

**RNA-Seq-basierte Untersuchung einer viralen Ätiologie und
Analyse der Prognose von Mundhöhlenkarzinomen junger Patienten**

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Hohen Medizinischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

Johannes Mathias Brägelmann
aus Köln

2014

Angefertigt mit Genehmigung der
Medizinischen Fakultät der Universität Bonn

1. Gutachter: Prof. Dr. med. Peter Brossart
2. Gutachter: Prof. Dr. med Sven Perner

Tag der Mündlichen Prüfung: 22. Mai 2014

Aus der Medizinische Klinik III für Hämatologie und Onkologie
Direktor: Prof. Dr. Peter Brossart

Meiner Familie

Inhaltsverzeichnis

Abkürzungsverzeichnis	7
1. Einleitung.....	8
1.1 Anatomie von Kopf-Hals-Tumoren	8
1.1.1 Anatomische Klassifikation von Kopf-Hals-Tumoren	8
1.1.2 Histologie von Kopf-Hals-Tumoren	9
1.1.3 Anatomische Klassifikation der Mundhöhlenkarzinome.....	9
1.2 Epidemiologie von Kopf-Hals-Tumoren	10
1.3 Ätiologie und Risikofaktoren von Kopf-Hals-Tumoren	13
1.3.1 „Klassische“ Risikofaktoren (Tabak und Alkohol)	13
1.3.2 Humane Papillomaviren (HPV)	15
1.3.3 Weitere Risikofaktoren	17
1.3.4 Ätiologie und Risikofaktoren für Tumore der Mundhöhle und der mobilen Zunge .	18
1.4 Klinische Aspekte von Kopf-Hals-Tumoren	19
1.4.1 Klinik	19
1.4.2 Diagnostik, Staging und Therapie.....	19
1.4.3 Einflussfaktoren auf die Prognose	21
1.5 Onkoviren.....	23
1.5.1 Bedeutung von Onkoviren	23
1.5.2 Virale Onkogenese.....	24
1.5.3 Detektion von Onkoviren.....	24
1.6 Next-Generation-Sequencing und RNA-Seq Technologie	25
1.7 Ziele dieser Arbeit.....	28
2. Methoden.....	29
2.1 Analyse der klinischen Daten	29
2.2 Gewebeproben von Tumoren der Mobilen Zunge.....	30
2.3 Positivkontrollen	30
2.3.1 HPV-16 positive Kopf-Hals-Tumorzelllinie (UM-SCC47).....	30
2.3.2 EBV-Transgenes Maus-Lymphom	31
2.3.3 HPV-16 positives Zervixkarzinom	31
2.4 Transcriptome Sequencing.....	32
2.4.1 Erstellung der „RNA-Seq Libraries“	32
2.4.2 Sequenzierung.....	35
2.5 Virusalgorithmus.....	35
2.5.1 Subtraktion nicht-viraler Sequenzen.....	37
2.5.2 Schritte des schnellen „Rapid-Algorithmus“	38
2.5.3 Schritte des spezifischeren „Discovery-Algorithmus“	39
2.5.4 Datenbanken mit Referenzsequenzen	41
2.5.5 Bewertungskriterien für die Ergebnisse der Algorithmen	42
2.5.6 In-silico Simulation eines unbekanntes Virus	44

2.6	CDKN2A/p16-Genexpressions-Analyse	45
3.	Ergebnisse.....	46
3.1	Auswertung der klinischen Daten	46
3.2	Tumorproben der oralen Zunge	55
3.3	Resultate des Virusalgorithmus	57
3.3.1	Effizienz der digitalen Subtraktion und Performance des Algorithmus	57
3.3.2	Ergebnisse des schnellen Virusalgorithmus.....	61
3.3.3	Ergebnisse des Discovery-Algorithmus.....	65
3.3.4	Ergebnis der Simulation eines unbekanntes Virus	69
3.4	Expression von CDKN2A/p16 ^{INK4a}	72
4.	Diskussion.....	73
4.1	Zur Analyse der klinischen Daten.....	73
4.1.1	Interpretation der Ergebnisse	73
4.1.2	Bewertung der Ergebnisse	74
4.2	Zum Virusalgorithmus	76
4.2.1	Stärken des Virusalgorithmus	77
4.2.2	Grenzen des Virusalgorithmus.....	81
4.2.3	Sensitivität des Verfahrens.....	85
4.3	Zur Expression von CDKN2A/p16 ^{INK4a}	86
4.4	Schlussfolgerungen und Ausblick.....	88
5.	Zusammenfassung	89
6.	Literaturverzeichnis	90
7.	Danksagung	106
8.	Lebenslauf	107

Abkürzungsverzeichnis

A	Adenosin
BLAST	Basic Local Alignment Search Tool
bp	base pairs
C	Cytosin
Ca	Karzinom
CDKN2A	Cyclin-dependent kinase inhibitor 2A (= p16 ^{INK4a})
cDNA	complementary DNA
DNA	Desoxyribonucleic Acid
EBV	Epstein-Barr-Virus
FBS	Fetale Bovine Serum
G	Guanosin
HHV	Humanes Herpes Virus
HNC	Head and Neck Cancer
HNSCC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papilloma Virus
HR	Hazard Ratio
IHC	Immunhistochemie
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NGS	Next-generation Sequencing
OC	Oral Cavity
OT	Oral tongue = orale/mobile Zunge
PBS	Phosphate Buffered Saline
PCR	Polymerasis Chain Reaction
PFS	Progression Free Survival
RCT	Radiochemotherapie
RNA	Ribonucleic Acid
RPKM	Reads Per Kilobase of exon model per million Mapped reads
T	Thymin
UICC	Union Internationale Contre le Cancer

1. Einleitung

1.1 Anatomie von Kopf-Hals-Tumoren

1.1.1 Anatomische Klassifikation von Kopf-Hals-Tumoren

Der Begriff der Kopf-Hals-Tumore (Head and Neck Cancer, HNC) umfasst eine heterogene Gruppe von Neoplasmen unterschiedlicher Ursprungsgewebe, deren Hauptgemeinsamkeit ihr Auftreten im Bereich des Halses und des Kopfes ist. Die Verwendung in der Literatur und die genaue Abgrenzung gegenüber anderen Tumorentitäten erfolgt allerdings uneinheitlich, was z.B. bei der Beurteilung berichteter Inzidenzraten zu beachten ist.

In seiner breitesten Definition beschreibt der Ausdruck „Kopf-Hals-Tumor“ Neubildungen der oberen Atem- und Speisewege inklusive Trachea und Ösophagus, der Schild- und Nebenschilddrüse, der Speicheldrüsen, der Haut, der Knochen und des mesenchymalen Weichgewebes dieser Region (Cooper et al., 2009; Kelley und Shah, 1998). In zahlreichen Berichten wird der Begriff Kopf-Hals-Tumor allerdings –implizit oder explizit– enger definiert, so dass er hauptsächlich epitheliale Neubildungen der Mundhöhle, der Nasenhaupthöhle, der nasalen und paranasalen Sinus, der Speicheldrüsen, des Pharynx und des Larynx einschließt, während okuläre und intrakranielle Neoplasmen und solche endokrinen bzw. lymphatischen Ursprungs ausgenommen sind (Argiris et al., 2008; Johnson und Amarasinghe, 2011). Dies ergibt sich daraus, dass einige Parallelen in Histologie, Risikofaktoren und Therapie innerhalb dieser Subgruppen zu verzeichnen sind, während z.B. Lymphome und Schilddrüsenkarzinome deutlich davon abweichen. Im Folgenden wird bei der Verwendung des Begriffs Kopf-Hals-Tumor von dieser umgrenzten Definition ausgegangen.

Da auch innerhalb dieser Tumorgruppe u.a. ätiologische, epidemiologische und prognostische Unterschiede zwischen verschiedenen Primärlokalisationen bestehen (Carvalho et al., 2004), wird die anatomische Einordnung in der Regel weiter differenziert (vgl. Abb. 1). So werden beispielsweise im Bereich des Larynx ein glottischer, ein supra- und ein subglottischer Anteil, im Bereich des Pharynx Oro-, Naso- und Hypopharynx unterschieden, wobei z.B. der Oropharynx wiederum in Tonsillen, Zungengrund (Radix linguae), weicher Gaumen und Rachenhinterwand subklassifiziert wird. Auch innerhalb der Mundhöhle erfolgt eine weitere Untergliederung (s.u.).

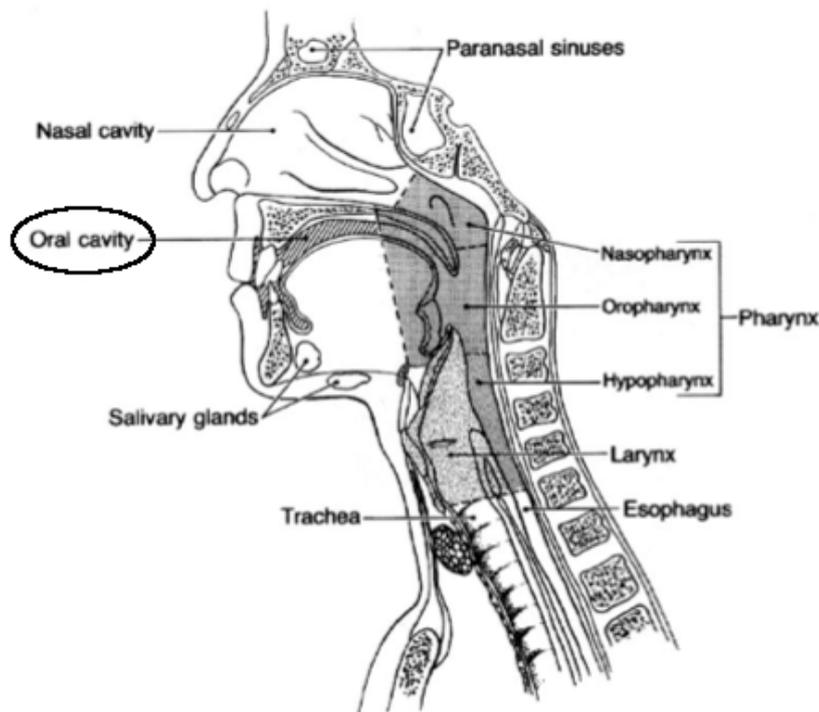


Abb. 1: Anatomie der Kopf-Hals-Region (nach Vokes et al., 1993).

1.1.2 Histologie von Kopf-Hals-Tumoren

Histopathologisch betrachtet handelt es sich bei >90% aller Kopf-Hals-Tumore um Plattenepithelkarzinome (HNSCC), die aus der Mukosa des oberen Aerodigestivtraktes hervorgehen (Koch und Nance, 2010; Sanderson und Ironside, 2002). Der restliche Anteil setzt sich vornehmlich aus Adenokarzinomen der Speicheldrüsen, Sarkomen und anderen mesenchymalen Tumoren sowie Melanomen zusammen (Shah und Lydiatt, 1995).

1.1.3 Anatomische Klassifikation der Mundhöhlenkarzinome

Auch innerhalb der Mundhöhle werden verschiedene Tumorlokalisationen unterschieden (DeLacure, 1998; Sasaki et al., 2005). Zu den Tumorlokalisationen der Mundhöhle werden im Allgemeinen die orale bzw. mobile Zunge, welche die vorderen 2/3 der Zunge ausmacht (von den Papillae vallatae bis zur Zungenspitze), der Mundboden, die Lippen, die Gingiva, die bukkale Mukosa, der harte Gaumen und das Trigonum retromolare gerechnet (DeLacure, 1998). Da Tumore der Mundhöhle und die der oralen Zunge, welche einen Großteil der Mundhöhlenkarzinome ausmachen (Chitapanarux et al., 2006), klinisch und histologisch von anderen Tumoren des

Kopf-Hals-Bereiches verschieden sind (Lingen et al., 2000; Rautava et al., 2007) und häufig einen aggressiven Phänotyp zeigen (Salem, 2010), werden sie als eigene Tumorentität aufgefasst.

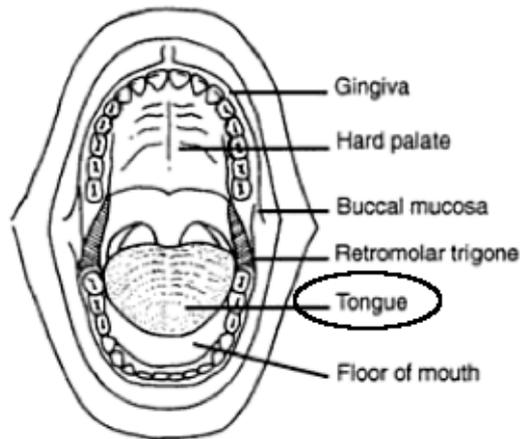


Abb. 2: Anatomie der Mundhöhle nach (modifiziert nach DeLacure, 1998).

1.2 Epidemiologie von Kopf-Hals-Tumoren

Global betrachtet stellen plattenepitheliale Hals- und Kopftumore insgesamt die sechsthäufigste aller Tumormanifestationen dar (Parkin et al., 2005). Schätzungen gehen davon aus, dass weltweit pro Jahr >500.000 neue Fälle auftreten und dass nur 40-50% dieser Patienten einen 5-Jahreszeitraum überleben (Kamangar et al., 2006; Leemans et al., 2011). Dabei ist allerdings eine erhebliche Variation der Inzidenz und Prävalenz in Abhängigkeit der jeweiligen Primärtumorlokalisation, der geographischen Region und des Geschlechts zu beobachten. Zusätzlich lassen sich im Zeitverlauf Veränderungen der Häufigkeiten verschiedener Tumore an der Gesamtzahl aller Fälle feststellen, was z.T. mit einer veränderten Exposition gegenüber bekannten Risikofaktoren erklärbar ist.

Unter den HNSCC sind Mundhöhlenkarzinome am häufigsten vertreten und waren im Jahr 2008 weltweit für ca. 263.000 Neuerkrankungen und 127.000 Todesfälle pro Jahr verantwortlich, womit sie insgesamt die zwölfte Stelle der globalen Krebsmortalität einnehmen. Larynxkarzinome sind am zweithäufigsten mit >150.000 Neuerkrankungen und 82.000 Todesfällen pro Jahr, während Oro- und Hypopharynxkarzinome 135.000 Neuerkrankungen und 95.000 Todesfälle pro Jahr verursachen. Die geringste Häufigkeit haben Nasopharynxkarzinome mit ca. 84.000 Neuerkrankungen und 51.000 Todesfällen pro Jahr (Ferlay et al., 2010). Ein Vergleich mit den Daten der Internationalen Agentur für Krebsforschung für Deutschland ergibt eine etwas andere Reihen-

folge, wobei auch hier der höchste Anteil für Mundhöhlenkarzinome beschrieben ist (Inzidenzrate 4,0/100.000). Am zweithäufigsten sind allerdings Tumore des Oro- und Hypopharynx (Inzidenzrate 3,4/100.000), danach folgen Larynxkarzinome (Inzidenzrate 2,7/100.000). Die Zahl der Nasopharynxkarzinome (Inzidenzrate 0,3/100.000) liegt erwartungsgemäß deutlich darunter (IARC, 2010).

Obwohl in absoluten Zahlen gesehen fast zwei Drittel aller Patienten mit Kopf-Hals-Tumoren aus wenig entwickelten Ländern stammen, ist die Inzidenz für Karzinome der Mundhöhle, des Pharynx (außer Nasopharynx) und Larynx in den Industriestaaten höher, während sich die Inzidenz von Nasopharynxkarzinomen hauptsächlich auf wenige Gebiete in Afrika und Südostasien konzentriert (Jemal et al., 2011). Insgesamt machten 2008 Plattenepithelkarzinome des Kopf-Hals-Bereichs in Deutschland 3,6% aller Krebsneuerkrankungen (Männer 5,3% , Frauen 1,8%) aus (RKI, 2012). Dies entspricht in etwa den Raten in den USA (Cooper et al., 2009) und Großbritannien (Sanderson und Ironside, 2002). Die weltweit höchste Inzidenz für HNSCC wurde mit 49,4/100.000 Einwohner bei Männern im Departement Bas-Rhin in Nordfrankreich gemessen, während die der Frauen dort mit 3,9/100.000 Einwohner signifikant niedriger lag (Franceschi et al., 2000). Dies passt zu dem beobachteten Trend, dass weltweit –abhängig von geographischer Region und Tumortyp– Männer im Verhältnis von 2:1 bis 15:1 häufiger betroffen sind als Frauen (Dwivedi et al., 2012).

Grundsätzlich sind Kopf-Hals-Tumore mit steigender Inzidenz bei älteren Menschen zu finden, mehr als 98% der Patienten in der Europäischen Union sind über 40 Jahre alt (Döbrossy, 2005). In den vergangenen Jahrzehnten allerdings hat sich besonders in den Industrieländern der Anteil junger Patienten erhöht (Toner und O'Regan, 2009). Außerdem veränderten sich die Anteile verschiedener Tumorlokalisationen an der Gesamtzahl aller Kopf-Hals-Tumore. So fanden Studien aus Skandinavien (Blomberg et al., 2010; Hammarstedt et al., 2007), England (Doobaree et al., 2009), den USA (Cooper et al., 2009; Ryerson et al., 2008) und Deutschland (Guntinas-Lichius et al., 2009) eine steigende Inzidenz von u.a. Oropharynx-, Tonsillen- und Mundhöhlenkarzinomen, während die Inzidenz anderer Subtypen (z.B. Larynx) sank und die Gesamtzahl aller Kopf-Hals-Tumore konstant blieb oder fallende Tendenzen zeigt. Parallel zu diesem Trend wurde besonders in den USA, aber auch in anderen Ländern, zusätzlich ein Anstieg der Inzidenz von Tumoren der oralen Zunge beobachtet, wobei vor allem junge Patienten und Frauen betroffen zu sein scheinen (Chitapanarux et al., 2006; Depue, 1986; Müller et al., 2008; Myers et al., 2000; Schantz und Yu,

2002; Shiboski et al., 2005). Einen Überblick gibt Abb. 3. Dabei ist zu beachten, dass der Anstieg oropharyngealer Tumore schwach dargestellt ist, da sie in eine Gruppe mit den Hypopharynxkarzinomen eingeordnet wurden, welche eine abnehmende Inzidenz zeigen. Zudem wird wohl ein Teil der Zungengrundkarzinome zu den Zungentumoren gezählt. Dennoch lassen sich die beschriebenen Trends erkennen.

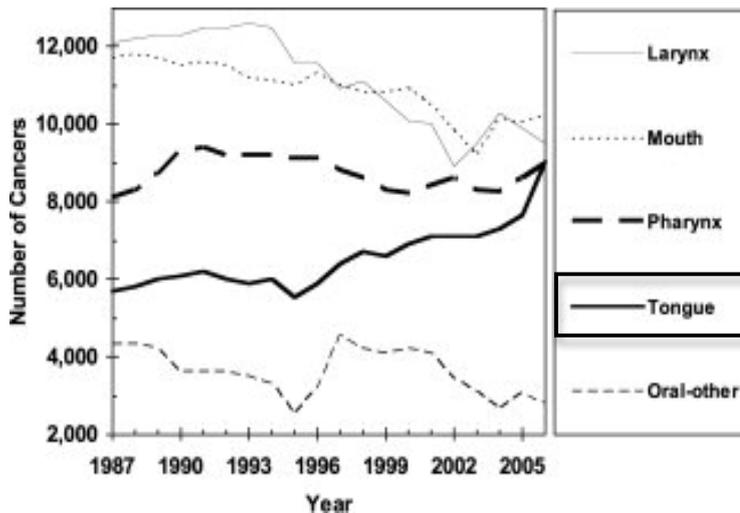


Abb. 3: Inzidenz von Kopf-Hals-Tumoren in den USA (modifiziert nach Sturgis und Cinciripini, 2007).

Für den Anstieg HPV-assoziiierter Karzinome (Oropharynx und Tonsillen) wird dabei eine vermehrte Verbreitung von HPV-Infektionen verantwortlich gemacht, während die sinkende Inzidenz anderer Subtypen durch die abnehmende Exposition gegenüber Tabak und Alkohol erklärt wird (vgl. Kapitel 1.3). Die Gründe für den Anstieg von Mundhöhlen- bzw. oralen Zungentumoren (welche den Großteil der Mundhöhlenkarzinome ausmachen) sind allerdings weiterhin unklar, da sie häufig bei Patienten ohne klassische Risikofaktoren auftreten (X-H Liang et al., 2008), aber auch eine geringe Häufigkeit von HPV zeigen (vgl. Kapitel 1.3.4). Ob diese Karzinome neben einer steigenden Inzidenz auch eine schlechtere Prognose zeigen ist wenig untersucht und wird u.a. in dieser Arbeit analysiert.

Insgesamt verdeutlichen diese Befunde die Notwendigkeit einer nach Subtypen differenzierten epidemiologischen Betrachtungsweise, da diese auch subtile Veränderungen abbildet und über die Epidemiologie hinaus auch die Suche nach ätiologischen Einflüssen befördern kann.

1.3 Ätiologie und Risikofaktoren von Kopf-Hals-Tumoren

Die Kenntnis ätiologischer Faktoren und der damit verbundenen Kausalzusammenhänge ist essentiell für die Entwicklung von Strategien für Prävention und Therapie und zur Prognoseabschätzung. Sie sollen deshalb im Folgenden dargestellt werden.

1.3.1 „Klassische“ Risikofaktoren (Tabak und Alkohol)

Die wichtigsten Risikofaktoren für die Entwicklung eines Hals-Kopf-Tumors sind der Konsum von Tabak und Alkohol (Leemans et al., 2011) und ein Großteil der im 20. Jahrhundert stark angestiegenen Inzidenz von Kopf-Hals-Tumoren lässt sich durch den erhöhten Alkohol- und Tabakgebrauch erklären (Gillison, 2007). Insgesamt lassen sich fast 80% aller Erkrankungen auf diese beiden Einflussgrößen zurückführen (Castellsagué et al., 2004; Negri et al., 1993).

Tabakrauchen hat sich dabei als besonders schädlich erwiesen und zahlreiche Studien haben epidemiologisch und kausal eine Rolle in der Karzinogenese nachgewiesen. So ist das Risiko für einen Kopf-Hals-Tumor bei Rauchern mehr als zehnmal höher als bei Menschen die niemals geraucht haben. Zusätzlich ist eine dosisabhängige Wirkung zu beobachten, so dass das Risiko mit Dauer und Ausmaß des Tabakkonsums ansteigt (Sturgis und Cinciripini, 2007). Eine internationale Meta-Analyse zeigte, dass nach Beendigung des Rauchens schon nach 1-4 Jahren eine signifikante Risikoreduktion messbar war und das Risiko der Ex-Raucher für Kopf-Hals-Tumore außerhalb des Larynx nach >20 rauchfreien Jahren das Risiko der „Niemals-Raucher“ annahm (Bosetti et al., 2006; Marron et al., 2010). Die schnelle Risikoreduktion nach Beendigung des Rauchens wurde in diesem Zusammenhang als Hinweis betrachtet, dass der Tabakkonsum auch an den späten Schritten bzw. kontinuierlich an der Karzinogenese beteiligt sein könnte.

Auch Passivrauchen (Zhang et al., 2000) und der Gebrauch nicht zum Rauchen bestimmter Tabakerzeugnisse („smokeless tobacco“) steigert signifikant das Risiko für Kopf-Hals-Tumore (Zhou et al., 2012). Eine Sonderrolle nimmt der Indische Subkontinent ein, wo der Konsum von Betelnüssen mit (und ohne) Tabak das Risiko für Mundhöhlenkarzinome potenziert (Znaor et al., 2003) und diese zur häufigsten Tumormanifestation bei Männern macht (Bray et al., 2012). In Europa und den USA ist der Gebrauch von rauchfreiem Tabak (z.B. Kautabak) deutlich weniger verbreitet und hat in den vergangenen Jahren stärker abgenommen als die Zahl der Raucher (Sturgis und Cinciripini, 2007).

Die Gründe dafür, warum nur ein verhältnismäßig kleiner Teil der Tabakexponierten jemals einen Tumor entwickelt, ist noch weitgehend unklar. In gerauchtem bzw. verbranntem Tabak finden sich mindestens 60 und in unverbranntem Tabak immerhin mindestens 16 bekannte karzinogene Stoffe, darunter u.a. polyzyklische Kohlenwasserstoffe und Nitrosamine. Es gibt Vermutungen, dass genetische Polymorphismen in Enzymen der Biotransformation bzw. der Entgiftung von Tabakrauch eine Rolle spielen könnten, aber belastbare Studien liegen bisher nicht vor (Lacko et al., 2009).

Alkoholkonsum steigert das Risiko für Kopf-Hals-Tumore bei Rauchern und ist sogar der größte Risikofaktor bei Nichtrauchern (Gillison, 2007). Dabei lässt sich auch hier eine dosisabhängige Wirkung feststellen, wobei die Alkoholkonzentration und der Typ der konsumierten Alkoholika (hochprozentige Getränke vs. Bier/Wein) einen größeren Einfluss auf das Risiko zu haben scheint als die absolut konsumierte Alkoholmenge (Castellsagué et al., 2004). Ein Alkoholverzicht führt auch hier zwar zu einer Risikoreduktion, besonders für Mundhöhlenkarzinome, aber der Effekt ist deutlich schwächer als für Tabak. So reduziert sich erst nach 20 Jahren Abstinenz das Risiko um lediglich 40%. Für Niemals- und Ex-Raucher war der Effekt sogar noch schwächer (Marron et al., 2010). Die Autoren vermuten, dass eine lange Latenz zwischen Alkoholkonsum und erhöhtem Krebsrisiko verantwortlich sein könnte, die z.B. auf einen frühen, nur inkomplett reversiblen Schaden schließen ließe.

Der molekulare Mechanismus der Risikoerhöhung ist nicht völlig geklärt, da Alkohol selber nicht direkt karzinogen wirkt. Vermutet wird die Bildung karzinogener Abbaustoffe wie Acetaldehyd, das DNA-Addukte formen kann (Gillison, 2007). Zusätzlich wirkt Alkohol als Lösungsmittel und könnte so die Durchlässigkeit der Mukosa für andere Karzinogene erhöhen. Dies könnte auch erklären, warum der kombinierte Gebrauch von Alkohol und Tabak nicht additiv, sondern synergistisch wirkt (Vokes et al., 1993). So ist das Tumorrisiko einer kombinierten Einnahme von Tabak und Alkohol deutlich höher als die Addition der Risiken, wenn beide Stoffe einzeln betrachtet werden. Starke Raucher mit gleichzeitig hohem Alkoholkonsum zeigten beispielsweise ein 50fach erhöhtes Risiko für einen Kopf-Hals-Tumor im Vergleich zu alkoholabstinenten Nichtrauchern (Castellsagué et al., 2004).

1.3.2 Humane Papillomaviren (HPV)

Erst seit relativ kurzer Zeit ist der kausale Einfluss einer Infektion mit Humanen Papillomaviren (HPV) auf die Pathogenese von Kopf-Hals-Tumoren nachgewiesen (D'Souza et al., 2007; Gillison et al., 2000). Besonders häufig finden sich dabei die Typen HPV-16 (90-95% aller Fälle) und –deutlich seltener– HPV-18 oder HPV-33 (Dayyani et al., 2010; Kreimer, 2005). Diese sog. „high-risk Typen“ können durch die Produktion der viralen Onkoproteinen E6 und E7 die Tumorsuppressorgene p53 und Rb inaktivieren. Auf diese Weise wird direkt in den Zellzyklus der Wirtszelle eingegriffen und so die Tumorgenese gefördert (vgl. Abb. 4). Als Nebeneffekt führt dies zu einer Überexpression des normalerweise von Rb negativ regulierten CDKN2A/p16^{INK4A}-Gens, welches deshalb auch als Surrogatparameter in der HPV-Diagnostik verwandt wird (Howard und Chung, 2012; Jordan et al., 2012).

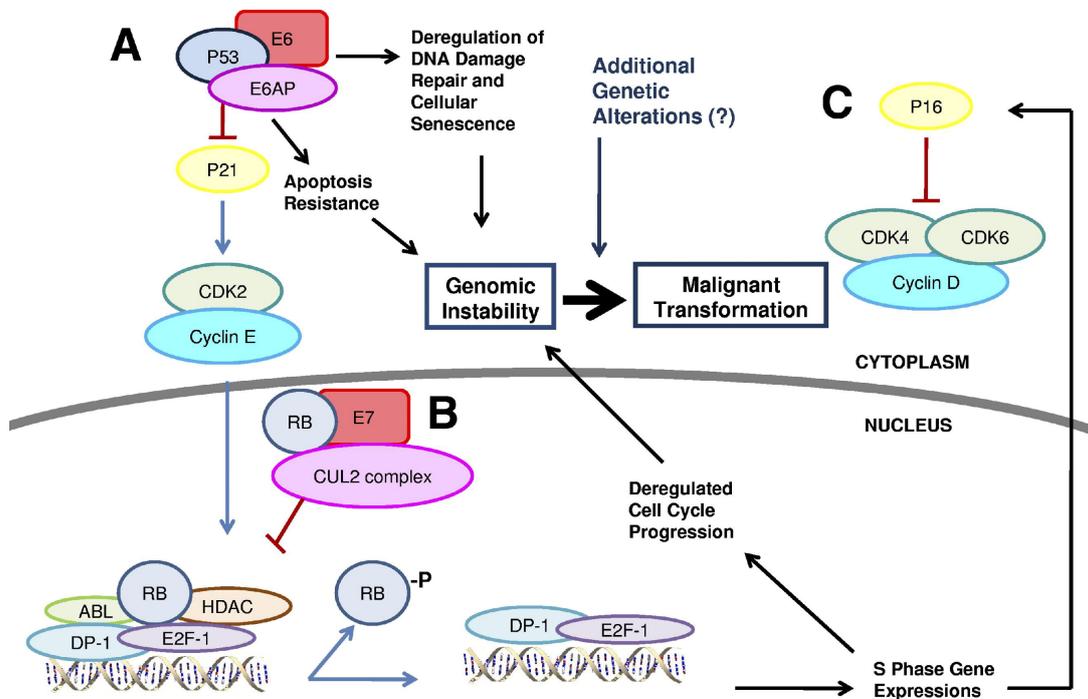


Abb. 4: Mechanismus maligner Transformation durch HPV. A) Abbau von p53 durch das virale E6-Onkogen B) Degradation des Rb-Proteins durch das virale E7-Onkogen C) Überexpression des zellulären p16-Proteins durch Wegfall von Feedback-Schleifen (modifiziert nach Howard und Chung, 2012).

Schätzungen gehen davon aus, dass insgesamt etwa 26% der Plattenepithelkarzinome des Kopf-Hals-Bereichs HPV-positiv sind (Kreimer, 2005). Allerdings ergeben sich signifikante Unterschiede zwischen verschiedenen Tumorlokalisationen, wobei Oropharynxkarzinome die stärkste Assoziation zu HPV-Infektionen aufweisen. Von diesen Tumoren sind 50-70% HPV-positiv, während die Infektionsraten in den übrigen Tumortypen signifikant niedriger sind (Ang et al.,

2010; Mork et al., 2001). Die stärkste Assoziation innerhalb der Oropharynxtumore weisen solche der Tonsillen und des Zungengrundes auf. Auch der serologische Nachweis einer Infektion mit HPV zeigt ein gesteigertes Risiko an. So ist das Risiko bei Vorliegen von Antikörpern gegen HPV-16 Kapsidprotein L1 3,5fach und bei Antikörpern gegen E6/E7-Proteine um das 9,2 fache erhöht (Herrero et al., 2003). Patienten mit HPV-positiven Tumoren zeigen dabei ein weniger ausgeprägtes Risikoprofil als andere Kopf-Hals-Tumorpatienten und als HPV-negative Patienten. Im Durchschnitt sind sie jünger, trinken weniger Alkohol und haben einen geringeren Tabakkonsum (Chaturvedi et al., 2011; Gillison et al., 2000; Haddad und Shin, 2008; Herrero et al., 2003; Ringström et al., 2002)

Zeitgleich zum Anstieg der Inzidenz HPV-assoziiierter Tumore stieg auch innerhalb dieser Gruppe der Anteil der Tumore, in denen HPV tatsächlich nachgewiesen werden konnte. Eine schwedische Studie berichtete, dass im Jahr 1970 nur 20,3% , im Jahr 2007 aber schon 93% der Tonsillenkarzinome HPV-positiv waren, wobei jeweils die gleiche Nachweismethode verwandt wurde (Näsman et al., 2009). Ähnliche Ergebnisse ergab eine Meta-Analyse, die vor dem Jahr 2005 40% HPV-positive Oropharynxkarzinome fand, aber einen steilen Anstieg auf 72% HPV-positive Tumore in Studien der Jahre 2005 - 2009 verzeichnete. Der Anteil an HPV-positiven Tumoren außerhalb des Oropharynx war während dieser Zeit konstant geblieben (Mehanna et al., 2012).

Der Einfluss von HPV auf Tumorlokalisationen außerhalb des Oropharynx ist umstritten, scheint in der Summe aber eine deutlich geringere Rolle zu spielen. So war in einer Metastudie der Anteil an HPV-positiven Tumoren in Larynxkarzinomen signifikant niedriger als in Oropharynxkarzinomen (Kreimer, 2005). Für Mundhöhlenkarzinome sind die Ergebnisse ähnlich (s. Abschnitt 1.3.4).

Als Gründe für den Anstieg HPV-assoziiierter Tumore werden hauptsächlich Veränderungen des sexuellen Verhaltens verantwortlich gemacht (Argiris et al., 2008). Dementsprechend steigt das Risiko für die Entwicklung eines HPV-16-positiven Oropharynxkarzinoms mit der Anzahl verschiedener Geschlechtspartner, bei Oralverkehr mit >6 verschiedenen Partnern, erstem Geschlechtsverkehr bei einem Alter unter 17 Jahren und seltenem Kondomgebrauch (D'Souza et al., 2007). Diese Risikofaktoren sind denen einer zervikovaginalen HPV-Infektion bei Frauen sehr ähnlich und machen eine sexuelle Übertragung wahrscheinlich (Hariri et al., 2011).

1.3.3 Weitere Risikofaktoren

Neben den oben beschriebenen Risikofaktoren Tabak, Alkohol und HPV-Infektionen, die für die überwiegende Mehrheit der Kopf-Hals-Tumore verantwortlich gemacht werden, lassen sich auch weitere Faktoren abgrenzen, die einen Einfluss auf das Tumorrisiko haben. Da ihr Effekt insgesamt aber deutlich schwächer ist, werden sie hier zusammengefasst dargestellt.

Positive Familienanamnese und genetische Prädisposition. Mehrere Studien haben eine Steigerung des Risikos für Kopf-Hals-Tumore um das 2- bis 4-fache bei Patienten mit positiver Familienanamnese, also dem Vorkommen eines Kopf-Hals-Tumors bei Verwandten 1. Grades, beschrieben (Gillison, 2007). Dieser Effekt ist besonders ausgeprägt, wenn Geschwister betroffen sind und unter Rauchern. Ob es sich dabei aber tatsächlich um eine direkt kausal wirkende genetische Komponenten –z.B. Polymorphismen in Enzymen der Biotransformation oder in Onko-/Tumorsuppressorgenen– handelt ist nicht klar. Denkbar ist auch, dass Verwandte ein ähnliches Risikoverhalten gegenüber Tabak und Alkohol oder eine ähnliche Neigung zu Nikotinabhängigkeit haben (Negri et al., 2009).

Verschiedene Tumorsyndrome wie z.B. das Li-Fraumeni-Syndrom oder auch Xeroderma pigmentosa sind mit HNSCC assoziiert worden. Am besten untersucht ist die Fanconi-Anämie, eine autosomal-rezessive Erberkrankung, die in Betroffenen ein erhöhtes Risiko für Leukämien und ein >500fach erhöhtes Risiko für HNSCC bewirkt, wobei 84% HPV assoziiert scheinen (Gillison, 2007). Diese Syndrome sind aber sehr selten und damit auf Populationsebene wenig relevant.

Orale Hygiene. Besonders für Mundhöhlenkarzinome hat sich eine schlechte orale Hygiene als Risikofaktor heraus gestellt. So steigt das Risiko mit der Zahl der schadhafte und fehlenden Zähne und ist umgekehrt proportional zur Zahl der Zahnarztbesuche. Dies ist unabhängig vom Tabak- und Alkoholkonsum (Rosenquist et al., 2005). Teilweise werden auch chronische Entzündungszustände in der Mundhöhle durch mangelnde orale Hygiene beschrieben, welche z.T. synergistisch mit weiteren Karzinogenen wirken könnten (Döbrossy, 2005).

Ernährung. Der Einfluss der Ernährung auf das Risiko von Kopf-Hals-Tumoren ist gut untersucht. So fand das „International Head and Neck Cancer Epidemiology Consortium (INHANCE)“ eine positive Korrelation zwischen HNSCC-Risiko und dem Verzehr von Fleisch,

wobei insbesondere „rotes“ Fleisch, also z.B. Rindfleisch, risikoe erhöhend war. Einen protektiven Effekt ergab ein erhöhter Verzehr frischer Früchte. (Chuang et al., 2011). Die Aufnahme von Nitrosaminen in gepökeltem Fisch erhöht vor allem in Asien das Risiko für Nasopharynxkarzinome (Vokes et al., 1993).

Belastungen durch Umwelt und Arbeitsplatz. Auch verschiedene Stoffe in der Luft erhöhen das Risiko für Kopf-Hals-Tumore. So wirken beispielsweise Stäube von Buchen- und Eichenholz sowie Nickel bei beruflich exponierten Personen als Karzinogene (Vokes et al., 1993). Zusätzlich sind Asbest und Luftverunreinigungen durch die Verbrennung fossiler Brennstoffe als unabhängige Risikofaktor für Larynxkarzinome bzw. HNSCC eingestuft worden (Sturgis und Cinciripini, 2007).

1.3.4 Ätiologie und Risikofaktoren für Tumore der Mundhöhle und der mobilen Zunge

Grundsätzlich gelten für Mundhöhlenkarzinome die gleichen Risikofaktoren wie für die anderen Kopf-Hals-Tumore wie sie weiter oben skizziert wurden, wobei vornehmlich Männer >60 Jahre betroffen sind und Tabak- und Alkoholkonsum die wichtigste Rolle spielen (Llewellyn et al., 2001).

Im Gegensatz zu anderen tabakassoziierten Tumoren hat die Inzidenz von Mundhöhlenkarzinomen in den letzten Jahren allerdings nicht deutlich abgenommen. Stattdessen wurden für Tumore der Mundhöhle und der mobilen Zunge, welche einen Großteil der Mundhöhlenkarzinome ausmachen, ein Anstieg der Inzidenz festgestellt (vgl. Abb. 3). Dieser Anstieg ist nicht hinreichend mit einer erhöhten Exposition gegenüber bekannten Risikofaktoren erklärbar. Ganz im Gegenteil ist er bei jungen Patienten ohne die klassischen Risikofaktoren und bei Frauen besonders ausgeprägt (Brown et al., 2012; Chitapanarux et al., 2006). Gehäuft treten Tumore der Mundhöhle auch bei immunsupprimierten Empfängern von Knochenmark- und Organtransplantationen auf (Curtis et al., 1997; Hasegawa et al., 2004). Da die Mundhöhle zusätzlich ein Ort des ständigen Kontakts zu Fremdmaterial (inkl. Körperflüssigkeiten) ist, legen diese Befunde eine infektiöse Ätiologie, z.B. durch ein Virus wie HPV, nahe.

Während der Einfluss von HPV auf die Pathogenese von Oropharynxkarzinomen allerdings durch zahlreiche Studien belegt und gesichert ist, sind die Ergebnisse für Mundhöhlenkarzinome und Tumore der oralen Zunge mindestens uneindeutig. Während ein positiver Titer für HPV-16 Antikörper in einer Studie ein unabhängiger Risikofaktor für die Entwicklung eines

Mundhöhlenkarzinoms war, stellten zwei weitere Studien keinen Zusammenhang fest (Gillison, 2007). Die Angaben zur Prävalenz einer HPV-Infektion liegen in der Mehrzahl der Studien allerdings deutlich unter denen des Oropharynx. Für die Gesamtheit aller Mundhöhlenkarzinome liegt die Rate HPV-positiver Tumore lediglich zwischen < 2 und 6% (El-Mofty und Lu, 2003; Herrero et al., 2003; Lingen et al., 2012; Machado et al., 2010; Ringström et al., 2002). Und auch bei Untersuchungen speziell zur HPV-Prävalenz in Tumoren der mobilen Zunge ergaben sich vergleichbare Werte (Dahlgren et al., 2004; Kabeya et al., 2011; X-H Liang et al., 2008; Salem, 2010; Siebers et al., 2008). Zusammengefasst spielt HPV also höchstens eine untergeordnete Rolle in der Pathogenese dieser Tumorentitäten. Stattdessen wurde eine Assoziation dieser Tumore mit anderen Viren (u.a. mit Herpes simplex Virus) postuliert, aber bisher nicht systematisch überprüft (Salem, 2010). Ein Ziel dieser Arbeit war es deshalb, ein nukleinsäurebasiertes Verfahren zu entwickeln, das die Suche nach allen bekannten und unbekanntem Viren, die eine Verwandtschaft zu schon bekannte Viren haben, ermöglicht. Anschließend sollte es zur Untersuchung einer potentiellen viralen Ätiologie oraler Zungentumoren in der Gruppe der Patienten bis 45 Jahre angewendet werden.

1.4 Klinische Aspekte von Kopf-Hals-Tumoren

1.4.1 Klinik

Die meisten Patienten mit Kopf-Hals-Tumoren stellen sich zunächst mit unspezifischen Symptomen vor, die z.T. von der jeweiligen Tumorlokalisation abhängig sind. Dazu gehören u.a. Heiserkeit, Dysphagie, Paukenergüsse, Ausfälle der Hirnnerven, Stridor und Ulzerationen der oralen Mukosa. Häufig sind auch zervikale Lymphadenopathien das erste Zeichen. Der extremste Fall ist das sog. CUP-Syndrom („cancer of unknown primary“), das bei ca. 10% der Patienten vorkommt. Es treten nodale Metastasen auf, während ein Primarius zunächst unbekannt ist und erst bei eingehender Diagnostik oder –bei 5% der Patienten– gar nicht zu finden ist. Eher selten sind Fernmetastasen zum Zeitpunkt der Diagnose (ca. 10% der Patienten), wobei die Lunge am häufigsten betroffen ist (Argiris et al., 2008; Sanderson und Ironside, 2002).

1.4.2 Diagnostik, Staging und Therapie

Diagnostische Schritte. Der erste Schritt bei Verdacht auf das Vorhandensein eines Kopf-Hals-Tumors besteht in der genauen klinischen Untersuchung gefolgt von einer indirekten Laryn-

goskopie bzw. einer Panendoskopie (Vokes et al., 1993). Um die genaue Ausdehnung zu ermitteln sind bildgebende Verfahren anzuschließen. Am häufigsten wird eine CT- und/oder eine MRT-Untersuchung durchgeführt, wobei letztere eine genauere Abgrenzung der Weichgewebe und damit der Eindringtiefe des Tumors zulässt. Ein PET-CT mit radioaktiv markierter 18-Fluordesoxyglucose erlaubt den Nachweis auch sehr kleiner Herde und eignet sich deshalb zur Suche nach Lymphknoten- und Fernmetastasen oder von Rezidiven (Argiris et al., 2008). Zusätzlich wird auch eine Röntgen- bzw. CT-Aufnahme des Thorax angefertigt, um Fernmetastasen und synchrone Lungentumore auszuschließen (Sanderson und Ironside, 2002).

Die Diagnose wird gesichert, indem Gewebe der verdächtigen Primärläsion per Biopsie und solches vergrößerter Lymphknoten per Feinnadelaspiration gewonnen und anschließend histopathologisch untersucht wird (Sanderson und Ironside, 2002).

Staging. Ein genaues Staging ist essentiell, um die Therapie und die Prognose der Patienten bestimmen zu können. Grundlage des Stagings ist die anatomische Ausdehnung des Tumors, eingeteilt nach dem TNM-System der UICC und des American Joint Committee on Cancer (AJCC). Kriterien sind die lokale Tumorausbreitung (T1-4), ein Befall von Lymphknoten (N0-3) und das Vorliegen von Fernmetastasen (M0-1). Die Definition der verschiedenen T- Stadien ist dabei abhängig von der Tumorlokalisation (Patel und Shah, 2005). Zur Vereinfachung werden verschiedene TNM-Kombinationen mit vergleichbarer Prognose zu vier Tumorstadien zusammengefasst (s. Tab. 1). Ein Nachteil dieses Klassifikationssystems ist die fehlende Einbindung weiterer Faktoren, z.B. pathologischer oder molekularbiologischer Art, die besonders für die Auswahl und Bedeutung nicht-chirurgischer Therapiemodalitäten relevant sein könnten (van der Schroeff und de Jong, 2009).

Tab. 1: UICC-Tumorstadien (nach van der Schroeff und de Jong, 2009).

UICC-Stadium	T	N	M
I	T1	N0	M0
II	T2	N0	M0
III	T1-2	N1	M0
	T3	N0-1	M0
IVa/b	jedes T	N2-3	M0
	T4	jedes N	M0
IVc	jedes T	jedes N	M1

Therapie. Der wichtigste Faktor für die Auswahl der geeigneten Therapie ist das Tumorstadium, wobei bis zum Vorliegen von Fernmetastasen (Stadium IVc) mit kurativer Intention behandelt wird (Seiwert und Cohen, 2005).

In lokal begrenzten Stadien (I und II) kann eine rein chirurgische Therapie meist mit gutem Erfolg durchgeführt werden. Besonders bei Tumoren der Mundhöhle und des Larynx ist die alleinige Bestrahlung eine Alternative, da sie die Organfunktion besser erhält. Dafür dauert sie länger, verhindert ein genaues Staging durch einen Pathologen und verursacht Strahlenschäden, z.B. Xerostomie (Haddad und Shin, 2008).

Mehr als 50% der Patienten haben zum Diagnosezeitpunkt lokal fortgeschrittene Tumore (Stadium III-IV M0). Die Standardtherapie ist in diesen Fällen eine Chemotherapie oder eine Operation gefolgt von einer Radiochemotherapie (Seiwert und Cohen, 2005). Auch haben sich Überlebensvorteile für Patienten mit Induktionschemotherapie und anschließender definitiver lokaler Therapie ergeben (Argiris et al., 2008). Für Patienten mit metastasierten oder rezidivierenden Tumoren gibt es wenige wirkungsvolle Behandlungsoptionen und sie werden in der Regel mit palliativer Chemotherapie, ggf. auch mit Bestrahlung behandelt. Zusätzlich wurde in Studien auch der gegen EGFR (epithelial growth factor receptor) gerichtete monoklonale Antikörper Cetuximab kombiniert und allein mit positiven Effekten auf die Prognose eingesetzt (Haddad und Shin, 2008).

Eine operative Therapie ist besonders auch für Patienten mit Tumoren der Mundhöhle und der mobilen Zunge belastend, da sehr häufig zu multimodalen Therapiekonzepten gegriffen werden muss und notwendige Tumorresektionen oft entstellend wirken und Sprache und Schlucken beeinträchtigen (Salem, 2010)

1.4.3 Einflussfaktoren auf die Prognose

Von der Vielzahl an pathologischen, klinischen und molekularen Faktoren, welche die Prognose beeinflussen und in Prognoseeinschätzung und Risikostratifizierung Verwendung finden, werden im Folgenden diejenigen besprochen, welche die größte Relevanz für die vorliegende Arbeit, aber auch für die klinische Praxis haben.

Tumorstadium. Der wichtigste Faktor zur Prognoseabschätzung ist das UICC-Tumorstadium. Darunter ist der nodale Status der einflussreichste einzelne Faktor. So verkürzt das Vorliegen einer Lymphknotenmetastase bei Erstdiagnose das mediane Überleben um 50% (Patel und Shah,

2005). In den lokal begrenzten Stadien I und II können 90% bzw. 70% der Patienten mit Bestrahlung oder Operation geheilt werden. Im Gegensatz dazu entwickeln >50% der Patienten mit lokal fortgeschrittenen Stadien (Stadium III und IV, M0) innerhalb von 2 Jahren Rezidive oder Fernmetastasen, ihr Gesamtüberleben liegt nach drei Jahren nur bei 30-50% (Argiris et al., 2008). Das mediane Überleben von Patienten mit Rezidiven oder Metastasen wiederum beträgt nur 6 bis 9 Monate (Sanderson und Ironside, 2002).

Alter und Geschlecht. Im Allgemeinen werden keine signifikanten Unterschiede z.B. im 5-Jahresüberleben zwischen den Geschlechtern festgestellt (Piccirillo et al., 2007). Die Daten für Überlebensunterschiede in Abhängigkeit vom Alter dagegen sind widersprüchlich (Massano et al., 2006). Besonders auch für Tumore der Mundhöhle und der oralen Zunge fanden einige Studie zwar ein verringertes Überleben für jüngere Patienten (Bello et al., 2010; Garavello et al., 2007), andere stellten aber keinen Unterschied fest (Sargeran et al., 2008; Sasaki et al., 2005).

Alkohol und Tabak. Neben ihrer Rolle in der Ätiologie von Kopf-Hals-Tumoren haben Alkohol- und Tabakkonsum auch einen negativen Einfluss auf die Prognose (Massano et al., 2006).

Tumorlokalisation. Innerhalb der Primärtumorlokalisationen der Kopf-Hals-Tumore gibt es insgesamt gesehen Unterschiede der Prognose, die sich u.a. aus der unterschiedlichen vaskulären und lymphatischen Versorgung und der Struktur der umgebenden Gewebe ergibt. Die 5-Jahresüberlebensrate fast aller Plattenepithelkarzinome liegt aber zwischen 50 und 60%. Lediglich der Hypopharynx ist mit einer 5-Jahresüberlebensrate von 30% deutlich schlechter (Piccirillo et al., 2007).

HPV-Status. Für Tumore des Oropharynx ist eine bessere Prognose und eine erhöhte 5-Jahresüberlebensrate beschrieben, wenn eine Infektion mit HPV vorliegt (Ang et al., 2010; Fakhry et al., 2008). Dies wird z.T. durch ein besseres Ansprechen auf eine Radiotherapie erklärt (Lindel et al., 2001). Außerhalb des Oropharynx sind die Befunde kontrovers. Einige Autoren beschreiben keine prognostischen Unterschiede, während bei anderen HPV-positive Tumore eine bessere Prognose haben (Massano et al., 2006; Ragin und Taioli, 2007; Schwartz et al., 2001). Zu beachten ist, dass die Infektionsrate außerhalb des Oropharynx deutlich geringer ist (s.o.) und dass die HPV-Nachweismethode eine Rolle spielt. So ist die erhöhte Expression von p16^{INK4A},

die häufig als Surrogatparameter einer HPV-Infektion verwendet wird (Jordan et al., 2012; Robinson et al., 2010) außerhalb des Oropharynx womöglich wenig verlässlich. Studien zeigen, dass p16-Immunhistochemie-Ergebnisse in Mundhöhlenkarzinomen eine hohe Falsch-Positiv-Rate haben und demnach auch die Prävalenz HPV-positiver Tumore überschätzen könnten (Kabeya et al., 2011; Lingen et al., 2012).

1.5 Onkoviren

1.5.1 Bedeutung von Onkoviren

Die Kenntnis infektiöser Ursachen in der Karzinogenese ist von herausragender Bedeutung, da sich daraus wichtige Implikationen für Diagnostik, Therapie und Prävention – z.B. im Rahmen von Impfprogrammen – ergeben. Schätzungen gehen davon aus, dass weltweit insgesamt etwa 2 Millionen (16,1%) aller Krebserkrankungen durch infektiöse Agentien kausal verursacht werden. Von diesen gehen 1,9 Millionen auf Kosten von *Helicobacter pylori*, Hepatitis B und C Viren und HPV (de Martel et al., 2012).

Für sechs Viren liegen ausreichende epidemiologische und mechanistische Daten vor, um von der „International Agency for Research on Cancer“ (IARC) als „Karzinogene in Menschen“ klassifiziert zu werden (Bouvard et al., 2009). Diese machen ca. 12% aller Krebserkrankungen weltweit aus und umfassen Epstein-Barr-Virus (EBV: Burkitt-Lymphom, Nasopharynxkarzinom, Hodgkin-Lymphom), das Humane Herpesvirus 8 (HHV-8: Kaposi-Sarkom), Hepatitis B und C Virus (HBV, HCV: Hepatozelluläres Karzinom), das Humane T-Zell-Leukämie Virus (HTLV-1: humane T-Zell Leukämie) und Humane Papillomviren (HPV: Zervixkarzinom, Oropharynxkarzinom, Analkarzinom) (Sarid und Gao, 2011). Für zahlreiche andere Viren ist eine Assoziation mit humanen Tumoren beschrieben, die Daten reichen aber nicht zum Nachweis einer klaren Kausalität oder die Ergebnisse sind (z.T. heftig) umstritten. Beispiele sind u.a. das Merkel-Zell Polyoma Virus (MCPyV: Merkelzell-Karzinom), das Xenotropic murine leukemia virus (XMRV: Prostatakarzinom) oder das Simian 40 Virus (SV40: Mesotheliom) (Feng et al., 2008; Schlaberg et al., 2009).

1.5.2 Virale Onkogenese

Die Mechanismen viraler Onkogenese sind vielfältig und von Virus zu Virus verschieden. Als obligat intrazelluläre Parasiten sind Viren auf den Stoffwechsel ihrer Wirtszellen angewiesen und haben deshalb Wege entwickelt um in den Zellzyklus einzugreifen und die Zellproliferation zu steigern. Grundsätzlich bewirken sie dazu entweder die Akkumulation von genetischen Veränderungen bzw. lassen diese zu oder exprimieren virale Onkogene, die die Kontrolle von Zellwachstum und Apoptose beeinflussen (Sarid und Gao, 2011). So inhibiert z.B. das Tax Protein des HTLV-1 verschiedene DNA-Reparaturmechanismen und führt so zu genetischer Instabilität (Martin und Gutkind, 2009), während die Latent membrane proteins (LMP) 1 und 2 des EBV direkte prostimulatorische Signale in Lymphozyten auslösen (Biegging et al., 2009) und HPV beide Wege geht (vgl. Abb. 4).

Ein gemeinsames Merkmal ist, dass alle bekannten Onkoviren mit der Expression spezifischer viraler regulatorischer und/oder strukturgebender Proteine assoziiert sind, welche deshalb zum Nachweis genutzt werden können (Poreba et al., 2011). Der Nachweis viraler Transkripte ist grundsätzlich einer der primären Wege um Onkoviren zu detektieren und kann zusätzlich verwendet werden, um einen aktiven viralen Einfluss auf den Tumor nachzuweisen und von einer zufälligen, koinzidentiellen viralen Infektion zu unterscheiden (Arron et al., 2011; Braakhuis et al., 2004; Wiest et al., 2002). Aus diesen Gründen ist der Nachweis viraler mRNA-Transkripte bzw. transkriptionaler Aktivität dem Nachweis viraler DNA in zahlreichen Studien überlegen, um die Bedeutung eines Virus in einem Tumor zu evaluieren (Attner et al., 2010; Jung et al., 2010; C Liang et al., 2012).

1.5.3 Detektion von Onkoviren

Die Identifikation neuer Onkoviren folgt in der Regel einem von zwei Ansätzen: entweder über eine Aufreinigung und Isolierung viraler Partikel mit Filtern, Zentrifugation, Zellkultursystemen und Nachweis z.B. per Elektronenmikroskopie oder über den Nachweis viraler Nukleinsäuren, die eine Homologie zu bekannten Viren aufweisen (Lipkin, 2010). Methoden der zweiten Art haben in den letzten Jahren erhebliche Verbreitung erfahren, da sie die Vorteile sehr hoher Sensitivität mit dem eines geringeren technischen Aufwands vereinen.

Bisher war dazu allerdings die Vorauswahl von „Kandidaten-Viren“ notwendig, um beispielsweise (ggf. degenerierte) PCR-Primer zur Vervielfältigung der Nukleinsäuren entwerfen zu können, was die Suche nach unbekanntem Viren erschwerte (Rota et al., 2003; Sarid und Gao, 2011). Die Entwicklung viraler Genchips, auf denen die Sequenzen mehrerer hundert Viren vertreten sind, hat das Spektrum zwar erweitert, aber dennoch dieselben prinzipiellen Limitationen in der Identifikation unbekannter Viren (Palacios et al., 2007; D Wang et al., 2002). Erst mit dem Aufkommen der Next-generation sequencing (NGS)-Technologie (s.u.) wurde die voraussetzungsfreie Entdeckung möglich, da die in einer Probe vorliegenden Nukleinsäuren amplifiziert und anschließend sequenziert werden können ohne im Voraus einen Teil dieser Sequenzen kennen zu müssen. Die praktische Umsetzbarkeit dieser Idee wurde u.a. mit der Identifikation des Merkel-cell Polyoma Virus in Merkel-Zell-Karzinomen demonstriert (Feng et al., 2008).

Während der Arbeit an den weiter unten beschriebenen Versuchen im Jahr 2009 gab es noch keine etablierten Methoden um die noch recht neue NGS-Technologie zur Suche nach Onkoviren einzusetzen und die großen Datenmengen dementsprechend auszuwerten und zu verarbeiten. In der Zwischenzeit ist zwar ein entsprechender Algorithmus öffentlich zugänglich geworden (Kostic et al., 2011). Dieser setzt allerdings die Benutzung aufwändiger und teurer Ressourcen (z.B. von Großrechner- bzw. Cloud-Computing-Kapazitäten) voraus, während der hier entwickelte Analyseansatz mit den Hilfsmitteln eines durchschnittlichen Labors und mit frei erhältlichen bzw. selber geschriebenen Programmen durchführbar ist.

1.6 Next-Generation-Sequencing und RNA-Seq Technologie

In den letzten 35 Jahren basierten fast alle Verfahren der DNA-Sequenzierung (inkl. der kapillarelektrophoretischen) auf der sog. Kettenabbruch-Methode, die Sanger 1977 entwickelte (Liu et al., 2012). Seit dem Jahr 2005 allerdings ist die Kapazität der DNA-Sequenzierung exponentiell gestiegen, während die Preise pro sequenziertem Nukleotid noch stärker gefallen sind (vgl. Abb. 5). Dies wurde ausgelöst durch die Einführung einer neuen Generation an Sequenziermaschinen (Next-Generation Sequencing, NGS), die zwar chemisch und technisch untereinander sehr verschieden sind, aber alle einem ähnlichen Prinzip folgen.

Anstatt nacheinander je ein möglichst langes DNA-Fragment (bei Sanger max. 600-1000bp) zu sequenzieren, werden DNA-Fragmente in kürzere Abschnitte zerlegt, anschließend aber mehrere Millionen dieser kürzeren Abschnitte im gleichen Experiment parallel ausgelesen (massively parallel sequencing). Die sich ergebenden kurzen Sequenzabschnitte werden Short-Reads (oder auch nur Reads) genannt, da sie z.B. im Fall des Illumina Genome Analyzer II lediglich 36 – 54 Basenpaaren lang sind. Dennoch ergibt sich eine hohe Gesamtmenge an ausgelesenen Nukleotiden, da diese eine Funktion der Länge der Reads (36bp) mal der Anzahl der sequenzierter Reads ($>10^6$) ist.

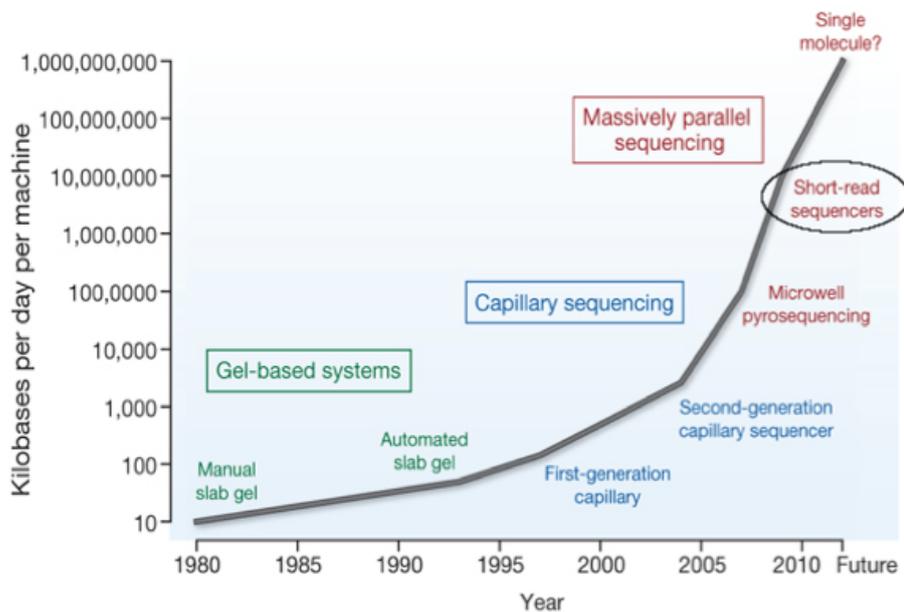


Abb. 5: Entwicklung der DNA-Sequenzierung der letzten 30 Jahre. Der steile Anstieg seit dem Jahr 2005 ist auf die Verwendung von Next-Generation-Sequencing Technologie begründet (modifiziert nach Stratton et al., 2009).

Eine Variante des NGS, die auch für dieses Projekt verwendet wurde, ist die sog. RNA-Seq-Technologie. Dabei werden möglichst viele –im Idealfall alle– mRNA-Moleküle einer Tumorseite isoliert, in cDNA überführt und anschließend sequenziert. Auf diese Weise erhält man einen Schnappschuss der zu diesem Zeitpunkt exprimierten Gene (des sog. Transkriptoms) und kann Aussagen sowohl über genetische Veränderungen als auch über die Genexpressionslevel machen (Z Wang et al., 2009).

Einer der wichtigsten Schritte in der Analyse von RNA-Seq Daten ist die Zuordnung der Reads zu Referenzsequenzen, z.B. zum humanen oder auch zu viralen Genomen, um den Ursprung des

jeweiligen Reads zu ermitteln. Dies geschieht über paarweise Vergleiche, sog. Alignments (s. Abb. 6).

```

Read:   A C G G T A T C T G
           |   | |   | | |   |
Referenz: A G G G A A T C G G T A ...

```

Abb. 6: Paarweiser Vergleich (Alignment) der Sequenz eines hier nur 10bp langen Reads mit einer Referenzsequenz; vertikale Striche zeigen Übereinstimmungen an.

Vereinfacht gesagt muss ein derartiges Alignment für jeden Read an jeder Stelle der Referenzsequenz durchgeführt und ein Bewertungsscore berechnet werden, bei dem für jedes übereinstimmende Nukleotid ein Wert addiert und für jeden Fehler („mismatch“) ein Wert subtrahiert wird. Die Position in der Referenzsequenz mit der höchsten Bewertungssumme stellt dann den wahrscheinlichsten Ursprung dar. An dieser Stelle wird deutlich, welchen Einfluss die Menge der Daten (ca. 1-2 Gigabyte große Textdateien mit Reads pro Tumorprobe) und die Größe der Referenzspielen.

Parallel zur ansteigenden Menge der pro Experiment gewonnenen Daten –und zur Größe der Referenzsequenz– steigt der Aufwand diese zu analysieren und damit die Notwendigkeit, sinnvolle und effiziente Analysestrategien zu entwickeln. Es gibt eine Vielzahl frei verfügbarer Programme die Alignments durchführen, sich aber erheblich in Laufzeitverhalten und im Umgang mit Fehlern und Lücken (gaps) im Alignment unterscheiden (Altschul et al., 1990; Ruffalo et al., 2011). Eine höhere Fehlertoleranz ist beispielsweise notwendig wenn Sequenzen verglichen werden sollen, die sich nicht sehr ähnlich sind (z.B. von miteinander verwandten Viren), erhöht aber drastisch die benötigte Rechenzeit. Die Auswahl der richtigen Parameter und der optimalen Programme für die jeweilige Aufgabe bei Durchführbarkeit auf einem Standardcomputer ist deshalb nicht trivial.

Die Kombination verschiedener bioinformatischer Werkzeuge und ihre Verknüpfung über selbstgeschriebene Programme als automatisierter Analysealgorithmus, um eine Suche nach bekannten und unbekanntem Viren in den RNA-Seq-Daten durchführen zu können, war deshalb eine der Zielsetzungen dieser Arbeit.

1.7 Ziele dieser Arbeit

Insgesamt gab es vier Ziele bei der Erstellung dieser Arbeit:

- 1) Da die Prognose von jungen Patienten (<45 Lebensjahre) mit Mundhöhlenkarzinomen im Vergleich zu Patienten mit anderen Kopf-Hals-Tumoren wenig erforscht ist, sollte eine Kohorte von insgesamt 748 Kopf-Hals-Tumorpatienten mit fortgeschrittenen Tumoren nach kombinierter Radiochemotherapie auf Unterschiede im Überleben analysiert werden.
- 2) Aufgrund der steigenden Inzidenz von Tumoren der Mundhöhle bzw. der mobilen Zunge bei jungen HPV-negativen Patienten ohne klassische Risikofaktoren ist eine virale Ätiologie durch ein ggf. noch unbekanntes Virus vermutet worden. Diese sollte anhand einer RNA-Seq-Analyse von sieben Gewebeproben aus Zungentumoren junger Patienten bestätigt bzw. ausgeschlossen werden.
- 3) Um Ziel 2) zu erreichen musste ein neuartiger Analysealgorithmus konzipiert und anschließend implementiert werden. Dieser sollte die RNA-Seq-Daten auf einem Standardcomputer verarbeiten und auf Sequenzen von bekannten und unbekanntem Viren (die mit bekannten Viren verwandt sind) untersuchen können.
- 4) Ein nachrangiges Ziel war die Bewertung der Expression von CDKN2A/p16^{INK4A} als Surrogatparameter einer HPV-Infektion in Mundhöhlenkarzinomen, da sich die p16-Expression dort –im Gegensatz zu Oropharynxkarzinomen– in verschiedenen Studien als unzuverlässiger Marker herausgestellt hat.

2. Methoden

2.1 Analyse der klinischen Daten

Um den möglichen Einfluss von Lebensalter, Lokalisation und klassischen Risikofaktoren wie z.B. Nikotinkonsum auf das Gesamtüberleben und das progressionsfreie Intervall von Patienten mit Hals- und Kopftumoren zu untersuchen, wurde eine retrospektive Analyse von Daten vorgenommen, die an der University of Chicago von Dr. Ezra Cohen und Dr. Dagogo-Jack erhoben worden waren. Ein besonderes Augenmerk wurde dabei auf die Gruppe jüngerer Patienten (≤ 45 Jahre) gelegt.

Insgesamt wurden anonymisierte Daten von 748 Patienten mit Plattenepithelkarzinomen der Hals- und Kopffregion (HNSCC) verwendet, die in den Jahren zwischen 1990 und 2007 im Rahmen prospektiver klinischer Studien unter kurativer Zielsetzung mit kombinierter Radiochemotherapie an den Kliniken für Hals-Nasen-Ohrenheilkunde („Otolaryngology“), Onkologie („Medical Oncology“) und Strahlenheilkunde therapiert wurden. In allen Fällen war die Präsenz von HNSCC histologisch und klinisch verifiziert worden, eine Erfassung des Tumorstadiums in der TNM- und der UICC-Einteilung, der anatomischen Lokalisation, des Geschlechts und des Alters erfolgte ebenfalls. Die Einordnung bezüglich des Nikotin- und Alkoholkonsums erfolgte über eine Selbsteinschätzung der Patienten zum jeweiligen Aufnahmezeitpunkt, wobei die Belastung semi-quantitativ mit Hilfe ordinalskaliert Variablen erfasst wurde. Nikotinkonsum: 1 = „niemals Raucher“, definiert als <100 Zigaretten im Leben geraucht, 2 = Pfeife/Zigarre, 3 = <20 pack years, 4 = 20 – 40 pack years. Alkoholkonsum: 1 = kein Alkoholkonsum, 2 = gelegentlicher Konsum, 3 = moderater Konsum und 4 = starker Alkoholkonsum.

Um Aussagen über die klinischen Unterschiede verschiedener Subpopulationen innerhalb der Kohorte treffen zu können wurden das progressionsfreie Überleben (PFS) und das Gesamtüberleben (Overall survival) berechnet, wobei der Zeitpunkt des Studieneintritts als Startpunkt fungierte, während die Endpunkte durch das Datum des letzten Nachbeobachtungstermins (Zensurereignis) und die relevanten Ereignisse durch Todeszeitpunkt bzw. Datum des Rezidivs gegeben waren. Überlebensanalysen wurden mittels Kaplan-Meier-Schätzern durchgeführt und mit dem Log-Rank-Test auf Signifikanz überprüft. Die Berechnungen wurden dabei nach den Studien und den jeweiligen Therapiegruppen stratifiziert um ungerichtete und unerwünschte Effekte zwischen verschiedenen Therapieprotokollen zu egalisieren. Um den Einfluss verschiedener Kovariate auf das Überleben zu untersuchen wurden mittels Cox-Regressions-Analyse Hazard ratios

berechnet, welche für die entsprechenden Variablen (Alter, Tumorstadium, Tabak- und Alkoholkonsum und anatomische Lokalisation) adjustiert waren (Cox, 1972; Ziegler et al., 2007).

Alle Analysen wurden in der statistischen Programmierumgebung „R - Version 2.13.1“ (R Development Core Team, 2011) in Verbindung mit dem Prozedurenpaket „survival“ (Therneau und Lumley, 2011) durchgeführt.

2.2 Gewebeproben von Tumoren der Mobilien Zunge

Aus der Tumorbank für Hals- und Kopftumore der University of Chicago („University of Chicago Head and Neck Cancer tissue bank“) wurden nach Genehmigung der vorliegenden Studie und Freigabe der Proben durch das „Institutional Review Board“ der University of Chicago (IRB approved protocol UCCCC#89800) sieben Tumorproben von Patienten mit Karzinomen der mobilen Zunge so ausgesucht, dass ihr Risikoprofil und ihre Charakteristika möglichst denen der jungen Patienten aus der klinischen Kohorte entsprechen sollten. Kriterien waren dabei das Alter, ein fehlender Tabak- und ein fehlender oder geringer Alkoholkonsum (≤ 1 Glas Wein oder Äquivalentes pro Tag).

Die gefrorenen Präparatschnitte wurden durch einen Pathologen begutachtet, der tumoröse Anteil markiert und anschließend eine Stanzbiopsie so durchgeführt, dass ein Tumoranteil von $\geq 70\%$ sichergestellt werden konnte. Die gewonnenen Proben wurden auf Trockeneis mit einem Mörser homogenisiert und DNA und RNA mit einem Qiagen AllPrep DNA/RNA MiniKit (Qiagen, USA) gemäß des beiliegenden Protokolls extrahiert. Mit einer etablierten nested-PCR-Methode (Sotlar et al., 2004) verifizierte ein Mitarbeiter des Labors den negativen HPV-Status.

2.3 Positivkontrollen

Zum Nachweis der korrekten Funktion des Virusalgorithmus unter verschiedenen Voraussetzungen wurden drei verschiedene Positivkontrollen verwandt, welche unterschiedliche Stärken des Verfahrens betonen.

2.3.1 HPV-16 positive Kopf-Hals-Tumorzelllinie (UM-SCC47)

Als erste Positivkontrolle wurde die Tumorzelllinie UM-SCC47 benutzt, da sie HPV-16 positiv ist (Gupta et al., 2009) und mehrere virale Genomkopien enthält (Bradford et al., 2003). Sie wurde von Dr. Thomas Carey (University of Michigan) zur Verfügung gestellt. Zellen wurden in

DMEM/F12 (Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12) (Thermo Scientific HyClone, USA) mit 10% FBS (Thermo Scientific HyClone, USA), 1% Penicillin-Streptomycin (Mediatech, USA) und 1% L-Glutamine (Mediatech, USA) kultiviert und regelmäßig auf Mycoplasmen getestet. Die Zellkultur wurde mit Einwegartikeln aus Plastik (Pipetten, Kulturflaschen, Filter) der Firmen Fisher Scientific und BD Falcon durchgeführt.

Inkubiert wurde die Kultur in Brutschränken bei 37,5 °C und 5% CO₂-Atmosphäre. Bei einer Zelldichte von 70-80% wurden die adhären wachsenden Zellen 2x mit PBS (Mediatech, USA) gewaschen und durch Zugabe von 2ml einer 1x Trypsin EDTA (Mediatech, USA) Lösung und 5-10 minütiger Inkubation bei 37 °C in Suspension gebracht. Anschließend wurde die Trypsinwirkung mit 8 ml DMEM gestoppt, die Suspension zentrifugiert und der Überstand verworfen. Das verbleibende Zellpellet wurden daraufhin 2x mit eisgekühltem PBS resuspendiert, zentrifugiert und die Überstände jeweils abpipettiert. Anschließend wurde die RNA mit Qiagen RNeasy columns (Qiagen, USA) dem Herstellerprotokoll folgend extrahiert.

2.3.2 EBV-Transgenes Maus-Lymphom

Als zweite Kontrolle wurde ein Mausmodell für Lymphome genutzt (Biegling et al., 2009), das an der University of Chicago sequenziert wurde und dessen Daten von Prof. Longnecker (Northwestern University, Chicago) zur Verfügung gestellt wurden. Dieses Mausmodell ist transgen für das onkogene EBV LMP2A-Gen (Latent Membrane Protein 2A), welches in B-Zellen einen Überlebensvorteil bewirkt (Mancao und Hammerschmidt, 2007) und dessen Expression in Burkitt Lymphomen detektierbar ist (AI Bell et al., 2006). Auf diese Weise sollte die Fähigkeit des Algorithmus gezeigt werden, in den Sequenzdaten ein einzelnes Virusgen zu detektieren und Daten eines nicht-humanen Wirts zu verarbeiten.

2.3.3 HPV-16 positives Zervixkarzinom

Als dritte Kontrolle wurde ein öffentlich zugänglicher Datensatz eines RNA-Seq-Experimentes, in dem u.a. die Sequenzen eines HPV-16 positiven Zervixkarzinoms enthalten waren (Arron et al., 2011), vom NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/Traces/sra>), einer öffentlichen Datenbank für genetische Experimente (Leinonen et al., 2010), heruntergeladen. Die Rohdaten enthielten Reads verschiedener Tumore, da sie in einem multiplex Ansatz erhoben

worden waren. Multiplex heißt in diesem Zusammenhang, dass mehrere Tumoren in einem einzelnen Reaktionsgefäß gleichzeitig sequenziert werden und die jeweiligen Reads im Anschluss wieder den entsprechenden Proben zugeordnet werden müssen. Diese Zuordnung geschieht computerbasiert anhand von 3bp langen Oligonukleotidabschnitten, den sog. Barcodes, die während der Probenvorbereitung zwischen Adapter (s.u.) und cDNA ligiert wurden, für jede Probe eindeutig sind und bei der Sequenzierung mit erfasst werden. Dies bedeutete, dass die in einer Textdatei vorliegenden Daten zunächst mit selbst programmierten Perl-Skripten und mittels Funktionen des FASTX-Toolkit (Blankenberg et al., 2010) sortiert, die Zervixkarzinomsequenzen extrahiert und vorformatiert werden mussten, um sie anschließend mit dem Virusalgorithmus verarbeiten zu können.

Dies diente der Demonstration, dass auch RNA-Seq-Datensätze verwendbar sind, die an anderen Institutionen und mit unterschiedlichen Techniken erzeugt wurden.

2.4 Transcriptome Sequencing

2.4.1 Erstellung der „RNA-Seq Libraries“

Die Erstellung der „RNA-Seq Libraries“ folgte einer modifizierten Version des paired-end Protokolls „mRNA Sequencing – Sample Preparation Guide“ des Herstellers der Sequenziermaschinen (Illumina, USA).

Qualität, Menge und Reinheit der aus den Tumorproben und der Zelllinie UM-SCC47 extrahierten totalen RNA wurde mittels NanoDrop (Thermo Scientific, USA) getestet, die Integrität mittels Agilent 2100 Bioanalyzer (Agilent Technologies, USA) sichergestellt und je 5 - 10µg RNA als Startmaterial verwendet.

Da die nicht-kodierende RNA (z.B. ribosomale RNA) über 90% der gesamten RNA einer Zelle ausmacht (Mattick, 2001) erfolgte zunächst eine Reinigung der messenger RNA (mRNA) mittels Ambion MicroPoly(A)Purist Kit (Invitrogen, USA), welcher mit an Säulen immobilisierten oligo-dT-Sequenzen polyadenylierte RNA bindet, während der Rest verloren geht. Da es während der reversen Transkription zur Überproportionalen Amplifikation des 5'-Endes kommen kann (sog. 5' bias) und Sekundärstrukturen innerhalb der mRNA-Moleküle die Effizienz mindern

(Hansen et al., 2010; Mortazavi et al., 2008) wurde die mRNA chemisch in ca. 300bp lange Fragmente zerlegt.

Anschließend erzeugte ein erster Reverse-Transkriptase-Schritt mit „Super Script II Reverse Transcriptase“ (Invitrogen, USA) und unspezifisch („random“) bindenden hexameren Primern (Random Primer 9, New England Biotechnology, USA) einsträngige cDNA Fragmente. Durch die Zugabe von RNaseH (Invitrogen, USA) wurde die verbliebene mRNA zerschnitten und so von der cDNA entfernt. An der dadurch frei vorliegenden einsträngigen cDNA wurden nun komplementäre Stränge mittels DNA Polymerase (DNA Pol I, Invitrogen, USA) und damit doppelsträngige DNA (dsDNA) synthetisiert, welche anschließend mit dem QIAquick PCR Purification Kit (Qiagen, USA) aufgereinigt wurde.

Um Überhänge an den Enden der dsDNA-Fragmente zu entfernen, die durch unterschiedlich lange Stränge entstanden sind und „stumpfe“ Enden zu erzeugen, erfolgte eine Reaktion mit T4 DNA Polymerase und Klenow DNA Polymerase (Invitrogen, USA). Die 3' zu 5' Exonukleaseaktivität dieser Enzyme entfernt 3'-Überhänge und die Polymeraseaktivität füllt 5'-Überhänge auf, so dass beide Stränge eines dsDNA-Fragments genau gleich lang werden. An die 3'-Enden dieser jetzt stumpfen DNA-Fragmente wurde dann mit der Polymeraseaktivität des Klenow-Fragment-Enzyms (3' zu 5' Exo-, Invitrogen, USA) jeweils eine einzelne Adenin-Base ligiert, die also daraufhin über das Ende hinausragt und die Ligation der Adapter vorbereitet, welche einen einzelnen komplementären 3'-Thymin-Überhang haben. Zweitens verhindert sie im nächsten Schritt die Bildung von Adapter-Adapter- sowie cDNA-cDNA-Dimeren. Die DNA wurde anschließend mit dem MinElute PCR Purification Kit (Qiagen, USA) gereinigt und die je 150bp langen, für die Sequenzierung notwendigen paired-end Adapter (PE Adapter p7 & p5, Illumina, USA) mit T4 DNA Ligase ligiert (vgl. Abb. 7) und anschließend erneut mit dem MinElute PCR Purification Kit (Qiagen, USA) gereinigt.

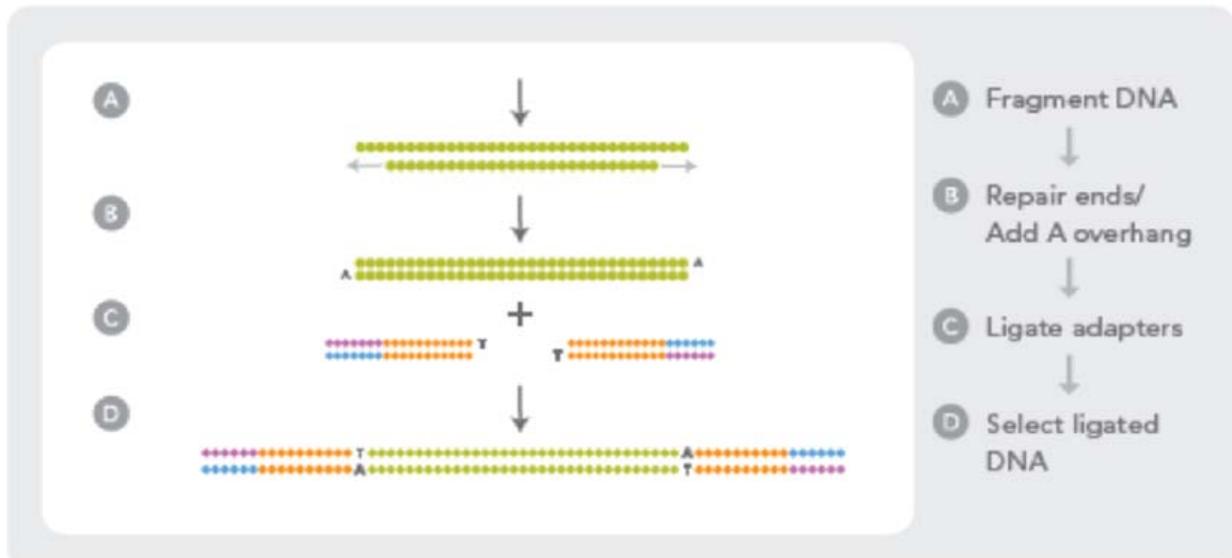


Abb. 7: Schema der cDNA Reperatur und Adapterligation (modifiziert nach Ansorge, 2009)

Da Fragmente mit einer Länge von ca. 500-700bp für die Sequenzierung ideal sind, wurden die entstandenen DNA-Produkte nach ihrer Größe sortiert, indem sie auf einem 2% NuSieve Agarose Gel (Fisher Scientific, USA) für 45min bei 120V aufgetrennt wurden. Die bei 600bp liegenden Banden wurden ausgeschnitten, die entsprechende DNA aus der Gelbande extrahiert (QiaQuick gel extraction kit, Qiagen, USA) und im Anschluss daran in 18 PCR-Zyklen mit Platinum *Pfx* DNA Polymerase (Invitrogen, USA) und einem adapterspezifischen Primerset (Illumina PE primer, Illumina, USA) amplifiziert. Nach einem letzten Reinigungsschritt mit dem QIAquick PCR Purification Kit (Qiagen, USA) folgte eine abschließende Kontrolle auf korrekte und ausreichende Länge, Reinheit und Konzentration der adapterligierten cDNA-Produkte mittels Messungen im Agilent 2100 Bioanalyser (Agilent Technologies, USA) und NanoDrop (Thermo Scientific, USA).

2.4.2 Sequenzierung

Für die Generierung der Sequenzinformationen wurden die vorbereiteten Proben an die Argonne National Laboratories (Bolingbrook, IL) versandt, wo sie auf einem Illumina Genome Analyzer II (Illumina, USA) sequenziert wurden. Dies geschah nach einem Standardprotokoll für entweder 36bp oder 54bp lange Reads mit je einer Spur (Lane) pro Probe in der GAII Flowcell (Illumina, USA). Die ausgelesenen Basen wurden nach Qualität gefiltert, zu Reads zusammengesetzt und die sich ergebenden Sequenzen als Textdateien zurück übermittelt.

2.5 Virusalgorithmus

Eine Übersicht über die verschiedenen Schritte und die beiden Hauptzweige des Virusalgorithmus, welche sich in einen schnellen aber unspezifischeren („rapid algorithm“) und einen komplexeren mit höherer Laufzeit aber größerer Sensitivität („discovery algorithm“) unterteilen, ist in Abb. 8 zu sehen. Ersterer wurde entwickelt um bekannte Viren zu detektieren, während letzterer noch unbekannte Viren identifizieren soll, die mit bekannten Viren verwandt sind und deshalb eine gewisse genetische Homologie zu jenen besitzen.

Alle Analysen wurden auf einem Standardcomputer vorgenommen (Apple iMAC, 2.4Ghz, 4Gb RAM, Apple, USA). Die verschiedenen Schritte wurden mit selbstgeschriebenen Programmen miteinander verknüpft und weitgehend automatisiert.

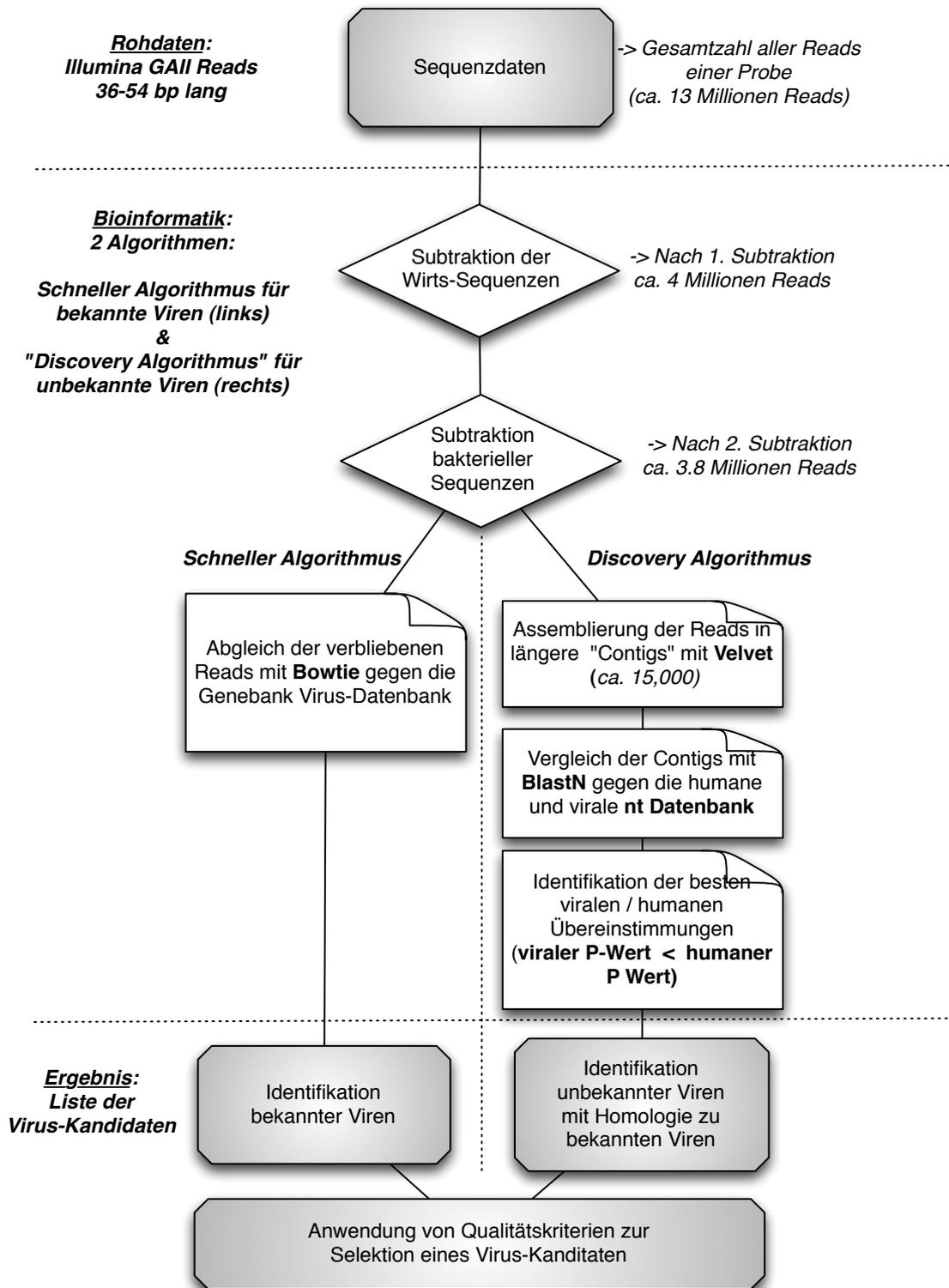


Abb. 8: Flussdiagramm, welches die verschiedenen Schritte des schnellen („Rapid“) und des komplexeren („Discovery“) Zweigs des Virusalgorithmus visualisiert

2.5.1 Subtraktion nicht-viraler Sequenzen

Um die Menge an Sequenzdaten zu vermindern und so die Suche nach viralen Transkripten zu beschleunigen, wurden in einem ersten Schritt möglichst viele Wirtsequenzen herausgefiltert und so virale Sequenzen in den verbleibenden Daten angereichert. Dazu wurde die Idee einer digitalen Subtraktion humaner Transkripte („digital transcript subtraction“, DTS), die in anderen experimentellen Szenarien erfolgreich angewendet worden war (Feng et al., 2007; Weber et al., 2002), aufgegriffen und so umgesetzt, dass sie auf die Datenmengen eines high-throughput RNA-Seq Experiments anwendbar wurde.

In einem ersten Schritt wurden deshalb die gewonnenen Sequenzdaten mit Referenzsequenzen des jeweiligen Wirts, also menschlichen bzw. murinen Genomen, verglichen (für Details zu den verwendeten Datenbanken und Referenzen s. Kapitel 2.5.4). Dies geschah mit Hilfe des open-source Programms „Bowtie“ (Version 0.10) (Langmead et al., 2009), welches ein hocheffizientes Mittel darstellt um Alignments von Reads an Referenzen zu ermitteln -also herauszufinden, von welcher Position innerhalb einer Referenzdatenbank (z.B. des menschlichen Genoms) einer der 36bp oder 54bp langen Reads am wahrscheinlichsten stammt. Verwendete Parameter waren: maximum number of mismatches ≤ 2 und seedlength = 28bp. Dies bedeutet, dass sich bei einem paarweisen Vergleich der 36 Nukleotide eines Reads mit einem 36bp langen Abschnitt im Referenzgenom maximal 2 Unterschiede ergeben dürfen und dass mindestens 28bp des entsprechenden Reads eine hohe Qualität aufweisen müssen. Diese Parameter hatten sich in Tests als optimal heraus gestellt, da sie falsche bzw. ungenaue Zuordnungen minimieren. Alle Reads wurden auf diese Weise mit dem jeweiligen Wirtsgenom verglichen. Wenn sie sich einer Stelle des Genoms zuordnen ließen und dieses Alignment den oben genannten Kriterien genügte, wurden sie von weiteren Analysen ausgeschlossen, da es sich höchstwahrscheinlich nicht um virale Sequenzen handelt.

In einem zweiten Schritt wurde äquivalent vorgegangen, um Sequenzen zu filtern, die am ehesten bakteriellen oder mikrobiellen Ursprungs sind. Dies wurde nötig, da sich durch die hohe Kontamination der Proben durch die orale Bakterienflora auf der einen und der Sequenzähnlichkeit von bakterieller DNA und von Plasmiden mit Viren (Paulsen, 2003) auf der anderen Seite zahlreiche falsch positive Ergebnisse ergaben, wenn später mit weniger stringenten Methoden nach viralen Sequenzen gesucht wurde. Dazu wurden verbliebene Reads mit bakteriellen und

nicht-viralen mikrobiellen Referenzsequenzen (s. Kapitel 2.5.4) verglichen und wiederum solche aussortiert, die eine gute Übereinstimmung zeigten.

Insgesamt wurden also nur solche Reads behalten, die weder einem Wirtsgenom noch einem bakteriellen oder mikrobiellen Genom zugeordnet werden konnten, was für virale Sequenzen der Fall sein sollte.

2.5.2 Schritte des schnellen „Rapid-Algorithmus“

Durch die digitale Subtraktion von wirtsspezifischen und nicht-viralen mikrobiellen Sequenzen aus den RNA-Seq-Daten (s.o.) blieb eine deutlich kleinere Zahl von Reads übrig. In diesen befanden sich nun zwar die potentiellen viralen Sequenzen, aber z.B. auch PCR-Artefakte wie Primer- und Adaptersequenzen oder humane Sequenzen, die –aufgrund normaler genetischer inter-individueller Unterschiede oder wegen Ablesefehler während der Sequenzierung– nicht ausreichend mit den Referenzen übereinstimmten und deshalb nicht aussortiert wurden.

In diesen verbliebenen Sequenzen wurden virale Kandidaten gesucht, indem diese restlichen Reads mit Referenzsequenzen viraler Einträge in den Datenbanken GenBank und RefSeq (s. Kapitel 2.5.4) verglichen wurde. Dies geschah wiederum mit „Bowtie“ (Langmead et al., 2009). Parameter waren die gleichen wie im Subtraktionsschritt, diesmal wurden allerdings diejenigen Reads behalten, die sich erfolgreich einer der Referenzsequenzen zuordnen ließen, also möglicherweise virale Kandidaten sein könnten.

Um einen Überblick über die Ergebnisse zu ermöglichen wurden die Reads, welche eine Übereinstimmung mit einem Virus zeigten, quantifiziert und nutzerfreundlich formatiert. Für jede Probe wurde errechnet, wie viele Reads insgesamt Übereinstimmungen mit Viren zeigten, wie viele Viren dabei insgesamt als möglicher Ursprung in Frage kommen und anschließend eine Liste ausgegeben, welche die wahrscheinlichsten viralen Kandidaten in abnehmender Häufigkeit enthielt. In dieser Liste stand also das Virus, welches die höchste Zahl an Reads auf sich vereinigte an erster Stelle. Diese Übersicht wurde jeweils in einer Textdatei mit der Endung .vir gespeichert, um so einen Vergleich zwischen verschiedenen Proben zu erleichtern. Die möglichen viralen Kandidaten wurden anschließend anhand festgelegter Kriterien bewertet (s. Kapitel 2.5.5

und Tab. 2) um einordnen zu können, ob es sich um falsch-positive oder richtig-positive Ergebnisse handelt.

2.5.3 Schritte des spezifischeren „Discovery-Algorithmus“

Um ein Virus zu detektieren, welches noch unbekannt ist und dessen Genom deshalb also nicht in der Referenzdatenbank enthalten sein kann, das aber mit einem bereits bekannten Virus verwandt ist, braucht man eine Vergleichsmethode für Sequenzen, die weniger stringent ist als der schnelle Algorithmus. „Bowtie“ erlaubt maximal 2 Unterschiede im Vergleich zweier Sequenzen was bedeutet, dass ein Read von 36bp Länge in 34bp (= 94,5%) mit einem Abschnitt einer Virusreferenz übereinstimmen muss um als von dieser abstammend eingeordnet zu werden. Verschiedene HPV-Spezies haben allerdings nur eine Sequenzhomologie von 60-70% (de Villiers et al., 2004). Andererseits erhöht eine größere Fehlertoleranz in solch kurzen Sequenzen die Wahrscheinlichkeit für falsche Zuordnungen. Gleichzeitig kommt es zu einem enormen Anstieg der Laufzeit, weil sich mit jedem erlaubten Unterschied mehr Möglichkeiten ergeben, von welcher Stelle innerhalb der Referenzsequenzen ein Read abstammen könnte. Diese müssen jeweils gespeichert und miteinander verglichen werden. Die Anzahl verbliebener Reads nach der Subtraktion der Wirtssequenzen war immer noch zu hoch um sie in zumutbarer Zeit mit einem weniger strengen, aber deutlich sensitiveren Alignmentprogramm wie BLAST (Altschul et al., 1990) verarbeiten zu können.

Um die Anzahl an Reads nach der Subtraktion der Wirtssequenzen weiter zu vermindern, wurden diese mit Hilfe des frei erhältlichen Programms „Velvet“ (Version 0.7, <http://www.ebi.ac.uk/~zerbino/velvet/>) (Zerbino und Birney, 2008) zu längeren Sequenzen zusammengesetzt („assembled“, engl. für zusammengebaut). Sehr vereinfacht gesagt ermittelt das Programm Überlappungen zwischen verschiedenen Reads und fügt sie daraufhin zusammen. Dadurch entstehen längere Abschnitte, sog. Contigs (von „contiguous Reads“, engl. für zusammenhängend), welche wiederum mit anderen Reads oder Contigs überlappen können und so weiter verlängert werden. Die tatsächliche Berechnung ist selbstverständlich komplexer und folgt einem Wahrscheinlichkeitsbaum, da häufig mehrere Reads ein Contig verlängern könnten, sich aber z.B. die Anzahl der überlappenden Nukleotide unterscheidet. Ein Maß dafür, ob verschiedene Reads zurecht zu einem Contig zusammengesetzt wurden, ergibt sich aus der Anzahl an Reads,

welche durchschnittlich die Schnittstellen überspannen (die sog. Coverage). Je höher also die Coverage, desto mehr Reads unterstützen diesen Contig.

Abhängig von der Längenverteilung der Contigs der verschiedenen Proben wurden Contigs mit einer Länge von $\geq 1,5x$ der Länge der Input-Reads (also 54bp bzw. 81pb) und solche mit einer Coverage ≥ 4 für die weiteren Analysen ausgewählt.

Die Contigs konnten anschließend mit dem Programm BlastN (Altschul et al., 1990) untersucht werden, welches eine Nukleotidsequenz gegen eine Nukleotidsequenzdatenbank vergleicht und durch paarweise Vergleiche die ähnlichste DNA-Sequenz ermittelt. Zu diesem Zweck wurde eine Stand-Alone-Version (BLAST Version 2.2.10, <http://www.ncbi.nlm.nih.gov>), also eine lokal installierte Variante, eingerichtet, da diese erstens internetunabhängig ist und zweitens eine automatisierten Verarbeitung im Rahmen der Analyseschritte ermöglicht. BlastN vergleicht die Nukleotide zweier Sequenzen paarweise miteinander („pairwise alignment“). Um Alignments bewerten und vergleichen zu können, wird jeweils ein Score errechnet, welcher für jedes übereinstimmende Nukleotid erhöht und für Fehler und Lücken („gaps“) vermindert wird. Je höher also der Score eines Alignments, desto besser stimmen die Sequenzen überein. Die benutzten BlastN-Parameter wurden so gewählt, dass Übereinstimmungen bis ca. 60% Sequenzhomologie gemeldet wurden. Sie waren: $r = 5$ (wird addiert, wenn Nukleotide übereinstimmen), $q = -4$ („mismatch penalty“, wird bei jedem Unterschied abgezogen), $W=11$ („word size“, Mindestlänge einer teilweisen Übereinstimmung, damit diese weiter verfolgt wird), $g = \text{True}$ („gapped alignments“ möglich), $G = 5$ (Strafe, die abgezogen wird, falls ein „Gap“ eröffnet werden muss), $E = -1$ (Strafe die abgezogen wird, falls ein „Gap“ erweitert werden muss), $e = 0.5$ („expectation value E“; akzeptierte Höchstwahrscheinlichkeit, dass in der Datenbank zufällig eine DNA-Sequenz vorkommt, die einen ähnlichen Score erreicht; Bei Überschreitung werden Alignments nicht gemeldet).

Als Referenzdatenbank wurden menschliche/murine und virale Einträge der „BLAST nt database“ (für Details s. Kapitel 2.5.4) verwandt. Dies war notwendig, da in den Contigs sowohl menschliche als auch virale Sequenzen vorhanden waren. Durch die erhöhte Toleranz von Abweichungen beim BLAST-Vergleich eines Contigs mit der Referenz wären bei der Verwendung einer rein viralen Datenbank aufgrund (partieller) Sequenzhomologie zahlreiche Contigs menschlichen Ursprungs als (falsch-positive) virale Treffer gemeldet worden. Indem jedes Contig aber

sowohl mit menschlichen als auch mit viralen Sequenzen verglichen wurde, ließ sich anhand der oben erklärten BLAST-Scores leicht abschätzen, ob ein menschlicher oder ein viraler Ursprung wahrscheinlicher war: Der BLAST-Score des Alignments eines Contigs an die Virusreferenz wäre zwar vielleicht signifikant, aber geringer als der beste BLAST-Score des Alignments des gleichen Contigs an eine humane Referenz – es handelte sich demnach am ehesten um ein Contig humanen Ursprungs und also ein falsch positives Ergebnis.

Um die BLAST-Verarbeitung der Contigs auf einem Standardcomputer zu ermöglichen ohne den Arbeitsspeicher zu überlasten wurden sie in Chargen von je 120 Contigs aufgeteilt und sequentiell abgearbeitet. Die Ergebnisse aller Contig-Chargen einer Probe wurden mit Perl-Skripten aufgearbeitet und in unterschiedlicher Detailtiefe dargestellt, um verschiedenen Fragen beantworten zu können. Die komplette Liste aller erfolgreicher Alignments von Contigs an humane und/oder virale Referenzsequenzen wurde in eine Textdatei mit der Endung „virlist.ext“ exportiert. Um eine etwas kürzere Übersicht zu ermöglichen wurden in eine zweite Datei mit der Endung „virlist“ für jedes Contig nur das jeweils beste Alignment und –falls vorhanden– die besten zwei viralen Alignments exportiert, was eine deutliche Längenreduktion der Liste ergab. In diesen beiden Dateien wurden Alignments mit bekannten tumorassoziierten Viren (HPV, Herpesviren, Polyomaviren) zur besseren Übersicht mit Sternchen (*****) markiert. Wenn das insgesamt beste Alignment eines Contigs (abhängig vom BLAST-Score und dem damit verbundenen E-Wert) mit einem Virus gefunden wurde, wurde dies mit einem Pfeil (→) markiert. Für eine schnelle Übersicht wurden zusätzlich in eine dritte Datei alle Treffer und Contigs exportiert, deren bestes Alignment an ein Virus war und deren E-Wert (die Wahrscheinlichkeit, dass eine Sequenz mit einem ähnlichen Alignment-Score zufällig in der Datenbank vorhanden ist) gleichzeitig kleiner als 10^{-25} war. Diese möglichen viralen Kandidaten wurden weiterhin nach den Bewertungskriterien für ein positives Ergebnis (s. Kapitel 2.5.5 und Tab. 2) beurteilt.

2.5.4 Datenbanken mit Referenzsequenzen

Um die aus den Proben gewonnenen Sequenzdaten einem Ursprungsorganismus zuzuordnen sind verschiedene Referenzsequenzen notwendig, mit denen sie verglichen werden können. Für die humanen Sequenzen wurde die Version Nr.18 des humanen Genoms der University of California Santa Cruz (UCSC hg18, NCBI Build 36.1, veröffentlicht März 2006) (Fujita et al., 2011; Lander et al., 2001) und für die Maus die 9. Version des *Mus musculus*-Genoms (UCSC mm9, NCBI

Build 37, veröffentlicht Juli 2007) (Karolchik et al., 2007) verwendet. Die bakteriellen und nicht-viralen mikrobiellen Genome wurden mit eigenen Pearl-Skripten aus entsprechenden Einträgen extrahiert, die sich in den GenBank-Verzeichnissen fanden (<ftp://ftp.ncbi.nih.gov/genbank/>, Stand Juli 2009), während eine virale Datenbank aus den Sequenzen viraler GenBank- und den viralen RefSeq-Einträgen (<ftp://ftp.ncbi.nih.gov/refseq/release/viral/>, Stand Juli 2009) gebildet wurde. Sowohl GenBank als auch RefSeq sind online verfügbare Speicherorte für Sequenzdaten, wobei GenBank Daten zu >380.000 Organismen enthält und alle veröffentlichten Sequenzen übernimmt, die von Autoren eingestellt werden (Benson et al., 2010), während diese in RefSeq überprüft und gegebenenfalls korrigiert bzw. aussortiert werden, um zum Beispiel Duplikate zu vermeiden. So enthält RefSeq weniger Informationen (für ca. 5300 Organismen), die aber eine höhere Qualität aufweisen (Pruitt et al., 2009).

Damit Referenzsequenzen von Bowtie effektiv verwendet werden können, müssen sie als sogenannter Index vorliegen, also eine bestimmte Formatierung und Struktur aufweisen. Die entsprechenden Indizes für *Homo sapiens* und *Mus musculus* konnten direkt von der Bowtie website (<http://bowtie-bio.sourceforge.net/index.shtml>) geladen und installiert werden, für die selbst erstellten Datenbanken mussten entsprechende Indizes erst erzeugt werden.

Als Grundlage für die Referenzen der spezifischeren BLAST-Suche im Discovery Algorithmus wurde die umfangreiche NCBI Blast nucleotide (nt) database (<http://www.ncbi.nlm.nih.gov/staff/tao/URLAPI/blastdb>, Stand Juli 2009) genommen, welche Nukleotidsequenzen verschiedener anderer Datenbanken (GenBank, EMBL, DDBJ) vereint (Tao et al., 2008). Um die Laufzeit zu verbessern wurden die sehr großen Textdateien (>10 Gigabyte) mit eigenen Programmen durchforstet, jeweils nur die viralen und die humanen bzw. die Mausequenzen extrahiert und daraus ein entsprechender Index gebildet. Versuchsweise wurden zunächst testweise alle Organismen in der Datenbank belassen, aber da –erwartungsgemäß– nur eine vernachlässigbare kleine Zahl von Contigs positive Alignments an diese Organismen zeigte, sie aber die Datenbankgröße vervielfachten, wurden sie entfernt.

2.5.5 Bewertungskriterien für die Ergebnisse der Algorithmen

Ein zentraler Punkt in der Entwicklung einer solchen Methodik ist die Frage, wie Ergebnisse zu bewerten sind und welche als relevant eingestuft werden. Aus diesem Grund wurden Kriterien

entwickelt, mit denen die Resultate des Algorithmus ausgewertet und eingeordnet werden sollten. Als Vergleichsmaßstab dienten dabei die Ergebnisse der Durchläufe der fünf positiven Kontrollen – jeweils UM-SCC47, das Mauslymphom und das Zervixkarzinom, sowie die Simulationen eines unbekanntes Virus mit den Daten von UM-SCC47 und denen des Mauslymphoms. Eine Übersicht über die Kriterien für ein positives Ergebnis finden sich in Tab. 2.

Tab. 2: Bewertungskriterien für Viruskandidaten; Beide Kriterien müssen jeweils erfüllt sein (BlastN E-Wert: Wahrscheinlichkeit, dass eine Sequenz mit einem ähnlichen Alignment-Score zufällig in der Datenbank vorhanden ist).

Definition eines pos. Virusergebnisses	1. Kriterium	2. Kriterium	Begründung
Rapid Algorithmus	> 5 virale Reads pro 100.000 Reads in den Rohdaten	>85% der viralen Reads passen zu einem einzelnen Virus	Die Schwellenwerte wurden so gewählt, dass virale Transkripte mit relevanter Expression detektiert, falsch positive Ergebnisse aber vermieden werden.
Discovery Algorithmus	Beste Übereinstimmung mit einem viralen Gen und nicht mit einem Gen des Wirts (Mensch bzw. Maus)	BLASTN E-Wert $<10^{-25}$	Um verbliebene humane Sequenzen zu filtern die auch eine Virushomologie aufweisen, werden virale und humane Sequenzen gemeinsam analysiert.

Die Schwellenwerte und Bedingungen um eine Übereinstimmung von Reads mit einem Virusgenom als relevantes Zeichen des Vorliegens viraler Transkription zu werten, wurden empirisch ermittelt und so festgesetzt, dass die Rate an falsch positiven Ergebnissen minimiert wird, während gleichzeitig die Zahl der falsch negativen Resultate möglichst konstant bleibt. Als Grundlage dienten dabei die Ergebnisse der Positivkontrollen, es müssen jeweils beide Bedingungen erfüllt sein.

Im Discovery-Algorithmus stellte sich als erstes sinnvolles Kriterium ein Mindestanteil von ≥ 5 viralen Reads pro 100.000 Reads innerhalb der Rohdaten heraus. Dies entspricht einem Anteil von 0,005% viraler Reads an der Gesamtheit aller Reads einer Probe und bedeutet somit auch, dass bis zu 99,995% aller Reads vom Wirtsorganismus stammen können und trotzdem ein Viruskandidat angezeigt würde. Ist die erste Bedingung erfüllt, dient die Verteilung der viralen Treffer

als zweites Kriterium. Dabei müssen $\geq 85\%$ dieser viralen Reads von einem einzelnen Virus stammen. Sie dürfen also nicht auf mehrere verschiedene Viren verteilt sein, da dies einer Mehrfachinfektion entspräche, welche biologisch wenig plausibel erscheint und in der Regel durch virale Homologe im menschlichen Genom verursacht ist.

Im Discovery-Algorithmus ist die erste Voraussetzung, um ein relevantes Virusergebnis zu erkennen, dass das entsprechende Contig eine bessere Übereinstimmung mit einem Virusgenom als mit einem Gen des jeweiligen Wirts zeigt. Auf diese Weise wird verhindert, dass Gene, welche Homologe in eukaryoten Spezies haben (z.B. v-src und c-src (Stehelin et al., 1976)) als signifikant erscheinen. Gleichzeitig muss der dazugehörige von BlastN errechnete E-Wert, welcher in seiner Bedeutung etwa einem p-Wert entspricht, $< 10^{-25}$ sein, also die Chance, dass sich in der Referenzdatenbank rein zufällig eine Sequenz befindet, welche ähnlich gut oder besser übereinstimmt, minimal sein. Um die Sensitivität zu überprüfen und die falsch negative Rate zu minimieren wurden auch Durchläufe mit weniger strengen E-Werten (z.B. $E = 10^{-20}$) durchgeführt, wobei sich in den Positivkontrollen und Proben lediglich ein Anstieg der falsch positiven Ergebnisse ohne Zugewinn an richtig positiven Resultaten zeigte.

Alle Ergebnisse wurden zusätzlich manuell gesichtet, Kandidaten-Contigs einem web-basierten BLAST-Vergleich unterzogen und die Kriterien so auf ihre Sensitivität und Spezifität überprüft.

2.5.6 In-silico Simulation eines unbekanntes Virus

Ein wichtiger Anspruch an den Algorithmus war es, neue unbekanntes Viren, die mit bekannten Viren verwandt sind, erfolgreich detektieren zu können.

Ein Test dieser Fähigkeit ist in-vivo/in-vitro naturgemäß schwierig zu bewerkstelligen, so dass diese Situation stattdessen „in-silico“ simuliert wurde. Dazu wurden aus den jeweiligen Referenzdatenbanken die Sequenzen für HPV-16 und EBV gelöscht und anschließend neue Indizes erstellt. Mit diesen veränderten Referenzindizes wurde daraufhin der Algorithmus für die Sequenzen und Contigs von UM-SCC47 und des transgenen Mauslymphoms wiederholt und die Resultate nach den Kriterien für ein positives Virusergebnis (s. Tab. 2) ausgewertet.

2.6 CDKN2A/p16-Genexpressions-Analyse

Um die Expression des Tumorsuppressorgens CDKN2A/p16 bestimmen zu können, wurden die entsprechenden mRNA-Transkripte in den RNA-Seq Daten quantifiziert. Dazu wurde das Programm „Cufflinks“ (Roberts et al., 2011; Trapnell et al., 2010) verwendet, welches die isoformspezifische Expression von Genen aus RNA-Seq-Daten berechnen kann. Als Maß für die Expression wurde das Konzept der RPKM (Reads per kilobases of exon model per million mapped reads) verwandt (Mortazavi et al., 2008). Dazu werden alle Reads, welche ein Alignment an die Referenzsequenzen eines Gens (z.B. des humanen CDKN2A-Gens) zeigen, gezählt. Voraussetzung ist dabei, dass diese Reads eindeutig sind, also nicht auch zu einer anderen Stelle des Genoms passen und ≤ 2 Abweichungen bzw. Fehler (mismatches) in der Nukleotidabfolge aufweisen.

Ein mRNA-Transkript liefert während des Fragmentierungsschritts in der Probenvorbereitung umso mehr 300bp lange Bruchstücke und damit später auch mehr Reads, je länger es ist. Dementsprechend sind lange Transkripte leichter zu detektieren als kurze und eine Normalisierung für die Länge des mRNA-Transkripts notwendig. Deshalb werden gezählte Reads durch die Anzahl der Nukleotide des verwendeten Referenztranskripts (als Kilobasen, kb) geteilt.

Da auch eine höhere absolute Zahl von Reads zu erwarten ist, je höher die Gesamtzahl aller Reads des Experiments ist, wird in einem zweiten Schritt auch für die Menge der Rohdaten normalisiert. Dazu werden die genspezifischen und für die mRNA-Länge normalisierten Read-Werte durch die Gesamtzahl aller Reads, welche erfolgreich dem Referenzgenom zugeordnet werden konnten, geteilt. Dieser Prozess wurde für alle Gene inklusive CDKN2A/p16 durchgeführt.

Die unterschiedliche Expression von CDKN2A/p16 zwischen Proben mit hoher und solcher mit niedriger Expression wurde mit einem 2-seitigen Student t-test anhand der RPKM-Werte auf Signifikanz überprüft.

3. Ergebnisse

3.1 Auswertung der klinischen Daten

Insgesamt wurden klinische Daten von 748 Patienten mit Kopf-Hals-Tumoren analysiert, die mit kombinierter Radiochemotherapie behandelt worden waren. Der mediane Nachbeobachtungszeitraum lag bei 41,4 Monaten. Für die Analyse wurde ein besonderes Augenmerk auf den Einfluss sowohl des Alters als auch der Tumorlokalisierung auf das progressionsfreie Intervall und das Gesamtüberleben gelegt. Dies war wichtig, da für junge Patienten (≤ 45 Lebensjahre) mit Tumoren der Mundhöhle aus der klinischen Erfahrung ein schlechteres klinisches Outcome vermutet wurde, welches nun systematisch untersucht werden sollte. Eine Übersicht zu den populationsbezogenen Daten und Informationen zur Exposition mit den bekannten Risikofaktoren Tabak- und Alkoholkonsum der Patientenkohorte findet sich in Tab. 3.

Tab. 3: Übersicht über das Patientenkollektiv

	Junge Patienten (≤ 45 Lebensjahre)		Ältere Patienten (> 45 Lebensjahre)		Gesamt
	Mundhöhlen- Ca	andere Ca- Lokalisation	Mundhöhlen- Ca	andere Ca- Lokalisation	
n =	24	111	83	530	748
mittleres Alter	38,9 Jahre	38,4 Jahre	61,3 Jahre	59,1 Jahre	55,7 Jahre
Tabakkonsum [%]					
„Nie-Raucher“	37,50	17,11	18,07	10,00	14,30
< 20 py	20,83	26,13	14,46	16,41	18,58
20 – 40 py	33,3	33,30	33,73	30,38	32,62
> 40 py	4,17	9,01	26,51	30,38	27,00
Pfeife/Zigarre	4,17	5,41	6,02	7,36	7,22
Alkoholkonsum [%]					
niemals	20,83	15,32	16,87	13,40	15,91
gelegentlich	33,3	22,52	24,10	18,87	21,91
moderat	12,50	15,32	18,10	22,26	21,39
stark	33,3	36,04	39,76	37,55	38,37
TNM-Stadium [%]					

T1	12,50	12,61	3,61	10,94	10,43
T2	20,83	19,82	16,87	18,49	18,58
T3	20,83	23,42	13,25	23,96	22,60
T4	45,83	35,14	65,06	40,94	43,04
N0	25,00	10,81	22,89	20,19	19,25
N1	29,17	17,12	22,89	17,55	19,12
N2	37,50	55,85	49,40	50,19	54,41
N3	8,33	7,21	3,61	6,98	7,22
M0	100,00	98,02	100,00	98,21	98,53
M1	0,00	0,00	0,00	0,80	0,53
Mx	0,00	1,98	0,00	0,99	0,94
UICC-Stadium [%]					
I	0	0	0	0	0
II	8,33	2,97	2,47	2,98	2,95
III	25,00	16,83	11,10	11,73	12,60
IV	66,67	80,20	86,42	85,29	84,45

Aus der Zusammenstellung in Tab. 3 werden einige wichtige Fakten ersichtlich. So fällt auf, dass der Anteil an jüngeren Patienten insgesamt kleiner ist als der der älteren Patienten und dass in beiden Gruppen die Patienten mit Mundhöhlenkarzinomen weniger häufig sind als solche mit Tumoren an anderen Lokalisationen des Hals- und Kopfbereichs. In der Kombination ergibt sich daraus, dass es sich bei Karzinomen der Mundhöhle bei Patienten ≤ 45 Lebensjahre immer noch um eine eher seltene Erkrankung handelt, obwohl in den letzten Jahrzehnten eine steigende Inzidenz und Prävalenz dieser Tumorentität nachweisbar war (Annertz et al., 2002; Chitapanarux et al., 2006). Es lässt sich ferner erkennen, dass das Durchschnittsalter der Patienten mit Mundhöhlenkarzinomen und jenen mit Tumoren anderer Lokalisation jeweils innerhalb der beiden Altersgruppen ähnlich ist, sich aber die verschiedenen Tumorlokalisationen in ihrer Exposition gegenüber bekannten Risikofaktoren zu unterscheiden scheinen. Besonders für die jungen Patienten mit Mundhöhlenkarzinom lässt sich dabei ein Trend zu einer insgesamt geringeren Exposition zeigen, welche auch in anderen Studien berichtet wurde (X-H Liang et al., 2008; Llewellyn et al.,

2004). Der Konsum von Tabak ist signifikant geringer in der Population junger Patienten mit Mundhöhlenkarzinom im Vergleich zum gesamten Rest der Kohorte ($p = 0,0017$; exakter Fisher-Test). Innerhalb der jüngeren Patienten ist ebenfalls eine geringere Exposition gegenüber Tabak bei den Patienten mit Mundhöhlenkarzinomen zu verzeichnen, wobei dieser Effekt hier – wahrscheinlich auf Grund deutlich kleinerer Fallzahlen – nicht statistisch signifikant ist. Der Anteil an Patienten welche niemals geraucht haben („Nie-Raucher“) beträgt beispielsweise 37,5% für die jungen Patienten mit Mundhöhlenkarzinom, während nur 17,11% der jungen Patienten mit Tumoren an anderen Lokalisationen des Hals- und Kopfbereichs angaben, noch niemals geraucht zu haben ($p=0,062$; 2x2 exakter Fisher-Test für „Nie-Raucher“ vs. Raucher innerhalb der jungen Patienten). Ein ähnlicher, aber ebenfalls nicht signifikanter Trend ergibt sich beim Vergleich des Alkoholkonsums. Dort zeigt sich, dass 54,1% der jungen Patienten mit Mundhöhlenkarzinom angaben, gar keinen oder wenig Alkohol zu konsumieren, während es bei den jungen Patienten mit anderen Tumoren des Hals- und Kopfbereichs nur 37,8% sind.

Insgesamt ist für junge Patienten mit Mundhöhlenkarzinomen also eine vergleichbare oder sogar etwas geringere Exposition gegenüber den klassischen Risikofaktoren Tabak und Alkohol im Vergleich zu den anderen jungen bzw. allen HNO-Tumorpatienten dieser Kohorte feststellbar.

Die UICC-Tumorstadien sind vergleichbar sowohl zwischen jungen und älteren Patienten (Exakter Fisher-Test $p=0,09$), als auch zwischen jungen Patienten mit und solchen ohne Mundhöhlenkarzinom (Exakter Fisher-Test $p=0,18$). Insgesamt zeigt sich ein hoher Anteil fortgeschrittener Tumorstadien. Außerdem ist eine nicht signifikante Tendenz hin zu niedrigeren Tumorstadien für jeweils jüngere Patienten bzw. für jüngere Patienten mit Mundhöhlenkarzinom erkennbar (vgl. Abb. 9 und Abb. 10).

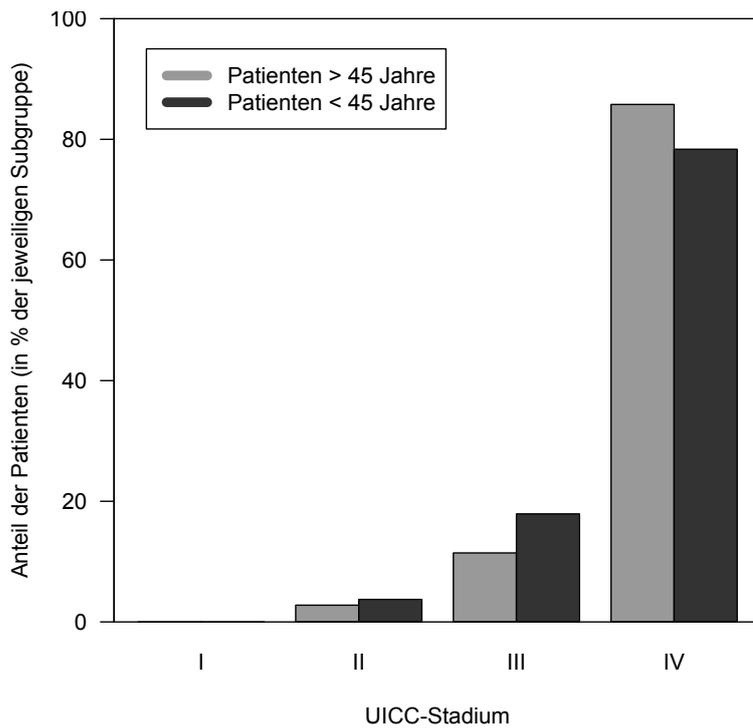


Abb. 9: Verteilung der Tumorstadien. Ein signifikanter Unterschied der Tumorstadien zwischen den Altersgruppen ist nicht gegeben (Exakter Fisher-Test $p=0,09$).

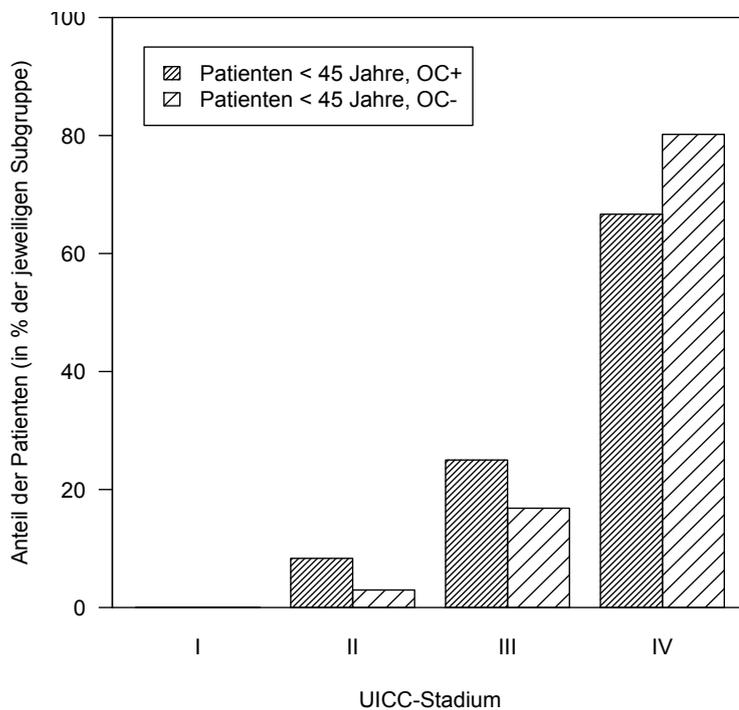


Abb. 10: Verteilung der Tumorstadien junger Patienten. Kein signifikanter Unterschied der Tumorstadien zw. Patienten mit Mundhöhlenkarzinomen und solchen mit anderen Tumorlokalisationen (Exakter Fisher-Test $p=0,018$; OC+ = Pat. mit Mundhöhlenkarzinom, OC- = Pat. mit Tumor außerhalb der Mundhöhle)

Bei der Untersuchung des Gesamtüberlebens zeigte sich für junge Patienten mit Mundhöhlenkarzinomen eine signifikant geringere 5-Jahres-Überlebensrate als für junge Patienten mit anderen HNO-Tumoren (s. Abb. 11): In der Gruppe der jungen Patienten (Alter ≤ 45 Jahre, $n=135$), hatten Patienten mit Primärtumoren im Bereich der Mundhöhle ($n= 24$) eine 5-Jahres-Überlebensrate von 42%, während diese für junge Patienten mit Tumoren anderer Lokalisation ($n=111$) bei 65% lag [Log-Rang Test $p=0,008$; univariate Cox-Regressionsanalyse HR=3,25 (95% CI: 1.50–7.05.)]. Auch das progressionsfreie Überleben unterschied sich signifikant zwischen diesen Gruppen: 35,7% der jungen Patienten mit Mundhöhlenkarzinomen waren progressionsfrei nach fünf Jahren, während dies für 64% der Patienten mit anderen Tumoren der Fall war [Log-Rang Test $p=0,002$; univariate Cox-Regressionsanalyse HR=2,96 (95% CI: 1.42-6.17)].

Für Patienten mit einem Alter > 45 Lebensjahre (s. Abb. 12) ist ein solcher Unterschied nicht feststellbar [Log-Rang test $p=0,66$, univariate Cox-Regressionsanalyse HR=1,08 (95% CI 0.76–1.55)].

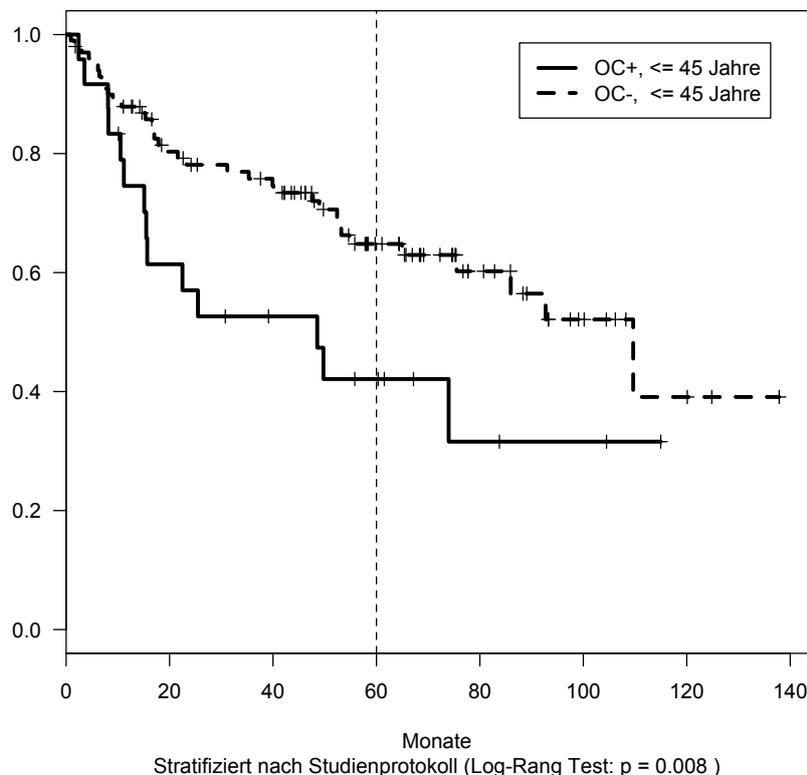


Abb. 11: Gesamtüberleben für Patienten < 45 Lebensjahre. Die 5-Jahres-Überlebensrate für junge Patienten mit Mundhöhlenkarzinomen ist signifikant schlechter als die 5-Jahres-Überlebensrate junger Patienten mit anderen Tumorlokalisationen (Log-Rang Test $p=0,008$; OC+ = Patienten mit Mundhöhlenkarzinom (OC = Oral Cavity), OC- = Pat. mit HNO-

Tumoren anderer Lokalisation)

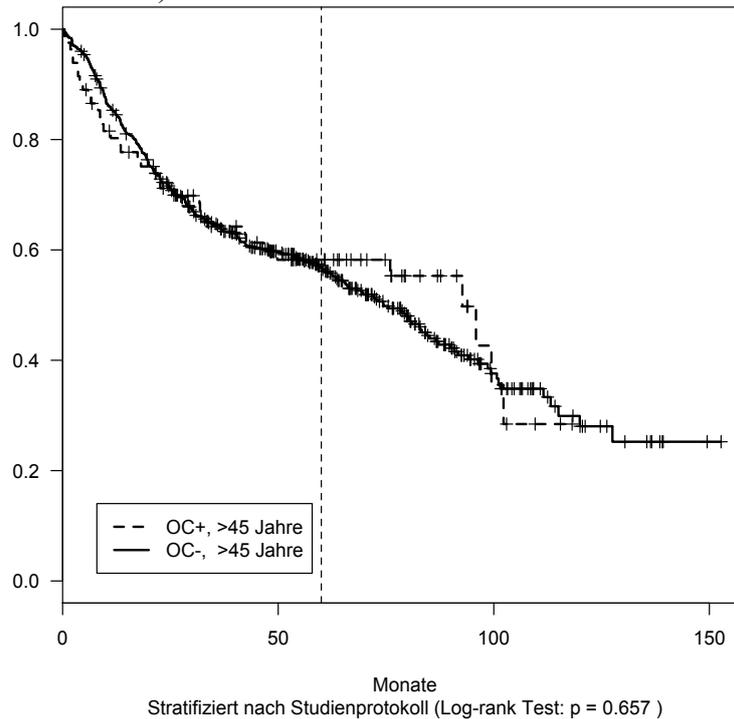


Abb. 12: Gesamtüberleben für Patienten > 45 Lebensjahre. Es zeigten sich keine signifikanten Unterschiede der 5-Jahres-Überlebensrate nach der Tumorlokalisation (Log-Rang Test $p=0,66$; OC+ = Patienten mit Mundhöhlenkarzinom; OC- = Pat. mit HNO-Tumoren anderer Lokalisation)

Es ist bekannt, dass HPV-positive Hals- und Kopftumore, welche sich gehäuft im Bereich des Oropharynx, der Tonsillen und des Zungengrundes finden, besonders bei jungen Patienten eine bessere Prognose zeigen als HPV-negative Tumore (Ang et al., 2010; Chaturvedi et al., 2011). Eine zufällige hohe Überrepräsentation HPV-positiver Tumore in der Gruppe der Karzinome außerhalb der Mundhöhle könnte also den beobachteten Effekt im Sinne eines Confounders hervorrufen. Da der HPV-Status für die Patienten der untersuchten Kohorte nicht bekannt ist, kann eine direkte Aussage über den Anteil an HPV-positiven Tumoren nicht getroffen werden. Der beschriebene Effekt bleibt allerdings auch bestehen, wenn das Überleben für junge Patienten mit Mundhöhlenkarzinomen, für junge Patienten mit Oropharynx Tumoren (inkl. Tonsillen- und Zungengrund) und für jene mit Tumoren, die weder zur Mundhöhle noch zum Oropharynx gehören, getrennt betrachtet wird (s. Abb. 13).

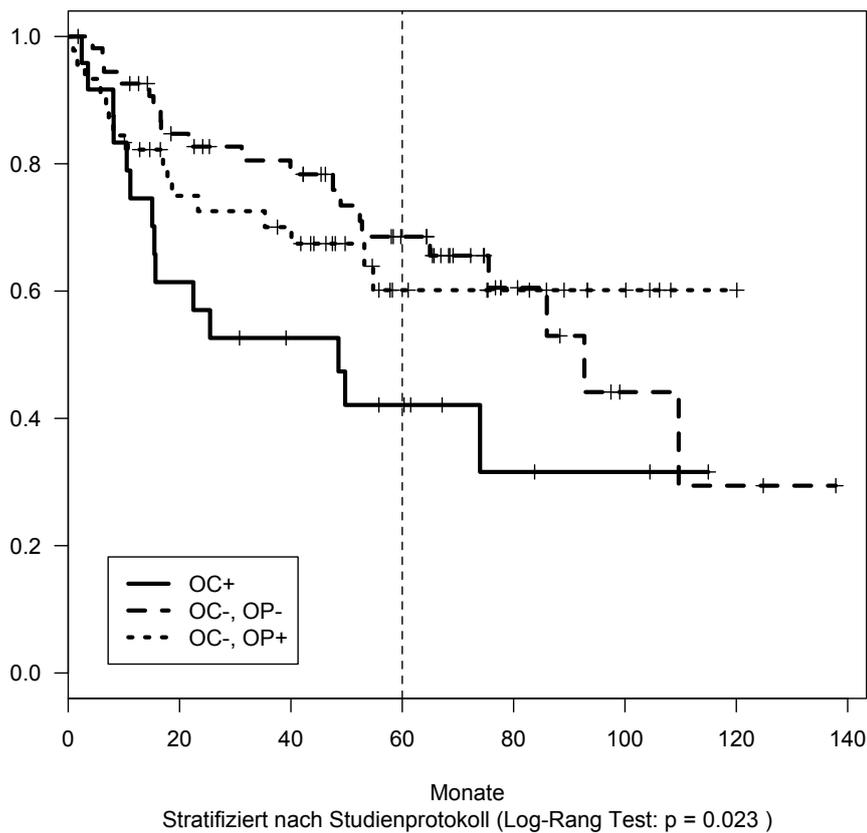


Abb. 13: Gesamtüberleben junger Patienten nach Tumorlokalisation. Insgesamt zeigt sich eine signifikant verringerte 5-Jahres-Überlebensrate für junge Patienten mit Mundhöhlenkarzinomen (Log-Rang Test für 3 Gruppen $p=0,023$; OC+ = Pat. mit Mundhöhlenkarzinom; OC-, OP- = Pat. mit einem Tumor weder in der Mundhöhle noch im Bereich des Oropharynx; OC-, OP+ = Pat. mit Oropharynxtumor, inkl. Zungengrund und Tonsillen)

Es zeigt sich auch hier ein signifikant verringertes Überleben für junge Patienten mit Mundhöhlenkarzinomen im Vergleich zu solchen mit anderen Tumorlokalisationen (Log-Rang Test für 3 Gruppen $p=0,023$). Ihre 5-Jahres-Überlebensrate liegt bei 42%, während sie für Oropharynxtumore 60% und für Tumore außerhalb von Mundhöhle und Oropharynx 68% beträgt. Der Unterschied zwischen Patienten mit Tumoren des Oropharynx (die die höchste Rate an HPV-positiven Tumoren enthalten sollten) und Patienten mit Tumoren an Lokalisationen außerhalb der Mundhöhle und des Oropharynx ist nicht signifikant. Die in Abb. 11 beobachtete Verminderung des Gesamtüberlebens junger Patienten mit Mundhöhlenkarzinomen wird also aller Wahrscheinlichkeit nach nicht durch HPV-positive Oropharynxtumore mit besonders guter Prognose hervorgerufen.

Innerhalb der Kohorte junger Patienten mit Mundhöhlenkarzinomen haben 50% ein Karzinom der mobilen/oralen Zunge, eine Übersicht zu den anatomischen Primärtumorlokalisationen findet sich in Abb. 14.

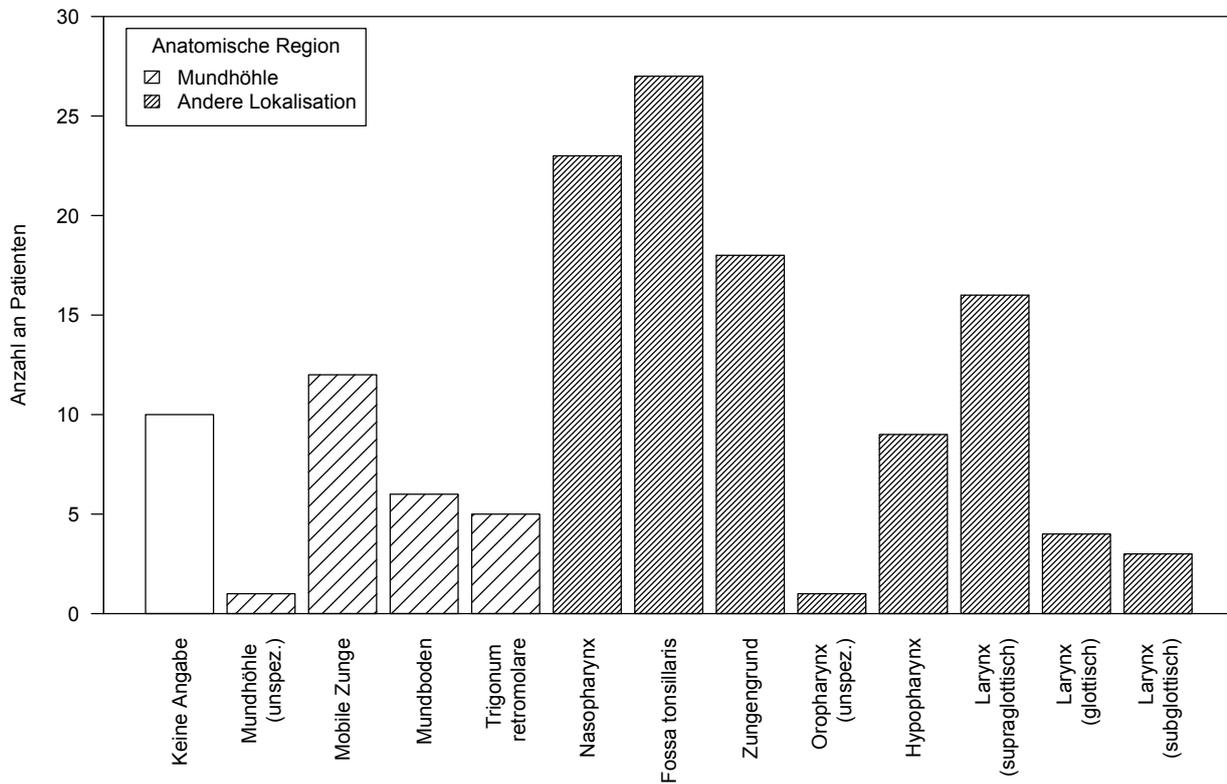


Abb. 14: Verteilung der Primärtumorlokalisationen junger Patienten mit Kopf-Hals-Tumoren der untersuchten Kohorte.

Innerhalb der Gruppe junger Patienten mit Karzinomen der Mundhöhle sind keine Unterschiede des Gesamtüberlebens in Abhängigkeit von der Tumorlokalisation vorhanden. Insbesondere nicht zwischen Patienten mit Tumoren der mobilen Zunge und solchen mit anderen Primärlokalisationen innerhalb der Mundhöhle (Log-Rang Test $p=0,70$; s. Abb. 15). Dies legt nahe, dass die verschiedenen Entitäten innerhalb der Mundhöhlenkarzinome insgesamt eine ähnliche Prognose aufweisen und in den vorangegangenen Analysen gerechtfertigt als eine einzelne Gruppe zusammengefasst wurden. Dies ist in Übereinstimmung mit anderen Studien, die keine Unterschiede des Überlebens zwischen verschiedenen Subgruppen innerhalb der Mundhöhle zeigten (RB Bell et al., 2007).

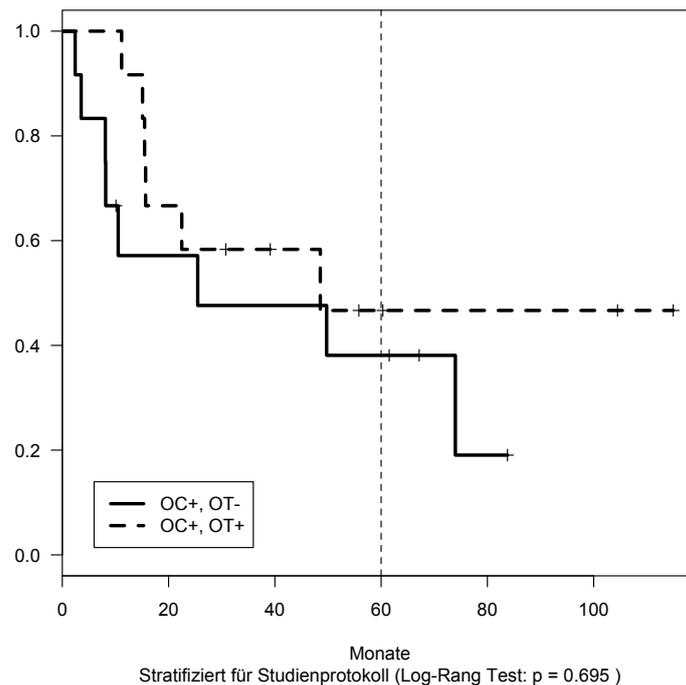


Abb. 15: Gesamtüberleben junger Patienten mit Mundhöhlenkarzinom. Zwischen Tumoren der mobilen Zungen und Tumoren anderer Lokalisation innerhalb der Mundhöhle zeigen sich keine signifikante Unterschiede der 5-Jahres-Überlebensrate (Log-Rang Test $p=0,695$; OC+, OT+ = Pat. mit Tumor der oralen Zunge (oral tongue); OC+, OT- Pat. mit anderer Lokalisation innerhalb der Mundhöhle)

Um zu testen, ob junge Patienten mit Mundhöhlenkarzinomen auch im Vergleich zur gesamten Patientenkohorte eine schlechtere Prognose haben und gleichzeitig den Einfluss verschiedener konfundierender Risikofaktoren und Variablen kontrollieren zu können, wurde ein multivariates Cox-Proportional-Hazards-Modell angepasst. In diesem wurde für die Kovariate Alter (≤ 45 Jahre vs. > 45 Jahre), Tumorlokalisierung (Mundhöhle (OC+) vs. andere Lokalisation (OC-)), Tabakkonsum, Alkoholkonsum und Tumorstadium adjustiert. Zusätzlich wurde eine Interaktion zwischen den Variablen Alter und Tumorlokalisierung eingeführt. Es zeigt sich, dass neben Tabak und Alkohol auch diese Interaktion signifikanten Einfluss auf das Überleben der Patienten hat ($p=0,015$, HR=2,47). Junge Patienten mit Mundhöhlenkarzinomen haben demnach ein 2,47-fach höheres Sterberisiko im Vergleich zur Gesamtkohorte, während diese beiden Faktoren einzeln betrachten keinen signifikanten Einfluss auf das Überleben zeigen (s. Abb. 16).

Variable	HR	95% CI	P-Wert ¹
Tumorlokalisation (OC+ vs OC-)	0.92	(0.64, 1.32)	0.65
Alter (≤45 Jahre vs. >45 Jahre)	0.86	(0.64, 1.23)	0.42
Alkoholkonsum (kein, gering, mittel, hoch)	1.12	(1.01, 1.24)	0.037
Tabakkonsum (Nie, <20py, 20-40py, >40py)	1.32	(1.19, 1.45)	8.8e-08
Interaktion: Junger Patient * OC+	2.47	(1.19, 5.10)	0.015
UICC Tumorstadium (II, III, IV)	1.31	(0.89, 1.94)	0.17

Abb. 16: Multivariates Cox Proportional Hazards Modell zur Untersuchung der Einflussgrößen auf das Überleben von Patienten mit und ohne Mundhöhlenkarzinomen. (OC+ = Mundhöhlenkarzinom (oral cavity), HR = Hazard Ratio, CI = Konfidenzintervall, ¹ P-Werte des Wald-Tests auf Signifikanz eines multivariaten Cox Regressionsmodells nach Adjustierung für Kovariate)

Insgesamt belegen diese Überlebensunterschiede den aggressiveren Phänotyp von Tumoren der Mundhöhle bzw. der mobilen Zunge bei jungen Patienten im Vergleich zu Kopf-Hals-Tumoren anderer Lokalisation. Bisher ist kein Risikofaktor bekannt, welcher für diese Unterschiede verantwortlich ist und die Analysen zeigten, dass sie durch klassische Variablen wie Tumorstadium und Tabak- bzw. Alkoholkonsum nicht hinreichend erklärt werden können. Dies macht zusätzliche Einflussfaktoren –z.B. ein infektiöses Agens wie ein Virus- wahrscheinlicher, so dass in einem folgenden Schritt Gewebeproben von Tumoren der oralen/mobilen Zunge auf eine mögliche virale Ätiologie untersucht wurden.

3.2 Tumorproben der oralen Zunge

Um eine mögliche virale Ätiologie von Tumoren der oralen Zunge zu untersuchen, wurden sieben Tumorproben aus der „University of Chicago Tissue Bank“ ausgewählt. Dabei sollten die Patienten der Risikostruktur der jungen Patienten mit Mundhöhlen-/Zungenkarzinomen in der klinischen Patientenkohorte möglichst ähnlich sein. Dementsprechend war das Durchschnittsalter 37 Jahre, bei fehlendem oder minimalem Tabak- und Alkoholkonsum. Die Patienten waren in

Abhängigkeit vom Stadium chirurgisch, mit Chemoradiotherapie oder mit kombinierten Ansätzen therapiert worden. Einen Überblick bietet Tab. 4, OT-1 bis OT-7 sind die anonymisierten Bezeichnungen der jeweiligen Patienten.

Tab. 4: Charakteristika der sequenzierten Tumorproben der oralen Zunge. (OT = oral tongue, CRT= Chemoradiotherapie, RT = Radiotherapie, py = pack years)

	OT-1	OT-2	OT-3	OT-4	OT-5	OT-6	OT-7
<u>Geschlecht</u>	M	W	W	M	M	M	M
<u>Alter</u>	42	44	33	21	33	44	42
<u>Tabak</u>	Nein	Minimal (<1 py)	Minimal (<5 py)	Nein	Nein	Nein	Nein
<u>Alkohol</u>	Minimal/ Sozial	Minimal/ Sozial	Minimal/ Sozial	Minimal/ Sozial	Minimal/ Sozial	Nein	Minimal/ Sozial
<u>Tumor-stadium</u>	T4N0	Initial T1N0, mehrere Rezidive	Initial T3N0 Tumor, mehrere Rezidive	T2N0	T4N2B	T2N0	Rezidiv post-op, initial T1N0
<u>Therapie</u>	CRT	Op, CRT, palliative Chemoth.	Op, RT, CRT, palliative Chemoth.	Op	CRT	Op, CRT	Op, Induktion-schemoth. CRT

Über den klinischen Verlauf der Patienten lässt sich sagen, dass bei drei der Patienten (43%) Rezidive auftraten, nachdem die Initialtherapie mit zufriedenstellendem Ergebnis durchgeführt worden war, während ein Patient einen Progress unter einer Induktionstherapie zeigte. Zum Zeitpunkt der Datenauswertung waren fünf der Patienten rezidivfrei, ein Patient war gestorben und ein weiterer Patient wurde auf Grund von Fernmetastasen mit einer palliativen Chemotherapie behandelt. Die Abwesenheit einer HPV-Infektion mit Viren der Gruppe I inkl. der in HNO-Tumoren vorherrschenden high-risk Typen HPV-16 und HPV-18 (Dayyani et al., 2010; Kreimer, 2005) wurde für alle Tumore mit einer etablierten nested-PCR-Methode (Sotlar et al., 2004) untersucht und bestätigt. Die mRNA aller Proben konnte erfolgreich extrahiert, in cDNA-Libraries umgewandelt und anschließend sequenziert werden. Die Menge gewonnener Daten – gemessen als Anzahl insgesamt ausgelesener/sequenzierter Nukleotide (= Read-Länge x Read-

Anzahl) – lag dabei für alle Proben höher als in den Positivkontrollen, da zum einen jeweils mehr Reads generiert wurden und zweitens für die Mehrheit der Proben im Verlauf des Untersuchungszeitraumes durch technische Verbesserungen eine größere Read-Länge von 56bp in ausreichender Qualität nutzbar wurde (vgl. Tab. 5). Eine größere sequenzierte Datenmenge bedeutet, dass auch Transkripte geringerer Häufigkeit – wie es für solche viralen Ursprungs denkbar ist – mit höherer Wahrscheinlichkeit erfasst und ausgelesen werden. Die Sensitivität wird also mit zunehmender Datenmenge gesteigert, was wiederum das Vertrauen in das Ergebnis bestärkt.

3.3 Resultate des Virusalgorithmus

Ein Grundgedanke in der Entwicklung dieses Algorithmus war es, die bei einem Next-Generation Sequencing Experiment anfallenden großen Datenmengen auf einem handelsüblichen Desktop-Computer und mit frei erhältlicher Software verarbeiten und auf Sequenzen bekannter und unbekannter Viren untersuchen zu können. Der erste Teil wurde durch die Implementation einer „digitalen Subtraktion“ gelöst, da diese Sequenzen des Wirts aus den Ergebnissen herausfiltert und so potentielle virale Sequenzen anreichert. Diese reduzierte Datenmenge wurde anschließend mit einem schnellen und einem zwar langsameren, aber sensitiveren Suchalgorithmus auf Viren untersucht. Für Details s. Kapitel 2.5 und für einen Überblick über den Ablauf des Algorithmus s. Abb. 8. Nachdem die im Methodikteil beschriebenen Analyseschritte hintereinander geschaltet, die Übergänge programmiert und damit weitgehend automatisiert waren, ließ sich der Algorithmus auf einem Apple iMAC mit 2.4Ghz Prozessorleistung und 4Gb RAM (Apple, USA) problemlos für die sieben Tumorproben und die drei Positivkontrollen (Daten eines HPV16-positiven Zervixkarzinoms, der ebenfalls HPV16-positiven HNC-Zelllinie UM-SCC47 und eines für EBV LMP2A transgenen Lymphom-Mausmodells), ausführen.

3.3.1 Effizienz der digitalen Subtraktion und Performance des Algorithmus

Einen detaillierten Überblick über die Menge an gewonnenen Daten der jeweiligen Proben, die Veränderung der Menge Sequenzen im Verlauf der Analyse und die Laufzeit des Algorithmus findet sich in Tab. 5. Dabei ist die „Anfangszahl an 36/56bp Reads“ die Anzahl an entweder 36bp oder 56bp langen Sequenzabschnitten (Reads), welche für eine Probe sequenziert werden konnte. Sie stellen, abgesehen von einer ersten Qualitätskontrolle, die unverarbeiteten Rohdaten dar. Je höher die Zahl, desto mehr mRNA-Moleküle einer Probe sind sequenziert worden und

desto mehr Informationen liegen für diese Probe vor. Die folgenden zwei Zeilen geben an, wie viele dieser Reads verbleiben, nachdem die wirtsspezifischen Reads und die Reads bakteriellen/mikrobiellen Ursprungs digital durch einen Vergleich mit den jeweiligen Referenzgenomen herausgefiltert (subtrahiert) wurden. Dies reduzierte die Anzahl an Reads um ca. 65-75% (s. Abb. 17), wobei der zweite Schritt primär dazu dient, die Zahl an falsch positiven Treffern durch Homologie zwischen Bakterien und Viren zu vermindern. Diese verbleibenden Daten wurden nun mit dem schnellen Teil des Algorithmus (dem Rapid-Algorithmus, s. Methoden) mit den Genomen aller bekannter Viren verglichen und Übereinstimmungen nach den Bewertungskriterien für die Beurteilung positiver Virustreffer (s. Kapitel 2.5.5) ausgewertet.

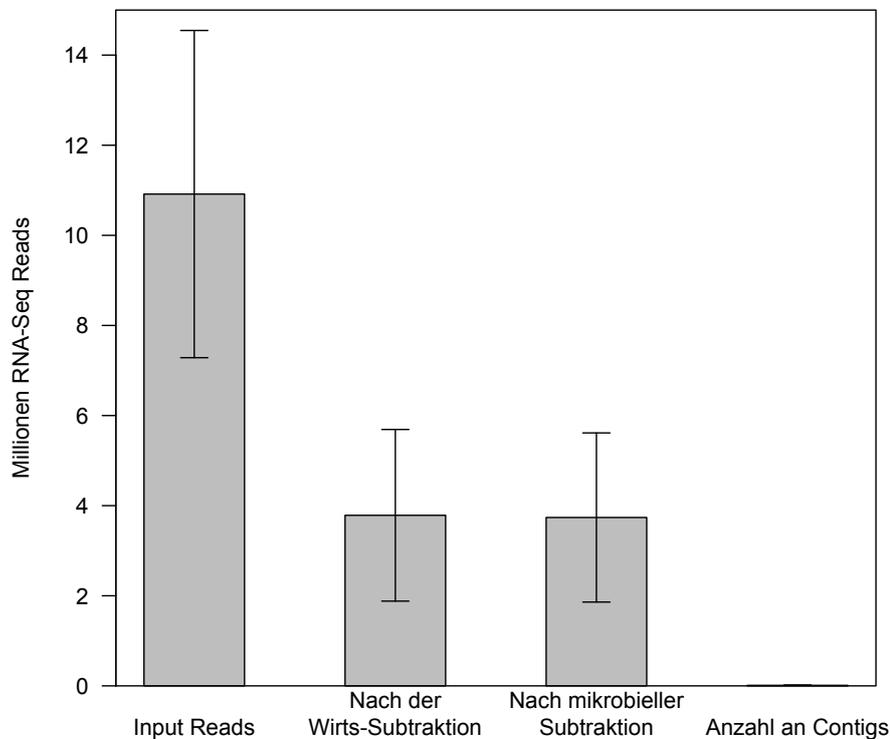


Abb. 17: Minderung der Anzahl an Reads während der Subtraktion, beginnend mit den Rohdaten (Input Reads) und nach der digitalen Subtraktion an wirtsspezifischen und mikrobiellen Sequenzen. Eine weitere deutliche Reduktion wird durch die Erstellung längerer Sequenzabschnitte (contiguous reads = „Contigs“) erreicht.

Tab. 5: Übersicht über die Algorithmus-Performance (*inkl. digitaler Subtraktion, †Ergebnisse des schnellen Algorithmus nach Beurteilung mit den Bewertungskriterien in Tab. 2)

	OT-1	OT-2	OT-3	OT-4	OT-5	OT-6	OT-7	UM-SCC47	EBV-LMP2A transgenes Lymphom	Zervixkarzinom
Anfangszahl an 36/56-bp Reads	14.580.891	15.624.000	12.543.623 (56bp)	13.757.855 (56bp)	12.019.919 (56bp)	8.868.629 (56bp)	11.803.357 (56bp)	8.630.847	7.366.864	3.962.855 (56bp)
Nach Wirts-Subtraktion	2.530.183	3.083.935	5.445.100	4.600.610	5.988.255	5.491.216	5.792.660	1.995.271	2.282.521	641.150
Reads nach mikrobieller Subtraktion	2.482.044	3.018.725	5.413.371	4.573.304	5.844.735	5.446.629	5.710.324	1.966.226	2.278.594	640.649
Anzahl an Contigs	11.849	10.006	20.876	22.992	14.528	10.684	15.186	5.657	2.441	857
Dauer des schnellen Algorithmus*	52min	56min	1h 25min	1h 20min	1h 45min	1h 35min	1h 55min	45min	38min	32min
Ergebnisse des Rapid Algorithmus†	-	-	-	-	-	-	-	HPV16	EBV (=HHV4)	HPV16
Dauer des Discovery Algorithmus	2h 50min	2h 35min	7h 13min	6h 50min	6h 30min	5h 7min	7h 11min	1h25min	30min	26min

Eine weitere deutliche Reduktion an zu analysierenden Sequenzen ergibt sich, wenn die verbleibenden kurzen Reads zu längeren Sequenzen (sog. Contigs, von contiguous = zusammenhängend) zusammengesetzt werden (s. Abb. 17). Diese konnten anschließend mit dem sensitiveren „Discovery-Algorithmus“ verarbeitet werden, welcher auch virale Sequenzen erkennen sollte die noch unbekannt sind, aber eine Verwandtschaft zu in der Datenbank vorhandenen Viren zeigen. Von diesen Contigs stammt wiederum der Großteil eindeutig vom Wirtsorganismus (s. Abb. 18). Nur ein kleiner Teil zeigt Übereinstimmungen mit einem Virus, wobei diese auch Contigs enthalten, die an humane und virale Sequenzen passen (z.B. bei humanen Homologen eines viralen Gens), so dass eine weitere Beurteilung dieser Treffer mit den Bewertungskriterien notwendig ist. Nur ein minimaler Teil der Contigs passt weder zu viralen noch humanen/murinen Genomen. Dies kann Indikator dafür gelten, dass die Zusammenfügung („Assembly“) der kurzen Sequenzen zu längeren Contigs zuverlässig zu funktionieren scheint.

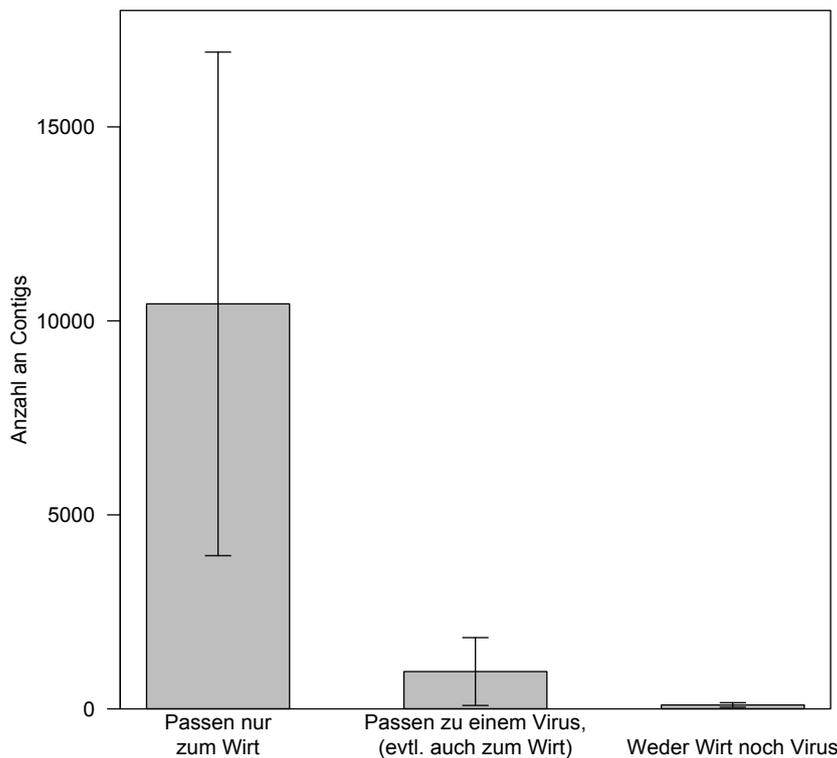


Abb. 18: Die Mehrheit der Contigs hat nur Übereinstimmungen mit Wirtssequenzen, ein kleiner Teil zeigt Übereinstimmungen mit einem Virus (z.T. auch zusätzlich Übereinstimmungen des gleichen Contigs mit Wirtsgenen) und nur eine Minderheit lässt sich gar nicht zuordnen.

3.3.2 Ergebnisse des schnellen Virusalgorithmus

Der schnelle Arm des Virusalgorithmus, in Verbindung mit den Beurteilungskriterien für Virus-treffer (Tab. 2), identifizierte in den drei Positivkontrollen zuverlässig die erwarteten Viren HPV16 und EBV (vgl. Abb. 19 und Abb. 20).

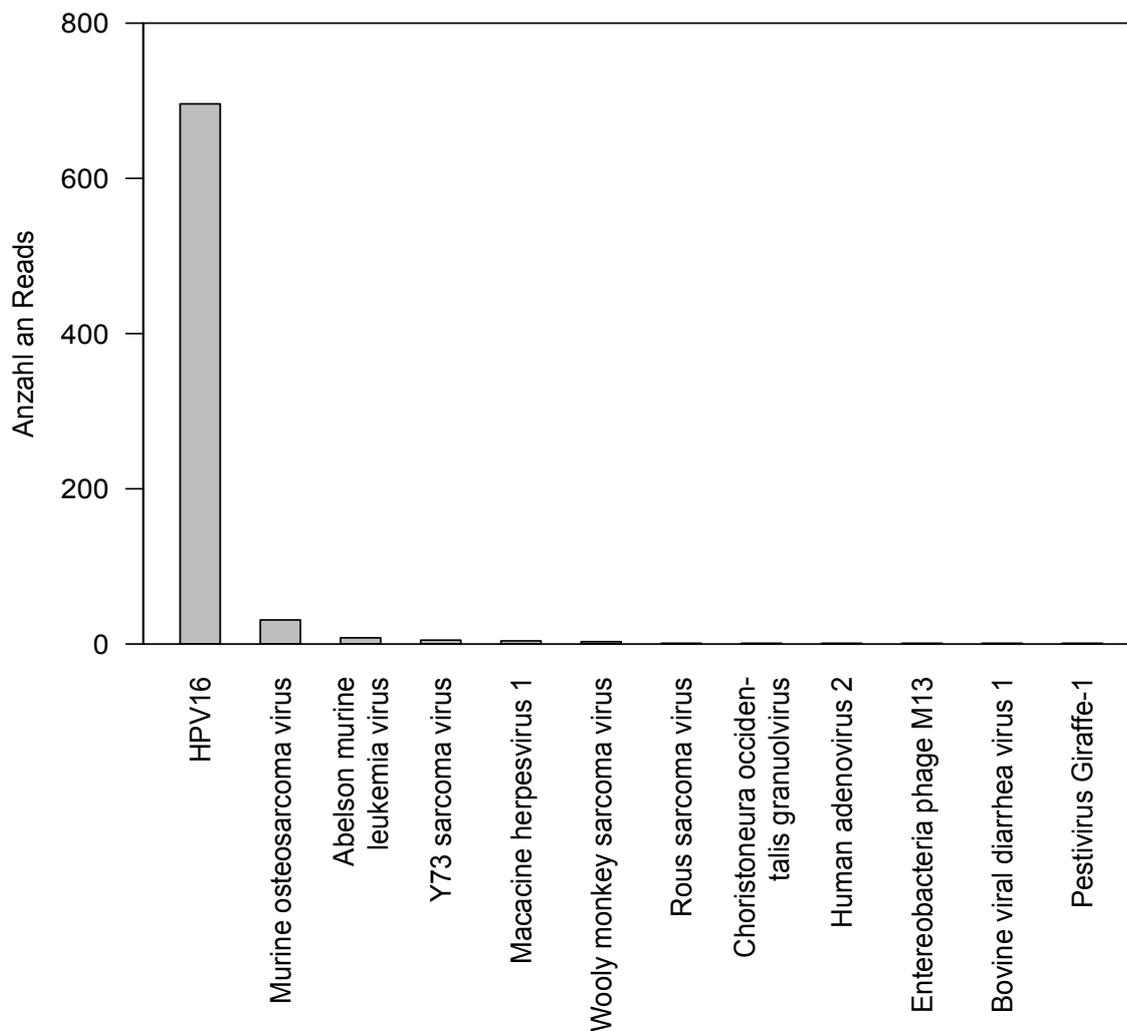


Abb. 19 Virustreffer für UM-SCC47: Im schnellen Algorithmus wird die Mehrheit viraler Reads korrekt HPV-16 zugeordnet. Die übrigen angezeigten Viren sind als falsch-positive Treffer zu bewerten, die größtenteils auf PCR-Artefakten und Homologie humaner und viraler Sequenzen beruhen.

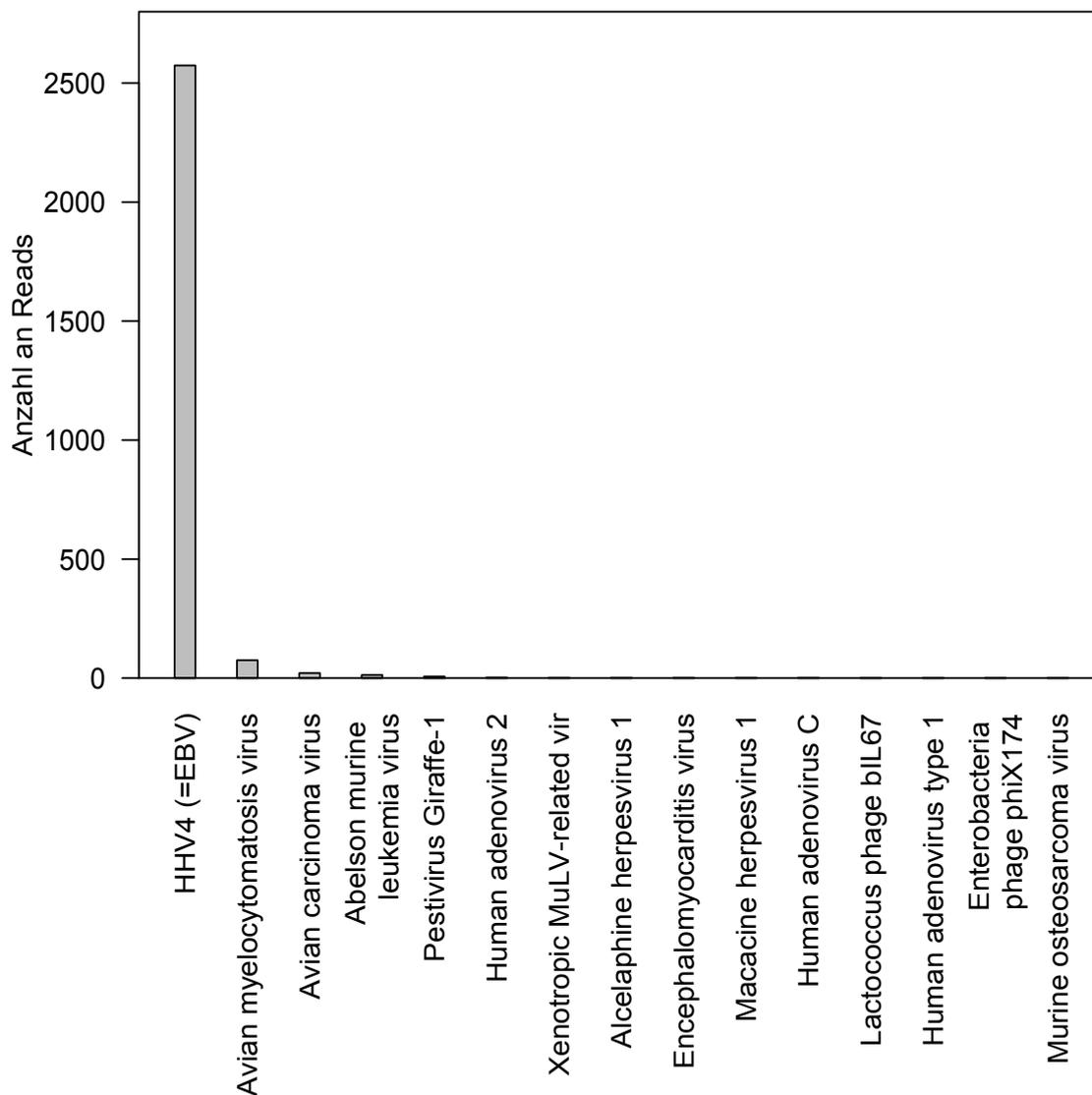


Abb. 20: Virustreffer für das EBV LMP2A-transgene Lymphom. Im schnellen Algorithmus wird die Mehrheit viraler Reads korrekt HHV4 (=EBV) zugeordnet. Die übrigen angezeigten Viren sind als falsch-positive Treffer zu bewerten, die größtenteils auf PCR-Artefakten und Homologie humaner und viraler Sequenzen beruhen.

Für die Tumorproben der oralen Zunge wurden keine nach den Bewertungskriterien signifikanten Übereinstimmungen mit bekannten Viren identifiziert.

Um die Zuverlässigkeit der Bewertungskriterien zu überprüfen und die Unterscheidung zwischen falsch-positiven und richtig-positiven Ergebnissen des Rapid-Algorithmus zu verbessern, wurde für jede Probe das Verhältnis von Reads mit Virusübereinstimmung zu allen sequenzierten Reads berechnet (s. Abb. 21). Obwohl auch in den Tumorproben (OT1 – 6) einige Reads Übereinstim-

mungen mit viralen Sequenzen zeigten, war die Anzahl an Reads mit viralem Ursprung für die Positivkontrollen, normalisiert für die Gesamtmenge an vorhandenen Daten (=Anzahl aller Reads) signifikant verschieden (Kruskal-Wallis-Test: $p=0,017$). Zusätzlich sind die viralen Reads in den Positivkontrollen wesentlich homogener verteilt: durchschnittlich gehören in ihnen 95,7% (95% CI des wahren erwarteten Mittelwertes 86,4-100%) der Reads mit einer Virusübereinstimmung von einem einzigen Virus (HPV16 bzw. EBV). Dies zeigt, dass die wahren positiven Treffer die Gesamtzahl aller viralen Treffer deutlich dominieren. Ähnliche hohe Werte werden für die Tumorproben nicht beobachtet (s. Abb. 22). Sowohl die absolute Zahl als auch die Verteilung viraler Reads in den Tumorproben erinnert vielmehr an die Verteilung in UM-SCC47, wenn man HPV16-Übereinstimmungen ignoriert und nur die übrigen Virustreffer betrachtet. Diese können als Maß für die erwarteten falsch-positiven Treffer dienen (vgl. Abb. 19, Abb. 21 und Abb. 22).

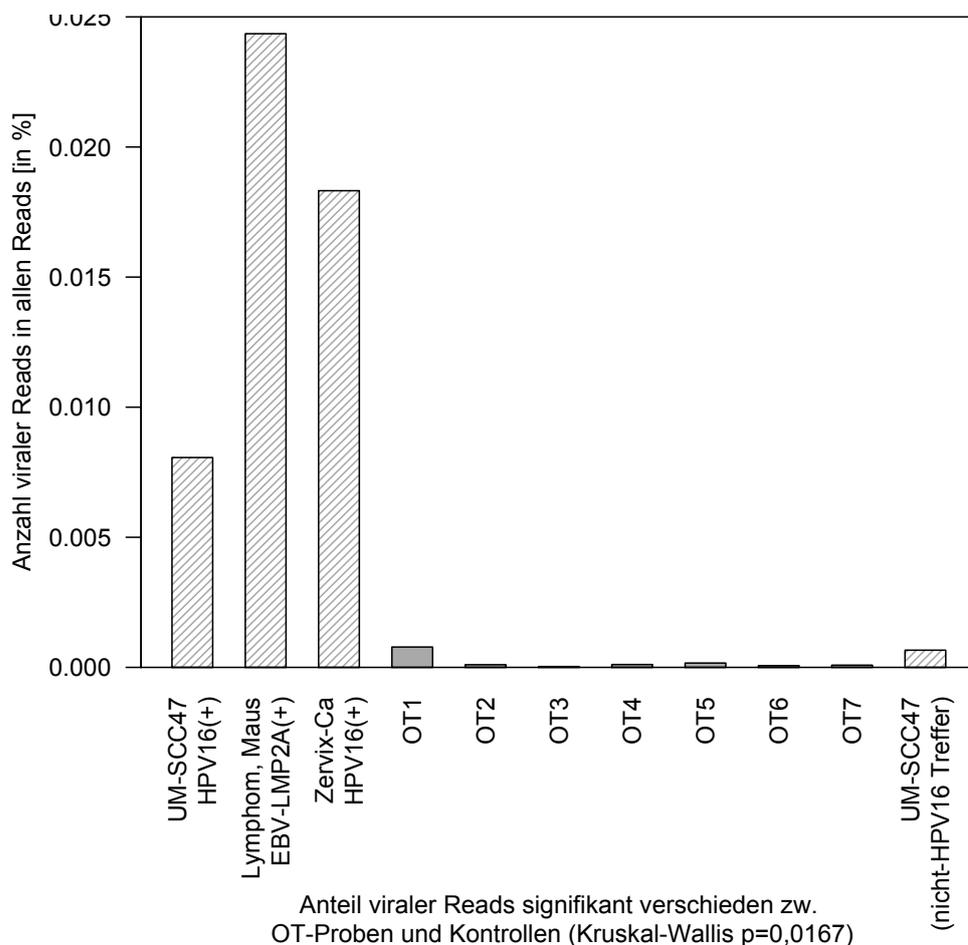


Abb. 21 Anzahl viraler Reads (Rapid-Algorithmus): In den Positivkontrollen stellen virale Reads einen größeren Anteil der Gesamtzahl aller Reads dar. In den Tumorproben sind es so wenige, wie die un spez. nicht-HPV Reads in der Positivkontrolle UM-SCC47 (rechts).

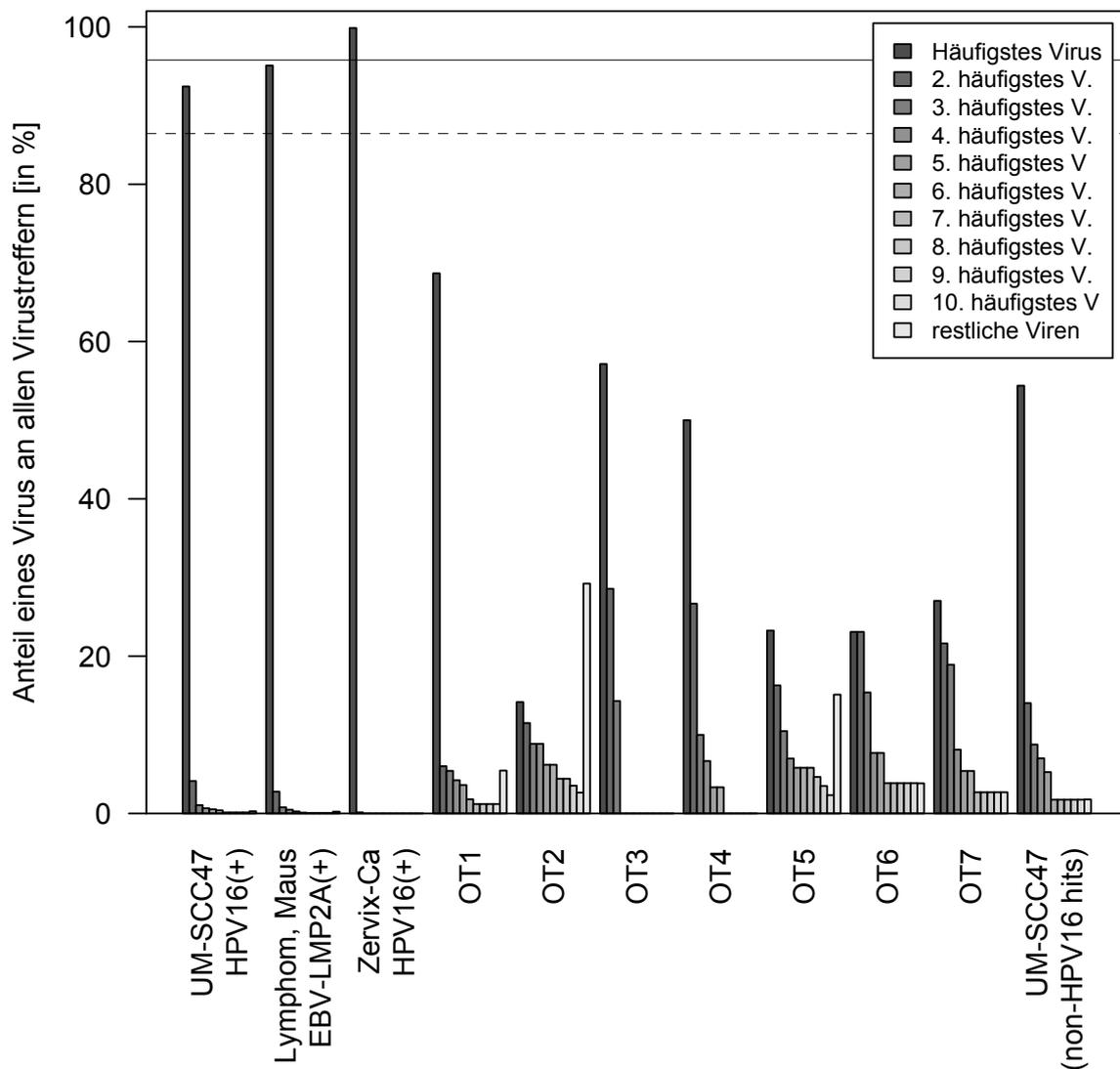


Abb. 22: Relativer Anteil eines Virus an allen Virustreffern (schwarze Linie = Mittelwert der Positivkontrollen, 95,7%; gestrichelte Linie = unteres Limit des 95% CI des Mittelwertes, 86,4%)

Sequenzen bzw. Reads aus den Proben der oralen Zunge, die zwar Übereinstimmungen mit Viren zeigten, aber nicht die Kriterien für positive Virusergebnisse erfüllten, wurden zusätzlich mit einer manuell web-basierten BLAST-Abfrage überprüft (Ye et al., 2006), da diese über eine noch umfangreichere Referenzdatenbank verfügt. Dabei zeigte sich, dass es sich bei diesen Reads entweder um Sequenzen humanen Ursprungs oder um repetitive bzw. wenig komplexe Reads handelte, welche am ehesten PCR-Artefakten –und damit falsch-positiven Ergebnissen– entsprechen.

Zusammengefasst finden sich in den Resultaten des schnellen Algorithmus also keinerlei Hinweise darauf, dass in den Tumorproben aktive Onkoviren vorhanden sind, während eine Identifikation der zu erwartenden Viren in den Positivkontrollen erfolgreich ist.

Die vergleichsweise hohe Zahl an Reads, die sowohl Übereinstimmungen mit humanen als auch mit viralen Sequenzen zeigen, ist am ehesten mit der Homologie menschlicher und viraler Gene zu erklären. So belegen Studien, dass bis zu 8% des menschlichen Genoms vermutlich aus endogenen Retroviren entstanden sind (Griffiths, 2001). Zusätzlich zeigen 13% aller Herpesvirusproteine (Holzerlandt et al., 2002), sowie einige humane Parvoviren (Kerr und Boschetti, 2006) eine deutliche Sequenzähnlichkeit zu Produkten menschlicher Gene. Für derartige, häufig stark konservierte Homologe wird eine Bedeutung für die Interaktion zwischen Wirt und Virus (z.B. im Zusammenhang mit Zelladhäsion, Wnt-Signalweg, Zytoskelett) postuliert (Kerr und Boschetti, 2006). Ein Beispiel dafür, wie Sequenzhomologie zu falsch positiven Ergebnissen führen kann, ist das Murine osteosarcoma-Virus, welches in allen untersuchten Proben -inklusive der Positivkontrollen- gefunden wurde (vgl. Abb. 19 und Abb. 20). Durch die hohe Ähnlichkeit des viralen v-Fos Gens zu seinem humanen Homolog c-Fos (van Straaten et al., 1983) werden im schnellen Algorithmus einige der kurzen Reads dem Virus zugeschrieben, die längeren Contigs in der spezifischeren BLAST-Suche des Discovery-Algorithmus dann aber als humane Variante erkannt.

3.3.3 Ergebnisse des Discovery-Algorithmus

Im Discovery-Algorithmus konnten in allen Positivkontrollen eindeutig die bekannten und erwarteten Viren identifiziert werden. Abb. 23 und Abb. 24 zeigen einen Bildschrifto/Screenshot der kürzeste der Ausgabedateien für UM-SCC47 und das EBV LMP2A-transgene Lymphom. Diese Datei zeigt für alle Contigs, die einen nach den Beurteilungskriterien für Virusübereinstimmungen signifikanten Virustreffer aufweisen, jeweils den wahrscheinlichsten Ursprungsorganismus an, so dass auf einen Blick alle signifikanten Viruskandidaten erkannt werden können. Die Struktur der BLAST-Ergebnis-Zeilen ist: Datenbank in der die Zielsequenz gefunden wurde (z.B. gb = GenBank) | Identifikationsnummer (Accession number) | Beschreibung des Organismus/Gens mit dem das Contig überein stimmt; Score(bits): errechneter Wert aus der Zahl an konkordanten Nukleotiden, Substitutionen und Gaps im paarweisen Alignment, höhere Werte bedeuten eine höhere Übereinstimmung; E-value: Wahrscheinlichkeit, dass eine Sequenz mit einem ähnlich

hohen Score zufällig in der Datenbank vorhanden ist (entspricht in seiner Bedeutung einem p-Wert, je kleiner der E-value desto signifikanter ist die Übereinstimmung).

```

5657 sequences/contigs in your input file
5598 were successfully matched to human and/or viral sequences,
  59 could not be matched at all.

5270 contigs produced only human hits
  328 contigs had a viral match (potentially together with human matches).

-----
Top scoring contigs for which a virus might be suspected are:
-----

Sequences producing significant alignments:
                                     Score   E
                                     (bits) Value
-----
1. gb|AF472509.1| Human papillomavirus type 16 variant (African typ...  497  e-138
2. gb|AF472509.1| Human papillomavirus type 16 variant (African typ...  263  3e-68
3. dbj|D00735.1|PPH16E Human papillomavirus type 16 genes for E6 pr...  213  2e-53
4. gb|AF402678.1| Human papillomavirus type 16 Asian-American varia...  132  3e-29

```

Abb. 23 Ergebnis des Discovery-Algorithmus für UM-SCC47 (hier als Screenshot der zusammenfassenden Ausgabedatei)

```

2441 sequences/contigs in your input file
2369 were successfully matched to human and/or viral sequences,
  72 could not be matched at all.

2224 contigs produced only human hits
  145 contigs had a viral match (potentially together with human matches).

-----
Top scoring contigs for which a virus might be suspected are:
-----

Sequences producing significant alignments:
                                     Score   E
                                     (bits) Value
-----
1. gb|M24212.1|HS4MEMBR Epstein-Barr virus membrane protein mRNA, c...  1515  0.0
2. gb|M24212.1|HS4MEMBR Epstein-Barr virus membrane protein mRNA, c...   512  e-142
3. emb|AJ507799.2| Human herpesvirus 4 complete wild type genome         323  5e-86

```

Abb. 24: Ergebnis des Discovery-Algorithmus für das EBV LMP2A-transgene Mauslymphom (hier als Screenshot der zusammenfassenden Ausgabedatei)

In drei der sieben Proben von Tumoren der oralen Zunge wurden ebenfalls Viruskandidaten angegeben. Bei Überprüfung der detaillierteren Ergebnisse stellten sich zwei davon allerdings als Antikörper humanen Ursprungs gegen Rotaviren heraus, die lediglich wegen des Begriffs „Rota-

virus“ in der jeweiligen Ergebniszeile als virale Kandidaten angezeigt worden waren. Ein dritter (in OT-2) zeigte bei einer web-basierten BLAST-Überprüfung die höchste Ähnlichkeit mit dem „Cloning vector pT2gammaCryK2“, welcher in der Datenbank des Discovery-Algorithmus nicht vorhanden war. Grundsätzlich wurden alle Ergebnisse manuell durchgesehen und per web-basierter BLAST-Überprüfung verifiziert (s. Tab. 6).

Insgesamt finden sich also für die Tumorproben der oralen Zunge auch in den Resultaten des wesentlich sensitiveren Discovery-Algorithmus keine überzeugenden Hinweise auf das Vorhandensein viraler Transkripte, während dies in den Positivkontrollen möglich war.

Tab. 6: Überprüfung signifikanter Viruskandidaten des Discovery-Algorithmus mittels BLAST-Abfrage (Auszug einer Excel-Tabelle; Vergleich zwischen Discovery-Algorithmus und der web-basierter Blast-Überprüfung: für die Positivkontrollen stimmen beide überein und werden als richtig pos. gewertet, für die Tumorproben ergeben sich diskordante Ergebnisse (2x humane Antikörper und ein Klonierungsvektor), sie werden deshalb als falsch pos. gewertet. In der Originaldatei ist jeweils noch die Nukleotidsequenz des entsprechenden Contigs angegeben, hier aus Platzgründen weg gelassen.)

Probe	Kandidaten im Discovery algorithm (nach den pos. Selektionskriterien)	Verifikation mittels web-basiertem BLAST-Abgleich (Jeweils die besten Übereinstimmungen)
UM-SCC47	1. gb AF472509.1 Human papillomavirus type 16 variant (African typ... 497 e-138 2. gb AF472509.1 Human papillomavirus type 16 variant (African typ... 263 3e-68 3. dbj D00735.1 PPH16E Human papillomavirus type 16 genes for E6 pr... 213 2e-53 4. gb AF402678.1 Human papillomavirus type 16 Asian-American varia... 132 3e-29	gb AF472509.1 Human papillomavirus type 16 variant (African type 2), complete genome (E = 0.0) gb AF472509.1 Human papillomavirus type 16 variant (African type 2), complete genome dbj D00735.1 PPH16E Human papillomavirus type 16 genes for E6 protein gb AY686582 Human papillomavirus type 16 isolate Qv15351AA, complete genome (E = 9e-40)
Mauslymphom	1. gb M24212.1 HS4MEMBR Epstein-Barr virus membrane protein mRNA, c... 1515 0.0 2. gb M24212.1 HS4MEMBR Epstein-Barr virus membrane protein mRNA, c... 512 e-142 3. emb AJ507799.2 Human herpesvirus 4 complete wild type genome 323 5e-86	gb M87779.1 Epstein-Barr virus latent membrane protein 2 (LMP2) mRNA, complete cds (E = 0.0) gb M24212.1 Epstein-Barr virus membrane protein mRNA, complete cds (E = 0.0) emb AJ507799.2 Human herpesvirus 4 complete wild type genome (E = 2e-113)
Zervixkarzinom	1. gb AF193428.1 AF193428 Human papillomavirus type 16 isolate DAG4... 331 2e-88 2. gb AF402678.1 Human papillomavirus type 16 Asian-American varia... 183 2e-44 3. gb AF402678.1 Human papillomavirus type 16 Asian-American varia... 143 2e-32 4. gb AF327851.1 AF327851 Human papillomavirus type 16 early transf... 130 6e-29	gb FJ610152.1 Human papillomavirus type 16 strain CU7, complete genome (E = 2e-117) gb FJ237041.1 Human papillomavirus isolate b00480 E6 (E6) and E7 (E7) genes, complete cds (E = 1e-59) gb HQ398224.2 Human papillomavirus type 16 isolate D7 1176 clone c2 E2 protein (E2) gene, complete cds (E = 3e-39) gb JN381067.1 Human papillomavirus type 16 strain 10-219 E6 protein (E6) gene, complete cds (E = 3e-39)
OT-1	NA	NA
OT-2	1. emb Z48163.2 Semliki forest virus genomic RNA for non-structura... 189 5e-46 2. emb Z48163.2 Semliki forest virus genomic RNA for non-structura... 137 6e-31 3. emb Z48163.2 Semliki forest virus genomic RNA for non-structura... 126 1e-27	gb JN418198.1 Cloning vector pT2gammaCryK2, complete sequence (E = 9e-26) gb JN580989.1 Cloning vector pCZ-NAT, complete sequence (E = 7e-42) gb JN580989.1 Cloning vector pCZ-NAT, complete sequence (E = 1e-37)
OT-3	NA	NA
OT-4	1. gb AF453139.1 Synthetic construct clone 6-64VL rotavirus VP6-specific antibody light ch	emb X87892.1 H.sapiens mRNA for light chain variable region, DO1-VL (E = 3e-29)
OT-5	NA	NA
OT-6	1. gb AF452982.1 Synthetic construct clone 6-52VH rotavirus VP6-specific antibody 131	gb AF452982.1 Synthetic construct clone 6-52VH rotavirus VP6-specific antibody heavy chain variable region mR
OT-7	NA	NA

3.3.4 Ergebnis der Simulation eines unbekanntes Virus

Um das Vorkommen eines unbekanntes Virus zu simulieren und gleichzeitig die Sensitivität in der Entdeckung noch unbekanntes Viren (die mit bekannten Viren verwandt sind) zu testen, wurden die Sequenzen von HPV16 und EBV aus allen Datenbanken gelöscht, wodurch sie quasi „unbekannt“ wurden. Anschließend wurden beide Algorithmen für UM-SCC47 und das EBV LMP2A-transgene Lymphom wiederholt.

Der schnelle Algorithmus produzierte dabei identische Resultate mit dem einzigen Unterschied, dass jeweils HPV16 bzw. EBV fehlten. Er würde aufgrund seiner recht hohen Intoleranz gegenüber Abweichung zweier Sequenzen verwandte Viren also übersehen.

Im Gegensatz dazu identifizierte der Discovery-Algorithmus in den aus UM-SCC47 gewonnenen Sequenzen allerdings mehrere verwandte E6-Transkripte (s. Abb. 25).

Top scoring contigs for which a virus might be suspected are:				
Sequences producing significant alignments:			Score	E
			(bits)	Value
1.	gb AF469198.1	Human papillomavirus E6 protein (E6) gene, comple...	373	e-100
2.	gb AF548023.1	Human papillomavirus from Thailand E6 protein (E6...	192	8e-47
3.	gb AF469198.1	Human papillomavirus E6 protein (E6) gene, comple...	161	2e-37

Abb. 25: Ergebnis des Discovery-Algorithmus für UM-SCC47 nachdem HPV-16 aus der Datenbank gelöscht wurde (hier als Screenshot der zusammenfassenden Ausgabedatei)

Zusätzlich fanden sich im detaillierteren Output eine Vielzahl weiterer HPV-Typen und HPV-Spezies (s. Abb. 26), welche per definitionem nur 60-70% Sequenzähnlichkeit haben (de Villiers et al., 2004). Dies zeigt eine hohe Sensitivität des Algorithmus bei der Detektion noch unbekanntes Viren an, die mit bekannten Viren verwandt sind.

Sequences producing significant alignments:	Score (bits)	E Value
**gb AF469198.1 Human papillomavirus E6 protein (E6) gene, comple...	373	e-100
** gb J04353.1 PPH31A Human papillomavirus type 31 (HPV-31) complet...	328	3e-87
** gb AF548023.1 Human papillomavirus from Thailand E6 protein (E6...	325	2e-86
** emb X74477.1 Human papillomavirus type 35H genomic DNA	322	2e-85
** gb M74117.1 PPH35CG Human papillomavirus type 35 complete genome	322	2e-85
** dbj D10597.1 PPHE67 Human papillomavirus gene for E6 and E7 prot...	293	1e-76
** gb EF422154.1 Human papillomavirus type 31 strain P378 E7 prote...	290	5e-76
** gb EF422153.1 Human papillomavirus type 31 strain P366 E7 prote...	290	5e-76
** gb EF422152.1 Human papillomavirus type 31 strain P253 E7 prote...	290	5e-76
...		
** emb X64004.1 Human Papillomavirus type 33 mRNA for pot. E6*I, E...	199	2e-48
** dbj D21208.1 Human papillomavirus type 67 complete genome	197	5e-48
** gb M12732.1 PPH33CG Human papillomavirus type 33, complete genome	197	1e-47
...		
** emb X94165.1 Human papillomavirus type 73 E6, E7, E1, E2, E4, L...	111	4e-22
** gb EU918768.1 Human papillomavirus type 11 isolate LZod45-11, c...	110	9e-22
...		
** gb EF202153.1 Human papillomavirus type 18 isolate Qv21444, com...	62	5e-07
** gb AY192156.1 Synthetic construct Human papillomavirus E7e pept...	61	1e-06
** gb EF177181.1 Human papillomavirus type 56 clone Qv25665, compl...	60	1e-06
** gb AF293961.1 Human papillomavirus type 82 subtype IS39/AE2, co...	60	1e-06
** gb AY395706.1 Human papillomavirus type 62 isolate Qv18091, com...	60	2e-06
...		
** gb M75123.1 PPHE6E7GEN Human papillomavirus E6 and E7 protein se...	56	4e-05
** gb M38198.1 PPH45E67A Human papillomavirus type 45 E6 and E7 gen...	55	5e-05
** gb M62877.1 PPHDNA Human papilloma virus type 51 genomic DNA, pa...	55	7e-05
** emb Y15175.1 Human papillomavirus type 77 E6, E7, E1, E2, E4, L...	54	1e-04
** emb X64006.1 Human Papillomavirus type 33 mRNA for pot. E6*II, ...	53	2e-04
** gb U31793.1 HPU31793 Human papillomavirus type 61, complete genome	51	7e-04
** emb X74474.1 Human papillomavirus type 30 genomic DNA	50	0.002
** gb Q091857.1 Human papillomavirus type 97 isolate Qv20597 spli...	46	0.030
** gb AF436128.1 Human papillomavirus - cand89, complete genome	45	0.079
** emb X74463.1 Human papillomavirus type 7 genomic DNA	44	0.12
** gb AF131950.1 Human papillomavirus candHPV85, complete genome	44	0.14
** emb X74478.1 Human papillomavirus type 40 genomic DNA	43	0.21

Abb. 26: Ergebnis des Discovery-Algorithmus für UM-SCC47 nachdem HPV-16 aus der Datenbank gelöscht wurde (Screenshot einer der ausführlicheren Ausgabedateien)

Auch im Fall des Mauslymphoms hätte der Discovery-Algorithmus das Vorhandensein eines Virus nahegelegt, obwohl ihm die genaue Sequenz von EBV nicht bekannt war (s. Abb. 27).

```

2441 sequences/contigs in your input file
2369 were successfully matched to human and/or viral sequences,
  72 could not be matched at all.

2225 contigs produced only human hits
 144 contigs had a viral match (potentially together with human matches).

-----
Top scoring contigs for which a virus might be suspected are:
-----

```

Sequences producing significant alignments:	Score (bits)	E Value
1. gb U35853.1 HPU35853 Herpesvirus papio latent membrane protein 2...	386	e-104

Abb. 27: Ergebnisse des Discovery-Algorithmus für das EBV LMP2A-transgene Lymphom nachdem EBV aus der Datenbank entfernt wurde (hier als Screenshot der zusammenfassenden Ausgabedatei)

Es wurde hier allerdings nur ein einzelner signifikanter Viruskandidat angezeigt, nämlich das „Herpesvirus papio latent membrane protein 2Da“, welches aus einem in Pavianen beschriebenen Alphavirus bekannt ist (Tyler und Severini, 2006). Da im LMP2A-transgenen Lymphom nur ein virales Gen vorhanden ist, war dies zu erwarten. Bei manueller Durchsicht der Ergebnisse fanden sich Übereinstimmungen anderer Contigs mit weiteren Herpesviren (u.a. Kaposi's sarcoma virus (HHV8) und Herpes simplex virus 2 (HHV2)), die allerdings nicht signifikant waren (s. Abb. 28).

Sequences producing significant alignments:	Score (bits)	E Value
**gb U35853.1 HPU35853 Herpesvirus papio latent membrane protein 2...	386	e-104
** gb AY037858.1 Cercopithecine herpesvirus 15 strain LCL8664, com...	158	1e-35
gb AC129597.5 Mus musculus BAC clone RP23-262117 from 13, compl...	43	0.44
gb AC134826.4 Mus musculus BAC clone RP24-459P6 from chromosome...	43	0.44
...		
...		
** gb DQ149153.1 Cercopithecine herpesvirus 16 strain X313, comple...	53	4e-04
** gb DQ017380.1 Trichoplusia ni single nucleopolyhedrovirus, comp...	52	7e-04
** emb AJ890364.1 Emiliana huxleyi virus 86 isolate EHV86	51	0.001
** gb AF533768.1 Cercopithecine herpesvirus 1 strain E2490, comple...	49	0.004
** dbj AB074432.1 Cercopithecine herpesvirus 1 US1, US2, US3, US4,...	49	0.004
** gb AF204951.2 Ectocarpus siliculosus virus, complete genome	48	0.010
** gb AF281817.1 AF281817 Tupaia herpesvirus strain 2, complete genome	46	0.025
** emb Z86099.2 Herpes simplex virus type 2 (strain HG52), complet...	46	0.037
** dbj D10471.1 HS2ULIR Herpes simplex virus type 2 genomic DNA for...	46	0.037
** gb FJ483967.1 Squirrel monkey cytomegalovirus 1 genomic sequence	45	0.079
** dbj AB232129.1 Cercopithecine herpesvirus 16 UL44 gene for glyc...	45	0.095
** gb AY261359.1 Bovine herpesvirus 5 strain SV507/99, complete ge...	44	0.17
** tpg BK001744.1 TPA: TPA_exp: Suid herpesvirus 1, complete genome	42	0.44
** gb AF232689.2 Rat cytomegalovirus Maastricht, complete genome	42	0.44
** emb X95710.1 Pseudorabies virus UL23, UL24, UL25, UL26, UL26.5,...	42	0.44
...		
...		
** gb AF360120.1 AF360120 Human herpesvirus 8 ORF73 gene, complete cds	41	0.10
** gb AF305694.1 AF305694 Kaposi's sarcoma-associated herpesvirus l...	41	0.10
** gb AF192756.1 AF192756 Kaposi's sarcoma-associated herpesvirus O...	41	0.10
** gb U93872.2 Kaposi's sarcoma-associated herpesvirus glycoprotei...	41	0.10
** gb AF148805.2 Human herpesvirus 8 type P isolate GK18, complete...	41	0.10
** gb U75698.1 KSU75698 Human herpesvirus 8 type M, partial genome	41	0.10
...		
...		
** emb Z86099.2 Herpes simplex virus type 2 (strain HG52), complet...	36	0.44
** gb FJ427299.1 Human herpesvirus 2 strain 333 ICP34.5 (RL1) mRNA...	36	0.44
** gb DQ149924.2 Human herpesvirus 2 strain 333 ICP34.5 (RL1) gene...	36	0.44
** dbj D10471.1 HS2ULIR Herpes simplex virus type 2 genomic DNA for...	36	0.44
...		

Abb. 28: Ergebnisse des Discovery-Algorithmus für das EBV LMP2A-transgene Lymphom nachdem EBV aus der Datenbank entfernt wurde (hier ein Screenshot der ausführlicheren Ausgabedatei; jeder Kasten entspricht Übereinstimmungen je eines Contigs, nur der oberste ist allerdings signifikant)

3.4 Expression von CDKN2A/p16^{INK4a}

In Tumoren des Oropharynx ist die Proteinexpression von CDKN2A/p16^{INK4a}, die i.d.R. durch Immunhistochemie an Präparatschnitten quantifiziert wird, ein etablierter und weit verbreiteter Surrogatparameter zum Nachweis einer HPV-Infektion (Jordan et al., 2012; Kuo et al., 2008; Shi et al., 2009).

Interessanterweise zeigte sich aber auch in drei der sieben Tumorproben der oralen Zunge eine deutliche Expression von p16-mRNA (s. Abb. 29), obwohl sie per nested-PCR und per Viralalgorithmus als HPV-negativ verifiziert wurden. Die Expression in den drei p16-positiven Proben war dabei signifikant höher als in den vier p16-negativen Proben ($p=0,01$; zweiseitiger Student-t-Test) und in seiner Ausprägung vergleichbar mit dem des HPV16-positiven Zervixkarzinoms. Die Expression von p16 in UM-SCC47 ist zwar größer, aber da es sich um eine Zelllinie handelt in welcher z.B. keine Stromazellen oder Normalgewebe vorhanden sind, scheint das Zervixkarzinom einen besseren und realitätsnäheren Vergleich darzustellen.

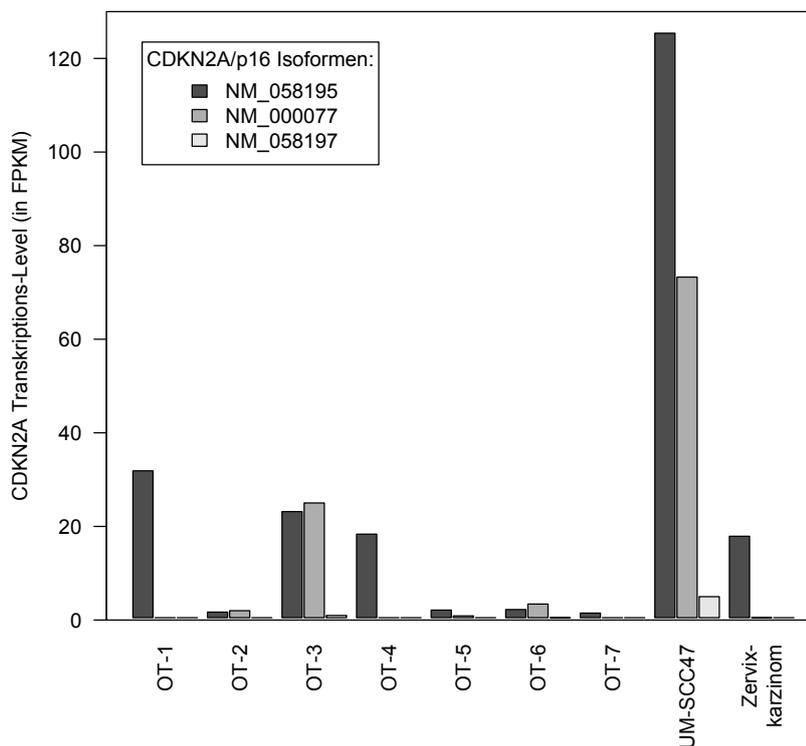


Abb. 29: Isoform-spezifische Expression von CDKN2A/p16 in den Tumorproben und den HPV-positiven Kontrollen (FPKM = Anzahl an p16-Reads in den Rohdaten, normalisiert für die Gesamtzahl an Reads und die Länge des jeweiligen p16-Transkriptes, s. Methoden)

4. Diskussion

4.1 Zur Analyse der klinischen Daten

4.1.1 Interpretation der Ergebnisse

Unter anderem wurde in dieser Arbeit der Einfluss verschiedener Faktoren auf das Überleben und die Prognose von Patienten mit Kopf-Hals-Tumoren untersucht. Da aus klinischen Beobachtungen ein aggressiverer Verlauf der Erkrankung besonders bei jüngeren Patienten mit Mundhöhlenkarzinomen vermutet wurde, lag ein Schwerpunkt auf den Unterschieden verschiedener Primärtumorlokalisation und Altersgruppen. Zu diesem Zweck wurden die Daten einer Kohorte von 748 Kopf-Hals-Tumorpatienten, welche im Rahmen von klinischen Studien an der University of Chicago behandelt worden waren, retrospektiv systematisch ausgewertet.

In einer in Altersgruppen getrennt durchgeführten Analyse zeigte sich, dass unter den jungen Patienten (≤ 45 Lebensjahre) solche mit Karzinomen der Mundhöhle eine deutlich schlechtere Prognose hatten als jene mit Tumoren an anderen Lokalisationen, was aus der signifikant verringerten 5-Jahresüberlebensrate und dem ebenfalls signifikant verringerten progressionsfreien Intervall hervor geht. Ein solcher Effekt war bei den älteren Patienten (>45 Lebensjahre) nicht nachweisbar. Auch bei Betrachtung der gesamten Kohorte hatten junge Patienten mit Mundhöhlenkarzinomen eine schlechtere Prognose im Vergleich zu allen übrigen Patienten. So zeigte die multivariate Analyse, dass junges Alter in Kombination mit dem Vorliegen eines Mundhöhlenkarzinoms einen signifikanten unabhängigen Risikofaktor darstellte, diese Patienten also ein erhöhtes Sterberisiko hatten. Weitere signifikante Risikofaktoren waren, übereinstimmend mit anderen Berichten (Massano et al., 2006), Alkohol- und Tabakkonsum. Weder Tumorstadium noch Alter bzw. Tumorlokalisation (allein betrachtet) hatten einen signifikanten Einfluss.

Diese Befunde belegen, dass in dieser Kohorte ein Tumor der Mundhöhle und jüngeres Alter Risikofaktoren für vermindertes Überleben sind und diese Patienten tatsächlich eine schlechtere Prognose zeigen. Zusätzlich verdeutlichen sie die Notwendigkeit einer nach anatomischen und altersabhängigen Subgruppen differenzierten Analyse, um korrekte Prognoseeinschätzungen treffen zu können, da diese unterschiedlich stark zum Gesamtüberleben aller Kopf-Hals-Tumorpatienten beitragen.

4.1.2 Bewertung der Ergebnisse

Auch einige frühere Studien haben eine schlechtere Prognose für jüngere Patienten mit oralen Tumoren beschrieben (Garavello et al., 2007; Liao et al., 2006), andere haben keinen derartigen Effekt nachweisen können (Pytynia et al., 2004; Sargeran et al., 2008; Sasaki et al., 2005). In allen Fällen war die untersuchte Kohorte allerdings kleiner als die hier verwendeten 748 Patienten. Mit einer univariaten Analyse wurde auch in der vorliegenden Arbeit kein Unterschied im Überleben zwischen älteren und jüngeren Patienten mit Mundhöhlenkarzinomen festgestellt. In einem multivariaten Cox-Regressions-Modell konnte allerdings für konfundierende Variablen (u.a. Tabak und Alkohol) kontrolliert und auf diese Weise das erhöhte Risiko junger Patienten mit Mundhöhlenkarzinomen herausgestellt werden.

Interessanterweise war das Tumorstadium kein unabhängiger Faktor in der Prognose der Kopf-Hals-Tumorpatienten, obwohl es als einer der wichtigsten prognostischen Faktoren gilt (Patel und Shah, 2005). Ein Grund dafür könnte die sehr ausgeglichene Verteilung der Tumorstadien zwischen älteren und jüngeren Patienten sein. Zudem liegen fast ausschließlich fortgeschrittene Stadien (III und IV) vor, während die deutlichsten Überlebensunterschiede zwischen frühen (I/II) und späten (III/IV) Stadien gesehen werden (Argiris et al., 2008). Zudem waren die Patienten mit Radiochemotherapie behandelt worden. Die Analysen wurden anhand der Therapieprotokolle stratifiziert durchgeführt, um mögliche Effekte zwischen verschiedenen Behandlungen auszugleichen. Ein stärkeres Ansprechen von Patienten mit Tumoren im Stadium IV im Vergleich zu solchen im Stadium III ist denkbar und würde mögliche Unterschiede im Überleben innerhalb eines Behandlungsarms verwischen. Insgesamt ist bei der Beurteilung der Ergebnisse der vorliegenden Untersuchung der hohe Anteil von Patienten mit fortgeschrittenen Tumorstadien zu berücksichtigen, so dass Aussagen über Patienten mit frühen Tumorstadien schwierig sind und diese gesondert analysiert werden sollten.

Es bleibt allerdings festzuhalten, dass dennoch ein nicht signifikanter Trend zu einer Erhöhung des Sterberisikos mit steigendem Tumorstadium erkennbar war. Im Gegensatz dazu zeigten die jungen Patienten mit Mundhöhlenkarzinomen trotz ihrer erhöhten Mortalität einen (ebenfalls nicht signifikanten) Trend zu niedrigeren Tumorstadien.

Insgesamt scheinen diese Befunde zu bestätigen, dass Mundhöhlenkarzinomen von Patienten <45 Jahre einen aggressiveren Phänotyp zeigen als andere Tumore und diese deshalb als eigenständige Tumorentität betrachtet werden sollten, wie es in der Literatur geäußert wird (Chitapanarux et

al., 2006; Salem, 2010). Bisher konnte keiner der bekannten Risikofaktoren dafür verantwortlich gemacht werden. Im Gegenteil ist in dieser Kohorte z.B. der Anteil an „Niemals-Rauchern“ unter den jungen Patienten mit Mundhöhlenkarzinom höher (37,1%) als bei den jungen Patienten mit anderen Tumoren (20,7%). Als möglichen ätiologischen und prognostischen Faktor haben verschiedenen Autoren den Einfluss von Onkoviren vorgeschlagen, da diese auch die steigende Inzidenz dieser Tumorentität erklären könnten (Rosenquist et al., 2005; Salem, 2010). Die weiteren Ergebnisse dieser Arbeit (s.u.) machen den Einfluss eines Virus aber unwahrscheinlich, so dass andere Gründe verstärkt erwogen und abgeklärt werden müssen.

Ein besonders hoher Anteil an HPV-positiven Tumoren bei den jungen Patienten mit Karzinomen außerhalb der Mundhöhle könnte die Ergebnisse natürlich zum Teil erklären, da für diese eine bessere Prognose und eine hohe Radiosensibilität beschrieben sind (Ang et al., 2010; Fakhry et al., 2008; Lindel et al., 2001). Leider ist der HPV-Status der Patienten nicht erfasst worden, so dass sich ein nennenswerter Einfluss von HPV nur indirekt ausschließen lässt. So ergaben sich für Tumore außerhalb der Mundhöhle keine Unterschiede zwischen solchen mit bekannter HPV-Assoziation (Oropharynx inkl. Zungengrund und Tonsillen) und solche mit fehlender bzw. geringerer HPV-Prävalenz. Beide zeigten aber ein signifikant besseres Überleben als Mundhöhlenkarzinome. Für zukünftige Studien wäre der Nachweis einer HPV-Infektion wünschenswert, da eine indirekte Analyse Annahmen über die HPV-Inzidenz machen muss, die auf epidemiologischen Daten basieren (Ang et al., 2010; Chaturvedi et al., 2011; Kreimer, 2005) und auf die untersuchte Kohorte nicht zutreffen müssen.

Zusätzlich gibt es einige weitere patientenbezogene Variablen, die für diese Kohorte nicht erhoben worden waren, aber in Folgestudien eingeschlossen werden sollten, um eine genauere Risikoeinschätzung zu ermöglichen. Ein negativer Einfluss auf das Überleben ist z.B. beschrieben für das Vorliegen von Komorbiditäten (unter denen aufgrund des Risikoprofils von Kopf-Hals-Tumorpatienten vor allem solche des kardiovaskulären Systems zahlreich sind (Piccirillo, 2000)), einen niedrigen sozioökonomischen Status (Döbrossy, 2005), unverheirateten Familienstand und afrikanische Herkunft (Rose et al., 2011).

Zahlreiche weitere Gründe für das verminderte Überleben der Patienten mit Mundhöhlenkarzinomen lassen sich vermuten und sollten in der Zukunft weiter abgeklärt werden. So ist für Mundhöhlenkarzinome in der Vergangenheit ein schlechteres Ansprechen auf eine RCT ange-

nommen worden (Brizel, 2006), wobei neuere Studien eine adäquate Wirksamkeit feststellen (Stenson et al., 2009). Zudem ist denkbar, dass zum Ziel der Organerhaltung besonders bei jungen Patienten und bei solchen mit Mundhöhlenkarzinomen eine weniger radikale chirurgische Therapie durchgeführt wird und ihnen mit einer RCT z.B. die Einschränkungen einer totalen Glossektomie erspart bzw. die Resektionssäume gering gehalten werden sollen. Auf molekularer Ebene sind u.a. Amplifikationen des EGFR-Gens bzw. Überexpression des EGFR-Proteins mit verringertem Überleben in kleinen Kohorten von Patienten mit Zungentumoren assoziiert worden (Huang et al., 2012; Nakata et al., 2011) und in einer Studie zur Wirksamkeit des EGFR-Inhibitors Cetuximab verlängerte sich das mediane Überleben bei Patienten mit Mundhöhlenkarzinomen signifikant (Vermorken et al., 2008). Verkürzte Überlebensintervalle finden sich darüber hinaus auch bei Überexpression von p53 (Tandon et al., 2010). Genetische Untersuchung dieser Art haben eine große praktische Bedeutung, da sie Ansätze für zielgerichtete Therapien bilden und Patientengruppen identifizieren können, die von neuen Therapien profitieren werden.

Zusammenfassend bleibt die Erkenntnis, dass es sich bei fortgeschrittenen Mundhöhlenkarzinomen junger Patienten um eine eigenständige Tumorentität zu handeln scheint, die aus bisher ungeklärten Gründen mit einer schlechteren Prognose einhergeht. Ausgehend von dieser Beobachtung sollte eine genauere epidemiologische, soziodemographische und besonders auch molekulare Charakterisierung dieser Patientenpopulation erfolgen, um die zugrunde liegende Tumorbio-logie verstehen und erfolgreichere Strategien für Prävention und Therapie entwickeln zu können.

4.2 Zum Virusalgorithmus

In den vergangenen Jahrzehnten ist aus bisher ungeklärter Ursache ein Anstieg der Inzidenz von Tumoren der mobilen Zunge beobachtet worden, der besonders bei jungen Patienten ohne klassische Risikofaktoren ausgeprägt ist und mit einer schlechteren Prognose einher geht. Da dieser Anstieg in etwa parallel zum Anstieg HPV-assoziiierter Tumore des Oropharynx verlief, Tumore der mobilen Zunge aber praktisch keine HPV-Infektionen aufweisen, ist die Beteiligung eines anderen, insbesondere auch eines bisher unbekanntes Virus als ätiologischer Faktor vermutet worden (Rosenquist et al., 2005; Salem, 2010). Diese Hypothese wurde in der vorliegenden Arbeit überprüft, indem sieben Gewebeproben aus Tumoren der mobilen Zunge junger Patienten und drei Positivkontrollen mittels RNA-Seq sequenziert und anschließend mit einem neu entwickelten digitalen Subtraktionsalgorithmus auf das Vorliegen viraler Transkripte von bekannten

und unbekanntem Viren untersucht wurden. Während in den Positivkontrollen die Identifikation der jeweiligen Viren eindeutig und problemlos gelang, waren in den Tumoren keine viralen Transkripte nachweisbar, was eine virale Ätiologie für diese Tumorentität unwahrscheinlich macht.

Zu beachten ist bei der Interpretation dieses Ergebnisses, dass die Kombination von RNA-Seq mit diesem Virusalgorithmus einige Stärken gegenüber anderen Verfahren hat und seine hohe Sensitivität und Spezifität bei der Untersuchung der Positivkontrollen und der Simulation eines unbekanntem Virus unter Beweis gestellt wurden. Es sind selbstverständlich aber auch Szenarien denkbar, die zu einem falsch negativen Befund führen könnten. Im Folgenden werden deshalb zunächst die Vorteile und anschließend die Grenzen bzw. potentielle Einschränkungen dieser Vorgehensweise besprochen.

4.2.1 Stärken des Virusalgorithmus

Einer der größten Vorteile der Verwendung von Next-Generation-Sequencing Technologie zur Identifikation viraler Nukleinsäuren ist ihre (relative) Voraussetzungslosigkeit, also die fehlende Notwendigkeit, a priori Annahmen über die vermuteten Sequenzen treffen zu müssen. Sie bedarf im Voraus keiner Auswahl einiger Virus-Kandidaten, wie sie z.B. zum Primerdesign für PCR-basierte Methoden üblich ist. Stattdessen ermöglicht der für diese Arbeit gewählte Ansatz die Suche nach allen bereits bekannten Viren durch einen Abgleich mit online verfügbaren Datenbanken. Da diese Datenbanken regelmäßige Updates mit neuen Sequenzen erhalten, wäre es möglich, diese nach einiger Zeit erneut herunter zu laden und für den Gebrauch mit dem Virusalgorithmus zu formatieren. Auf diese Weise könnte eine erweiterte und aktualisierte Analyse durchgeführt werden, ohne den ursprünglichen Datensatz verändern oder neu generieren zu müssen.

Die Suche nach bekannten viralen Sequenzen im Datensatz geschieht mit Hilfe von frei verfügbaren Programmen, die so in einen automatisierten Analyseablauf eingeflochten sind, dass sich auch auf einem Standardcomputer vertretbare Laufzeiten ergeben. Auf diese Weise entfällt die Notwendigkeit von Großrechner- oder Cloud-Computing-Kapazitäten, wie er für vergleichbare Ansätze Bedingung ist (Kostic et al., 2011). Der erste Analyseschritt ist dabei das Filtern von

Sequenzen, die zum Wirt gehören (sog. „digitale Subtraktion“) um virale Sequenzen anzureichern.

In allen drei Positivkontrollen (einer HPV-16 positiven Kopf-Hals-Tumorzelllinie, einem veröffentlichten RNA-Seq Datensatz eines HPV-16 positiven Zervixkarzinoms (Arron et al., 2011) und einem für das EBV-Protein LMP2A transgenen Mausmodells) wurde das entsprechende Virus zweifelsfrei nachgewiesen.

Durch die Auswahl der Positivkontrollen sollte neben der prinzipiellen Funktionsfähigkeit auch die Flexibilität des Algorithmus demonstriert werden. So zeigt sich, dass die digitale Subtraktion sowohl bei einem humanen als auch bei einem murinen Wirt funktioniert und dass das Vorhandensein eines einzelnen Gens (LMP2A) ausreicht um einen Virusnachweis zu führen. Zudem wurde deutlich, dass sich das Verfahren erfolgreich sowohl auf Daten einer Zelllinie, eines Lymphommodells in der Maus, aber auch auf solche eines „echten“ menschlichen Tumors anwenden lässt. Dazu kommt, dass die Daten des Zervixkarzinoms in einer anderen Institution mit einer anderen Probenvorbereitung und einer anderen Technik (als sog. Multiplex-Experiment) generiert wurden und zusätzliche Vorformatierungen benötigten. Der erfolgreiche HPV-16-Nachweis war von diesen Unterschieden allerdings völlig unabhängig.

Grundsätzlich ist der Algorithmus, unabhängig von der verwendeten Technik, auf alle NGS-Daten anwendbar solange es sich um kurze Reads (< 80-100bp lang) handelt. Es könnten also auch DNA-basierte Daten verwendet werden, die z.B. einer kompletten Genomsequenzierung („whole genome sequencing“ WGS) entstammen. Für die vorliegende Arbeit wurde allerdings die Technologie des „transcriptome sequencing“ in Form von RNA-Seq benutzt, da diese einige Vorteile aufweist. Zunächst ist fest zu halten, dass alle bekannten humanen Onkoviren eine Produktion viruspezifischer Proteine aufweisen (Poreba et al., 2011). Die Fokussierung auf virale mRNA-Transkripte ist in einigen Studien als essentiell gedeutet worden, um einen aktiven –und möglicherweise kausalen– Einfluss eines Virus auf einen Tumor vom bloßen Vorliegen einer viralen Infektion bzw. von inaktiver viraler DNA unterscheiden zu können (Arron et al., 2011; Braakhuis et al., 2004; Jung et al., 2010). Zudem bestehen 8% des menschlichen Genoms aus Sequenzen, die sich von integrierten Retroviren abgeleitet haben. Dazu gehört auch das „Murine leukemia virus“, welches in fast allen Proben identifiziert wurde und in dieser Untersuchung den häufigsten falsch positiven Befund darstellt. Die meisten dieser ehemals viralen Gene werden

aber nicht mehr als RNA exprimiert (Griffiths, 2001). Daraus folgt, dass die erwartete Rate an falsch positiven Ergebnissen bei Sequenzierung des gesamten Genoms deutlich höher wäre als bei Verwendung des Transkriptoms. Aktuelle Berechnungen gehen davon aus, dass für die Detektion eines infektiösen Agens per WGS die 10-fache Datenmenge im Vergleich zu RNA-Seq notwendig wäre, was sowohl die Kosten für die Sequenzierung als auch die Laufzeiten der Analyse stark nach oben treiben würde (Moore et al., 2011).

Als Nebeneffekt ließ sich per RNA-Seq auch das Expressionslevel von CDKN2A/p16 und damit seine Bedeutung als Surrogatmarker einer HPV-Infektion in Zungentumoren bestimmen.

In der Vergangenheit waren mit NGS-Experimenten auch Nachweise von unbekanntem Viren gelungen, die zwar eine genetische Verwandtschaft mit bereits bekannten Viren zeigten, aber die per Viruskultur, PCR und viralen Genchips nicht identifizierbar waren. Dazu gehörten das Merkel-Zell-Polyoma Virus in Merkelzell-Karzinomen (Feng et al., 2008) und Arenaviren in akut erkrankten Empfängern von Organtransplantationen (Palacios et al., 2008).

Da auch in Zungentumoren junger Patienten die Beteiligung eines bisher unbekanntem Virus vermutet wurde, sollte der Algorithmus in der Lage sein, unbekanntem Viren zu identifizieren, sofern diese eine genetische Ähnlichkeit mit bereits bekannten Viren haben. Dies wurde ermöglicht, indem die subtrahierten Reads digital zu längeren Abschnitten (Contigs) zusammengesetzt wurden. Auf diese Weise ließ sich eine weitere deutliche Reduktion der Datenmenge erreichen. Dies erlaubte die Verwendung eines fehlertoleranteren Alignment-Programms wie BLAST auch auf einem Standardcomputer.

Das Funktionieren dieser Methode wurde demonstriert, indem die HPV-16 und EBV-Sequenzen aus den Referenzdatenbanken gelöscht wurden. Dennoch wurden verschiedene andere Herpesviren und HPV-Typen als potentielle Viruskandidaten identifiziert. Da die verschiedenen HPV-Typen und -Spezies untereinander definitionsgemäß 60-70% Sequenzähnlichkeit haben (de Villiers et al., 2004), lässt dies einen Schluss auf die Sensitivität des Algorithmus zur Detektion abweichender Virusgenome zu. Zu beachten ist, dass die Stärke der Konservierung nicht für alle HPV-Gene konstant sein muss (Narechania et al., 2005). Eine genauere Quantifizierung des benötigten Verwandtschaftsgrades der Sequenzen könnte erreicht werden, indem in einer bekannten Virussequenz systematisch ein Anteil der Nukleotide (z.B. je 10% , 20% , 30% usw.) verändert

und der Algorithmus jedes Mal durchlaufen würde. Auf diese Weise ließe sich eingrenzen, welcher Grad an Übereinstimmung zwischen zwei Sequenzen für eine erfolgreiche Detektion mindestens notwendig ist. Die Identifikation eines alternativen Herpesvirus machte deutlich, dass auch eine Homologie in einem einzelnen Gen ausreichend ist, um ein verwandtes Virus zu finden.

Ausgehend von den Ergebnissen der drei Positivkontrollen sind Bewertungskriterien entwickelt worden, die die Unterscheidung zwischen falsch-positiven und richtig-positiven Ergebnissen erleichtern und systematisieren sollten. Als geeignetes Kriterium stellte sich für die Suche nach bekannten Viren der Anteil registrierter viraler Transkripte an der Gesamtzahl aller Reads heraus, wobei von den viralen Reads fast alle von einem einzelnen Virus stammen sollten, um falsch positive Ergebnisse abzugrenzen. Die zweite Bedingung geht dabei von der Annahme aus, dass eine simultane Infektion mit zwei Viren biologisch unwahrscheinlich ist. Sonst würden sich die viralen Reads eher zu je ca. 50% auf beide Viren aufteilen, die zweite Bedingung nicht erfüllt und ein falsch-negatives Ergebnis die Folge sein.

Bei der Suche nach „unbekannten“ Viren war eine Voraussetzung, dass das entsprechende Contig seine beste Übereinstimmung mit einer viralen Sequenz zeigt, da sich durch Homologien zwischen humanen und viralen ORFs (s.o.) zahlreiche Contigs finden, die sowohl mit humanen als auch mit viralen Sequenzen übereinstimmen. Zudem wurde ein maximaler Wert für die Wahrscheinlichkeit einer zufälligen Übereinstimmung (der sog. E-Wert, der in seiner Bedeutung einem p-Wert entspricht) definiert, der unterschritten werden muss.

Da die jeweiligen Grenzwerte empirisch anhand der Ergebnisse der Positivkontrollen ermittelt wurden, ist ihre konkrete Festlegung kritisch zu betrachten. Die Beurteilungen der Analysen wurden mehrfach durchgeführt, wobei die Grenzwerte in weitem Rahmen variiert wurden. Dabei zeigten sich keine bedeutsamen Unterschiede der Ergebnisse, insbesondere auch nicht in den Tumorproben, so dass Werte gewählt wurden, die die Zahl an falsch-negativen Ergebnissen minimierten, aber die Rate falsch-positiver Resultate begrenzen. Dennoch wäre eine größere Eichstichprobe, welche auch mehr Proben aus Tumorgewebe einschließen sollte, sicherlich wünschenswert, um ihre Korrektheit zu bestätigen. Zusätzlich wurden alle Ergebnisse auch manuell durchgesehen, um Fehler der automatisierten Beurteilung (bes. in den Tumorproben) auszuschließen.

Während in allen drei Positivkontrollen eindeutig die erwarteten bekannten und auch simulierte „unbekannte“ Viren identifiziert wurden, ließen sich in den sieben Tumoren der mobilen Zunge von jungen Patienten keine belastbaren Belege für bekannte oder unbekannte Viren finden. Nach automatisierter Anwendung der Bewertungskriterien wurden zwar in drei Tumorproben mögliche Viruskandidaten angezeigt, bei zwei davon handelte es sich allerdings lediglich um humane Antikörper gegen Rotaviren. Die Sequenz der dritten Probe zeigte bei einem web-basierten BLAST-Vergleich mit einer Datenbank aller bekannter Nukleotidsequenzen die höchste Übereinstimmungen mit verschiedenen Klonierungsvektoren, deren Sequenzen in den Datenbanken des Virusalgorithmus nicht vorhanden waren. Diese Übereinstimmungen waren signifikant besser als die mit dem berichteten Virus. Deshalb wurden sie als falsch-positive Resultate gewertet, die am ehesten in einer molekularbiologischen Kontamination von Material, Proben oder Sequenziereinheit ihren Ursprung haben dürften. Dies macht zwei Dinge deutlich: erstens bedarf die Interpretation der Ergebnisse auch nach automatisierter Auswahl möglicher Kandidaten einer manuellen Überprüfung auf Plausibilität, zweitens wird aber auch erfolgreich eine enorme Eingrenzung möglicher Kandidaten vorgenommen, die eine manuelle Überprüfung erst zulässt.

Insgesamt ließen sich also in den Tumoren der mobilen Zunge keine Viruskandidaten finden. Da die Verbindung aus RNA-Seq und digitaler Subtraktion in anderen Studien erfolgreich war und auch der hier entwickelte Virusalgorithmus erfolgreich bekannte und „unbekannte“ Viren in den Positivkontrollen identifizierte, macht dies eine Genese von Tumoren der mobilen Zunge junger Patienten durch transkriptionell aktive Viren unwahrscheinlich.

4.2.2 Grenzen des Virusalgorithmus

Da der Virusalgorithmus eine hohe Sensitivität und Spezifität in den Kontrollproben aufwies, erscheint die tatsächliche Abwesenheit eines Virus in den Tumorproben als wahrscheinlichstes und auch korrektes Ergebnis. Dennoch sind auch einige Szenarien denkbar, die zu diesem negativen Ergebnis führen könnten, obwohl ein Virus verantwortlich bzw. vorhanden ist.

Als potentieller Grund ist wohl zunächst die recht geringe Zahl von sieben Tumorproben zu nennen. Diese eröffnet natürlich die Möglichkeit, dass zufällig nur virusnegative Beispiele gewählt wurden, insbesondere falls ein Virus nur eine geringe Inzidenz aufweisen sollte. Es wurde bei der Auswahl allerdings darauf geachtet, dass diese Tumorproben möglichst genau dem Risikoprofil

der Kohorte junger Patienten mit Tumoren der mobilen Zunge entsprachen, wie es sich aus den klinischen Daten ergab. Es handelt sich also um junge Nichtraucher mit geringem Alkoholkonsum, da besonders für diese Patientengruppe eine virale Ätiologie vermutet wurde.

Die Verwendung einer aufwändigen, teuren und damals recht wenig etablierten Technologie wie RNA-Seq verlangte zwar die Beschränkung auf wenige, für das Kollektiv aber repräsentative Patienten. Im Gegenzug ermöglichte es aber eine wesentliche tiefere und breitere Suche nach viralem Material, da der Algorithmus prinzipiell alle bekannten und auch noch unbekannte Viren identifizieren kann, während herkömmliche Methoden mehr Patienten auf weniger Viren testen. Dennoch ist eine Validierung der Ergebnisse anhand einer größeren Zahl an Zungenkarzinomen sicherlich anzustreben. Durch die Flexibilität des Virusalgorithmus ergäbe sich zudem die Möglichkeit, öffentlich zugängliche NGS-Daten von Kopf-Hals-Tumoren zu untersuchen, die in der Zukunft z.B. im Rahmen des „The Cancer Genome Atlas“-Projekts generiert werden.

Verschiedene nukleinsäurebasierte Verfahren zur Identifikation von Viren sind weit verbreitet und werden erfolgreich eingesetzt (Lipkin, 2010; Sarid und Gao, 2011; Weber et al., 2002). Dennoch haben sie einige prinzipielle Limitationen, die auch für die vorliegende Untersuchung gelten. Zunächst einmal setzen sie natürlich das Vorliegen von Nukleinsäuren voraus, so dass DNA- und RNA-freie infektiöse Agentien, wie z.B. Prionen, nicht erfasst würden. Da Prionen allerdings bisher nur für sporadisch auftretende neurodegenerative Erkrankungen verantwortlich scheinen (Kaski und Mead, 2009), sollte dies keine große Rolle spielen.

Zusätzlich sind die meisten dieser Verfahren darauf angewiesen, dass sich die Virus- und Wirtssequenzen ausreichend unterscheiden, um sie differenzieren zu können und um z.B. PCR-Primer zu entwerfen. Im Virusalgorithmus würde eine zu große Ähnlichkeit humaner und viraler Sequenzen dazu führen, dass sie im Subtraktionsschritt als „human“ verworfen würden. Da jedoch alle bekannten humanen Onkoviren virusspezifische Proteine exprimieren, die für Struktur und Funktion notwendig sind (Poreba et al., 2011), sollte dies eher von untergeordneter Bedeutung sein. Eine zu geringe Verwandtschaft eines unbekanntes Virus mit bekannten Viren wiederum würde dazu führen, dass der BLAST-Teil des Algorithmus diese nicht identifizieren kann. Aus den Ergebnissen der Simulation „unbekannter“ Viren lässt sich ableiten, dass im Idealfall Sequenzähnlichkeiten bis zu 60 – 70% und auch kurze Übereinstimmungen des Genoms zur Detektion ausreichen könnten (s.o.). Aufgrund der hohen Sequenzkonservierung innerhalb viraler Familien in Verbindung mit der hohen Zahl bereits bekannter und sequenzierter Viren (de Vil-

liers et al., 2004; Woolhouse et al., 2008) und der erfolgreichen Anwendung ähnlicher Methoden zur Virusdetektion (Feng et al., 2007; Palacios et al., 2008) stellt dies deshalb ein geringes, aber dennoch vorhandenes Risiko dar.

Während die Verwendung von RNA-Seq-Technologie zur Untersuchung der Tumorproben ausgewählt wurde, da sie einige erhebliche Vorteile gegenüber z.B. kompletter Genomsequenzierung bietet, hat sie auch einige inhärente Einschränkungen. So ist sie auf die Produktion von spezifischen viralen RNA-Transkripten angewiesen. Diese ist zwar für die bekannten Onkoviren beschrieben (Martin und Gutkind, 2009), muss aber nicht notwendigerweise für noch unbekannte Viren gelten. Da als erster Schritt in der Probenvorbereitung eine Anreicherung von mRNA über eine Selektion mit an Säulen gebundenen oligo-dT-Sequenzen erfolgt, ist eine Polyadenylierung der RNA-Moleküle Voraussetzung. Ein derartiger poly-A-Schwanz innerhalb der untranslatierten Region am 3'-Ende der RNA (3'UTR) ist bei fast allen Viren zu finden, da sie zur erfolgreichen Translationsinitiation notwendig ist. Es wurden aber auch einzelne Ausnahmen innerhalb der Familie der Flaviviridae beschrieben (Polacek et al., 2009), die demnach nicht erfasst würden.

Klassischerweise wird angenommen, dass die kontinuierliche Expression viraler Onkogene notwendig ist, um den transformierten Phänotyp der Zellen aufrecht zu erhalten (Nevels et al., 2001). Ein Vorteil der Verwendung von RNA-Seq ist deshalb die Abgrenzung eines aktiven viralen Einflusses auf die Tumorerhaltung vom bloßen Vorliegen viraler DNA, die sich in verschiedenen Studien als vorteilhaft erwiesen hat (s.o.). Auf der anderen Seite würde mit dem hier entwickelten Verfahren allerdings die virale Onkogenese von Tumoren einem Nachweis entgehen, deren Transformation bzw. Tumoringenese durch ein Virus verursacht wurde, das anschließend aber nicht mehr aktiv ist. Als möglicher Weg wurde beispielsweise erstmals in den 1970er Jahren ein sog. „hit-and-run“-Mechanismus vorgeschlagen (Skinner, 1976). Dabei wird vermutet, dass ein Virus mutagen wirken könnte, indem es z.B. die Proliferation steigert, DNA-Reperaturmechanismen inhibiert oder für genetische Instabilität sorgt. Auf diese Weise könnten ausreichend somatische Mutationen angesammelt werden („hit“) die einen weiteren Einfluss des Virus auf den Tumor unnötig machen, woraufhin das virale Genom verloren gehen könnte („run“). Manche Autoren vermuten sogar einen Vorteil für die klonale Expansion der Tumorzellen durch den Verlust der viralen Sequenzen, da so eine Immunevasion der Tumorzellen erleichtert würde (Stevenson et al., 2010). Obwohl für verschiedene Viren derartige Möglichkeiten er-

wogen werden (u.a. Adenoviren und Gammaherpesviren, (Nevels et al., 2001; Stevenson et al., 2010)), gibt es bisher allerdings nur Hinweise für eine solche Form viraler Transformation unter experimentellen Bedingungen in Zellkultur- und Tiermodellssystemen. Ob sie einen tatsächlichen Einfluss auf die Bildung humaner Tumore haben könnten, ist nicht bewiesen und die praktische Relevanz dadurch höchst unklar (Niller et al., 2011). Aus diesen Gründen ist nicht von einer großen Bedeutung eines „hit-and-run“-Mechanismus in der Genese von Tumoren der mobilen Zunge auszugehen.

Ein ähnliches Konzept ist das der „Onkomodulation“, wie es für das humane Zytomegalievirus (CMV) vorgeschlagen wurde. Dabei wird davon ausgegangen, dass ein Virus nicht direkt die Initiation des Tumors bewirkt, sondern einer bereits veränderten Zelle zu einem maligneren Phänotyp verhilft, indem es beispielsweise die Proliferation weiter steigert, die Immunevasion der Tumorzellen verbessert oder die Angiogenese anregt (Michaelis et al., 2009). Es wird vermutet, dass diese Effekte wenigstens z.T. durch nicht-kodierende RNAs (z.B. miRNA) vermittelt werden (Söderberg-Nauclér, 2006). Allerdings ist auch die Validität dieses Konzepts in Menschen bisher nicht bewiesen und Studien an Medulloblastomen und in kolorektalen Tumoren zeigten neben einer Assoziation von CMV mit diesen Tumoren auch jeweils die Expression von CMV-Proteinen (Baryawno et al., 2011; Harkins et al., 2002), welche per RNA-Seq detektierbar sein sollten.

Deutlich größere praktische Relevanz hat die Unfähigkeit des Algorithmus, mögliche nicht-virale mikrobielle Einflussfaktoren wie z.B. Bakterien und Pilzinfektionen zu entdecken, da derartige Sequenzen im Subtraktionsschritt gefiltert wurden, um die Zahl falsch-positiver Ergebnisse zu vermindern. Im Gegensatz zu ansonsten weitgehend sterilen Tumorproben (z.B. aus Mammakarzinomen, Gliomen etc.) enthalten solche der mobilen Zunge erhebliche Kontaminationen mit Mikroorganismen der oralen Flora, die mehr als 600 verschiedene bakterielle Spezies umfasst (Dewhirst et al., 2010). Zwischen zahlreichen viralen Genen und bakteriellen Genomen bzw. Plasmiden wiederum bestehen Sequenzähnlichkeiten. Dies führte dazu, dass der weniger stringente Teil des Virusalgorithmus bei der Suche nach unbekanntem Viren zahlreiche Viruskandidaten meldete, welche sich bei manueller BLAST-Überprüfung als Genprodukte bakteriellen Ursprungs heraus stellten.

Eine Lösungsmöglichkeit wäre gewesen, bakterielle Sequenzen in die Referenzdatenbank des „Discovery-Algorithmus“ aufzunehmen. Dies hätte aufgrund der enormen zusätzlich zu testenden

Datenmenge allerdings die Laufzeit deutlich verlängert. Stattdessen wurden im wesentlich effizienter ablaufenden Subtraktionsschritt nicht-virale mikrobielle Sequenzen gefiltert. Dies ergibt natürlich zumindest theoretisch das Risiko, Sequenzen eines Virus mit einer Homologie zu bakteriellen Sequenzen fälschlicherweise auszusortieren. Da die Subtraktion recht stringent vorgeht und > 95% Übereinstimmung eines Reads mit der Referenz voraussetzt, Virusgene aber durch eine hohe Mutationsrate recht variabel sind (Elena und Sanjuan, 2005), sollte dieses Risiko in der Praxis vernachlässigbar sein. In den Positivkontrollen veränderte sich die Zahl der Virustreffer durch diese zusätzliche Subtraktion nicht.

Da die gefilterten mikrobiellen Sequenzen nicht verworfen, sondern gespeichert werden, wäre die Suche nach einem verantwortlichen Bakterium theoretisch möglich. Da aber, im Gegensatz zu Viren, zahlreiche Bakterien sequenziert und identifiziert wurden, ließe sich allenfalls eine statistische Assoziation eines Bakteriums mit Zungentumoren herstellen, wenn mit der Flora gesunder Patienten verglichen würde. Entsprechende Daten liegen nicht vor. Zudem ist die orale mikrobielle Flora durch äußere Einflüsse leicht störbar und könnte in Tumorpatienten auch durch eine Chemotherapie etc. verändert sein, so dass Kausalzusammenhänge sehr komplex zu beweisen wären. Aus diesen Gründen unterblieb die Analyse bakterieller Einflüsse.

4.2.3 Sensitivität des Verfahrens

Ein weiterer Gesichtspunkt in der Interpretation der Ergebnisse ist, dass für die Identifikation eines Virus eine ausreichend Zahl sequenzierter Reads vorliegen muss. Dies liegt daran, dass die absolute Zahl an mRNA-Molekülen einer Zelle sich nicht gleichmäßig auf die Transkripte verschiedener Gene verteilt. Stattdessen machen einige wenige Gene (darunter klassische „house-keeping genes“ wie β -Actin) einen Großteil der mRNA-Moleküle aus. Wird insgesamt nur eine geringe Menge sequenziert, werden überwiegend diese Gene erfasst. Mit steigender Datenmenge steigt auch die Anzahl an identifizierten Transkripten, die in geringer Konzentration vorliegen (Mortazavi et al., 2008). Schätzungen gehen davon aus, dass ab etwa 4 Millionen Reads ca. 80% aller Transkripte einer Zelle erfasst werden. Mit weiter steigender Datenmenge werden auch die restlichen 20% sequenziert, die Kurve flacht aber deutlich ab (Z Wang et al., 2009).

Eine genaue Quantifizierung der Anzahl an benötigten Reads zur Virusidentifikation hängt dementsprechend von vielen Faktoren (u.a. auch von der transkriptionellen Aktivität des Virus) ab

und kann für die vorliegenden Proben nur geschätzt werden. Eine systematische Analyse der Sensitivität könnte z.B. mit „spike-in Experimenten“ erfolgen, in denen durch die Zugabe viraler RNA in verschiedenen bekannten Konzentrationen zu virusfreien Proben mit anschließender Sequenzierung und Analyse die untere Nachweisgrenze bestimmt würde. Dies wäre wünschenswert, ist bisher aber nicht erfolgt.

Einen Anhalt für die Sensitivität bietet die Virusidentifikation in den Positivkontrollen, die problemlos möglich war, obwohl für das Zervixkarzinom weniger als 4 Millionen Reads sequenziert worden waren. Aus den Tumorproben wurde jeweils das 3- bis 4-fache an Reads gewonnen, so dass eine Detektion bei Vorliegen eines Virus wahrscheinlich wäre. RNA-Seq hat zudem in anderen Studien eine hohe Sensitivität für die Identifikation von Viren gezeigt, wo wenige virale mRNA-Moleküle pro Zelle ausreichen (Moore et al., 2011).

4.3 Zur Expression von CDKN2A/p16^{INK4a}

Da RNA-Seq die Möglichkeit bietet, die Transkriptionslevel von mRNA-Molekülen direkt zu messen, wurde auch die Expression des Proteins CDKN2A/p16^{INK4a} (im Folgenden p16 genannt) bestimmt. Die Proteinexpression dieses zellulären Gens, welches eine wichtige Rolle in der Zellzykluskontrolle spielt, wird durch HPV gesteigert und bei Oropharynx Tumoren deshalb als etablierter Surrogatparameter einer HPV-Infektion angesehen (Jordan et al., 2012; Robinson et al., 2010). In Tonsillenkarzinomen haben manche Autoren p16-Immunhistochemie sogar als alleinigen Marker für eine HPV-Infektion vorgeschlagen, da es sich um eine kostengünstige und weit verbreitete Methode handelt (Park et al., 2012). Für Mundhöhlenkarzinome bzw. für Tumore der mobilen Zunge haben in der letzten Zeit verschiedenen Studien allerdings eine begrenzte Aussagekraft und eine hohe falsch-positive Rate, also eine hohe p16-Expression bei negativem HPV-Status, festgestellt (Kabeya et al., 2011; Lingen et al., 2012).

Die Ergebnisse der vorliegenden Arbeit deuten in eine ähnliche Richtung. So wurden in 3 der 7 (43%) der Tumorproben hohe Expressionslevel von CDKN2A/p16 gemessen, während die übrigen vier Proben deutlich niedrigere Level aufwiesen. Dabei war sowohl mit dem Virusalgorithmus als auch mit einer davon unabhängigen HPV-PCR ein Ausschluss einer HPV-Infektion erfolgt. Die hohe p16-Expression ist hier folglich nicht durch eine HPV-Infektion erklärbar und bestätigt demnach die hohe Rate falsch-positiver Ergebnisse anderer Studien bei Mundhöhlen-

karzinomen. Dies spricht dafür, dass die Expression von p16 als alleiniger Marker nicht zuverlässig ist.

Es ist natürlich zu beachten, dass hier die Expression von p16 anhand des Vorkommens von mRNA-Molekülen ausgewertet wurde, während in den meisten anderen Untersuchungen Proteinexpression mittels Immunhistochemie nachgewiesen wird. Grundsätzlich zeigt sich eine gute Korrelation zwischen der Höhe der Level an mRNA- und der Proteine einer Zelle (Gry et al., 2009) und mRNA-basierte Techniken sind in Form von Microarray-Experimenten und quantitativer „Reverser Transkriptase“ (RT)-PCR-Techniken weit verbreitet. Dennoch bildet mRNA die Konzentration der translatierten Proteine längst nicht perfekt ab. Dementsprechend haben im Bezug auf p16 verschiedene Studien eine gute Korrelation zwischen der Expression von p16-mRNA und der p16-Proteinexpression nachgewiesen (Kim et al., 2005; Omura-Minamisawa et al., 2001), andere dagegen fanden nur einen schwachen Zusammenhang (Quentin et al., 2004). Als mögliche Ursachen kommen beispielsweise posttranslationale Modifikationen in Betracht, welche die Degradation beschleunigen (Li et al., 2011). Außerdem wurde die Wirkung bestimmter microRNAs auf die p16-mRNA beschrieben, die z.B. die Translation bremsen (Lal et al., 2008). Aus diesen Gründen wäre die Bestätigung einer ebenfalls erhöhten Expression des p16-Proteins per Immunhistochemie oder Westernblot in den drei Tumorproben mit hoher p16-mRNA-Expression empfehlenswert.

Eine weitere wichtige Frage wäre zudem, ob es auch falsch-negative Ergebnisse dieses Surrogatparameters gibt – also HPV-positive Proben mit fehlender p16-Expression. Ansonsten wäre eine Verbindung von p16-IHC mit folgender HPV-PCR bei p16-positiven Patienten, wie sie von manchen Autoren vorgeschlagen wird (Smeets et al., 2007), ein zielführendes Verfahren, um Patienten mit transkriptionell aktiven Viren zu identifizieren. Insgesamt muss die Wertigkeit als HPV-Surrogatparameter zudem in einem erweiterten Patientenkollektiv geklärt werden, das aufgrund seiner Größe verlässliche statistische Aussagen in Bezug auf die Gesamtpopulation ermöglicht.

Für Oropharynxkarzinomen ist bei Patienten mit vorhandener p16-Expression ein besseres Ansprechen auf Radiotherapie und eine bessere Prognose auch bei HPV-negativen Tumoren beobachtet worden (Kumar et al., 2008; Witkiewicz et al., 2011). Die Aussagekraft bei Mundhöhlenkarzinomen ist unklar. Bei einem so hohen Anteil HPV-negativer, aber p16-positiver Patienten,

wie es die vorliegende (und andere) Arbeiten vermuten lassen, sollte dieser Aspekt weiter untersucht werden, da sich daraus ggf. bedeutende Konsequenzen für Risikostratifizierung und Therapieentscheidung ergeben könnten.

4.4 Schlussfolgerungen und Ausblick

Zusammenfassend lassen sich aus den Ergebnissen der vorliegenden Arbeit verschiedene Schlüsse ziehen, aber auch zahlreiche Ansätze für weiterführende Untersuchungen entwickeln, von denen die meisten bereits im Rahmen der Diskussion genannt wurden.

Anhand der retrospektiven Auswertung klinischer Daten von 748 Patienten mit fortgeschrittenen Kopf-Hals-Tumoren ließ sich eine signifikant schlechtere Prognose für junge Patienten mit Mundhöhlenkarzinomen nachweisen, obwohl sie eine geringere Exposition gegenüber klassischen Risikofaktoren zeigten als andere Patienten. Da junge Patienten mit Mundhöhlenkarzinomen aber noch immer einen recht kleinen Anteil der Gesamtzahl aller Patienten ausmachen, sollte dieses Ergebnis, idealerweise prospektiv und unter Einbeziehung weiterer Risikofaktoren und molekularer Marker (z.B. des p16- und des HPV-Status) in einem größeren Kollektiv bestätigt werden.

Die Gründe sowohl für die steigende Inzidenz dieser Tumorentität und die schlechtere Prognose dieser Subgruppe ist unklar, eine virale Ätiopathogenese wurde vermutet. Mit Hilfe von RNA-Seq und einem neu entwickelten Analysealgorithmus, dessen Funktionsfähigkeit an verschiedenen Positivkontrollen demonstriert wurde, konnte die Abwesenheit eines transkriptionell-aktiven Virus in sieben Tumoren der oralen Zunge belegt werden. Dies macht eine virale Ätiologie dieser Tumorentität unwahrscheinlicher.

Die Suche nach ätiologischen Faktoren sollte sich in Zukunft deshalb vermehrt auf andere Gebiete fokussieren. Beispiele sind die Exposition der Mundhöhle gegenüber toxischen Stoffen oder mikrobiellen Faktoren, aber auch genetische Varianten wie z.B. Single nucleotide polymorphisms (SNPs), die für diese Krankheit prädisponieren könnten. Eine systematische Analyse der hier gewonnenen Sequenzdaten auf Mutationen und Veränderungen der Genexpression könnte zudem helfen, die genetischen Hintergründe zu entschlüsseln, die für die eingeschränkte Prognose dieser Patienten verantwortlich sind.

5. Zusammenfassung

Für Karzinome der Mundhöhle und der mobilen Zunge wurde in den letzten Jahren eine steigende Inzidenz verzeichnet. Besonders betroffen sind junge Patienten ohne die „klassischen“ Risikofaktoren Alkohol- und Tabakkonsum. Die Prognose dieser Patienten ist bislang aber nicht gut untersucht. Der Anstieg der Inzidenz verlief parallel zu dem der Oropharynxkarzinome, die ein ähnliches Risikoprofil aufweisen und mit einer Infektion durch humane Papillomaviren (HPV) assoziiert sind. Tumore der Mundhöhle und der mobilen Zunge sind aber mehrheitlich HPV negativ. Aus diesem Grund ist eine virale Ätiologie durch ein anderes, ggf. noch unbekanntes Virus vermutet worden.

Zunächst wurde in dieser Arbeit retrospektiv in einer Kohorte von 748 mit Radiochemotherapie behandelten Patienten mit fortgeschrittenen Kopf-Hals-Tumoren die Prognose in Abhängigkeit von Alter und anatomischer Lokalisation analysiert. Dabei zeigten junge Patienten mit Mundhöhlenkarzinomen ein signifikant schlechteres Überleben als die übrigen Patienten.

Als zweiter Punkt wurden sieben Gewebeproben aus Tumoren der mobilen Zunge von jungen, HPV-negativen Patienten ohne Alkohol- und Tabakkonsum auf eine virale Ätiologie untersucht. Dazu wurde die gesamte mRNA der Proben isoliert und per RNA-Seq sequenziert. Für die weitere Analyse musste ein neuartiger bioinformatischer Algorithmus entwickelt werden, der auf einem Standardcomputer die Suche nach allen bekannten, aber auch unbekanntem Viren ermöglichen sollte. Die Fähigkeiten des Algorithmus wurden anhand der Daten einer Tumorzelllinie, eines Zervixkarzinoms (beide HPV-16 positiv) und eines EBV-transgenen murinen Lymphommodells nachgewiesen. Die Detektion unbekannter Viren wurde erfolgreich simuliert. In den Tumorproben wurden keine viralen Transkripte identifiziert, was den Einfluss eines transkriptionell aktiven Virus in der Ätiologie dieser Tumorentität unwahrscheinlich macht.

Als dritter Punkt wurde in den RNA-Seq Daten die Expression des zellulären Gens CDKN2A/p16^{INK4a} quantifiziert. Es gilt bei Oropharynxkarzinomen als diagnostischer Surrogatmarker einer HPV-Infektion. Drei der sieben Tumorproben zeigten aber hohe p16-Level trotz negativem HPV-Status, was die in der Literatur berichtete hohe Rate an falsch-positiven Ergebnissen und die zweifelhafte Wertigkeit dieses Markers in Mundhöhlenkarzinomen bestätigt.

6. Literaturverzeichnis

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; 215: 403–410

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

Annertz K, Anderson H, Biörklund A, Möller T, Kantola S, Mork J, Olsen JH, Wennerberg J. Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults. *Int J Cancer* 2002; 101: 95–99

Ansorge WJ. Next-generation DNA sequencing techniques. *N Biotechnology* 2009; 25: 195–203

Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *The Lancet* 2008; 371: 1695–1709

Arron ST, Ruby JG, Dybbro E, Ganem D, Derisi JL. Transcriptome sequencing demonstrates that human papillomavirus is not active in cutaneous squamous cell carcinoma. *J Invest Dermatol* 2011; 131: 1745–1753

Attner P, Du J, Näsman A, Hammarstedt L, Ramqvist T, Lindholm J, Marklund L, Dalianis T, Munck-Wikland E. The role of human papillomavirus in the increased incidence of base of tongue cancer. *Int J Cancer* 2010; 126: 2879–2884

Baryawno N, Rahbar A, Wolmer-Solberg N, Taher C, Odeberg J, Darabi A, Khan Z, Sveinbjörnsson B, Fuskev-Åg OM, Segerström L, Nordenskjöld M, Siesjö P, Kogner P, Johnsen JI, Söderberg-Nauclér C. Detection of human cytomegalovirus in medulloblastomas reveals a potential therapeutic target. *J Clin Invest* 2011; 121: 4043–4055

Bell AI, Groves K, Kelly GL, Croom-Carter D, Hui E, Chan ATC, Rickinson AB. Analysis of Epstein-Barr virus latent gene expression in endemic Burkitt's lymphoma and nasopharyngeal carcinoma tumour cells by using quantitative real-time PCR assays. *J Gen Vir* 2006; 87: 2885–2890

Bell RB, Kademani D, Homer L, Dierks EJ, Potter BE. Tongue Cancer: Is There a Difference in Survival Compared With Other Subsites in the Oral Cavity? *J Maxillofac Oral Surg* 2007; 65: 229–236

Bello IO, Soini Y, Salo T. Prognostic evaluation of oral tongue cancer: means, markers and perspectives (I). *Oral Oncol* 2010; 46: 630–635

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2010; 39: D32–D37

Biegging KT, Amick AC, Longnecker R. Epstein-Barr virus LMP2A bypasses p53 inactivation in a MYC model of lymphomagenesis. *Proc Natl Acad Sci USA* 2009; 106: 17945–17950

Blankenberg D, Gordon A, Kuster Von G, Coraor N, Taylor J, Nekrutenko A. Manipulation of

FASTQ data with Galaxy. *Bioinformatics* 2010; 26: 1783–1785.

Blomberg M, Nielsen A, Munk C, Kjaer SK. Trends in head and neck cancer incidence in Denmark, 1978-2007: Focus on human papillomavirus associated sites. *Int J Cancer* 2010; 129: 733–741

Bosetti C, Garavello W, Gallus S, Vecchia CL. Effects of smoking cessation on the risk of laryngeal cancer: An overview of published studies. *Oral Oncol* 2006; 42: 866–872

Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, Ghissassi El F, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Cogliano V, WHO International Agency for Research on Cancer Monograph Working Group. A review of human carcinogens--Part B: biological agents. *Lancet Oncol.* 2009; 10: 321–322

Braakhuis BJM, Snijders PJF, Keune W-JH, Meijer CJLM, Ruijter-Schippers HJ, Leemans CR, Brakenhoff RH. Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. *J Natl Cancer Inst* 2004; 96: 998–1006

Bradford CR, Zhu S, Ogawa H, Ogawa T, Ubell M, Narayan A, Johnson G, Wolf GT, Fisher SG, Carey TE. P53 mutation correlates with cisplatin sensitivity in head and neck squamous cell carcinoma lines. *Head Neck* 2003; 25: 654–661

Bray F, Ren J-S, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* 2012

Brizel DM. Concurrent Chemoradiotherapy for Locally Advanced, Nonmetastatic, Squamous Carcinoma of the Head and Neck: Consensus, Controversy, and Conundrum. *J Clin Oncol* 2006; 24: 2612–2617

Brown LM, Check DP, Devesa SS. Oral Cavity and Pharynx Cancer Incidence Trends by Subsite in the United States: Changing Gender Patterns. *J Onc* 2012; 2012: 1–10

Carvalho AL, Nishimoto IN, Califano JA, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: A site-specific analysis of the SEER database. *Int J Cancer* 2004; 114: 806–816

Castellsagué X, Quintana MJ, Martínez MC, Nieto A, Sánchez MJ, Juan A, Monner A, Carrera M, Agudo A, Quer M, Muñoz N, Herrero R, Franceschi S, Bosch FX. The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis. *Int J Cancer* 2004; 108: 741–749

Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, Wentzensen N, Jordan RC, Altekruze S, Anderson WF, Rosenberg PS, Gillison ML. Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States. *J Clin Oncol* 2011; 29: 4294–4301

Chitapanarux I, Lorvidhaya V, Sittitrai P, Pattarasakulchai T, Tharavichitkul E, Sriuthaisiriwong P, Kamnerdsupaphon P, Sukthomya V. Oral cavity cancers at a young age: analysis of patient, tumor and treatment characteristics in Chiang Mai University Hospital. *Oral Oncol* 2006; 42: 83–88

Chuang S-C, Jenab M, Heck JE, Bosetti C, Talamini R, Matsuo K, Castellsagué X, Franceschi S, Herrero R, Winn DM, La Vecchia C, Morgenstern H, Zhang Z-F, Levi F, Dal Maso L, Kelsey K, McClean MD, Vaughan T, Lazarus P, Muscat J, Ramroth H, Chen C, Schwartz SM, Eluf-Neto J, Hayes RB, Purdue M, Boccia S, Cadoni G, Zaridze D, Koifman S, Curado MP, Ahrens W, Benhamou S, Matos E, Lagiou P, Szeszenia-Dabrowska N, Olshan AF, Fernández L, Menezes A, Agudo A, Daudt AW, Merletti F, Macfarlane GJ, Kjaerheim K, Mates D, Holcatova I, Schantz S, Yu GP, Simonato L, Brenner H, Mueller H, Conway DI, Thomson P, Fabianova E, Znaor A, Rudnai P, Healy CM, Ferro G, Brennan P, Boffetta P, Hashibe M. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. *Cancer Causes Control* 2011; 23: 69–88

Cooper JS, Porter K, Mallin K, Hoffman HT, Weber RS, Ang KK, Weber RS, Ang KK, Gay EG, Langer CJ. National Cancer Database report on cancer of the head and neck: 10-Year update. *Head Neck* 2009; 31: 748–758

Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol* 1972: 187–220

Curtis RE, Rowlings PA, Deeg HJ, Shriner DA, Socie G, Travis LB, Socie G, Travis LB, Horowitz MM, Witherspoon RP, Hoover RN, Sobocinski KA, Fraumeni JF, Boice JD. Solid cancers after bone marrow transplantation. *N Engl J Med* 1997; 336: 897–904

D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, Westra WH, Gillison ML. Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* 2007; 356: 1944–1956

Dahlgren L, Dahlstrand HM, Lindquist D, Högmö A, Björnestål L, Lindholm J, Lundberg B, Dalianis T, Munck-Wikland E. Human papillomavirus is more common in base of tongue than in mobile tongue cancer and is a favorable prognostic factor in base of tongue cancer patients. *Int J Cancer* 2004; 112: 1015–1019

Dayyani F, Etzel CJ, Liu M, Ho CH, Lippman SM, Tsao AS. Research Meta-analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC). *Head Neck Oncol* 2010; 2:15

de Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, Plummer M. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol* 2012; 13: 607–615

de Villiers E-M, Fauquet C, Broker TR, Bernard H-U, Hausen zur H. Classification of papillomaviruses. *Virology* 2004; 324: 17–27

DeLacure MD. Lip and Oral Cavity. In: Close LG, Larson DE, Shah JP, Hrsg: *Essentials of Head and Neck Oncology*. New York: Thieme Medical Publishers, Inc; 1998

Depue RH. Rising mortality from cancer of the tongue in young white males. *N Engl J Med* 1986

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu WH, Lakshmanan A, Wade WG. The Human Oral Microbiome. *J Bacter* 2010; 192: 5002–5017

Doobaree IU, Landis SH, Linklater KM, El-Hariry I, Moller H, Tyczynski J. Head and neck cancer in South East England between 1995–1999 and 2000–2004: An estimation of incidence and distribution by site, stage and histological type. *Oral Oncol* 2009; 45: 809–814

Döbrossy L. Epidemiology of head and neck cancer: magnitude of the problem. *Cancer Metastasis Rev* 2005; 24: 9–17

Dwivedi RC, Chisholm C, Kanwar N, Komorowski A, Kazi R. Epidemiology, aetiology and natural history of head and neck cancer. In: Staffieri A, Sebastian P, Kapran M, Varghese BT, Kazi R, Hrsg: *Essentials of Head and Neck Cancer*. Delhi: Byword Books; 2012

El-Mofty SK, Lu DW. Prevalence of Human Papillomavirus Type 16 DNA in Squamous Cell Carcinoma of the Palatine Tonsil, and Not the Oral Cavity, in Young Patients: A Distinct Clinicopathologic and Molecular Disease Entity. *Am J Surg Pathol* 2003; 27: 1463

Elena SF, Sanjuan R. Adaptive Value of High Mutation Rates of RNA Viruses: Separating Causes from Consequences. *J Virol* 2005; 79: 11555–11558

Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, Forastiere A, Gillison ML. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst* 2008; 100: 261–269

Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008; 319: 1096–1100

Feng H, Taylor JL, Benos PV, Newton R, Waddell K, Lucas SB, Chang Y, Moore PS. Human transcriptome subtraction by using short sequence tags to search for tumor viruses in conjunctival carcinoma. *J Virol* 2007; 81: 11332–11340

Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; 127: 2893–2917

Franceschi S, Bidoli E, Herrero R, Muñoz N. Comparison of cancers of the oral cavity and pharynx worldwide: etiological clues. *Oral Oncol* 2000; 36: 106–115

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 2011; 39: D876–82

Garavello W, Spreafico R, Gaini RM. Oral tongue cancer in young patients: A matched analysis. *Oral Oncol* 2007; 43: 894–897

Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV, Sidransky D. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst* 2000; 92: 709–720

Gillison ML. Current topics in the epidemiology of oral cavity and oropharyngeal cancers. *Head*

Neck 2007; 29: 779–792

Griffiths DJ. Endogenous retroviruses in the human genome sequence. *Genome Biol* 2001; 2: 1017

Gry M, Rimini R, Strömberg S, Asplund A, Pontén F, Uhlén M, Nilsson P. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics* 2009; 10: 365

Guntinas-Lichius O, Wendt T, Buentzel J, Esser D, Lochner P, Mueller A, Schultze-Mosgau S, Altendorf-Hofmann A. Head and neck cancer in Germany: a site-specific analysis of survival of the Thuringian cancer registration database. *J Cancer Res Clin Oncol* 2009; 136: 55–63

Gupta A, Lee JH, Wilke W, Quon H, Smith G, Maity A, John Buatti J, Spitz DR. Radiation Response in Two HPV-Infected Head-and-Neck Cancer Cell Lines in Comparison to a Non-HPV-Infected Cell Line and Relationship to Signaling Through AKT. *Int J Radiat Oncol Biol Phys* 2009; 74: 928–933

Haddad RI, Shin DM. Recent advances in head and neck cancer. *N Engl J Med* 2008; 359: 1143–1154

Hammarstedt L, Dahlstrand H, Lindquist D, Onelöv dahlgren L, Ryott M, Luo J, Dalianis T, Ye W, Munck-Wikland E. The incidence of tonsillar cancer in Sweden is increasing. *Acta Otolaryngol* 2007; 127: 988–992

Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res* 2010; 38: e131

Hariri S, Unger ER, Sternberg M, Dunne EF, Swan D, Patel S, Markowitz LE. Prevalence of Genital Human Papillomavirus Among Females in the United States, the National Health and Nutrition Examination Survey, 2003–2006. *J Infect Dis* 2011; 204: 566–573

Harkins L, Volk AL, Samanta M, Mikolaenko I, Britt WJ, Bland KI, Cobbs CS. Specific localisation of human cytomegalovirus nucleic acids and proteins in human colorectal cancer. *The Lancet* 2002; 360: 1557–1563

Hasegawa W, Pond GR, Rifkind JT, Messner HA, Lau A, Daly AS, Kiss TL, Kotchetkova N, Galal A, Lipton JH. Long-term follow-up of secondary malignancies in adults after allogeneic bone marrow transplantation. *Bone Marrow Transplant* 2004; 35: 51–55

Herrero R, Castellsagué X, Pawlita M, Lissowska J, Kee F, Balaram P, Rajkumar T, Sridhar H, Rose B, Pintos J, Fernández L, Idris A, Sánchez MJ, Nieto A, Talamini R, Tavani A, Bosch FX, Reidel U, Snijders PJF, Meijer CJLM, Viscidi R, Muñoz N, Franceschi S. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst* 2003; 95: 1772–1783

Holzerlandt R, Orengo C, Kellam P, Albà MM. Identification of new herpesvirus gene homologs in the human genome. *Genome Res* 2002; 12: 1739–1748

Howard JD, Chung CH. Biology of Human Papillomavirus–Related Oropharyngeal Cancer. *Se-*

min Radiat Oncol 2012; 22: 187–193

Huang S-F, Cheng S-D, Chien H-T, Liao C-T, Chen I-H, Wang H-M, Chuang W-Y, Wang C-Y, Hsieh LL. Relationship between epidermal growth factor receptor gene copy number and protein expression in oral cavity squamous cell carcinoma. *Oral Oncol* 2012; 48: 67–72

IARC. GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. Available from: <http://globocan.iarc.fr>, Zugriff am 03/11/2012

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; 61: 69–90

Johnson NWW, Amarasinghe H. Epidemiology and Aetiology of Head and Neck Cancers. In: Bernier J, Hrsg: *Head and Neck Cancer: Multimodality Management*. New York: Springer; 2011

Jordan RC, Lingen MW, Perez-Ordóñez B, He X, Pickard R, Koluder, Jiang B, Wakely P, Xiao W, Gillison ML. Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. *Am J Surg Pathol* 2012; 36: 945–954

Jung AC, Briolat J, Millon R, de Reyniès A, Rickman D, Thomas E, Abecassis J, Clavel C, Wasylyk B. Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. *Int J Cancer* 2010; 126: 1882–1894

Kabeya M, Furuta R, Kawabata K, Takahashi S, Ishikawa Y. Prevalence of human papillomavirus in mobile tongue cancer with particular reference to young patients. *Cancer Sci.* 2011; 103: 161–168

Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* 2006; 24: 2137–2150

Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, Kober KM, Miller W, Pedersen JS, Pohl A, Raney BJ, Rhead B, Rosenbloom KR, Smith KE, Stanke M, Thakkapallayil A, Trumbower H, Wang T, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* 2007; 36: D773–D779.

Kaski D, Mead S. Prion diseases. *Medicine* 2009; 37: 579–581

Kelley DJ, Shah JP. Global Incidence and Etiology. In: Close LG, Larson DE, Shah JP, Hrsg: *Essentials of Head and Neck Oncology*. New York: Thieme Medical Publishers, Inc; 1998

Kerr JR, Boschetti N. Short regions of sequence identity between the genomes of human and rodent parvoviruses and their respective hosts occur within host genes for the cytoskeleton, cell adhesion and Wnt signalling. *J Gen Virol* 2006; 87: 3567–3575

Kim BN, Yamamoto H, Ikeda K, Damdinsuren B, Sugita Y, Ngan CY, Fujie Y, Ogawa M, Hata T, Ikeda M, Ohue M, Sekimoto M, Monden T, Matsuura N, Monden M. Methylation and ex-

pression of p16INK4 tumor suppressor gene in primary colorectal cancer tissues. *Int J Oncol* 2005; 26: 1217–1226

Koch WM, Nance M. Classification, Clinical Features, and Molecular Genetic Models. In: Olschan AF, Hrsg: *Epidemiology, Pathogenesis, and Prevention of Head and Neck Cancers*. New York: Springer; 2010

Kostic AD, Ojesina AI, Peadarallu CS, Jung J, Verhaak RGW, Getz G, Meyerson M. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* 2011; 29: 393–396

Kreimer AR. Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide: A Systematic Review. *Cancer Epidemiology Biomarkers & Prevention* 2005; 14: 467–475

Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, Wolf GT, Urba SG, Chepeha DB, Teknos TN, Eisbruch A, Tsien CI, Taylor JMG, D'Silva NJ, Yang K, Kurnit DM, Bauer JA, Bradford CR, Carey TE. EGFR, p16, HPV Titer, Bcl-xL and p53, Sex, and Smoking As Indicators of Response to Therapy and Survival in Oropharyngeal Cancer. *J Clin Oncol* 2008; 26: 3128–3137

Kuo K-T, Hsiao C-H, Lin C-H, Kuo L-T, Huang S-H, Lin M-C. The biomarkers of human papillomavirus infection in tonsillar squamous cell carcinoma—molecular basis and predicting favorable outcome. *Mod Pathol* 2008; 21: 376–386

Lacko M, Oude Ophuis MB, Peters WHM, Manni JJ. Genetic polymorphisms of smoking-related carcinogen detoxifying enzymes and head and neck cancer susceptibility. *Anticancer Res.* 2009; 29: 753–761

Lal A, Kim HH, Abdelmohsen K, Kuwano Y, Pullmann R, Srikantan S, Subrahmanyam R, Martindale JL, Yang X, Ahmed F, Navarro F, Dykxhoorn D, Lieberman J, Gorospe M. p16(INK4a) translation suppressed by miR-24. *PLoS ONE* 2008; 3: e1864

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson L, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier

P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Feder-spiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Ol-son MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, R Schultz R, Roe BA, Chen F, Pan H, Ramser, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nord-siek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wag-ner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guy-er MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; 10: R25

Leemans CR, Braakhuis BJM, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer* 2011; 11: 9–22

Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. *Nucleic Acids Res* 2010; 39: D19–D21

Li J, Poi MJ, Tsai M-D. Regulatory Mechanisms of Tumor Suppressor P16 INK4A and Their Relevance to Cancer. *Biochemistry* 2011; 50: 5566–5582

Liang C, Marsit CJ, McClean MD, Nelson HH, Christensen BC, Haddad RI, Clark JR, Wein RO, Grillone GA, Houseman EA, Halec G, Waterboer T, Pawlita M, Krane JF, Kelsey KT. Biomarkers of HPV in Head and Neck Squamous Cell Carcinoma. *Cancer Res* 2012; 72: 5004–5013

Liang X-H, Lewis J, Foote R, Smith D, Kademani D. Prevalence and significance of human papillomavirus in oral tongue cancer: the Mayo Clinic experience. *J Oral Maxillofac Surg* 2008; 66: 1875–1880

Liao C-T, Wang H-M, Hsieh L-L, Chang JT-C, Ng S-H, Hsueh C, Lee L-Y, Lin CH, Chen I-H, Kang C-J, Huang S-F, Yen T-C. Higher distant failure in young age tongue cancer patients. *Oral Onc* 2006; 42: 718–725

Lindel K, Beer KT, Laissue J, Greiner RH, Aebersold DM. Human papillomavirus positive squamous cell carcinoma of the oropharynx: a radiosensitive subgroup of head and neck carcinoma. *Cancer* 2001; 92: 805–813

Lingen MW, Chang KW, McMurray SJ, Solt DB, Kies MS, Mittal BB, Haines GK, Pelzer PJ. Overexpression of p53 in squamous cell carcinoma of the tongue in young patients with no known risk factors is not associated with mutations in exons 5-9. *Head Neck* 2000; 22: 328–335

Lingen MW, Xiao W, Schmidt A, Jiang B, Pickard R, Kreinbrink P, Perez-Ordóñez B, Jordan RC, Gillison ML. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. *Oral Onc* 2012

Lipkin WI. Microbe Hunting. *Microbiology and Molecular Biology Reviews* 2010; 74: 363–377

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology* 2012; 2012: 1–11

Llewellyn CD, Johnson NW, Warnakulasuriya KA. Risk factors for squamous cell carcinoma of the oral cavity in young people--a comprehensive literature review. *Oral Onc* 2001; 37: 401–418

Llewellyn CD, Linklater K, Bell J, Johnson NWW, Warnakulasuriya S. An analysis of risk factors for oral cancer in young people: a case-control study. *Oral Onc* 2004; 40: 304–313

Machado J, Reis PP, Zhang T, Simpson C, Xu W, Perez-Ordóñez B, Goldstein DP, Brown DH, Gilbert RW, Gullane PJ, Irish JC, Kamel-Reid S. Low prevalence of Human Papillomavirus in oral cavity carcinomas. *Head Neck Oncol* 2010; 2: 6

Mancao C, Hammerschmidt W. Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. *Blood* 2007; 110: 3715–3721

Marron M, Boffetta P, Zhang Z-F, Zaridze D, Wünsch-Filho V, Winn DM, Wie Q, Talamini R, Szeszenia-Dabrowska N, Sturgis EM, Smith E, Schwartz SM, Rudnai P, Purdue MP, Olshan AF, Eluf-Neto J, Muscat J, Morgenstern H, Menezes A, McClean M, Matos E, Mates IN, Lissowska J, Levi F, Lazarus P, La Vecchia C, Koifman S, Kelsey K, Herrero R, Hayes RB, Franceschi S, Fernández L, Fabianova E, Daudt AW, Dal Maso L, Curado MP, Cadoni G, Chen C, Castellsagué X, Boccia S, Benhamou S, Ferro G, Berthiller J, Brennan P, Moller H, Hashibe M. Cessation of alcohol drinking, tobacco smoking and the reversal of head and neck cancer risk. *Int J of Epid* 2010; 39:182-196

Martin D, Gutkind JS. Human tumor-associated viruses and new insights into the molecular mechanisms of cancer. *Oncog* 2009; 27: S31–S42

Massano J, Regateiro FS, Januário G, Ferreira A. Oral squamous cell carcinoma: Review of prognostic and predictive factors. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2006; 102: 67–76

Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2001; 2: 986–991

Mehanna H, Beech T, Nicholson T, El-Hariry I, McConkey C, Paleri V, Roberts S. Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer-systematic review and meta-analysis of trends by time and region. *Head Neck* 2012

Michaelis M, Doerr HW, Cinatl J. The story of human cytomegalovirus and cancer: increasing evidence and open questions. *Neoplasia* 2009; 11: 1–9

Moore RA, Warren RL, Freeman JD, Gustavsen JA, Chénard C, Friedman JM, Chénard C,

Friedman JM, Suttle CA, Zhao Y, Holt RA. The Sensitivity of Massively Parallel Sequencing for Detecting Candidate Infectious Agents Associated with Human Tissue. *PLoS ONE* 2011; 6: e19838

Mork J, Lie AK, Glattre E, Hallmans G, Jellum E, Koskela P, Møller B, Pukkala E, Schiller JT, Youngman L, Lehtinen M, Dillner J. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N Engl J Med* 2001; 344: 1125–1131

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* 2008; 5: 621–628

Müller S, Pan Y, Li R, Chi AC. Changing Trends in Oral Squamous Cell Carcinoma with Particular Reference to Young Patients: 1971–2006. The Emory University Experience. *Head Neck Pathol* 2008; 2: 60–66

Myers JN, Elkins T, Roberts D, Byers RM. Squamous cell carcinoma of the tongue in young adults: Increasing incidence and factors that predict treatment outcomes. *Otolaryngol Head Neck Surg* 2000; 122: 44–51

Nakata Y, Uzawa N, Takahashi K-I, Sumino J, Michikawa C, Sato H, Sato H, Sonoda I, Ohyama Y, Okada N, Amagasa T. EGFR gene copy number alteration is a better prognostic indicator than protein overexpression in oral tongue squamous cell carcinomas. *Eur J Cancer* 2011; 47: 2364–2372

Narechania A, Chen Z, Desalle R, Burk RD. Phylogenetic Incongruence among Oncogenic Genital Alpha Human Papillomaviruses. *J Virol* 2005; 79: 15503–15510

Näsman A, Attner P, Hammarstedt L, Du J, Eriksson M, Giraud G, Ahrlund S, Richter S, Marklund L, Romanitan M, Lindquist D, Ramqvist T, Lindholm J, Sparén P, Ye W, Dahlstrand H, Munck-Wikland E, Dalianis T. Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? *Int J Cancer* 2009; 125: 362–366

Negri E, Boffetta P, Berthiller J, Castellsagué X, Curado MP, Maso LD, Daudt AW, Fabianova E, Fernández L, Wünsch-Filho V, Franceschi S, Hayes RB, Herrero R, Koifman S, Lazarus P, Lence JJ, Levi F, Mates D, Matos E, Menezes A, Muscat J, Eluf-Neto J, Olshan AF, Rudnai P, Shangina O, Sturgis EM, Szeszenia-Dabrowska N, Talamini R, Wie Q, Winn DM, Zaridze D, Lissowska J, Zhang Z-F, Ferro G, Brennan P, La Vecchia C, Hashibe M. Family history of cancer: Pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Int J Cancer* 2009; 124: 394–401

Negri E, La Vecchia C, Franceschi S, Tavani A. Attributable risk for oral cancer in northern Italy. *Cancer Epidemiol. Biomarkers Prev* 1993; 2: 189–193

Nevels M, Täuber B, Spruss T, Wolf H, Dobner T. „Hit-and-run“ transformation by adenovirus oncogenes. *J Virol* 2001; 75: 3089–3094

Niller HH, Wolf H, Minarovits J. Viral hit and run-oncogenesis: genetic and epigenetic scenarios. *Cancer Lett* 2011; 305: 200–217

Omura-Minamisawa M, Diccianni MB, Chang RC, Batova A, Bridgeman LJ, Schiff J, Cohn SL, London WB, Alice LY. p16/p14ARF Cell Cycle Regulatory Pathways in Primary Neuroblastoma p16 Expression Is Associated with Advanced Stage Disease. *Clin Cancer Res* 2001; 7: 3481–3490

Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan P-I, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh W-J, Goldsmith CS, Zaki SR, Catton M, Lipkin WI. A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *N Engl J Med* 2008; 358: 991–998

Palacios G, Quan P-L, Jabado OJ, Conlan S, Hirschberg DL, Liu Y et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerging Infect Dis* 2007; 13: 73–81

Park K, Cho KJ, Lee M, Yoon DH, Kim J, Kim SY, Zhai J, Renwick N, Hui J, Hegyi H, Grolla A, Strong JE, Towner JS, Geisbert TW, Jahrling PB, Büchen-Osmond C, Ellerbrok H, Paz Sanchez-Seco M, Lussier Y, Formenty P, Nichol ST, Feldmann H, Briese T, Lipkin WI. p16 immunohistochemistry alone is a better prognosticator in tonsil cancer than human papillomavirus in situ hybridization with or without p16 immunohistochemistry. *Acta Otolaryngol* 2012: 1–8

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005; 55: 74–108

Patel SG, Shah JP. TNM staging of cancers of the head and neck: striving for uniformity among diversity. *CA Cancer J Clin* 2005; 55: 242–264

Paulsen IT. Role of Mobile DNA in the Evolution of Vancomycin-Resistant *Enterococcus faecalis*. *Science* 2003; 299: 2071–2074

Piccirillo JF, Costas I, Reichman ME. Chapter 2: Cancers of the Head and Neck. In: Ries L, Young LF, Keel GE, Eisner MP, Lin YD, Horner M-J, Hrsg: SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics. Bethesda, MD: National Cancer Institute, SEER Program; 2007

Piccirillo JF. Importance of comorbidity in head and neck cancer. *Laryngoscope* 2000; 110: 593–602

Polacek C, Friebe P, Harris E. Poly(A)-binding protein binds to the non-polyadenylated 3' untranslated region of dengue virus and modulates translation efficiency. *J Gen Vir* 2009; 90: 687–692

Poreba E, Broniarczyk JK, Gozdzicka-Jozefiak A. Epigenetic mechanisms in virus-induced tumorigenesis. *Clin Epigenet* 2011; 2: 233–247

Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* 2009; 37: D32–6

Pytynia KB, Grant JR, Etzel CJ, Roberts D, Wei Q, Sturgis EM. Matched analysis of survival in patients with squamous cell carcinoma of the head and neck diagnosed before and after 40 years of age. *Arch. Otolaryngol. Head Neck Surg* 2004; 130: 869–873

Quentin T, Henke C, Korabiowska M, Schlott T, Zimmerman B, Kunze E. Altered mRNA expression of the Rb and p16 tumor suppressor genes and of CDK4 in transitional cell carcinomas of the urinary bladder associated with tumor progression. *Anticancer Res* 2004; 24: 1011–1023

R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2011. Available from: <http://www.R-project.org/>

Ragin CCR, Taioli E. Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: Review and meta-analysis. *Int J Cancer* 2007; 121: 1813–1820

Rautava J, Luukkaa M, Heikinheimo K, Alin J, Grenman R, Happonen R-P. Squamous cell carcinomas arising from different types of oral epithelia differ in their tumor and patient characteristics and survival. *Oral Oncol* 2007; 43: 911–919

Ringström E, Peters E, Hasegawa M, Posner M, Liu M, Kelsey KT. Human papillomavirus type 16 and squamous cell carcinoma of the head and neck. *Clin Cancer Res* 2002; 8: 3187–3192

RKI. Krebs in Deutschland 2007/2008. 8. Aufl. Berlin: Robert Koch-Institut und Gesellschaft der deutschen Krebsregister in Deutschland eV; 2012

Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 2011; 27: 2325–2329

Robinson M, Sloan P, Shaw R. Refining the diagnosis of oropharyngeal squamous cell carcinoma using human papillomavirus testing. *Oral Oncol* 2010; 46: 492–496

Rose BS, Jeong J-H, Nath SK, Lu SM, Mell LK. Population-based study of competing mortality in head and neck cancer. *J Clin Oncol* 2011; 29: 3503–3509

Rosenquist K, Wennerberg J, Schildt E-B, Bladström A, Göran Hansson B, Andersson G. Oral status, oral infections and some lifestyle factors as risk factors for oral and oropharyngeal squamous cell carcinoma. A population-based case-control study in southern Sweden. *Acta Otolaryngol* 2005; 125: 1327–1336

Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Peñaranda S, Bankamp B, Maher K, Chen M-H, Tong S, Tamin A, Lowe L, Frace M, Derisi JL, Chen Q, Wang D, Erdman DD, Peret TCT, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Günther S, Osterhaus ADME, Drost C, Pallansch MA, Anderson LJ, Bellini WJ. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003; 300: 1394–1399

Ruffalo M, LaFramboise T, Koyuturk M. Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* 2011; 27: 2790–2796

Ryerson AB, Peters ES, Coughlin SS, Chen VW, Gillison ML, Reichman ME, Wu X, Chaturvedi AK, Kawaoka K. Burden of potentially human papillomavirus-associated cancers of the oropharynx and oral cavity in the US, 1998-2003. *Cancer* 2008; 113: 2901–2909

- Salem A. Dismissing links between HPV and aggressive tongue cancer in young patients. *Ann Onc* 2010; 21: 13–17
- Sanderson RJ, Ironside JAD. Squamous cell carcinomas of the head and neck. *BMJ* 2002; 325: 822–827
- Sargeran K, Murtomaa H, Safavi SMR, Vehkalahti MM, Teronen O. Survival after diagnosis of cancer of the oral cavity. *Br J Oral Maxillofac Surg* 2008; 46: 187–191
- Sarid R, Gao S-J. Viruses and human cancer: From detection to causality. *Cancer Lett* 2011; 305: 218–227
- Sasaki T, Moles DR, Imai Y, Speight PM. Clinico-pathological features of squamous cell carcinoma of the oral cavity in patients. *J Oral Pathol Med*. 2005; 34: 129–133
- Schantz SP, Yu G-P. Head and Neck Cancer Incidence Trends in Young Americans, 1973-1997, With a Special Analysis for Tongue Cancer. *Arch Otolaryngol Head Neck Surg* 2002; 128: 268–274
- Schlaberg R, Choe DJ, Brown KR, Thaker HM, Singh IR. XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. *Proc Natl Acad Sci USA* 2009; 106: 16351–16356
- Schwartz SR, Yueh B, McDougall JK, Daling JR, Schwartz SM. Human papillomavirus infection and survival in oral squamous cell cancer: A population-based study. *Head Neck Surg* 2001; 125: 1–9
- Seiwert TY, Cohen EEW. State-of-the-art management of locally advanced head and neck cancer. *Br J Cancer* 2005; 92: 1341–1348
- Shah JP, Lydiatt W. Treatment of cancer of the head and neck. *CA Cancer J Clin* 1995; 45: 352–368
- Shi W, Kato H, Perez-Ordóñez B, Pintilie M, Huang S, Hui A, O'Sullivan B, Waldron J, Cummings B, Kim J, Ringash J, Dawson LA, Gullane P, Siu L, Gillison ML, Liu FF. Comparative Prognostic Value of HPV16 E6 mRNA Compared With In Situ Hybridization for Human Oropharyngeal Squamous Carcinoma. *J Clin Oncol* 2009; 27: 6213–6221
- Shiboski CH, Schmidt BL, Jordan R. Tongue and tonsil carcinoma. *Cancer* 2005; 9: 1843-1849
- Siebers TJH, Merkx MAW, Slootweg PJ, Melchers WJG, Cleef P, Wilde PCM. No high-risk HPV detected in SCC of the oral tongue in the absolute absence of tobacco and alcohol—a case study of seven patients. *Oral Maxillofac Surg* 2008; 12: 185–188
- Skinner GR. Transformation of primary hamster embryo fibroblasts by type 2 simplex virus: evidence for a „hit and run“ mechanism. *Br J Exp Pathol* 1976; 57: 361
- Smeets SJ, Hesselink AT, Speel E-JM, Haesevoets A, Snijders PJF, Pawlita M, Meijer CJLM, Braakhuis BJM, Leemans CR, Brakenhoff RH. A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. *Int J Cancer* 2007; 121:

2465–2472

Sotlar K, Diemer D, Dethleffs A, Hack Y, Stubner A, Vollmer N, Menton S, Menton M, Dietz K, Wallwiener D, Kandolf R, Bültmann B. Detection and typing of human papillomavirus by e6 nested multiplex PCR. *J Clin Microbiol* 2004; 42: 3176–3184

Söderberg-Nauclér C. Does cytomegalovirus play a causative role in the development of various inflammatory diseases and cancer? *J Intern Med* 2006; 259: 219–246

Stehelin DD, Varmus HEH, Bishop JMJ, Vogt PKP. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* 1976; 260: 170–173

Stenson KM, Kunnavakkam R, Cohen EEW, Portugal LD, Blair E, Haraf DJ et al. Chemoradiation for patients with advanced oral cavity cancer. *Laryngoscope* 2009; 93–99

Stevenson PG, May JS, Connor V, Efstathiou S. Vaccination against a hit-and-run viral cancer. *J Gen Vir* 2010; 91: 2176–2185

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009; 458: 719–724

Sturgis EM, Cinciripini PM. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? *Cancer* 2007; 110: 1429–1435

Tandon S, Tudur-Smith C, Riley RD, Boyd MT, Jones TM. A Systematic Review of p53 as a Prognostic Factor of Survival in Squamous Cell Carcinoma of the Four Main Anatomical Subsites of the Head and Neck. *Cancer Epidemiol Biomarkers Preve* 2010; 19: 574–587

Tao T, Madden T, Camacho C. Blast Ftp Site. In: BLAST Help. Bethesda (MD): National Center for Biotechnol Information; 2008 Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1762/> Zugriff am: 08/06/2012

Therneau T, Lumley T. „survival“: Survival analysis, including penalised likelihood [Internet]. 2011. Available from: <http://CRAN.R-project.org/package=survival> Zugriff am: 12/08/2009

Toner M, O’Regan EM. Head and Neck Squamous Cell Carcinoma in the Young: A Spectrum or a Distinct Group? Part 1. *Head Neck Pathol* 2009; 3: 246–248

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; 28: 511–515

Tyler SD, Severini A. The complete genome sequence of herpesvirus papio 2 (Cercopithecine herpesvirus 16) shows evidence of recombination events among various progenitor herpesviruses. *J Virol* 2006; 80: 1214–1221

van der Schroeff MP, de Jong RB. ScienceDirect.com - Oral Onc - Staging and prognosis in head and neck cancer. *Oral Oncol* 2009; 45: 356-360

van Straaten F, Müller R, Curran T, Van Beveren C, Verma IM. Complete nucleotide sequence

of a human c-onc gene: deduced amino acid sequence of the human c-fos protein. *Proc Natl Acad Sci USA* 1983; 80: 3183–3187

Vermorken JB, Mesia R, Rivera F, Remenar E, Kawecki A, Rottey S, Erfan J, Zabolotnyy D, Kienzer H-R, Cupissol D, Peyrade F, Benasso M, Vynnychenko I, De Raucourt D, Bokemeyer C, Schueler A, Amellal N, Hitt R. Platinum-based chemotherapy plus cetuximab in head and neck cancer. *N Engl J Med* 2008; 359: 1116–1127

Vokes EE, Weichselbaum RR, Lippman SM, Hong WK. Head and neck cancer. *N Engl J Med* 1993; 328: 184–194.

Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, Dirisi JL. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 2002; 99: 15687–15692

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; 10: 57–63

Weber G, Shendure J, Tanenbaum DM, Church GM, Meyerson M. Identification of foreign gene sequences by transcript filtering against the human genome. *Nat Genet* 2002; 30: 141–142

Wiest T, Schwarz E, Enders C, Flechtenmacher C, Bosch FX. Involvement of intact HPV16 E6/E7 gene expression in head and neck cancers with unaltered p53 status and perturbed pRb cell cycle control. *Oncogene* 2002; 21: 1510–1517

Witkiewicz AK, Knudsen KE, Dicker AP, Knudsen ES. The meaning of p16ink4a expression in tumors: Functional significance, clinical associations and future developments. *Cell Cycle* 2011; 10: 2497–2503

Woolhouse MEJ, Howey R, Gaunt E, Reilly L, Chase-Topping M, Savill N. Temporal trends in the discovery of human viruses. *Proc Biol Sci* 2008; 275: 2111–2115

Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis. *Nucleic Acids Res* 2006; 34: W6–W9

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; 18: 821–829

Zhang ZF, Morgenstern H, Spitz MR, Tashkin DP, Yu GP, Hsu TC, Schantz SP. Environmental tobacco smoking, mutagen sensitivity, and head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 2000; 9: 1043–1049

Zhou J, Michaud DS, Langevin SM, McClean MD, Eliot M, Kelsey KT. Smokeless tobacco and risk of head and neck cancer: Evidence from a case-control study in New England. *Int J Cancer* 2012

Ziegler A, Lange S, Bender R. Überlebenszeitanalyse: Die Cox-Regression. *Dtsch med Wochenschr* 2007; 132: e42–e44

Znaor A, Brennan P, Gajalakshmi V, Mathew A, Shanta V, Varghese C, Boffetta P. Independent

and combined effects of tobacco smoking, chewing and alcohol drinking on the risk of oral, pharyngeal and esophageal cancers in Indian men. *Int J Cancer* 2003; 105: 681–686

7. Danksagung

Mein besonderer Dank gilt Prof. Dr. med. Tanguy Seiwert für die Überlassung des Themas, seinen Hinweisen bei der Bearbeitung des Projekts und die großartige Möglichkeit, in seinem Labor an der University of Chicago forschen zu dürfen.

Herrn Prof. Dr. med. Peter Brossart danke ich sehr für die Bereitschaft, meine Dissertation zu betreuen und die Hilfsbereitschaft und Geduld bei der Erstellung dieser Arbeit. Ebenso danke ich ihm für die Durchsicht des Manuskripts und für die Vorschläge, die zur Vollendung desselben hilfreich waren.

Besonderer Dank gilt außerdem allen Mitarbeitern der Labore von Prof. Tanguy Seiwert, Prof. Kevin White und Prof. Ezra Cohen, die mir bei meinen Experimenten mit Rat und Tat zur Seite standen. Stellvertretend sind vor allem Dr. med. Mohamed El Dinali, Wanda Shen, Dr. rer. nat. Casey Brown, Dr. med. Tom Stricker, cand. med. Abir Kundu und Dr. med. Essam El Hassani zu nennen.

Nicht zuletzt gebührt auch meiner Freundin und meinen Eltern großer Dank, da sie mit ihrer moralischen Unterstützung einen großen Teil zur Fertigstellung dieser Arbeit beigetragen haben.

Publikationen/Präsentationen

Brägelmann J, Dagogo-Jack I, El Dinali M, Stricker T, Brown CD, Zuo Z, Khattri A, Keck M, McNerney ME, Longnecker R, Bieging K, Kocherginsky M, Alexander K, Salgia R, Lingen MW, Vokes EE, White KP, Cohen EE, Seiwert TY. Oral cavity tumors in younger patients show a poor prognosis and do not contain viral RNA. *Oral Oncol* 2013 Mar 9; pii: S1368-8375(13)00035-3. doi: 10.1016/j.oraloncology.2013.02.003. [Epub ahead of print]

Catenacci DV, Cervantes G, Yala S, Nelson EA, El-Hashani E, Kanteti R, El Dinali M, Hasina R, Brägelmann J, Seiwert T, Sanicola M, Henderson L, Grushko TA, Olopade O, Karrison T, Bang YJ, Kim WH, Tretiakova M, Vokes E, Frank DA, Kindler HL, Huet H, Salgia R. RON (MST1R) is a novel prognostic marker and therapeutic target for gastroesophageal adenocarcinoma. *Cancer Biol Ther* 2011 Jul 1;12(1):9-46

Oral plenary session presentation beim „22nd AACR – NCI – EORTC Joint-Symposium on Molecular Targets and Cancer Therapeutics 2010“ in Berlin:

Brägelmann J, Stricker T, Brown C, El Dinali M, Zou X, Vokes E, White KP, Seiwert T.

Transcriptome sequencing of upper aerodigestive tract cancer cell lines to reveal potential therapeutic targets. *EJC Supplements* 2010; 8(7):84

Posterpräsentation beim „Chicago Lakeshore Symposium 2009“:

Brägelmann J, El Dinali M, Brown C, Salgia R, Vokes E, Cohen E, White KP, Seiwert T.

„Screening for oncoviruses using Solexa high-throughput transcriptome sequencing.“