCA TTGAACTGGAGGATTCTGGGTATGGAAACACAGAAGTG GAGACTGAACAGATGGCACATCAGGTCGAGAGTCAGA ACGGCGACGCCGATCTGAATGACTCAGAAAGCTCCGG CGTGGGGGCCTCTAGTGATGTCTCAAGCGAGACCGAC GTGGATTCCTGTAATACAGTCCCTCTGCAGAACATCTC CAATATTCTGCACAACTCTAATACAAAGGCCACTCTGCT GTACAAATTCAAGGAGGCTTATGGCGTGTCTTTCATGG AACTGGTCAGACCCTTCAAGAGTGACAAGACCTCATGC ACAGATTGGTGTATCACAGGATACGGCATTAGTCCCTC AGTGGCCGAGTCTCTGAAAGTCCTGATCAAGCAGCACA GTATCTACACACATCTGCAGTGCCTGACTTGTGACAGG GGGATCATTCTGCTGCTGCTGATCAGGTTCAAATGCAG CAAGAACCGCCTGACAGTGGCCAAACTGATGAGCAAT CTGCTGTCCATTCCCGAGACTTGTATGATCATTGAACC ACCTAAGCTGCGCAGCCAGGCATGCGCACTGTACTGG TTTCGAACCGCCATGTCAAACATCAGCGACGTGCAGG GCACTACCCCTGAGTGGATTGATCGGCTGACAGTCCT GCAGCACTCATTCAACGACGATATCTTTGACCTGAGCG AAATGATTCAGTGGGCCTACGACAATGATATCACCGAC GATAGCGACATTGCTTACAAATATGCACAGCTGGCCGA TGTGAACAGCAATGCCGCTGCATTCCTGCGATCCAACG CTCAGGCAAAAATCGTGAAGGACTGCGGCGTCATGTG CCGGCACTACAAACGGGCCGAGAAGAGAGGGATGACT ATGGGACAGTGGATTCAGAGCCGGTGCGAAAAGACCA ACGATGGCGGGAATTGGCGACCAATCGTGCAGTTTCT GCGGTATCAGAATATTGAGTTCACAGCTTTTCTGGTGG CATTCAAACAGTTTCTGCAGGGCGTCCCCAAGAAATCC TGCATGCTGCTGTGTGGCCCTGCCAACACTGGGAAGT CTTACTTCGGAATGAGTCTGATCCACTTTCTGAAAGGAT GTATCATTAGCTATGTGAATAGCAAGTCCCATTTCTGGC TGCAGCCCCTGTCCGACGCTAAGCTGGGCATGATCGA CGATGTGACCGCAATCTCTTGGACATACATTGACGATT ATATGCGGAACGCACTGGACGGGAATGATATCAGTATT GACGTGAAACACAGAGCCCTGGTCCAGCTGAAGTGCC CACCCCTGATCATTACTAGCAACACCAATGCTGGAAAG GATAGTAGATGGCCTTACCTGCATTCAAGGCTGACCGT GTTCGAGTTTAACAATCCTTTCCCATTTGACGCAAACG GCAACCCAGTGTACAAAATCAACGATGAAAACTGGAAG AGCTTCTTCAGCCGGACTTGGTGTAAACTGGGCCTGAT CGAGGAAGAGGACAAGGAGAACGATGGAGGCAATATT TCAACCTTTAAGTGCAGCGCAGGACAGAACCCAAGGC ACATCCGCAGC	MDDPEGTNGVGAGCT GWFEVEAVIERRTGDNI SDDEDETADDSGTDLIE FIDDSVQSTTQAEAEAA RALFNVQEGVDDINAVC ALKRKFAACSESAVED CVDRAANVCVSWKYKN KECTHRKRKIIELEDSG YGNTEVETEQMAHQVE SQNGDADLNDSESSGV GASSDVSSETDVDSCN TVPLQNISNILHNSNTKA TLLYKFKEAYGVSFMEL VRPFKSDKTSCTDWCIT GYGISPSVAESLKVLIKQ HSIYTHLQCLTCDRGIIL LLLIRFKCSKNRLTVAKL MSNLLSIPETCMIIEPPK LRSQACALYWFRTAMS NISDVQGTTPEWIDRLT VLQHSFNDDIFDLSEMI QWAYDNDITDDSDIAYK YAQLADVNSNAAAFLR SNAQAKIVKDCGVMCR HYKRAEKRGMTMGQWI QSRCEKTNDGGNWRPI VQFLRYQNIEFTAFLVA FKQFLQGVPKKSCMLL CGPANTGKSYFGMSLI HFLKGCIISYVNSKSHF WLQPLSDAKLGMIDDV TAISWTYIDDYMRNALD GNDISIDVKHRALVQLK CPPLIITSNTNAGKDSR WPYLHSRLTVFEFNNP FPFDANGNPVYKINDEN WKSFFSRTWCKLGLIEE EDKENDGGNISTFKCSA GQNPRHIRS	26.5%
HPV58 E2	ATGGAGGAAATATCAGCACGTTTAAGTGCAGTGCAGGA CAAAATCCTAGACATATACGAAGCTGATAAAAATGATTT AACATCACAAATTGAACATTGGAAACTAATACGCATGGA GTGTGCTATAATGTATACAGCCAGACAAATGGGAATAT CACATTTGTGCCACCAGGTGGTGCCGTCATTGGTAGCA TCAAAGACTAAAGCGTTTCAAGTAATTGAACTGCAAATG GCATTAGAGACATTAAATGCATCACCATATAAAACAGAT GAATGGACATTGCAACAAACAAGCTTAGAAGTGTGGTT ATCAGAGCCACAAAAATGCTTTAAAAAAAAAGGCATAAC AGTAACTGTACAATATGACAATGATAAAGCAAACACAAT GGATTATACAAATTGGAGTGAAATATATATTATTGAGGA AACAACATGTACTTTGGTAGCAGGAGAAGTTGACTATG TGGGGTTGTATTATATACATGGCAATGAAAAGACGTATT TTAAATATTTTAAAGAGGATGCAAAAAAGTACTCTAAAA CACAATTATGGGAGGTACATGTGGGTAGTCGGGTAATT GTATGTCCTACATCTATACCTAGTGATCAAATATCCACT ACTGAAACTGCTGACCCAAAGACCACCGAGGCCACCA ACAACGAAAGTACACAGGGGACAAAGCGACGACGACT CGATTTACCAGACTCCAGAGACAACACCCAGTACTCCA CAAAGTATACAGACTGCGCCGTGGACAGTAGACCACG AGGAGGAGGACTACACAGTACAACTAACTGTACATACA AAGGGCGGAACGTGTGTAGTTCTAAAGTTTCACCTATC GTGCATTTAAAAGGTGACCCAAATAGTTTAAAATGTTTA AGATATAGATTAAAACCATTTAAAGACTTATACTGTAATA TGTCATCCACATGGCATTGGACCAGTGATGACAAAGGT GACAAAGTAGGAATTGTTACTGTAACATACACAACGGA AACACAACGACAACTGTTTTTAAACACTGTTAAAATACC ACCCACTGTGCAAATAAGTACTGGTGTTATGTCATTGTA A	ATGGAGGAAATCTCCGCAAGACTGAGTGCCGTCAGG ACAAAATTCTGGACATCTATGAAGCCGACAAAAATGAC CTGACATCACAGATCGAGCACTGGAAACTGATTAGGAT GGAATGCGCCATCATGTACACCGCTCGCCAGATGGGC ATTTCTCACCTGTGTCATCAGGTGGTCCCCTCCCTGGT CGCATCTAAAACAAAGGCCTTCCAGGTCATCGAGCTGC AGATGGCTCTGGAAACTCTGAACGCAAGTCCCTACAAG ACCGATGAGTGGACACTGCAGCAGACTAGCCTGGAGG TCTGGCTGTCCGAACCTCAGAAATGCTTTAAGAAAAAG GGGATTACAGTGACTGTCCAGTATGACAACGATAAGGC AAATACAATGGACTACACAAACTGGAGCGAAATCTACA TCATTGAGGAAACCACATGTACCCTGGTGGCCGGAGA AGTGGATTACGTCGGCCTGTACTATATTCACGGGAACG AGAAGACATACTTCAAGTACTTCAAGGAAGACGCTAAA AAGTACTCCAAGACCCAGCTGTGGGAGGTGCATGTCG GCAGCAGAGTGATCGTCTGCCCAACCTCAATCCCCAG CGATCAGATTAGCACTACCGAAACTGCCGACCCTAAAA CAACTGAGGCTACCAACAATGAATCCACTCAGGGAACC AAGCGGAGAAGGCTGGACCTGCCAGACAGCCGGGAC AACACACAGTACAGTACAAAGTATACTGATTGTGCAGT GGACTCACGCCCCCGAGGCGGAGGACTGCACAGCAC CACAAACTGCACCTACAAAGGAAGGAATGTGTGTAGCT CCAAGGTGAGTCCTATTGTCCATCTGAAAGGCGATCCA AACTCACTGAAGTGCCTGCGGTACAGACTGAAACCATT CAAGGACCTGTATTGTAATATGTCTAGTACTTGGCATTG GACCTCCGACGATAAAGGCGATAAAGTGGGGATCGTG ACCGTCACATATACTACCGAGACCCAGCGGCAGCTGTT TCTGAATACAGTGAAGATTCCCCCTACAGTCCAGATCA GTACTGGCGTGATGTCACTG	MEEISARLSAVQDKILDI YEADKNDLTSQIEHWKL IRMECAIMYTARQMGIS HLCHQVVPSLVASKTKA FQVIELQMALETLNASP YKTDEWTLQQTSLEVW LSEPQKCFKKKGITVTV QYDNDKANTMDYTNW SEIYIIEETTCTLVAGEV DYVGLYYIHGNEKTYFK YFKEDAKKYSKTQLWE VHVGSRVIVCPTSIPSD QISTTETADPKTTEATN NESTQGTKRRRLDLPD SRDNTQYSTKYTDCAV DSRPRGGGLHSTTNCT YKGRNVCSSKVSPIVHL KGDPNSLKCLRYRLKPF KDLYCNMSSTWHWTS DDKGDKVGIVTVTYTTE TQRQLFLNTVKIPPTVQI STGVMSL	26.6%
HPV58 E4	TTGTATGTCCTACATCTATACCTAGTGATCAAATATCCA CTACTGAAACTGCTGACCCAAAGACCACCGAGGCCACC CAACAACGAAAGTACACAGGGGACAAAGCGACGACGA CTCGATTTACCAGACTCCAGAGACAACACCCAGTACTCC CACAAAGTATACAGACTGCGCCGTGGACAGTAGACCA CGAGGAGGAGGACTACACAGTACAACTAACTGTACATA CAAAGGGCGGAACGTGTGTAGTTCTAAAGTTTCACCTA TCGTGCATTTAA	ATGTATGTCCTGCACCTGTACCTGGTCATCAAGTATCC ACTGCTGAAACTGCTGACCCAGAGACCCCCACGCCCT CCAACAACTAAAGTGCATCGGGGCCAGAGCGACGATG ACTCCATCTACCAGACCCCAGAAACCACACCCTCTACA CCTCAGAGTATTCAGACAGCCCCCTGGACTGTCGATCA CGAGGAAGAGGACTATACTGTGCAGCTGACTGTCCATA CCAAGGGCGGGACATGCGTGGTCCTGAAATTCCACCT GTCCTGTATC	LYVLHLYLVIKYPLLKLL TQRPPRPPTTKVHRGQ SDDDSIYQTPETTPSTP QSIQTAPWTVDHEEED YTVQLTVHTKGGTCVVL KFHLSCI	22.1%
HPV58 E5	ATGATATTACCTATTTTTGTTGTTTGTTTTATACTGTTTTT ATGCTTGTGCATTTTTTTGCGGCCATTGGTGCTATCTAT TTCTATATATGCTTGGTTGCTGGTGTTGGTGTTGCTGCT TTGGGTGTCTGTGGGGTCGGCTCTACGAATTTTTTTCT GTTACTTAATATTTTTATATATACCAATGATGTGTATTAA TTTTCATGCACAATACTTAACCCAACAAGACTAA	ATGATTCTGCCCATTTTTGTGGTGTGCTTTATTCTGTTC CTGTGCCTGTGTATCTTTCTGCGGCCCCTGGTCCTGTC TATTTCTATTTACGCCTGGCTGCTGGTGCTGGTCCTGC TGCTGTGGGTGAGCGTCGGCTCCGCTCTGCGGATCTT CTTTTGCTACCTGATCTTCCTGTATATTCCCATGATGTG TATTAACTTTCACGCCCAGTATCTGACCCAGCAGGAC	MILPIFVVCFILFLCLCIFL RPLVLSISIYAWLLVLVL LLWVSVGSALRIFFCYLI FLYIPMMCINFHAQYLT QQD	25.5%

HPV58 E6	ATGTTCCAGGACGCAGAGGAGAAACCACGGACATTGC ATGATTTGTGTCAGGCGTTGGAGACATCTGTGCATGAA ATCGAATTGAAATGCGTTGAATGCAAAAAGACTTTGCA GCGATCTGAGGTATATGACTTTGTATTTGCAGATTTAAG AATAGTGTATAGAGATGGAAATCCATTTGCAGTATGTAA AGTGTGCTTACGATTGCTATCTAAAATAAGTGAGTATAG ACATTATAATTATTCGCTATATGGAGACACATTAGAACA AACACTAAAAAAGTGTTTAAATGAAATATTAATTAGATGT ATTATTTGTCAAAGACCATTGTGTCCACAAGAAAAAAAA AGGCATGTGGATTTAAACAAAAGGTTTCATAATATTTCG GGTCGTTGGACAGGGCGCTGTGCAGTGTGTTGGAGAC CCCGACGTAGACAAACACAAGTGTAA	ATGTTCCAGGATGCCGAAGAAAAACCCCGAACTCTGCA CGATCTGTGTCAGGCTCTGGAGACCTCTGTCCATGAGA TTGAACTGAAATGCGTGGAGTGTAAGAAAACACTGCAG CGGAGCGAAGTGTACGACTTCGTCTTTGCCGATCTGC GCATCGTCTATCGAGACGGAAACCCATTCGCTGTGTGC AAGGTCTGTCTGCGCCTGCTGAGCAAAATTTCCGAGTA CCGGCACTACAACTATAGTCTGTATGGCGATACCCTGG AGCAGACACTGAAGAAATGCCTGAATGAAATCCTGATT AGGTGCATCATTTGTCAGCGCCCCCTGTGTCCTCAGGA AAAGAAACGACACGTGGACCTGAACAAGAGGTTTCATA ATATCTCCGGCCGGTGGACTGGAAGATGCGCAGTGTG TTGGAGGCCCCGGAGAAGGCAGACCCAGGTC	MFQDAEEKPRTLHDLC QALETSVHEIELKCVEC KKTLQRSEVYDFVFADL RIVYRDGNPFAVCKVCL RLLSKISEYRHYNYSLY GDTLEQTLKKCLNEILIR CIICQRPLCPQEKKRHV DLNKRFHNISGRWTGR CAVCWRPRRRQTQV	27.7%
HPV58 E7	ATGAGAGGAAACAACCCAACGCTAAGAGAATATATTTT AGATTTACATCCTGAACCAACTGACCTATTCTGCTATGA GCAATTATGTGACAGCTCAGACGAGGATGAAATAGGCT TGGACGGGCCAGATGGACAAGCACAACCGGCCACAGC TAATTACTACATTGTAACTTGTTGTTACACTTGTGGCAC CACGGTTCGTTTGTGTATCAACAGTACAACAACCGACG TACGAACCCTACAGCAGCTGCTTATGGGCACATGTACC ATTGTGTGCCCTAGCTGTGCACAGCAATAA	ATGCGGGGAAATAATCCTACCCTGAGAGAGTACATCCT GGACCTGCACCCTGAGCCCACCGACCTGTTCTGTTAC GAGCAGCTGTGTGACAGCTCCGACGAGGATGAAATCG GCCTGGACGGACCAGATGGACAGGCACAGCCTGCAAC CGCTAACTACTATATCGTGACATGCTGTTACACTTGCG GCACCACAGTCCGGCTGTGTATTAATTCTACTACCACA GATGTGAGAACACTGCAGCAGCTGCTGATGGGGACTT GCACCATTGTCTGCCCCAGCTGTGCCCAGCAG	MRGNNPTLREYILDLHP EPTDLFCYEQLCDSSD EDEIGLDGPDGQAQPA TANYYIVTCCYTCGTTV RLCINSTTTDVRTLQQL LMGTCTIVCPSCAQQ	22.9%
HPV58 L1	ATGTCCGTGTGGCGGCCTAGTGAGGCCACTGTGTACC TGCCTCCTGTGCCTGTGTCTAAGGTTGTAAGCACTGAT GAATATGTGTCACGCACAAGCATTTATTATTATGCTGGC AGTTCCAGACTTTTGGCTGTTGGCAATCCATATTTTTCC ATCAAAAGTCCCAATAACAATAAAAAAGTATTAGTTCCC AAGGTATCAGGCTTACAGTATAGGGTCTTTAGGGTGCG TTTACCTGATCCCAATAAATTTGGTTTTCCTGATACATC TTTTTATAACCCTGATACACAACGTTTGGTCTGGGCATG TGTAGGCCTTGAAATAGGTAGGGGACAGCCATTGGGT GTTGGCGTAAGTGGTCATCCTTATTTAAATAAATTTGAT GACACTGAAACCAGTAACAGATATCCCGCACAGCCAG GGTCTGATAACAGGGAATGCTTATCTATGGATTATAAAC AAACACAATTATGTTTAATTGGCTGTAAACCTCCCACTG GTGAGCATTGGGGTAAAGGTGTTGCCTGTAACAATAAT GCAGCTGCTACTGATTGTCCTCCATTGGAACTTTTTAAT TCTATTATTGAGGATGGTGACATGGTAGATACAGGGTT TGGATGCATGGACTTTGGTACATTGCAGGCTAATAAAA GTGATGTGCCTATTGATATTTGTAACAGTACATGCAAAT ATCCAGATTATTTAAAAATGGCCAGTGAACCTTATGGG GATAGTTTGTTCTTTTTTCTTAGACGTGAGCAGATGTTT GTTAGACACTTTTTTAATAGGGCTGGAAAACTTGGCGA GGCTGTCCCGGATGACCTTTATATTAAAGGGTCCGGTA ATACTGCAGTTATCCAAAGTAGTGCATTTTTTCCAACTC CTAGTGGCTCTATAGTTACCTCAGAATCACAATTATTTA ATAAGCCTTATTGGCTACAGCGTGCACAAGGTCATAAC AATGGCATTTGCTGGGGCAATCAGTTATTTGTTACCGT GGTTGATACCACTCGTAGCACTAATATGACATTATGCA CTGAAGTAACTAAGGAAGGTACATATAAAAATGATAATT TTAAGGAATATGTACGTCATGTTGAAGAATATGACTTAC AGTTTGTTTTTCAGCTTTGCAAAATTACACTAACTGCAG AGATAATGACATATATACATACTATGGATTCCAATATTTT GGAGGACTGGCAATTTGGTTTAACACCTCCTCCGTCTG CCAGTTTACAGGACACATATAGATTTGTTACCTCCCAG GCTATTACTTGCCAAAAAACAGCACCCCCTAAAGAAAA GGAAGATCCATTAAATAAATATACTTTTTGGGAGGTTAA CTTAAAGGAAAAGTTTTCTGCAGATCTAGATCAGTTTCC TTTGGGACGAAAGTTTTTATTACAATCAGGCCTTAAAGC AAAGCCCAGACTAAAACGTTCGGCCCCTACTACCCGTG CACCATCCACCAAACGCAAAAAGGTTAAAAAATAA	ATGGTGCTGATCCTGTGCTGTACCCTGGCTATCCTGTT TTGTGTCGCCGATGTCAATGTGTTTCATATTTTCCTGCA GATGTCCGTGTGGAGGCCCTCTGAGGCCACCGTCTAT CTGCCCCCTGTGCCTGTCTCAAAAGTGGTCAGCACCG ACGAATACGTGAGCAGGACATCCATCTACTATTACGCA GGAAGCTCCCGCCTGCTGGCTGTCGGCAACCCTTATTT CAGCATCAAGAGTCCAAACAATAACAAGAAAGTGCTGG TCCCCAAGGTGAGTGGGCTGCAGTATCGGGTGTTTAG GGTCCGCCTGCCAGATCCCAACAAGTTTGGATTCCCC GACACTTCCTTCTACAATCCTGATACCCAGCGACTGGT GTGGGCTTGCGTCGGACTGGAGATCGGACGAGGACAG CCACTGGGAGTGGGCGTCTCAGGACACCCCTATCTGA ACAAATTTGACGATACAGAGACTAGCAATAGATACCCC GCACAGCCTGGCAGTGACAACAGAGAATGTCTGTCAAT GGATTACAAGCAGACTCAGCTGTGCCTGATTGGCTGTA AACCACCCACCGGGGAGCATTGGGGGAAGGGAGTGG CTTGCAATAACAATGCCGCTGCAACTGACTGTCCTCCA CTGGAGCTGTTCAATAGCATCATTGAAGACGGCGATAT GGTGGACACCGGCTTTGGGTGCATGGATTTCGGGACA CTGCAGGCCAACAAGTCTGACGTGCCTATCGATATTTG CAACAGTACCTGTAAGTACCCTGACTACCTGAAGATGG CTTCCGAGCCCTACGGCGACTCTCTGTTCTTTTTCCTG CGGAGAGAACAGATGTTTGTGAGACACTTTTTCAACAG GGCAGGGAAACTGGGAGAGGCCGTCCCTGACGATCTG TACATCAAGGGAAGCGGCAATACCGCTGTGATTCAGTC TAGTGCATTTTTCCCTACACCATCAGGCAGCATCGTGA CTTCCGAATCTCAGCTGTTTAACAAGCCATACTGGCTG CAGCGAGCACAGGGACATAACAATGGGATTTGCTGGG GAAACCAGCTGTTCGTGACAGTGGTGGACACCACAAG ATCCACTAATATGACCCTGTGTACAGAGGTCACTAAGG AAGGCACTTACAAGAACGACAACTTCAAGGAGTACGTG AGACACGTCGAGGAATACGATCTGCAGTTTGTGTTCCA GCTGTGCAAGATCACCCTGACAGCAGAGATCATGACCT ACATTCATACAATGGACTCTAATATTCTGGAAGATTGGC AGTTTGGGCTGACCCCCCCTCCAAGTGCCTCACTGCA GGACACATATAGGTTCGTGACTAGCCAGGCAATCACTT GTCAGAAAACCGCCCCCCCCTAAGGAGAAAGAAGATCC CCTGAACAAGTACACATTTTGGGAAGTGAATCTGAAGG AAAAATTCTCCGCTGACCTGGATCAGTTTCCACTGGGG CGCAAGTTCCTGCTGCAGTCTGGACTGAAGGCAAAAC CCCGACTGAAACGGAGCGCACCAACTACCCGCGCTCC CTCCACCAAGCGAAAGAAAGTGAAGAAA	MSVWRPSEATVYLPPV PVSKVVSTDEYVSRTSI YYYAGSSRLLAVGNPY FSIKSPNNNKKVLVPKV SGLQYRVFRVRLPDPN KFGFPDTSFYNPDTQR LVWACVGLEIGRGQPL GVGVSGHPYLNKFDDT ETSNRYPAQPGSDNRE CLSMDYKQTQLCLIGCK PPTGEHWGKGVACNN NAAATDCPPLELFNSIIE DGDMVDTGFGCMDFG TLQANKSDVPIDICNST CKYPDYLKMASEPYGD SLFFFLRREQMFVRHFF NRAGKLGEAVPDDLYIK GSGNTAVIQSSAFFPTP SGSIVTSESQLFNKPYW LQRAQGHNNGICWGN QLFVTVVDTTRSTNMTL CTEVTKEGTYKNDNFK EYVRHVEEYDLQFVFQ LCKITLTAEIMTYIHTMD SNILEDWQFGLTPPPSA SLQDTYRFVTSQAITCQ KTAPPKEKEDPLNKYTF WEVNLKEKFSADLDQF PLGRKFLLQSGLKAKPR LKRSAPTTRAPSTKRKK VKK	30.1%
HPV58 L2	ATGAGACACAAACGGTCTACAAGGCGCAAGCGTGCAT CTGCTACACAACTTTACCAAACATGCAAGGCCTCAGGC ACCTGCCCACCTGATGTTATACCCAAAGTTGAAGGCAC TACTATAGCAGATCAAATATTACGATATGGTAGCTTAGG GGTGTTTTTTGGAGGTTTAGGCATTGGTACAGGGTCGG GTACAGGTGGCAGGACTGGATATGTGCCCCTTGGTAG TACCCCACCGTCTGAGGCTATACCTTTACAGCCCATAC GTCCCCCAGTTACCGTTGATACTGTGGGGCCTTTGGAT TCTTCTATTGTATCTTTAATAGAGGAATCTAGTTTTATAG ACGCCGGTGCACCAGCCCCATCAATTCCCACTCCATCT GGTTTTGATATTACCACCTCTGCAGATACTACACCTGCA ATACTTAATGTTTCCTCTATTGGAGAATCATCTATACAA ACTGTTTCTACACATTTAAATCCCTCCTTTACTGAGCCA TCCGTACTCCGCCCTCCTGCACCTGCAGAGGCCTCTG GACATTTAATATTTTCCTCTCCTACTGTTAGCACACATA GTTATGAAAACATACCAATGGATACCTTTGTTATTTCTA CTGACAGTGGCAATGTCACGTCTAGCACACCCATTCCA GGGTCTCGCCCTGTGGCACGCCTTGGTTTATACAGTC GCAACACCCAACAAGTTAAGGTTGTTGACCCTGCTTTT TTAACATCTCCTCATAGACTTGTAACATATGATAATCCA GCATTTGAAGGCTTTAACCCTGAGGACACATTGCAGTT TCAACATAGTGACATATCGCCTGCTCCTGATCCTGATTT TCTAGATATTGTTGCATTACACAGACCTGCATTAACCTC TCGCAGGGGTACTGTACGTTATAGTAGGGTTGGGCAAA AGGCTACACTTCGTACTCGCAGTGGAAAGCAAATAGGG GCTAAAGTACATTACTACCAAGACTTAAGTCCCATACAG CCTGTCCAGGAACAGGTACAACAGCAGCAACAATTTGA ATTACAATCTTTAAATACTTCTGTTTCTCCCTATAGTATT AATGATGGACTTTATGATATTTATGCTGACGATGCTGAT ACTATACATGATTTTCAGAGTCCTCTGCACTCACATACG TCCTTTGCCACCACACGTACCAGTAATGTGTCCATACC ATTAAATACTGGATTTGACACTCCTCTTGTGTCATTGGA ACCTGGTCCAGACATTGCATCTTCTGTAACATCTATGTC TAGTCCATTTATTCCTATATCTCCACTAACTCCTTTTAAT ACCATAATTGTGGATGGTGCTGATTTTATGTTGCACCCT AGCTATTTTATTTTGCGTCGCAGACGTAAACGTTTTCCA TATTTTTTTGCAGATGTCCGTGTGGCGGCCTAG	ATGAGGCATAAGAGGAGCACCAGAAGAAAAAGAGCAT CCGCAACCCAGCTGTATCAGACCTGTAAAGCATCCGG CACCTGTCCACCTGACGTGATCCCCAAGGTCGAGGGC ACCACAATCGCCGATCAGATTCTGAGATACGGATCTCT GGGCGTGTTCTTTGGAGGACTGGGAATTGGAACCGGC AGTGGGACAGGAGGCCGGACTGGATATGTGCCACTGG GGAGTACCCCCCCTTCAGAAGCCATCCCACTGCAGCC CATTAGACCACCCGTGACTGTGGACACCGTGGGCCCT CTGGATAGCTCCATCGTCAGCCTGATTGAGGAATCTAG TTTCATCGACGCAGGAGCACCAGCTCCTAGCATCCCAA CACCATCCGGGTTTGACATTACTACCTCTGCTGATACA ACTCCAGCAATTCTGAACGTGTCAAGCATCGGGGAGTC CTCTATTCAGACAGTCAGCACTCACCTGAATCCTTCTTT CACCGAGCCAAGTGTGCTGAGACCTCCAGCACCAGCA GAAGCTAGCGGACACCTGATCTTCAGTTCACCCACCGT GAGTACACATTCATACGAAAACATCCCTATGGACACATT TGTGATTAGCACTGATTCCGGAAATGTCACTAGCTCCA CCCCTATCCCAGGAAGCCGGCCTGTGGCAAGACTGGG ACTGTACTCCCGAAACACTCAGCAGGTCAAGGTGGTG GACCCCGCTTTTCTGACTTCCCCTCATCGCCTGGTGAC CTATGATAACCCAGCATTCGAGGGCTTTAATCCCGAAG ACACCCTGCAGTTCCAGCACTCTGATATCAGTCCCGCC CCTGACCCAGATTTTCTGGACATTGTGGCCCTGCATAG GCCCGCTCTGACCTCACGGAGAGGGACAGTGCGCTAC AGCCGAGTCGGACAGAAAGCAACACTGAGAACTAGGA GCGGGAAGCAGATCGGAGCCAAAGTGCACTACTATCA GGATCTGTCCCCTATTCAGCCAGTGCAGGAGCAGGTC CAGCAGCAGCAGCAGTTCGAACTGCAGTCCCTGAACA CCTCCGTGTCTCCTTATTCTATCAATGACGGCCTGTAC GATATCTACGCAGACGATGCCGACACAATCCATGATTT CCAGTCCCCACTGCACTCACATACCAGCTTCGCAACCA CACGCACTTCCAACGTGTCTATCCCCTCTGAATACCGGA TTTGACACACCACTGGTGTCTCTGGAGCCCGGCCCTG ATATTGCTTCTAGTGTCACCAGTATGTCAAGCCCCTTCA TCCCTATTTCACCACTGACTCCCTTTAATACCATCATTG TGGACGGCGCCGATTTCATGCTGCACCCAAGCTACTTT ATCCTGAGGCGCCGACGGAAAAGGTTCCCCTATTTCTT TGCTGACGTGCGCGTCGCCGCT	MRHKRSTRRKRASATQ LYQTCKASGTCPPDVIP KVEGTTIADQILRYGSL GVFFGGLGIGTGSGTG GRTGYVPLGSTPPSEAI PLQPIRPPVTVDTVGPL DSSIVSLIEESSFIDAGA PAPSIPTPSGFDITTSAD TTPAILNVSSIGESSIQT VSTHLNPSFTEPSVLRP PAPAEASGHLIFSSPTV STHSYENIPMDTFVIST DSGNVTSSTPIPGSRPV ARLGLYSRNTQQVKVV DPAFLTSPHRLVTYDNP AFEGFNPEDTLQFQHS DISPAPDPDFLDIVALHR PALTSRRGTVRYSRVG QKATLRTRSGKQIGAKV HYYQDLSPIQPVQEQV QQQQQFELQSLNTSVS PYSINDGLYDIYADDAD TIHDFQSPLHSHTSFAT TRTSNVSIPLNTGFDTP LVSLEPGPDIASSVTSM SSPFIPISPLTPFNTIIVD GADFMLHPSYFILRRRR KRFPYFFADVRVAA	28.5%

APPENDIX B

NUCLEOTIDE SEQUENCES OF MODIFIED CAS9 GENES, GRNAS, AND PRIMERS USED IN CHAPTER 5

Cas9-β2:

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAG
ACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTG
TGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTG
CTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAG
AAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAG
CAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTT
CCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGT
GGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAA
ACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGG
CCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCG
ACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGC
TGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGT
CTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCC
GGCGAGAAGAAGAATGGCCTGTTCGGAAACCTTATTGCCCTGAGCCTGGGCCTG
ACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTG
AGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGA
CCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCT
GAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCT
CTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCA
AGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTAC
AAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTG
AAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAG
CATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGAC
CTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGC
CTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGT
GGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGA
TAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTA
CTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAG
AAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTT
CAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGC
CTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTG
GACAATGAGGAAAACGAGGACATTGGTGAAGATATCGTGCTGACCCTGACACTG
TTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTC
GACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAG
GCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAA
TCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGAT
CCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGG
CCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCA
TTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGA
TGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC

CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATAT
GTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCA
TATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGAC
CAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCG
TGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCC
AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG
GATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAG
CACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGAT
TTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACG
CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACC
CTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGA
AGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCT
TCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA
GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGT
GGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA
GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCT
ATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGAC
CCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTG
GTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTG
CTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTC
TGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTA
AGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCG
GCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCC
TGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGC
AGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGA
ATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGA
CGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCT
GTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAG
GCAAAAAAGAAAAAGTAA

Cas9-α2:

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAG
ACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTG
TGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTG
CTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAG
AAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAG
CAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTT
CCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGT
GGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAA
ACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGG
CCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCG
ACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGC
TGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGT

CTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCC
GGCGAGAAGAAGAATGGCCTGTTCGGAAACGGTATTGCCCTGAGCCTGGGCCT
GACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCT
GAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCG
ACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGC
TGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCT
CTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCA
AGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTAC
AAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTG
AAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAG
CATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGAC
CTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGC
CTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGT
GGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGA
TAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTA
CTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAG
AAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTT
CAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGC
CTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTG
GACAATGAGGAAAACGAGGACATTCTTGAAGATATCGTGCTGACCCTGACACTGT
TTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCG
ACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGG
CTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAAT
CCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATC
CACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGC
CAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCAT
TAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGAT
GGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGA
CCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACC
CAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATG
TACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCAT
ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACC
AGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGT
GAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCA
GAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGG
ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGC
ACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACA
AGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATT
TCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGC
CCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCC
TAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAA
GATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAG
ATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTG
GGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGT

GAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTAT
CCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACC
CTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGG
TGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGC
TGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCT
GGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAA
GTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGG
CGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCT
GTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCA
GAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA
GATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGA
ATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGA
CGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCT
GTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAG
GCAAAAAAGAAAAAGTAAG

Cas9-α2-β2:
ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAG
ACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTG
TGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTG
CTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAG
AAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAG
CAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTT
CCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGT
GGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAA
ACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGG
CCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCG
ACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGC
TGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGT
CTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCC
GGCGAGAAGAAGAATGGCCTGTTCGGAAACGGTATTGCCCTGAGCCTGGGCCT
GACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCT
GAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCG
ACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGC
TGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCT
CTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTC
TCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCA
AGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTAC
AAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTG
AAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAG
CATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGAC
CTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGC
CTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGT
GGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGA

TAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTA
CTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAG
AAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTT
CAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGC
CTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTG
GACAATGAGGAAAACGAGGACATTGGTGAAGATATCGTGCTGACCCTGACACTG
TTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTC
GACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAG
GCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAA
TCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGAT
CCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGG
CCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCA
TTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGA
TGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC
CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATAT
GTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCA
TATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGAC
CAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCG
TGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCC
AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG
GATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAG
CACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGAC
AAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGAT
TTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACG
CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACC
CTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGA
AGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCT
TCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA
GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGT
GGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA
GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCT
ATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGAC
CCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTG
GTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTG
CTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTC
TGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTA
AGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCG
GCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCC
TGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGC
AGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAG
TGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGA
ATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTA
CTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGA
CGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCT

GTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAG
GCAAAAAAGAAAAAGT

Sequences of gRNAs used in Chapter 5:

MIAT-14bp gRNA	GAGGCTGAGCGCAC
TTN-14bp gRNA	GGAAGTCTCCTTTG
Reporter2-20bp gRNA	GTCCCCTCCACCCCACAGTG
CR10-14bp-gRNA	GCATCAGGAACATGT
EMX1- 20bp gRNA	CACC GAGTCCGAGCAGAAGAAGAA

Sequences of primers used in Chapter 5:

Cas9 fragment1- FW	ttttGGTCTCTAGGTCCACCATGGACTATAAGGACCACGA
Cas9 fragment1- RV	tttggtctcaGAACAGCTGGTTGTAGGTCTGCA
Cas9 fragment2-FW	ttttGGTCTCTACCAACCGGAAAGTGACCGTGAAG
Cas9 fragment2-RV	ttttGGTCTCAAAGCTTACTTTTTCTTTTTTGCC
qPCRMIAT-FW	TGGCTGGGGTTTGAACCTTT
qPCR-MIAT RV	AGGAAGCTGTTCCAGACTGC
qPCRTTN FW	TGTTGCCACTGGTGCTAAAG
qPCR-TTN-RV	ACAGCAGTCTTCTCCGCTTC
PCR-EMX1-FW	CCATCCCCTTCTGTGAATGT
PCR-EMX1-RV	GGAGATTGGAGACACGGAGA