Genetic and functional analysis of head and neck carcinogenesis

Serge Jean Smeets
The work described in this thesis was performed at the Section Tumor Biology, Department of Otolaryngology/Head-Neck Surgery. The research project was carried out within the framework of the VUmc Institute for Cancer and Immunology (VUmc CCA/V-ICI).


Printed by PrintPartners Ipskamp, Enschede, The Netherlands.

Publication of this thesis was financially supported by MRC-Holland BV, Greiner Bio-one, Enzo Life Sciences, Netherlands Bioinformatics Centre (NBIC) and Merck Serono Oncology.

ISBN: 9789086593262

©Copyright 2009, S.J. Smeets

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system or transmitted in any form or means without the prior permission in writing from the author.
Genetic and functional analysis of head and neck carcinogenesis

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op maandag 22 juni 2009 om 15.45 uur
in de aula van de universiteit,
De Boelelaan 1105

doctor

Serge Jean Smeets

geboren te Amsterdam
promotoren: prof.dr. R.H. Brakenhoff  
prof.dr. C.R. Leemans

copromotoren: dr. B.J.M. Braakhuis  
dr. B. IJlstra
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>Genome wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene expressing human papillomavirus</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td><em>Oncogene</em> (2006);25(17): 2558–64</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Genetic classification of oral and oropharyngeal carcinomas identifies subgroups with a different prognosis</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td><em>Cellular Oncology</em> (2009); <em>in press</em></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Genomic profiling identifies common HPV-associated chromosomal alterations in squamous cell carcinomas of cervix and head and neck</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td><em>Submitted</em></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Functional and molecular consequences of p53 and pRb pathway inactivation in oral keratinocytes</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td><em>Submitted</em></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Figures in color</td>
<td>139</td>
</tr>
<tr>
<td>9</td>
<td>Summary and Future perspectives</td>
<td>157</td>
</tr>
<tr>
<td>10</td>
<td>Samenvatting en Toekomst perspectief</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>Curriculum Vitae and Publications</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>Dankwoord</td>
<td>179</td>
</tr>
</tbody>
</table>
CHAPTER 1

General introduction
HEAD AND NECK SQUAMOUS CELL CARCINOMA

Epidemiology

Head and neck cancer consists of a heterogeneous group of lesions that arise in the upper aerodigestive tract, which includes the oral cavity, pharynx (nasopharynx oropharynx and hypopharynx) and larynx (Figure 1) [1]. The majority of these neoplasms (90 percent) are head and neck squamous cell carcinomas (HNSCC) that develop in the squamous layer of the mucosal lining. HNSCC develops in the middle-aged population, with the highest incidence in the sixth decade of life. Even though our knowledge about HNSCC has increased significantly, the survival rate has improved modestly during the last 20 years [2].

The incidence of HNSCC varies globally. Although the precise reasons are not known, this variation likely reflects worldwide ethnic, cultural and socio-economical differences [3]. Exemplary is the incidence of nasopharyngeal cancer (NPC), which is highest in Southeast Asia, whereas, in Western countries the oral cavity, the pharynx and the larynx are the most preferential sites [4;5]. All together, there are approximately 500,000 new cases of HNSCC worldwide [4]. In the Netherlands approximately 2,400 cases are yearly diagnosed. Particularly in females, the incidence of HNSCC in the Netherlands is still rising, presumably related to an increase in tobacco-smoking, one of the major risk factors of HNSCC (see also below). Vice versa, in the United States a decrease in tobacco-smoking seems to cause a reduction of the incidence of HNSCC [6], except for tonsillar cancer of which the occurrence is increasing particularly among men.
Figure 1. The upper aerodigestive tract

Risk Factors

Tobacco smoking and Alcohol
The major risk factors for the development of HNSCC are life-style factors, most notably tobacco smoking and alcohol consumption [7]. Tobacco smoking and alcohol consumption are independent risk factors, but have a synergistic effect when combined [8-12]. The relative risk appears to be related to the time of exposure [13]. Nicotine and tar are not the only harmful ingredients; over 4,000 chemicals have been identified in cigarette smoke. Carcinogens from tobacco smoking, for example benz-(a)-pyrene and nitrosamines, are known to produce the precise types of guanine nucleotide transversions causing mutations in cancer genes that are involved in HNSCC development [13-15].

Together with tobacco, alcohol is the most abundantly consumed noxious compound worldwide [16]. There is convincing evidence that acetaldehyde, the first metabolite produced during alcohol degradation, is responsible for the carcinogenic effect of ethanol on the upper aerodigestive tract due to its multiple mutagenic effects on DNA [17;18]. Ethanol may also stimulate carcinogenesis by inhibiting DNA methylation and by interacting with retinoid metabolism [19;20]. Another source of potent carcinogens originates from chewing tobaccos or betel nuts, but these habits are rare in Western countries [21].
**Human papillomavirus**

In recent years human papillomavirus (HPV) was found to be involved in a subset of HNSCC [22;23], and sexual transmission is suggested to cause the infection [24-27]. In cervical carcinomas HPV is found in nearly all cases [28] and its causal involvement is widely accepted [29], but the situation in HNSCC is less clear.

HPV is a DNA virus with a genome of approximately 7.9 kilobase pairs in size, with more than 100 subtypes identified. Only 14 types are oncogenic, referred to as high-risk types (hrHPV). HrHPV produces two oncoproteins, encoded by the E6 and E7 genes, that inactivate the p53 and pRb proteins, respectively, allowing cell cycle entry and a DNA synthesis environment needed for viral replication. HPV16-E6 interacts besides p53 with a large number of other cellular proteins [33]. The direct targeting of the p53 protein via E6AP by HPV16-E6 for proteasome-mediated degradation is well characterized, but the direct binding to proteins with a PDZ-binding motif like Scribble, MAGI-1,2,3, MUPP1 and Dlg, is less investigated. These proteins are mainly involved in cell signaling, cell adhesion and possibly tumor suppressor activity [30-32]. HPV-E7 interacts besides pRb, also with its family members p107 and p130. HPV-E7 further interacts with other proteins such as HDACs 1, 2, 3, 8, the cyclin-dependent kinase inhibitors p21^{CIP1} and p27^{kip1}, and histone acetylase p300 [33]. The intensive research on the role of HPV in cervical carcinomas has instigated studies on the prevention or treatment of HPV–induced cancers using preventive and therapeutic vaccines [34]. Particularly the preventive vaccination seems very successful in reducing HPV16 infection in young women [35;36].

The reported prevalence of high-risk HPV DNA in HNSCC tumor specimens varies tremendously between studies ranging from 0 to 100% [27;37]. The most commonly detected HPV-type in HNSCC is HPV16, which has been demonstrated in 90-95% of all HPV-positive HNSCC, followed by HPV18 and HPV33. Part of the variation in HPV prevalence can be explained by differences with regard to the location of the tumor, i.e. it is relatively low in the oral cavity and high in the tonsil [22;23;38;39]. More importantly, variations in the type of tissue material studied and the HPV detection methods used, seem to have a major impact on the discrepancy in reported prevalence rates [22;37;40]. Frequently PCR methods are employed that detect HPV DNA. A major problem with these assays is that they are extremely sensitive causing false positive results. Sample to sample contaminations and transient infections might all cause positive HPV DNA assays, while they do not reflect a true oncogenic infection. Based on this problem we focused previously on the detection of HPV E6
and E7 transcripts, and showed that of 24 HPV DNA positive cases only 12 reflected an oncogenic infection as determined by E6 RT-PCR [41]. This approach was substantiated by the recent finding that tumors containing transcriptionally active HPV are genetically different from those that do not contain HPV and that arose most likely as a result from exposure to environmental carcinogens [23]. Tumors with transcriptionally active HPV did not show any TP53 mutation and displayed very limited allelic losses at 3p, 9p and 17p. In contrast, in tumors not containing HPV a TP53 mutation had occurred in 75% of the cases and very frequent allelic losses were found [23]. These results suggest that an infection with HPV is an early event in HNSCC development that persists throughout the entire period of cancer progression. Moreover, the disruption of the pRb and p53 pathways by HPV oncoproteins is reflected in a unique genotype of the resulting tumors. Although indirectly, these data support the view that particularly tumor with transcriptionally active HPV are a separate entity, an observation in line with the known biological properties of the virus in cervical carcinogenesis. There is evidence suggesting that patients with HPV-positive HNSCC have a more favourable prognosis, than those with HPV-negative HNSCC [40;42-44]. However, in a recent meta-analysis showed this was still not clear. A major problem in this respect is that there is no consensus on how HPV should be detected.

**Genetic predisposition**
Defects in DNA repair processes are associated with a high HNSCC cancer risk [45]. One example is Fanconi Anemia (FA). Patients suffering from FA, as a result of a defect in one of the FANC genes, have a very high risk to develop oral and oropharyngeal carcinoma at early age [46]. Some studies have shown that HNSCC shows a familial clustering, indicative for a susceptibility to HNSCC with a genetic basis [47;48]. The search for cancer-predisposing genes is ongoing, with functional tests and large scale genome-wide association studies as tools [47]. It is very likely that a combination of certain alleles of these cancer predisposing genes determines the risk for cancer.
Clinical and histopathological staging

Head and neck squamous cell carcinomas are classified according to the TNM system of the ‘International Union Against Cancer’ (UICC) [50]. The TNM anatomical staging is based on the assessment of three components: the extent of the primary tumor (T), the absence or presence and extent of regional lymph node metastases (N), and the absence or presence of distant metastases (M). Two classifications are used: (a) clinical classification (pre-treatment clinical classification), based on evidence acquired before treatment, e.g. from physical examination, imaging, endoscopy, biopsy and ultrasound-guided fine needle aspiration cytology (USgFNAC), and (b) pathological classification (post-surgical histopathological classification), based on histopathological examination of the resection specimen. Based on the TNM classification, a clinical stage is used to group tumors with comparable clinical prognosis. Patients are classified into four clinical stages I-IV. Stages I and II represent early stages of the disease, with relative small tumors and no lymph node metastases. Stages III and IV represent more advanced stages of HNSCC with larger tumors that have spread to regional lymph nodes or secondary organs. For HNSCC patients TNM staging is strongly associated with survival although individual patients with similar TNM stage may differ significantly in clinical outcome and response to specific treatment. This indicates room for an improved staging system, one that is perhaps based on molecular characteristics, to more accurately predict disease progression and treatment response.

Treatment and prognosis

In general, one-third of the HNSCC patients presents with early stage (I and II) disease, while two-third presents with advanced disease (stage III and IV) [51]. The vast majority of patients with early stage HNSCC can be cured with single modality treatment (surgery or radiotherapy) focussed only at the tumor site. Patients with advanced stages are mostly treated with a combination of surgery and (chemo-)radiotherapy or increasingly with combined chemo-radiation, but the prognosis remains relatively modest or poor.

Despite major improvements in surgical techniques and radiotherapy in HNSCC treatment, the long term survival of HNSCC patients has only moderately improved during the last 20 years [2]. An important reason for this lack of progress is the relatively high locoregional recurrence rates observed in these patients. Local recurrences occur in about 10 to 30% of the cases, even when the surgical margins are histologically tumor-free [52]. Moreover, 10 to 20% of the patients develop
regional recurrences and 15 to 25% distant metastases [53]. Finally, HNSCC patients frequently develop second primary tumors (SPT) in the upper aerodigestive tract [54]. These SPT develop on average with a constant rate of 2 to 3% per year [55]. Disease staging is strongly associated with survival, but tumors with similar TNM stage may differ significantly in response to treatment. Currently, systemic therapy including chemotherapy and molecularly targeted agents, like the epidermal growth factor receptor inhibitors, angiogenesis inhibitors and other molecularly targeted agents, has been successfully integrated into curative treatment of locoregionally advanced HNSCC [56]. One shortcoming is that the majority of the clinical trails are based on an ‘one size fits all’ approach and treatment is not tailored to the patient. Although many clinical-pathological variables and molecular markers (e.g., EGFR expression and HPV) are of prognostic value, wide heterogeneity in clinical outcomes is seen, implicating the necessity of standardization of established and additional biomarkers that will guide treatment decisions. Advances in basic research and applications of genomic profiling are expected to provide powerful methods for the individualisation of treatment approaches in HNSCC patients.

**HNSCC carcinogenesis**

It is generally accepted that HNSCC develops by an accumulation of genetic alterations. These genetic alterations result in inactivation of tumour suppressor genes and activation of proto-oncogenes by deletions, point mutations, promoter methylation, or gene amplification. These genetic changes have been identified in HNSCC by a variety of methods, including conventional cytogenetic karyotyping, loss of heterozygosity (LOH) analysis [57] and comparative genomic hybridization (CGH) [58;59]. HNSCC develops through a number of precursor stages. Most premalignant changes can be recognized histologically as dysplasia in the mucosal epithelium graded as mild, moderate and severe. By genetic analysis of these histologically defined preneoplastic changes and invasive tumours Califano et al. presented the first genetic progression model for HNSCC [60] This model was adapted and simplified by Braakhuis et al. [61] (Figure 2).
Allelic loss of the chromosomal regions 3p, 9p21 and 17p13 are considered to be altered at an early phase in HNSCC carcinogenesis, as they are found in 60–80% of the preneoplastic lesions and invasive HNSCC [61-65]. Amplification of chromosomal region 11q13 is found with a lower frequency in preneoplastic lesions, but is present in 70% of primary HNSCC tumors. Some alterations of chromosomal regions are identified as prognosticator of outcome in HNSCC, including amplification at 11q13, gain of 12q24, and losses at 5q11, 6q14, and 21q11 [66].

Many potential cancer genes are suggested to be localized at the frequently altered chromosomal regions. The gene locus found in chromosome 9p21 encodes two different transcripts, p16\textsuperscript{Ink4A} and p14\textsuperscript{ARF}, which are responsible for G1 cell cycle regulation and MDM2 mediated degradation of p53, respectively. P16 is often inactivated in HNSCC through homozygous deletion, by promoter methylation, and less commonly by point mutations [67]. On 3p it is difficult to pinpoint a potential cancer gene, since in most cases a large part or whole chromosomal p-arm is lost, making characterization of the specific locus responsible for the tumour suppressor phenotype almost impossible. One candidate tumor suppressor gene at this chromosomal area is \textit{FHIT}, which has been found to be inactivated by exonic deletions in many tumour types, including a small percentage of HNSCC [68:69]. The \textit{TP53} gene is located on the short arm of chromosome 17 (17p13) and is important for maintaining genomic stability. Mutation of \textit{TP53}, often accompanied with allelic loss of 17p13, is the most common genetic event in human tumors, including head and neck cancer [70-72]. Inactivation of p53 leads to a loss of the G1/S phase cell

\begin{figure}
\centering
\includegraphics[width=\textwidth]{hnscc_progression_model.png}
\caption{HNSCC progression model. Braakhuis \textit{et al.} suggested an adapted progression model for cancer in the head and neck. A stem cell acquires one or more genetic alterations and forms a patch, a small cluster of less than 200 cells in diameter that can be detected by p53 immunostaining. Eventually the stem cell escapes normal growth control. A precursor lesion arises and laterally displaces the normal mucosa. Additional genetic hits give rise to different subclones within a field and eventually a subclone evolves into invasive cancer [61].}
\end{figure}
cycle checkpoint and as a result cells continue dividing without pausing to repair DNA damage [73]. Besides increasing the risk for developing HNSCC, inactivation of p53 makes the tumor less sensitive to radiotherapy and chemopreventive strategies [74;75]. P63, a p53 homologue located on 3q28 with oncogenic properties is frequently amplified in HNSCC [76]. Mutation of TP53 can lead to overexpression and accumulation of truncated p53 within the cells, making detection of the mutated form with immunostaining possible. Sequence analysis and immunostaining on dysplastic lesions have shown that the gene is often mutated before the invasive cancer develops, making it an early marker for progression.

In addition, amplification of 11q13 is linked with overexpression of Cyclin D1 and has been associated with an increased rate of lymph node metastases and overall poor prognosis [77;78]. Cyclin D1 induces phosphorylation of Rb, thus enabling progression from G1 to S phase.

Other potential cancer genes in HNSCC are DCC and SSCRO. DCC, located on 18q21, is commonly deleted in colorectal cancer and is inactivated in HNSCC by hypermethylation of its promoter [79;80]. SSCRO, also known as DCUN1D1, is located on 3q26.3 and often amplified in HNSCC. Amplification and overexpression of SSCRO correlates with poor clinical outcome and the oncogenic potential of SCCRO is underscored by its ability to transform fibroblasts (NIH-3T3 cells) in vitro and in vivo [81]. Finally, overexpression of epidermal growth factor receptor (EGFR) and its ligands is an important feature of HNSCC cancer with up to 80% of HNSCC showing elevated EGFR mRNA levels compared to normal mucosa [81;82].

Taken together, the current findings provide evidence that many genetic alterations are necessary to drive the progression from a normal cell to a cancer cell. To pinpoint the causative genes located in the altered chromosomal regions is hampered severely by a number of reasons. First, the majority of the HNSCC have a chromosomal instability phenotype, resulting in gains and losses of large parts or even whole chromosomes. As a consequence, many DNA changes or differentially expressed genes may be considered as a result and not a cause of carcinogenesis. It is therefore of utmost importance to discriminate the carcinogenic ‘driving’ from the ‘passenger’ genetic events [83]. Second, the data provided is based on the findings with ‘low resolution methods’ as LOH analysis, multiplex ligation-dependent probe amplification (MLPA) or conventional CGH that do not allow to identification of causal genes.
Micro Array Comparative genomic hybridization
The recent completion of the Human Genome Project combined with polymerase chain reaction (PCR)-based genomic evaluation techniques has contributed tremendously to the analysis of chromosomal aberration in solid tumors [84]. Lately, much knowledge is gained about chromosomal copy number changes or allelic imbalance in HNSCC, using low resolution techniques focusing on pre-selected genomic regions like LOH and MLPA analysis. A revolution in cancer genetics was the development of conventional CGH, as it allowed the genome wide analysis of genetic changes with a 5 Megabase (Mb) resolution. CGH is based on hybridization of equal amounts of tumor and normal reference DNA labelled with different fluorescent compounds to metaphase spreads. By quantitative fluorescent microscopy the numerical changes can be determined and summarized. Both the resolution and speed of genome-wide analysis of genetic profiles has been improved immensely by the introduction of microarray-based analysis (maCGH). The technique is explained in Figure 3. The first arrays consisted of large-insert genomic clones, such as BACs (Bacterial Artificial Chromosomes) or PACs (P-1 derived Artificial Chromosomes), yielding an average resolution of 1 Mb [85;86]. The use of BAC clones with known genomic positions allowed direct identification of numerical changes throughout the genome. Nowadays, high resolution tiling arrays exist, containing more than 30,000 overlapping BAC clones [87].
**Figure 3. Principle of array CGH.** This figure shows the steps in BAC array CGH. (A) BAC clones are selected from a physical map of the genome. (B) DNA samples are extracted from selected BAC clones and their identity is confirmed by DNA fingerprinting or sequence analysis. (C) A multi-step amplification process generates sufficient material from each clone for array spotting. Each clone is spotted in replicate onto a solid support. (D) Reference DNA and test DNA are differentially labeled with cyanine 3 and cyanine 5 respectively. (E) The two labeled products are combined and hybridized onto the spotted slide. (F) Images from hybridized slides are obtained by scanning in two channels. Signal intensity ratios from individual spots can be displayed as a simple plot (G) or by using more complex software such as Imagene, which can display copy number alterations throughout the whole genome [88].

In addition to BAC arrays, oligonucleotide arrays (oligo arrays) have been developed. Oligo arrays have major advantages over BAC arrays, as their manufacturing is
cheaper, since BAC-DNA isolation and PCR amplification are not necessary [89-91]. In addition, they can be custom-made with a much higher resolution and are commercially available. Since the introduction of array CGH, many studies reported DNA copy number aberrations in HNSCC and frequent gains and losses of almost every chromosome have been described [59;92-94]. Unfortunately, most of the data is descriptive as most studies comprise small numbers of patients and data is not related to prognostic factors or clinical outcome.

**Functional analysis of cancer genes and their pathways.**

Based on epidemiological data and *in vitro* transformation experiments, it has been estimated that four to six genetic events are required in humans to transform a normal cell into a malignant cell [95;96]. As delineated above, HNSCC carcinogenesis is mainly driven by a few genetic events that involve the impairment of the P53 and the pRb pathway, two pathways that are involved in cell cycle regulation and apoptosis [97]. Yet, in some HNSCC these pathways seem not to be disrupted and other genes may be involved. However, the genetic instability of tumors is a factor of importance. Many DNA changes or differentially expressed genes could be considered as a consequence and not a cause of carcinogenesis. To discriminate the carcinogenic ‘driving’ from the ‘passenger’ genetic events, functional characterization of candidate genes *in vivo* or *in vitro* is therefore essential to elucidate causative involvement.

**In vivo models**

Mouse models play an important role in understanding the mechanisms of human carcinogenesis and have accelerated the search for finding new molecular targets for cancer therapy. However, suitable transgenic and knock-out mouse models for HNSCC are still not available. Some initial successes have been reported [98], but the models show a low rate of tumorigenesis and are late at development, hampering efficient investigation. Moreover it is a major task to generate them and for this reason not suitable for the investigation of candidate genes. Initial functional data obtained in *in vitro* models should be available before a mouse model is worth the effort.
**In vitro models**

Cell lines have frequently been used as a tool for the investigation of pathways involved in HNSCC [99]. Unlike tumors, cell lines are generally clonal and therefore it is much easier to examine their underlying genetic abnormalities [100]. However the majority of the cancer cell lines are extracted from invasive carcinomas that may have already acquired numerous mutations and genetic changes, which hampers the investigation of early phenotypic changes. Moreover, the process of adapting to culture likely selects for pheno- and genotypes important for *in vitro* culture, causing artifacts [101;102].

A more compelling way to characterize novel cancer genes, would be to investigate the transforming ability of these genes in normal cells [103;104]. The introduction of viral oncogenes in combination with long term *in vitro* cultivation allowed the outgrowth of rare immortalized clones [105-107]. More interestingly are the modern approaches in which specific cancer-associated pathways are manipulated using genetic modifications. These models have been used to study the malignant transformation of human fibroblasts *in vitro* [1]. Similar *in vitro* models have been exploited using primary oral mucosa cells, which demonstrated that an immortal phenotype can be achieved by introduction of a combination of Cyclin D1 overexpression and p53 inactivation [108], or deficient p16 expression, in combination with ectopic expression of the telomerase catalytic subunit, hTERT [109]. Unfortunately, the use of primary cells as *in vitro* model has important limitations. First of all, experimental transformation of primary human oral keratinocytes is very difficult due to the rapid premature senescence of oral keratinocytes in culture [110]. Due to the limited lifespan of primary cells, these types of studies must be carried out at a small scale and in a short time-frame. It is therefore difficult if not impossible to reproduce experiments, as individual primary cultures contain a too low number of cells to accomplish this. Secondly, primary cultures consist of mixtures of various types of cells as stem cells, transit amplifying cells and differentiated cells that cannot be identified separately in culture while the cellular origin might have major impact on the obtained phenotypes. Finally, large genetic screens will not be possible in primary cultures [111-113]. Above reasons show that there is a clear need for a reproducible and robust *in vitro* model of oral squamous cells.
Aim and outline of this thesis

Improvements in the long-term survival of HNSCC can be achieved if the generic signalling pathways involved in malignant transformation are elucidated, and when tumors can be classified in subgroups with a comparable prognosis, treatment response and therapy choice. At present two different subgroups of HNSCC can be discriminated on the basis of the etiological factor involved. In one group the human papillomavirus (HPV) is the causative factor, while in the other group predominantly smoking and alcohol consumption seems to play a role. These subgroups are genetically and clinically distinct, suggesting that a different route of carcinogenesis has been followed. A remaining question is the choice of a suitable assay to detect an high-risk HPV infection which is necessary to assess the exact prevalence of HPV-positive HNSCC and allow a reliable classification.

Based on descriptive genetic data and experimental models, convincing evidence is present that alterations in the genes $TP53$, $p16^{INK4A}$ as well as $CCND1$ are causative, suggesting an important role for the p53 and pRb pathways. This is supported by the notion that in HPV-infected tumors these same pathways are inactivated by the viral oncogenes E6 and E7. Many other potential genes and signalling pathways have been indicated, including involvement of the genes $DCC$, $SSCRO$ and $FHIT$, but the evidence is less compelling. Moreover, even for the well established cancer genes, $TP53$, $p16^{INK4A}$ and $CCND1$, many questions have remained such as: 1) in what order does the abrogation of the p53- and pRb-pathways take place, 2) is there an additional effect of combined $p16^{INK4A}$ and $CCND1$ abrogation, etcetera.

Furthermore, what is the driving molecular event in the 40% of the HNSCC that are $TP53$ wild type? As described above, the latter question may be partly explained by the presence of HPV in a subgroup of these $TP53$ wild-type tumors. However, still approximately 30% of the tumors are apparently $TP53$ wild type and negative for HPV. Are other genes in the p53 pathway altered in these tumors, or are we dealing with a separate carcinogenic route?

Besides these issues on the established (viral) cancer genes we are facing three major problems that hamper head and neck cancer research. First, the chromosomal regions harbouring potential cancer genes have been identified by low resolution methods such as microsatellite PCR or conventional CGH analysis, and these data is not stratified for the presence of HPV or biologically relevant outcome parameters, such as development of lymph node metastases, distant metastases or second field tumors. Second, the genetic instability of tumors is a significant problem, because
many DNA changes comprise large parts or whole chromosomal arms harbouring thousands of genes. Third, there are no models to test in particular the genes causing the earliest changes in carcinogenesis. The only models available at present are established HNSCC cell lines that do not allow analysis of phenotypes of early carcinogenesis such as immortalization, anoikis, and others.

In this thesis the use of high-density genome wide profiling of HNSCC by microarray comparative genomic hybridization (maCGH) and development of algorithms to decipher these data is described. Moreover, a number of approaches and assays to distinguish different patient groups have been developed and evaluated. In addition an in vitro model was established using conditionally immortalized primary oral squamous keratinocytes, which is not only suited for functional analysis of candidate cancer genes but also allows high throughput functional screening to identify these.

In Chapter 2, we address the question on the discrepancy in literature about the HPV prevalence in HNSCC, and evaluate relevant assays that are applicable for detection on archival formalin-fixed paraffin-embedded specimen. In Chapter 3 [114] a bioinformatic tool is described that allows comparative analysis using CGH data. In Chapter 4 both the maCGH platform as the developed bioinformatic tools are exploited to detect the genome wide genetic alterations that are in common and those that differ between tumors caused by HPV infection and tumors caused by other etiological factors such as alcohol consumption and tobacco smoking. In Chapter 5 we exploited a novel algorithm to classify HNSCC on basis of genome wide chromosomal aberrations, and showed that within the group of HPV-negative HNSCC at least two, possibly three genetically distinct subgroups of tumors exist associated with both etiological, clinical and molecular characteristics. In Chapter 6 the genetic data of HPV-positive and HPV-negative tumors was compared with that of HPV-positive cervical squamous cell carcinomas. Both genetic changes associated with HPV infection as well as changes associated with the tissue origin could be distinguished, an observation that will help to focus the attention on the most relevant chromosomal locations harbouring cancer genes.

Additionally, to solve the critical lack of relevant models to study the earliest cancer-associated phenotypes, we generated an in vitro transformation model using conditionally immortalized primary oral squamous keratinocytes. The establishment, validation and application of this model to answer a number of questions on the role of the p53 and pRb pathways in HNSCC are described in Chapter 7.

In Chapter 8 a general discussion of the results presented in this thesis is provided.
REFERENCES


A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen


Serge J. Smeets
Albertus T. Hesselink
Ernst-Jan M. Speel
Annick Haesevoets
Peter J.F. Snijders
Michael Pawlita
Chris J.L.M. Meijer
Boudewijn J.M. Braakhuis
C. René Leemans
Ruud H. Brakenhoff
ABSTRACT

Human papillomavirus type 16 (HPV16) plays a role in the development of a subgroup of head and neck squamous cell carcinomas (HNSCC). However, uncertainty exists about the true impact of HPV in this tumor type as conflicting reports have been published with prevalence rates from 0 to 100%. We aimed to find a detection algorithm of a biologically and thus clinically meaningful infection, applicable for high-throughput screening of frozen and formalin-fixed paraffin embedded (FFPE) specimens.

By considering detection of HPV E6 oncogene expression in frozen biopsies as gold standard for a meaningful HPV infection, the value of several assays was evaluated on FFPE tumor specimens and sera of 48 HNSCC patients. The following assays were evaluated on FFPE tissue samples: HPV DNA general primer (GP)5+/6+ PCR, viral load analysis, HPV16 DNA FISH detection, HPV16 E6 mRNA RT-PCR, p16 immunostaining, and on corresponding serum samples detection of antibodies against the HPV16 proteins L1, E6 and E7. Comparing single assays on FFPE tissue samples detection of E6 expression by RT-PCR was superior, but application remains at present limited to HPV16 detection. Most suitable algorithm with 100% sensitivity and specificity appeared p16 immunostaining followed by GP5+/6+ PCR on the p16-positive cases.

We show that clinically meaningful viral HPV infections can be more reliably measured in FFPE HNSCC samples in a standard and high throughput manner, paving the way for prognostic and experimental vaccination studies, regarding not only HNSCC, but possibly also cancer types with HPV involvement in subgroups such as penile and anal cancer.
INTRODUCTION

Tobacco smoking and alcohol consumption are the main risk factors in the etiology of head and neck squamous cell carcinomas (HNSCC). In addition, over the last 15 years, infection with high-risk human papillomavirus (HPV) types has also been etiologically linked with a subset of HNSCCs [1-7] and sexual transmission is suggested to cause the infection [5;8;9].

High risk HPV type 16 (HPV16) has been identified in approximately 90% of HPV-positive HNSCC, and HPV18, -31 and -33 in the remaining cases [10]. These high risk viruses play a role in the carcinogenic process by producing two oncoproteins encoded by the viral E6 and E7 genes. These oncoproteins directly inactivate p53 and pRb, respectively, and promote cell cycle entry and DNA synthesis as well as blockade of apoptosis, cellular conditions favoring viral replication [11].

At present, the causal role of HPV in cervical carcinomas, has been well established. High-risk HPV DNA is detected in almost all cervical carcinomas [12-14] and this finding has instigated studies on the prevention or treatment of HPV infections using vaccines [15]. As for etiology of HNSCC, the importance of HPV infection is more controversial, mainly because no unanimously accepted detection method exists. The reported prevalences of high-risk HPV DNA tremendously vary per study from 0 to 100% [10;16]. Part of the variation in HPV prevalence can be explained by differences with regard to the location of the tumor, i.e. it is relatively low in the oral cavity and high in the tonsil [1;4;7] More importantly, variations in the type of tissue material studied and the HPV detection methods used, likely have a major impact on the discrepancy in reported prevalence rates. The most widely applied detection methods are based on PCR amplification of viral DNA. However, these methods are extremely sensitive and can even detect a few DNA copies per sample, which might yield false-positive results [4;10;17]. Moreover, HPV-DNA presence does not per se indicate viral involvement in the carcinogenic process and may reflect a transient infection that does not carry any risk of neoplastic transformation [18]. It is our opinion that the lack of agreement of what to consider a HPV-involvement is for a large part the cause of confusion in the HNSCC literature. As an example, some studies reported an association between HPV infection and improved survival [4;19-21] whereas other studies failed to show this [22-24].

Recently, we demonstrated HPV DNA by PCR in 24/143 (16.7%) frozen oral and oropharyngeal tumors, but could confirm viral involvement by E6/E7 expression analysis in only 12/24 (50%) of these samples [22]. This observation shows that presence of HPV DNA solely does not necessarily mean that the virus is biologically
active in these tumors. In addition, it was found that tumors containing transcriptionally active HPV can be genetically distinguished as a specific subgroup [1], strongly supporting the hypothesis that HPV only plays a carcinogenic role when viral oncogenes are expressed. This group of tumors with transcriptionally active HPV showed absence of TP53 mutations and a limited number of genetic abnormalities, whereas their counterparts without transcriptionally active HPV showed TP53 mutations in 75% of cases and many genomic abnormalities [1;2;25;26]. Given this concept, measurement of E6/E7 expression seems to be the most reliable way to detect a biologically relevant association between HPV and malignancy, but a precondition to measure E6/E7 mRNA expression in a user-friendly reliable manner, is the availability of frozen material. A method feasible for more accessible routinely collected material, such as formalin-fixed, paraffin-embedded (FFPE) tissue is still missing.

Commonly used HPV detection methods for FFPE tissue are known from studies on cervical cancer, and include, besides HPV consensus PCR methods, type-specific HPV-DNA detection by Fluorescence In Situ Hybridization (FISH) [27] and real-time PCR assays, the latter allowing viral load analysis [28;29]. Also the use of surrogate biomarkers such as p16 immunostaining can be considered for studying FFPE specimens of HNSCC [7;21;30]. Alternatively, detection of serum antibodies [31] directed against HPV epitopes might be an option when serum has been collected.

In this study we aimed to find a reliable and easily implementable test algorithm to assess a directly carcinogenesis-related HPV involvement in both frozen and FFPE tumor specimens of HNSCC patients.
MATERIAL AND METHODS

Patients, tissue samples and group definitions

For this study we used tumor specimens and serum samples of 48 patients who underwent surgical treatment for HNSCC at the VU University Medical Center. The study was approved by the Institutional Review Board, and informed consent was obtained from all patients. Regarding the tumors a dual work-up was followed: a part of the tumorous tissue was snap-frozen in liquid nitrogen and stored for research purposes [32]. The remaining part of each tumor was processed for routine histopathology and the tissue detection methods that are the subject of this study were performed on this routinely processed material.

Group classification. The frozen tumor specimens were analyzed for the presence of HPV-DNA and E6/E7 mRNA as described previously [1;2]. All HPV DNA-positive carcinomas contained high risk HPV16. For this study we considered presence of HPV16 E6/E7 mRNA in the frozen specimens as 'gold standard' reflecting direct viral involvement in carcinogenesis and this was used as selection criterion for the case group. Among 24 HPV16 DNA-positive tumors, 12 (50%) samples were positive for E6 and E7 transcripts and were considered true HPV-positive [1;2] and further referred to as HPV D+/R+ (DNA+/RNA+). The group of twelve HPV16 DNA-positive, but E6/E7 mRNA-negative tumors was also included in the comparative analysis as an as yet undefined group regarding the HPV status, and was further referred to as the HPV D+/R- group. A third HPV-negative (DNA-/RNA-) control group of 24 tumors were selected in such a way that a similar distribution over the groups was ensured for clinical parameters that might in theory confound the analysis (i.e., age, gender, tumor site and stage, tobacco and alcohol consumption, and histology). After analysis, the HPV D+/R- group showed in general negative test results and together with the negative E6-expression in frozen tissue, this group was scored as being HPV-negative and was included as such in the calculation of sensitivity and specificity.

Preparation of paraffin sections and isolation of nucleic acids

Paraffin sections were prepared according to the sandwich method: the first and last section were stained by haematoxylin and eosin to check for tumor presence and to guide microdissection. One 4 µm section was used for FISH analysis and one 5 µm section for p16 immunohistochemical staining. Routinely 10-20 sections were used for microdissection of the neoplastic cells for viral load analysis. From 10-20 subsequent sections RNA and DNA was isolated simultaneously for GP5+/6+ PCR
and HPV16 E6* RT-PCR (see below). To avoid cross-contamination, a new microtome blade was used each time a new case was sectioned, aerosyl tips were used for all pipetting steps, and separate laboratories were used for pre- and post-PCR handling.

Detection of High Risk HPV DNA by GP5+/6+/PCR
Detection of high risk HPV was performed by general primer GP5+/6+/PCR on 50 ng of DNA, quantified by the Quant-It Picogreen dsDNA assay kit (Invitrogen, Breda, The Netherlands), followed by reverse line blot genotyping [33;34]. Serial dilutions of DNA isolated from cervical carcinoma cell line SiHa (ATCC; HTB35; 1–2 integrated HPV16 copies), and reactions without template were run in parallel as controls.

Viral load analysis
Quantification of HPV16 DNA copy numbers per cell was performed by real-time PCR using the LightCycler® technology (Roche Molecular Biochemicals, Mannheim, Germany) as described previously [35], except that for PCR detection of both HPV16 and beta-globin other primers were employed to shorten the amplicon lengths to 114 and 111 base pairs, respectively (Table 1). Based on the assumption that at least one viral copy is needed per cell for clonal expansion, tumors with > 0.5 copies per cell were scored as positive, allowing some normal tissue contamination in the microdissected tumor samples.

Table 1. Primer and probe sequences used for viral load and E6*I expression analysis.

<table>
<thead>
<tr>
<th>Assay (amplicon length in basepairs)</th>
<th>Target</th>
<th>Sequence 5’ → 3’</th>
<th>Primer location in genome sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral load analysis (114)</td>
<td>HPV16 E7</td>
<td>fw: GAGGAGGAGATGAAATGATGGT</td>
<td>658-681 (24)†</td>
</tr>
<tr>
<td></td>
<td>(K02718)</td>
<td>re: TTGTACGCAACGGAGGAC</td>
<td>722-753 (19)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dp: CAAAAGTACAAATTTGTAATGGC</td>
<td>735-706 (30)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ap: CCGGGTTCTGTGTCAGCTGG</td>
<td>703-682 (22)‡</td>
</tr>
<tr>
<td>Viral load Analysis (111)</td>
<td>Human β-globin</td>
<td>fw: GGAGAAGTCGCCCCTTACTGC</td>
<td>62207-62227 (21)†</td>
</tr>
<tr>
<td></td>
<td>(U01317)</td>
<td>re: TTGTACGCTACAGTCGCTTTG</td>
<td>62318-62295 (24)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dp: CCAGGGCTTACACACCAACTTC</td>
<td>62275-62254 (22)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ap: CCAGGTTACCTCTGCCCAGC</td>
<td>62251-62230 (22)‡</td>
</tr>
<tr>
<td>E6*I expression analysis (86)</td>
<td>HPV16 E6</td>
<td>fw: TTACTGCACGTGAGGTTAGTA</td>
<td>212-226^409-413^409 (20)‡</td>
</tr>
<tr>
<td></td>
<td>(K02718)</td>
<td>re: GGAATCTTTTTTTTTTTTTC</td>
<td>479-460 (19)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pr: AAGCCACTCCTGCTGTCTGGAGAAAGC</td>
<td>425-451 (26)‡</td>
</tr>
<tr>
<td>E6*I expression analysis (100)</td>
<td>BGUS*</td>
<td>fw: GAAAAAATATGGGGAGACCTCGTATT</td>
<td>3462209-3462184 (26)‡</td>
</tr>
<tr>
<td></td>
<td>(NT007758.11)</td>
<td>re: CCAGGTTAGATGCCCCCTTTTTA</td>
<td>3458871-3458850 (22)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tp: CCAAGACCTCTGCGGTCCTGACTG</td>
<td>3458876-3458902 (26)‡</td>
</tr>
</tbody>
</table>

Abbreviations used: fw, forward primer; re, reverse primer; dp, donor probe (3’end labelled with fluorescein); ap, acceptor probe (3’end phosphorylated, for HPV16 5’end labelled with LightCycler-Red-640 and for β-globin 5’end labelled with LightCycler-Red-705); pr, Enzyme-Immuno-Assay (EIA)-probe; tp, TaqMan-probe (5’end labelled with FAM and 3’end labelled TAMRA). †Genbank accession number, ‡Length in basepairs, § This gene is located at chromosome 11, ¶ This gene is located at chromosome 7, ^HPV16 E6*I is spliced between genome position 226 and 409, ¶¶ HPV16 E6 reverse primer is 5’-biotinylated.
Detection of HPV16/18 DNA by FISH
FISH was performed on paraffin-embedded tissue sections as described previously [7]. Controls included hybridizations on FFPE sections of known HPV16- and 18-positive human cervical carcinoma cell lines (CaSki [ATCC; CRL1550; 500 integrated HPV16 copies], HeLa [ATCC; CCL2; 20–50 integrated HPV 18 copies] and SiHa [ATCC; HTB35; 1–2 integrated HPV16 copies]) as well as hybridizations on tissue sections of cervical lesions with proven integration or episomal presence (replication) of HPV genomic DNA to guarantee probe specificity, sensitivity and interpretation accuracy [36]. Negative controls consisted of HPV PCR- and FISH-negative cell lines and tissue sections and hybridizations omitting the viral probe. Evaluation of nuclear hybridization signals was performed by three investigators according to the criteria described by Cooper et al. [37] without information on the HPV-status of the samples. Both staining intensity (0 to 3) and punctate and/or diffuse signals throughout the nucleus indicating integrated and episomal HPV DNA, respectively, were evaluated. The level of inter-observer agreement was determined by calculating Cohen's kappa values [38]. A definitive consensus score was determined by mutual agreement in a separate session.

Immunohistochemical staining of p16
For p16INK4a or short p16 (the protein encoded by CDKN2A) immunohistochemistry the CINtecTM Histology Kit (DakoCytomation B.V., Heverlee, Belgium) was used. For every case analyzed, an extra tissue section was stained with a mouse IgG, as a negative control. Staining intensity as a result of this mouse antibody was considered background and all samples with staining intensity above that background were scored as positive. An extra tissue section of an HPV-positive tumor with high p16 expression was included as positive control. Both the staining intensity (graded 0-3 proportional to staining intensity) and the percentage of the tumor cells positively stained per slide were assessed independently by three investigators. The level of inter-observer agreement was determined by calculating Cohen’s kappa values. A definitive consensus score was determined by mutual agreement in a separate session. The eventual decision if tumor was analyzed was made after consultation of an experienced pathologist (CJLMM).
Detection of antibodies against the proteins HPV16 L1, E6 and E7
Antibodies against L1, E6 and E7 of HPV16 were measured in a sandwich ELISA using a glutathione S-transferase capture method with native recombinant L1-tag, E6-tag and E7-tag proteins [39;40]. All sera were measured at least twice and the median of the absorbance values was taken as the final read out.

Detection of HPV16 E6*I mRNA on paraffin embedded tissue
We developed an RT-PCR assay to detect the most abundant splice variant within the HPV16 E6 open reading frame, namely E6*I, in FFPE specimen [41;42]. For further details about the primer and assay design: see Supplementary Information. Primer sequences are listed in Table 1.

PCR products were detected using an enzyme immunoassay (EIA) as described previously [43] with a HPV16 E6*I specific probe (Table 1) according to a method that was described previously [33]. Samples were scored positive when the EIA signal was above the threshold value of three times the average of the EIA signals of four negative PCR controls. RNA from cells of the HPV16 containing SiHa cell line, that were formalin-fixed and embedded in paraffin, was run in parallel as positive control. For each clinical sample a parallel RT-PCR without reverse transcriptase was used as a control for possible amplification of contaminating HPV16 DNA. RNA integrity as assayed by detection of β-glucuronidase transcripts.
RESULTS
All results are summarized in Table 2. Three groups of carcinomas have been tested, divided according to the HPV DNA and E6/E7 mRNA status as determined in frozen tissue [22]. In general, all tests scored positive in the carcinomas of the D+/R+ group and negative in the D-/R- group. The D+/R- group showed in most cases negative test results and was also negative for E6/E7-expression in frozen tissue. For these reasons this group was scored as being HPV-negative and was included as such in the calculation of sensitivity and specificity.

Of all PCR-based detection methods in FFPE specimen, only E6* mRNA detection showed both a sensitivity and specificity of 100%. Typical examples are shown in Figure 1. FISH analysis showed to be very specific, but was less sensitive: In 10 out of 12 HPV D+/R+ tumors HPV16 DNA was detected. In all positive tumors a punctuated hybridization pattern was observed consistent with viral DNA integration, although some tumors in addition showed patterns suggestive of the presence of episomal viral DNA as well (Figure 2). Three samples could not be analyzed, because of a high background. Between the three investigators, there was total agreement on FISH scoring in 43 of the 45 cases (96%). When the inter-observer agreement was analyzed, an average kappa score of 0.85 was observed. This can be interpreted as a very good strength of agreement [38].
All HPV D+/R+ tumors showed expression of p16 by immunohistochemistry. One sample could not be analyzed as the tissue section appeared not to contain neoplastic cells anymore. Between the three investigators, there was full agreement on p16-IHC scoring, with respect to positive staining intensity (graded 0-3 proportional to staining intensity) in 34 of the 44 cases (77%), with an average kappa score of 0.81. This can be interpreted as a very good strength of inter-observer agreement [38]. Figure 3 shows typical examples of tumors with different staining intensities. We measured HPV16 L1, E6 and E7 antibody levels above the cut-off values in the serum of patients in the HPV D+/R+ group, the HPV D+/R- and the HPV-negative group (Table 2). Regarding the E6 antibodies, a statistically significant difference (P<0.05, Fisher’s exact test) was observed between the HPV D+/R+ group and the HPV D+/R- group as well as the HPV-negative group. The significant difference is based on the number of samples that were above the cut-off level, defined as specified in the legend of Table 2.
Figure 2. Typical examples of HPV16 FISH on sections of paraffin embedded HNSCC specimens. All three tumors were positive for HPV DNA and E6/E7 expression. The images A and B show punctuate nuclear FISH signals indicating HPV16 DNA integrated into the host genome. Image C shows an area with diffuse nuclear FISH staining indicative for episomal HPV16 DNA.

Figure 3. Representative examples of p16 immunostaining on three tumors of the HPV D+/R+ group (A-C) and one of the HPV-negative group (D). A = Tumor 1 (table 2) with 100% of the cells stained with high intensity in both nucleus and cytoplasm; B = Tumor 5 (table 2) with approximately 50% of the cells positively stained with high intensity; C = Tumor 11 (table 2) with 100% of the cells stained, but with a lower intensity; D = Tumor 26 (table 2) negative for p16 immunostaining.
Table 2. Different Methods for Detecting Clinically Relevant HPV Infections in Paraffin embedded Head and Neck Carcinomas and Patient Sera

<table>
<thead>
<tr>
<th>Tumor Site</th>
<th>p16 IHC&lt;sup&gt;1&lt;/sup&gt;</th>
<th>E6*I mRNA&lt;sup&gt;3&lt;/sup&gt;</th>
<th>GP5+/6+ Virgin load&lt;sup&gt;4&lt;/sup&gt;</th>
<th>FISH&lt;sup&gt;5&lt;/sup&gt;</th>
<th>16L1</th>
<th>16E6</th>
<th>16E7</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV DNA- and RNA- positive group&lt;sup&gt;6&lt;/sup&gt; (D+/R+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 OC</td>
<td>3</td>
<td>90</td>
<td>+</td>
<td>+</td>
<td>87.0</td>
<td>2</td>
<td>NA</td>
</tr>
<tr>
<td>2 OP</td>
<td>NE</td>
<td>NE</td>
<td>+</td>
<td>+</td>
<td>0.2</td>
<td>0</td>
<td>383</td>
</tr>
<tr>
<td>3 OP</td>
<td>2</td>
<td>90</td>
<td>+</td>
<td>+</td>
<td>4.5</td>
<td>3</td>
<td>400</td>
</tr>
<tr>
<td>4 OP</td>
<td>3</td>
<td>80</td>
<td>+</td>
<td>+</td>
<td>8.0</td>
<td>2</td>
<td>1509</td>
</tr>
<tr>
<td>5 OP</td>
<td>3</td>
<td>50</td>
<td>+</td>
<td>+</td>
<td>156.6</td>
<td>2</td>
<td>255</td>
</tr>
<tr>
<td>6 OP</td>
<td>2</td>
<td>90</td>
<td>+</td>
<td>+</td>
<td>2.1</td>
<td>2</td>
<td>717</td>
</tr>
<tr>
<td>7 OP</td>
<td>3</td>
<td>80</td>
<td>+</td>
<td>+</td>
<td>316.3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8 OC</td>
<td>3</td>
<td>90</td>
<td>+</td>
<td>+</td>
<td>1.3</td>
<td>2</td>
<td>71</td>
</tr>
<tr>
<td>9 OC</td>
<td>3</td>
<td>80</td>
<td>+</td>
<td>+</td>
<td>10.1</td>
<td>2</td>
<td>55</td>
</tr>
<tr>
<td>10 OC</td>
<td>2</td>
<td>90</td>
<td>+</td>
<td>+</td>
<td>1.6</td>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>11 OC</td>
<td>1</td>
<td>90</td>
<td>+</td>
<td>+</td>
<td>5.3</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>12 OC</td>
<td>2</td>
<td>60</td>
<td>+</td>
<td>+</td>
<td>41.0</td>
<td>2</td>
<td>451</td>
</tr>
<tr>
<td>13 OP</td>
<td>NE</td>
<td>NE</td>
<td>-</td>
<td>-</td>
<td>NE</td>
<td>NE</td>
<td>96</td>
</tr>
<tr>
<td>14 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>15 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16 OC</td>
<td>1</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>17 OP</td>
<td>2</td>
<td>40</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>NE</td>
<td>NA</td>
</tr>
<tr>
<td>18 OC</td>
<td>1</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>135</td>
</tr>
<tr>
<td>19 OC</td>
<td>3</td>
<td>90</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>20 OC</td>
<td>NE</td>
<td>NE</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>21 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>22 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>86</td>
</tr>
<tr>
<td>23 OC</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>24 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>25 OP</td>
<td>1</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>26 OP</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>27 OC</td>
<td>NE</td>
<td>NE</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>68</td>
</tr>
<tr>
<td>28 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>29 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>30 OP</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>NE</td>
<td>40</td>
</tr>
<tr>
<td>31 OP</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>32 OP</td>
<td>1</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>33 OP</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>34 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>173</td>
</tr>
<tr>
<td>35 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>90</td>
</tr>
<tr>
<td>36 OC</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>96</td>
</tr>
<tr>
<td>37 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>38 OC</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>39 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>40 OP</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>41 OP</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>42 OC</td>
<td>3</td>
<td>90</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>43 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>44 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>45 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>46 OC</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>53</td>
</tr>
</tbody>
</table>

<sup>1</sup>p16 IHC: Intensity Percentage
<sup>2</sup>Sera
<sup>3</sup>E6*I mRNA
<sup>4</sup>Viral load
<sup>5</sup>FISH
<sup>6</sup>HPV DNA- and RNA- positive group
<sup>7</sup>HPV DNA-positive, but RNA-negative group
<sup>8</sup>HPV-negative group

CHAPTER 2
Different methods for detecting clinically relevant HPV infections in paraffin embedded head and neck carcinomas and patient sera. Tumors are grouped on basis of the presence of HPV DNA and HPV E6/E7 transcripts as determined in the frozen samples (gold standard); #HPV DNA- and E6/E7 RNA-positive (D+/R+), ##HPV DNA-positive, but E6/E7 RNA-negative (D+/R-), and ###HPV-negative (D-/R-). † p16 IHC staining intensity: 0-3 and the percentage of tumor cells positively stained. ‡ (+) = Positive or (-) negative respectively for E6*I expressing according to the criteria as described in the material and methods. § Viral DNA copy number per cell ‖ FISH staining intensity 0-3 was scored according to Cooper et al. [37]. ¶ The cut-off value to define HPV-antibody-positive sera was calculated separately for each antigen as the median of the specific absorbance values of control sera from individuals who did not have cancer plus three standard deviations excluding positive outliers, as described elsewhere [47]. The samples were scored positive when the optical density (OD)-value was above the following levels: 16L1: 100; 16E6: 150 and 16E7: 60 and are in bold. OC = Oral cavity. OP = Oropharynx. NE = Not evaluable. Some tumors could not be analyzed reliably, as the tissue section did not contain enough neoplastic cells. NA= Not available for serologic detection of HPV16 L1, E6 and E7 antibodies. Sensitivity: Relative number of positive samples detected in the HPV D+/R+ group. Specificity: Relative number of negative samples detected in the HPV DNA+/R- and HPV-negative group together. Sera results: positivity was scored when any of the three markers was scored positive.

Since all single methods showed limitations with respect to sensitivity and specificity, and E6* mRNA detection is not yet available for HPV-types other than 16, combinations of high throughput methods were considered. Our results showed that when p16 immunochemistry is followed by GP5+/6+-PCR on the p16-positive cases, 100% sensitivity and specificity might be reached (Figure 4).

![Proposed flowchart for high throughput identification of HNSCC with a clinically relevant HPV infection on paraffin-embedded tissue sections with 100% sensitivity and specificity.](image-url)
DISCUSSION

The reported large variation of HPV prevalence in HNSCC has complicated understanding of the role of HPV in this tumor type [4;10;16;17] The high sensitivity of the widely used PCR methods for HPV-DNA amplification is a likely explanation for this variation, and a HPV DNA-positive assay may not always reflect biologically meaningful viral involvement [18]. Recent findings point out that the only way to support a conclusive viral involvement is measuring levels of E6/E7 RNA, thus far necessitating frozen tissue material [1;2;26;32]. We aimed to find a sensitive and specific high throughput HPV detection algorithm for FFPE specimen.

The D+/R- group was included with an unknown definitive HPV status and the present results were interpreted to be additional arguments to consider this group of tumors HPV negative and to include this as such in the test performance assessment. This decision was motivated initially by the argument that E6 RT-PCR was negative in frozen tissue. All other data supported this decision: most tests showed negative results, these tumors showed genetic profiles indistinguishable from HPV negative tumors [1;2], and the GP5+/6+-PCR signal intensities in the frozen tissues, the starting situation, were just above background. Regarding this last item, details on how this relatively low level was detected and interpreted has been mentioned in a previous report [22]. Nevertheless, we consider samples that are positive for the DNA assay but negative for the E6/E7 assay as false positive. The presence of HPV DNA in these cases is not relevant for the carcinogenic process. This is supported by all other assays that we performed. How the presence of HPV DNA should be explained is not easy to answer. HPV DNA could be present in the saliva or mucosa without influencing the physiology of the host. It has to be stressed that no matter how this virus ended up in these samples, no viral oncogene expression was detected, making it unlikely that the virus is biologically active in carcinogenesis. The E6*I mRNA detection designed for FFPE material showed a hundred percent sensitivity and specificity and seems very well applicable for large retrospective studies, i.e. the readout is performed by PCR-EIA resulting in a basically dichotomous output. This assay also does not require laborious microdissection. A limitation is that the HPV-type should be known, and similar assays for types other than 16 need to be developed. Nevertheless, this novel assay may already solve current clinical controversies about HPV involvement [44;45].

All methods based on PCR amplification of viral DNA showed limitations with respect to sensitivity and specificity. As expected, the GP5+/6+ PCR assay detected HPV in all tumors of the D+/R+ group. However, only a single case was GP5+/6+-PCR
positive in the D+/R- group, while the corresponding frozen samples were all positive, although with relative low values [22], demonstrating that HPV is more readily detectable in frozen tumor tissue than in paraffin embedded tissue, which is not unexpected. PCR amplification is more efficient on frozen than on FFPE material, since it is known that the fixation procedure leads to DNA fragments that are often shorter than 200 bp [46]. In the group of 24 tumors that were HPV-negative on basis of the frozen biopsy, three corresponding FFPE tumor samples were GP5+/6+ PCR positive. An explanation for this phenomenon might be related for instance to tumor heterogeneity and sampling error. The pieces that were selected for cryopreservation were often small biopsies, while the FFPE samples encompassed transsectional cuts through the whole tumor. It has to be emphasized that the E6/E7 mRNA expression remained negative in these samples, indicating that the virus was not biologically active and it demonstrates that PCR-based HPV DNA detection methods can lead to falsely positive results in terms of biologically active HPV detection.

Quantitative assessment of the viral copy number per cell might improve specificity, but resulted in one false negative observation, and it is very labor-intensive as the neoplastic cells need to be enriched by microdissection [17].

FISH allows for the direct visualization of down to 1 copy of viral DNA per cell and in addition the discrimination between integrated and replicative (episomal) HPV as a punctuated or diffuse hybridization signal, respectively [47]. It is unclear why FISH was unable to identify the virus in the D+/R+ cases two and eleven. The viral load values were relatively low in these cases, suggesting a low viral copy number. In a second instance, sample 11 was repeated on a different section and came out positive. This outcome was not included in the results as we decided to present the first unbiased results. Thus, to increase sensitivity of FISH one could think of testing multiple sections to minimize a possible sampling error. The sensitivity of FISH was 85% (10/12), but the specificity was optimal (100%) in our data set.

Another potentially interesting method, exploiting a surrogate marker to identify clinically relevant HPV infections, is detection of antibodies to E6 and E7 in sera, chosen for its easy applicability and used in several large clinical studies with cervical cancer patients [48-50]. We exploited the potential value of serology and were able to analyze serum samples of most of the patients. There was indeed a significant difference between the HPV D+/R+ group and the HPV D+/R- group, in particular for antibodies against HPV16 E6. Highest sensitivity was reached with positive serology with any of the three antibodies (91%), but the specificity was then limited (74%).
A second method using a surrogate marker is p16 over-expression as it is strongly related to active HPV infection based on the concept that functional inactivation of Rb by E7 induces p16 up-regulation. P16 immunohistochemistry is a relatively standardized technique and easy applicable on FFPE samples. Analysis of p16 over-expression can have clinical value [21] and although we and other investigators found an association with viral oncogene activity [7;51], p16 overexpression is certainly not limited to HPV D+/R+ samples only. Some cases in the HPV D+/R- and the HPV-negative group showed very high p16-expression and these were further analyzed in detail for the possibility that we failed to detect the virus by GP5+/6+ PCR due to viral integration into the host genome in the L1 gene. This was performed by type-specific HPV DNA PCR using primers in the E6/E7 region that also enables detection of all high-risk HPV types when integrated. However, these assays did not yield positive results, indicating that these cases should be considered as truly false-positive cases for p16 immunostaining.

As we found that each single assay seems to meet limitations, which is in line with observations of others [52], we decided to investigate algorithms based on the combination of assays. We could extract an algorithm with a satisfactory performance and allowing high-throughput analysis (Figure 4). P16-immunostaining is upfront and when positive staining is observed GP5+/6+ PCR is used for confirmation. In our series the sensitivity and specificity of this approach is 100%. P16 immunohistochemistry can easily be combined with standard histology when a hematoxylin/eosin (HE)-stained tissue section is prepared for examination by a pathologist. Preselection by p16-staining reduces the workload and the combination gives a dramatic decrease of the number of false-positive observations by either of these assays. Based on the test outcome and the easy applicability we judge this as a reliable algorithm for HPV detection in FFPE specimen. One possible limitation of this algorithm is that both assays may yield a false-positive result in the same sample. However, based on the presently available data the chance that this occurs is around 2%. It should be emphasized that the proposed algorithm should be validated in subsequent studies, particularly in multi-center designs. Our data clearly show that HPV DNA detection by PCR only overestimates the number of samples with a clinically relevant infection.

In conclusion, we propose an algorithm for detecting a clinically relevant HPV infection, applicable for high-throughput analysis of paraffin embedded material, using a combination of p16 immunohistochemistry and GP5+/6+ PCR. This algorithm may provide the opportunity to assess the role of HPV in large retrospective studies,
and could be helpful to select patients who may benefit from immunotherapeutic strategies.

ACKNOWLEDGEMENTS
We thank Prof. Dr J.J. Manni for continuous support, and Dr. A.H.N. Hopman for support and evaluation of the FISH experiments. We also thank J. de Vries for help with the p16-immunostaining and H. Verdurmen and N. Fransen-Daalmeijer for their assistance in the PCR-based detection methods.
SUPPLEMENTARY INFORMATION

Detection of HPV16 E6*I mRNA on paraffin embedded tissue

Because of the potential value of measuring HPV-oncogene expression in FFPE specimens, we set out to develop an RT-PCR assay to detect the most abundant splice variant within the HPV16 E6 open reading frame, namely E6*I [41;42]. Last few years it has become increasingly clear that RNA can be isolated from paraffin material and analyzed by RT-PCR [22;47]. DNA and RNA can be isolated simultaneously from paraffin sections using proteinase K digestion, followed by phenol extraction and ethanol precipitation as described previously.[34] Based on our own experience RT-PCR on the RNA in these preparations works well and reliably, but only when the amplicons are approximately 100 bp (data not shown) and when the primers have been designed to prevent PCR amplification of the co-isolated DNA templates. The latter issue is in particular a problem for HPV genes as the open reading frames are usually unspliced transcripts from the viral genome. There is an important exception for a splice variant of the E6 gene, named E6*I. This splice variant misses 162 base pairs of the full length transcript. In an attempt to solve the problem of unwanted HPV DNA amplification, one of the primers was designed over the splice-site location of the E6*I splice variant of the E6 gene. A schematic view of the primer and probe design is shown in supplementary Figure 1. The primer set was analysed on a mixture of RNA and DNA isolated from frozen samples of CaSki and SiHa HPV16-positive cervical cancer cell line with high and low viral copy number (CaSki [ATCC; CRL1550; 500 integrated HPV16 copies] and SiHa [ATCC; HTB35; 1–2 integrated HPV16 copies]). The detection limit was determined by running a serial dilution of RNA isolated from SiHa cells in a background of normal human RNA. In total 100 pg of SiHa RNA (20 cells) in a background of 100 nanograms normal human RNA (20,000 cells) was still detectable (Data not shown). The results found with frozen and FFPE samples of these cell lines are shown in supplementary Figure 2 and explained in the legend.
Supplementary Figure 1. Schematic view of primer and probe design for E6*I mRNA detection. The genomic sequence of HPV 16 is shown from basepair 200 to 500. Exons sequences are in CAPITAL letters, intron sequences are in italics. The forward primer is spanning the splice-site specific for the E6*I transcript with exactly 10 bases located on each different exon. Homology of at least 5 bases at the 3'-end of the forward primer is indicated above the intron sequence.

Despite positioning of one of the primers over the splice site, the primer set did not detect E6*I only, but also the full-length E6 transcript and consequently also viral DNA in both cell lines (Figure 2). This is likely caused by a large sequence homology of 5 bases at the 3'-end of the forward primer to the 5' beginning of the intron sequence (supplementary Figure 1). In an attempt to reduce this unwanted amplification of unspliced E6 and viral DNA we evaluated two other primers located two and four base pairs further over the splice-site, but these gave no improvement. The type of the material analysed (frozen or FFPE) strongly influences the amplimers found. In RNA of frozen samples both the 248 (full length E6) and 86 bp (E6*I) base pairs fragment are seen. In FFPE samples only the E6*I amplimer (86 bp) is detectable. Moreover, also the contaminating DNA in the RNA preparations still gives rise to amplification of the 248 bp fragment, and we introduced RNAsase-free DNAse treatment in the isolation procedure to avoid this further.
Supplementary Figure 2. Validation of the new primer set designed for detection of E6*I mRNA in FFPE specimen. The Figure shows the PCR amplimer separated on an agarose gel and the corresponding EIA results, obtained from CaSki and SiHa RNA/DNA with different combinations of DNAse treatment and Reverse Transciptase. C = CaSki DNA/RNA; S = SiHa DNA/RNA; (-) = Negative control; The -DNAse/ -RT reaction represents a PCR reaction without DNase treatment and reverse transcriptase, and shows viral DNA amplification (248 bp). CaSki contains a much higher viral copy number compared to SiHa, which results in a more intense amplimer band of 248 basepairs compared to SiHa. The -DNAse/ +RT represents a PCR reaction without DNase, but with reverse transcriptase and shows the amplification of the E6*I mRNA at 86 basepairs, the full-length E6 transcript and viral DNA of 248 base pairs.

The +DNAse/ -RT represents the reaction with DNAse, but without reverse transcriptase and shows that viral DNA amplification was completely abolished. The +DNAse/ +RT shows the reaction with DNAse treatment and reverse transcriptase. Only the E6*I (86 bp) and full length E6 mRNA (248 bp) were detected in frozen material. In FFPE only the small fragment of 86 bp of E6*I is detected. The specificity of the generated amplimer was checked with the use of a specific probe in an EIA setting.

The quality and quantity of the RNA preparation of the tumor samples was tested by quantitative RT-PCR (TaqMan, Applied Biosystems, The Netherlands) for the housekeeping gene BGUS (β-glucuronidase). Primers are listed in table 1 and PCR conditions were as described previously [53]. The standard assay for the FFPE was as follows. A minimum of 1 ng (relative to BGUS expression of cell line UM-SCC-22A) of extracted tumor RNA was used for E6*I cDNA synthesis. To eradicate the unwanted amplification of co-isolated DNA we pre-treated the sample with 0.5 U RQ1 DNAse (Promega, Leiden, The Netherlands). In a reaction without reverse transcriptase this treatment is checked. The reverse transcriptase reaction was performed with 20 pmol biotinylated HPV16 E6*I reverse primer (Table 1). PCR analysis of the cDNAs was performed in 50 μl containing 50 mM KCl, 10 mM Tris HCl pH 8.3, 200 μM of each dNTP, 1.5 mM MgCl₂, 1 unit Taq polymerase, and 25 pmol of the forward and reverse primer (Table 1). Cycling conditions were 4 min at 95°C, followed by 40 cycles of 30 seconds at 95°C, 60 seconds at 55°C and 90 seconds at 72°C followed by 4 minutes of elongation at 72°C.
REFERENCES


CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data


Mark A. van de Wiel
Serge J. Smeets
Ruud H. Brakenhoff
Bauke Ylstra
SUMMARY
We compute P-values, based on the Wilcoxon test with ties, to compare two conditions with array comparative genomic hybridization data, and we provide a simple interface to export and plot these P-values.
Array comparative genomic hybridization (array CGH) is applied to the detection of genomic abnormalities in cancer and inheritable DNA copy number aberrations that cause genetic disorders. It is a high-resolution, high-throughput technique that allows for genome-wide measurement of chromosom al DNA copy number changes and determination of the associated breakpoints along the chromosomes [1].

Software such as aCGHsmooth [2] and similar programs [3] enables the visualization and identification of aberrated chromosomal regions by individual separately. CGH-Miner [4] has additional features to summarize alteration information over groups. We developed CGHMultiArray to integrate array CGH data over individuals by computing P-values per clone and visualizing these to find generic patterns. The program deals with the most common situation: comparison of two conditions.

When considering a suitable statistic to measure generic DNA copy number changes among individuals, we have to consider the nature of array CGH data. Although technical errors may disperse the data somewhat, the data in reality represent discrete levels of genetic aberrations. The normal DNA copy number of mammalian clones is two: one from both the paternal and maternal chromosomes. In particular diseases, such as cancer, changes in the DNA copy number with respect to the ‘normal’ value may occur as a ‘deletion’ (at least one copy is lost) or a ‘gain’ (at least one additional copy is present). These non-normal levels may be further detailed, e.g. by including ‘amplification’, which is a high level of copy number gains.

The granularity of (discretized) CGH data makes the t-statistic, or variations thereof, unsuitable. Moreover, the discrete levels possess a natural ordering, which rules out the Fisher exact test. The Wilcoxon test makes explicit use of both features. However, the data naturally contain many ties, i.e. sets of equal observations. The distribution of the Wilcoxon statistic, and consequently the P-values, depends on the tie structure [5]. Hence, it has to be re-computed for each new case.

Define the Wilcoxon statistic $W$ as the sum of the mid-ranks assigned to the smallest sample. The observed value of $W$ is denoted by $w$. The two-sided P-value is then defined by

$$2P(W \leq w)$$

if $w \leq E(W)$, and $2P(W \geq w)$ otherwise, where probabilities and expected values are computed under the null hypothesis of equally likely permutations of the mid-ranks. Since the number of tests is of the order of thousands, one needs a fast calculation method. Moreover, asymptotic theory is often not applicable, because the number of biological replicates per condition is small and the presence of ties worsens the accuracy of asymptotic approximations. For example, when both sample sizes equal eight, cases with asymptotic P-value approximations in the range
0.0001–0.05 correspond to 2–3 times larger exact (true) P-values, which leads to more than a doubling of the number of false calls when using the approximations. Therefore, a fast algorithm to compute exact P-values is needed. The relevance of such algorithms to solve bioinformatics problems was recently shown by Bejerano et al. [6]. We developed the split-up algorithm [7], which suits the requirements well: it is fast, exact and deals with ties. The algorithm represents the probability distribution of the test statistic under the null hypothesis ('no change between two conditions') as a generating function. Baglivo et al. [8] showed that generating functions are powerful tools to represent null distributions of discrete test statistics. We used the generating function introduced by Streitberg and Röhmel [9]. This generating function is a polynomial in product form, expansion of which would reveal the entire null distribution, but this may be time consuming. The split-up algorithm splits the product into two parts and requires the expansion of these two smaller parts, which are several orders of magnitude faster than full expansion. Then, these two results are efficiently combined to compute the P-value. CGHMultiArray is written in Mathematica [10]. The basic algorithm is also available as R code and as an executable. To make the algorithm easily accessible, we provide a web implementation too.

The website provides a tool to convert smoothed log2-ratios from other software to input data for CGHMultiArray. First, it transforms observed array CGH log2 values to discretized data: ‘1’ for gains, ‘−1’ for losses and ‘0’ for normals. It allows for the introduction of extra levels for amplifications or double deletions. Next, it counts occurrences of all levels for both conditions. The input for CGHMultiArray then consists of a simple text file, the rows of which represent clones in the chromosomal order. When \( \ell \) is the number of levels used (three in this example), the first (second) \( \ell \) columns represent counts of the number of control (treatment) samples attaining level \( j \), \( j = 1, \ldots, \ell \). Optional columns may provide name, chromosomal information and base pair position information, which allows for separate P-value plots by chromosome. The algorithm stores all count configurations, so P-value computations for clones with the same configuration are not repeated. More details are provided in the manual. Note that when one would like to use only \( \ell = 2 \) levels, e.g. for comparing gains, the Wilcoxon statistic with tie correction is equivalent to the hypergeometric statistic.

We have used CGHMultiArray to analyze the genomic changes of two groups of 12 head and neck tumors. Illustrative data shown here (also available on the website) are identical to the real data except for permutations of chromosomes.
CGHMultiArray generates an exportable list of univariate P-values, a DNA view of these (Fig. 1) and, optionally, a view by chromosome (data not shown). These views are useful to identify regions with unusually many differential aberrations. One may wish to perform a multiple testing correction to the P-values afterwards, such as the Benjamini and Yekutieli [11] FDR rule. In this case, one may want to focus on a limited number of clones or consider chromosomal regions instead of separate clones. An implementation for the latter option is provided on the website.

Figure 1. DNA view of P-values by position for an illustrative dataset

ACKNOWLEDGEMENTS
Part of this research has been funded by the Dutch BSIK/BRICKS project. We thank Marko Boon, Kyung In Kim and Marcel van Verk for their software support.
REFERENCES


CHAPTER 4

Genome wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene expressing human papillomavirus

Oncogene (2006); 25(17):2558–64

Serge J. Smeets
Boudewijn J.M. Braakhuis
Saman Abbas
Peter J.F. Snijders
Bauke Ylstra
Mark A. van de Wiel
Gerrit A. Meijer
C. René Leemans
Ruud H. Brakenhoff
ABSTRACT
Oncogene expressing human papillomavirus type 16 (HPV16) is found in a subset of head and neck squamous cell carcinomas (HNSCC). HPV16 drives carcinogenesis by inactivating p53 and pRb with the viral oncoproteins E6 and E7, paralleled by a low level of mutations in $TP53$ and allelic loss at 3p, 9p and 17p, genetic changes frequently found in HNSCCs of non-viral etiology. We hypothesize that two pathways to HNSCC exist: one determined by HPV16 and one by environmental carcinogens. To define the critical genetic events in these two pathways, we now present a detailed genome analysis of HNSCC with and without HPV16 involvement by employing high resolution micro-array comparative genomic hybridization. Four regions showed alterations in HPV-negative tumors that were absent in HPV-positive tumors: losses at 3p11.2-26.3, 5q11.2-35.2, 9p21.1-24, and gains/amplifications at 11q12.1-13.4. Also, HPV16-negative tumors demonstrated loss at 18q12.1-23, in contrast to gain in HPV16-positive tumors. Seven regions were altered at high frequency (>33%) in both groups: gains at 3q22.2-qter, 5p15.2-pter, 8p11.2-qter, 9q22-34.1, 20p-20q and losses at 11q14.1-qter and 13q11-33. These data show that HNSCC arising by environmental carcinogens are characterized by genetic alterations that differ from those observed in HPV16-induced HNSCC, and most likely occur early in carcinogenesis. A number of genetic changes are shared in both tumor groups and can be considered crucial in the later stages of HNSCC progression.
INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) comprises about 5% of all newly diagnosed cancer cases in the Northern and Western European countries and the United States, and is the fifth most common cancer worldwide [1]. Despite advances in local tumor control, the five-year survival rates of approximately 50% have only moderately improved during the last 20 years, and identification of the cancer genes causally involved in carcinogenesis seems crucial to enable development of new therapeutic agents aimed at improving survival. HNSCC arises by the accumulation of genetic and epigenetic changes in oncogenes, tumor suppressor genes and/or DNA stability genes [2]. Using cytogenetics, loss of heterozygosity (LOH) analysis, and comparative genomic hybridization (CGH), over 20 recurrent chromosomal alterations have been found in invasive HNSCCs, including losses at 3p, 9p and 17p, and gains at 3q26 and 11q13 [3-7]. The relatively high frequencies of these chromosomal aberrations, which occur usually in over 30% of carcinomas, or the association with prognosis, argue for a role in head and neck carcinogenesis.

HNSCC arises by a chemical etiology encompassing well-established causative life style related agents, like tobacco smoking and alcohol abuse. Recently, the role of the human papillomavirus (HPV) in head and neck carcinogenesis as a separate etiologic factor has been firmly established [8-10]. Human papillomaviruses are epitheliotropic DNA viruses with a genome of approximately 8 kb. More than 100 subtypes are distinguished at present, and some subtypes have been assigned as high-risk types, as they are the main causative factors in the development of cervical cancer [11]. The virus produces two oncoproteins, encoded by the E6 and E7 genes, which inactivate the p53 and pRb proteins, respectively, providing cell cycle entry and DNA synthesis needed for viral replication. It was recently shown that tumors containing oncogene expressing HPV are genetically different from those that do not contain HPV and arose most likely as a result from exposure to environmental carcinogens. Tumors with oncogene expressing HPV (in particular HPV16) did not show any TP53 mutation and very limited allelic losses at 3p, 9p and 17p, while in 75% of tumors not containing HPV a TP53 mutation had occurred and very frequent allelic loss was found [12]. These results suggest that an infection with HPV is an early event in HNSCC development that persists throughout the entire period of cancer progression. Moreover, the disruption of the pRb and p53 pathways by the viral oncoproteins is reflected in a unique genotype of tumors with oncogene expressing HPV.
These two different genetic pathways caused by chemical carcinogens and HPV leading to one single tumor type form a unique opportunity to detect the relevant genetic events in head and neck cancer. We now present for the first time a detailed and comprehensive genetic analysis of HNSCCs of different etiology. With a genome wide high-resolution method we determined the chromosomal regions that show different alterations between tumors with and without oncogene expressing HPV, and those that are shared, and combined the data to propose an integrated HNSCC progression model.

RESULTS
CGH profiles of HPV-positive and HPV-negative tumors
Twelve tumors that were HPV DNA-positive with E6/E7 expression and twelve HPV-negative tumors were selected from a previously published cohort of 143 tumors that had been analyzed for HPV and TP53 mutations [12], and maCGH was carried out. The two groups of tumors did not differ statistically with regard to patient age at diagnosis or tumor site (Table 1). The consumption of alcohol and tobacco was also not different between the groups (details in Table 1). The maCGH profiles were smoothed to define the regions that show copy number changes. The significance of the data was checked by calculating the 99% confidence interval on the log2 ratios of non-altered chromosomes for each sample. All values indicated as gain or loss by the smoothing algorithm exceeded the 99% confidence interval level, an indication for the reliability of the dataset. The total number of BAC-clones showing an alteration varied considerably between tumors: from 886 to 3,969 clones for HPV-negative tumors, and from 256 to 2,831 BAC clones for HPV-positive tumors. The frequency plot of alterations per BAC-clone of chromosome 1 to 22 is shown in Fig. 1. On average, the HPV-positive tumors showed a significantly (p=0.04, normal scores two-sample rank test) lower total number of alterations, as compared with HPV-negative tumors, i.e., 28% versus 44% of the BAC-clones studied, respectively. One HPV-positive tumor showed a remarkably low number of copy number changes (alterations for only 256 BAC clones), showing only gain of a small region of 11q and an extra copy of chromosome 20.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Case group HPV-positive* (n=12)</th>
<th>Control group HPV-negative* (n=12)</th>
<th>Difference between cases and controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>75</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Age at diagnosis, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-50</td>
<td>4</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>51-60</td>
<td>4</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>61-75</td>
<td>4</td>
<td>33</td>
<td>4</td>
</tr>
<tr>
<td>Tobacco use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Former</td>
<td>4</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>Current</td>
<td>7</td>
<td>58</td>
<td>8</td>
</tr>
<tr>
<td>Total amount of tobacco consumed,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pack years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 10</td>
<td>1</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>≥10 to &lt;40</td>
<td>9</td>
<td>75</td>
<td>4</td>
</tr>
<tr>
<td>≥40</td>
<td>2</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Alcohol use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Former</td>
<td>3</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Current</td>
<td>9</td>
<td>75</td>
<td>9</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral cavity</td>
<td>6</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>6</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>Tonsil</td>
<td>5</td>
<td>42</td>
<td>4</td>
</tr>
<tr>
<td>Nontonsil</td>
<td>1</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>TP53 gene mutation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>no</td>
<td>12</td>
<td>100</td>
<td>2</td>
</tr>
</tbody>
</table>

*HPV-positive: positive for HPV DNA and expression of E6/E7; HPV-negative: negative for HPV DNA.
†For the 2 x 2 comparison the Fisher’s exact test was used; for the other comparisons the Chi-square test was applied. A P-value <0.05 was considered to reflect a significant difference.
§The frequency of tonsillar cancers was compared with cancer at all other sites.
Genetic alterations different between HPV-positive and HPV-negative tumors
We analyzed the chromosomal regions that showed significant differences between the HPV-positive and HPV-negative tumors. Significantly (p<0.05) differential genetic alterations were found in nine regions (Table 2). Each significant chromosomal region consisted of spatially consecutive BAC clones and all regions encompassed in total 505 (12.6%) of the 3,969 clones analyzed. All significant regions spanned large parts of chromosomes or contained an entire chromosomal arm. The smallest different region just spanned four megabases and was located on chromosome 11q12.1 - 11q13.4. For some chromosomal regions the significance was based on a small number of tumors with opposite alterations. For example, the region 1p31.1 - 13.1 showed gains in two HPV-positive cases compared to losses in four cases of the HPV-negative group. In this example the opposite value (-1 vs. +1) of losses versus gains largely determined the significance. We therefore calculated the false discovery rate (FDR) on our data to adjust for multiple testing. FDR correction is likely to be conservative considering the relatively small number of cases, but several
differentially altered regions at various chromosomes remained highly significant, as indicated by relatively low FDR-values. The FDR value of 0.3 as for region 1p31.1 - 13.1 indicates that the relevance of this finding should be interpreted with caution, and we therefore focused particularly on the regions with p values < 0.05 and low FDR values.

Table 2. Chromosomal areas that are significantly different between HPV-positive and HPV-negative HNSCC

<table>
<thead>
<tr>
<th>Region</th>
<th>Size (Mbase)</th>
<th>HPV-positive (N)</th>
<th>HPV-negative (N)</th>
<th>Difference between HPV-positive and -negative</th>
<th>P-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Gain</td>
<td>Ampl</td>
<td>Loss</td>
<td>Gain</td>
<td>Ampl</td>
</tr>
<tr>
<td>1p31.1 - 31.1</td>
<td>39</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3p11.2 – 26.3</td>
<td>87</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5q11.2 – 35.2</td>
<td>126</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7p14.3 - 21.3</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7q13.22 - 21.11</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>9p21.1 – 24</td>
<td>34</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>11q12.1 – 13.4</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>18q12.1 – 23</td>
<td>52</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21p - 21q</td>
<td>62</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

NOTE: Frequency of chromosomal regions with significantly different (P≤0.05; see Materials and Methods for the statistical test) alterations between HPV-positive and HPV-negative tumors are depicted. Amplifications were scored, when the 2-log value of the peak exceeded 1. Most significant regions, based on the criteria of p<0.05 and a relatively low FDR value are indicated in bold. Ampl. = Amplification. FDR = False Discovery Rate *Significance of the region 9p21.1 – 24 was reached only when the data were dichotomized in ‘losses’ and ‘no-losses’.

Based on FDR correction five chromosomal areas were identified that displayed a very pronounced difference of alterations between the groups fulfilling the significance criterion (p<0.05). Four of these chromosomal regions were altered in a relatively large number of HPV-negative tumors, while hardly changed in the HPV-positive group. First, loss of 3p11.2–26.3 was observed in more than 10 HPV-negative cases, whereas in HPV-positive cases once a gain and twice a loss were found. Region 5q11.2–35.2 was lost in 8 cases of the HPV-negative group, whereas the 5q-arm was lost in just one case and gained in one case in the HPV-positive group. In addition, loss of 9p24-9p21.1 was seen in six cases of the HPV-negative group while this was not observed in the HPV-positive cases. Most striking observation was the gain or amplification of the 11q12.1–11q13.4 region in ten cases of the HPV-negative group while this gain was completely absent in the HPV-positive group. Furthermore, the difference of the region 18q12.1-18q23 was remarkable as it was lost in seven HPV-negative tumors and gained in five HPV-positive tumors (Table 2).
**Genetic alterations common HPV-positive and HPV-negative tumors.**

Besides differential changes between HPV-negative and HPV-positive tumors also seven genomic regions were found with a high frequency of common gains and losses. Regions that showed common gains in more than 4 of 12 cases in each group were on chromosome arms 3q, 5p, 8p, 8q, 9q, 20p and 20q (Table 3). Regions that showed common chromosomal losses in more than 4 of 12 cases in each group were detected on chromosome arms 11q and 13q (Table 3).

**Table 3.** Regions with frequent chromosomal alterations common in HPV-positive and HPV-negative HNSCC

<table>
<thead>
<tr>
<th>Region</th>
<th>Size (Mbases)</th>
<th>Alteration</th>
<th>HPV(-) N (%)</th>
<th>HPV(+) N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3q22.2-qter</td>
<td>98</td>
<td>Gain</td>
<td>9 (75)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>5p15.2-pter</td>
<td>36</td>
<td>Gain</td>
<td>6 (50)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>8p11.2-qter</td>
<td>111</td>
<td>Gain</td>
<td>10 (83)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>9q22-34.1</td>
<td>49</td>
<td>Gain</td>
<td>6 (50)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>11q14.1-qter</td>
<td>54</td>
<td>Loss</td>
<td>8 (66)</td>
<td>8 (66)</td>
</tr>
<tr>
<td>13q11-33</td>
<td>95</td>
<td>Loss</td>
<td>6 (50)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>20p – 20q</td>
<td>47</td>
<td>Gain</td>
<td>6 (50)</td>
<td>5 (42)</td>
</tr>
</tbody>
</table>

**NOTE:** Genetic alterations common among HPV-positive and HPV-negative HNSCC. A region was scored as being altered when at least 4 out of 12 tumors in both groups showed alteration.
DISCUSSION

The role of HPV as an important carcinogenic agent in a subset of HNSCCs has been well-established at present. Very recently Braakhuis et al. [12] suggested, based on the analysis of HPV status, TP53 mutation and loss of heterozygosity (LOH), that two genetic routes of multi-step head and neck carcinogenesis exist. According to their hypothesis the first and most frequent route results from chemical carcinogen exposure and is characterized by frequent TP53 mutations and allelic loss of large chromosomal regions at 3p, 9p, and 17p, which are considered early events in HNSCC development [9;13-18]. The second route is associated with active HPV infection and characterized by lack of TP53 mutations and a low level of allelic loss. These data were interpreted as proof for the concept that HPV16 is actively involved in the early steps of the development of a subgroup of HNSCC and that the virus remains active during the entire carcinogenic process [12;19]. We now show the common and differential genetic alterations between the two groups of tumors.

Four chromosomal regions showed more alterations in HPV-negative tumors (either gains or losses) than in HPV-positive tumors (hardly any or no changes): loss at 3p11.2-26.3, 5q11.2-35.2, 9p21.1-24, and gain/amplification at 11q12.1-13.4. Since there is evidence that an active HPV infection occurs early in cervical and HNSCC carcinogenesis [11;12], the genetic alterations that have been found presently in the HPV-negative tumors only, are likely to reflect the early carcinogenesis. These alterations can be considered an alternative for the biological effects of the viral oncogenes E6 and E7. The frequent gain/amplification at 11q13 and the frequent loss at the region 9p21 in HPV-negative tumors, while showing no alteration in HPV-positive tumors seem biologically plausible in this respect. The most likely candidate oncogene at the highly significantly amplified region 11q13 is CCND1 (Cyclin D1), a component of the CDK/Cyclin complex that phosphorylates pRb causing S-phase entry. In addition, the region 9p21 encompasses the tumor suppressor gene CDKN2A (p16), an important inhibitor of the CDK/Cyclin complex preventing S-phase entry, that is often altered in HNSCC. In HPV-infected tumors the pRb pathway is disrupted by the oncoprotein E7 while apparently in HPV-negative tumors the pathway is impaired by mutation or loss of p16 and/or amplification of CCND1. In the present study 5/12 tumors showed both loss at 9p21 and gain/amplification at 11q13, suggesting that in some cases it seems necessary to alter two components of the pRb pathway to be at least functionally equivalent to pRb binding by HPV E7. This hypothesis needs to be proven by in vitro or mouse models.
In contrast to the biologically plausible findings concerning the pRb pathway, we did not observe a difference at a region we also expected to find, namely, the region 17p13, encompassing the tumor suppressor gene TP53. The viral oncogene E6 binds and inactivates p53, while in tumors without HPV involvement TP53 gene is mostly inactivated by mutation, often accompanied with allelic loss [12;20-22]. In a previous study, we showed that a significant difference in allelic loss frequency at 17p was observed between HPV-positive and HPV-negative tumors as determined by microsatellite markers [12]. In the HPV-negative HNSCC frequent TP53 mutations and allelic loss at 17p was found, while absent in the HPV-positive tumors. With maCGH we found frequent copy number gains at this locus in the HPV-negative group, while expecting copy number losses. This paradox can be explained by the fact that using microsatellite markers allelic imbalance is measured, either gain or loss of an allele. In the present maCGH analysis a mixture of gains and losses was observed in the HPV-negative tumors, making it impossible to detect the difference.

Some of the loci that show significant differences between HPV-positive and HPV-negative tumors have been reported in other studies that focused on the identification of genetic events associated with prognosis of HNSCC. Loss of 3p and gain or amplification of 11q13 were previously identified by Bockmuhl et al. [3] as alterations, associated with poor clinical outcome. Also in the study of Wreesmann et al. [7], amplification of 11q13 and deletion of 5q11-15 amongst other changes were shown to have a prognostic value in HNSCC. These results are in line with previous observations that HPV-positive tumors have in general a better prognosis [10]. An intrinsic problem with these approaches, however, remains that genetic changes that are specifically linked to clinical failures, may only reflect late progression and do not allow identification of chromosomal loci that play a role in the earlier phase of carcinogenesis. Moreover, the different outcomes, distant metastasis, locoregional recurrence as result of residual cancer cells or second field tumors [22;24], are very different biologically entities and most likely associated with different genetic changes.

The majority of loci presented here showed losses or gains in HPV-negative tumors and no apparent changes in HPV-positive tumors. There appeared to be one exception. Chromosomal region 18q12.1-23 was lost in 7 of 12 (58%) HPV-negative tumors, but appeared to be gained in 5 of 12 (41%) HPV-positive tumors. SMAD4 has been proposed as a candidate cancer gene at 18q21.1, as deletion and mutation of this gene in a subset of HNSCC was observed [25;26]. Whether this or another gene is involved in HPV-induced HNSCC remains to be determined.
Besides differences, both tumor groups were shown to share several altered regions. Gains at 3q22.2-qter, 5p15.2-pter, 8p11.2-8qter, 9q22-34.1, 20p-20q, and loss of 11q14.1-qter (telomeric from the 11q13 amplicon) and 13q11-33 were detected in often more than 50% of the tumors in both groups. These can be considered genetic changes that are important in head and neck carcinogenesis as they frequently occur independent from the etiological factors causing these tumors. The question can be asked how these common alterations should be placed in the time-frame of HNSCC multistep carcinogenesis. The most likely option is that these common alterations occur late in progression.

The results of the present study are consistent with and provide further evidence for the hypothesis that HNSCCs develop by two different etiologies: one driven by exposure to environmental carcinogens (i.e. tobacco and alcohol) without HPV involvement and the other involving infection with oncogene expressing HPV16 [10;27;28]. However, we now also provide evidence that the two carcinogenic routes – one with HPV and one without HPV involvement – partly overlap. As stated above, this finding may implicate that these common alterations are necessary events in HNSCC, irrespective the etiological factor. It can not be excluded, however, that the overlapping genetic events are related to the exposure to cigarette smoke. Most of our HPV-positive patients have smoked or are current smokers, making this group not different from the normal HNSCC patient population in that respect. Our present results show that although the two carcinogenic routes, virus and exposure to environmental carcinogens, differ in multiple ways, particularly related to the early genetic events and the activity of the known viral oncogenes E6 and E7, the late genetic events are common. The results of the present study indicate that to describe the development of HNSCC an integrated genetic progression model needs to be considered. This model presented in Fig. 2 builds upon the previous existing models but now includes the HPV route and the different and common genetic events [9;18].
Figure 2. A genetic progression model of multi-step head and neck carcinogenesis is proposed. The two etiological factors, smoking and HPV16 are incorporated into the “patch-field-cancer” model that in essence has previously been published [9]. In this model the development of a field with genetically altered cells play a central role. In the initial phase a “patch” develops, a clonal unit in which the stem cell and its daughter cells acquire a genetic alteration. Some evidence point to an alteration in p53/MDM2 pathway as a likely first event. This pathway is disrupted by a mutation of TP53 in case of smoking as a causative factor, or alternatively, by the HPV effect of E6, that results in a degradation of p53. The conversion of a patch into a field is the next step in this progression model, and this field expands at the expense of normal epithelium. An important event in the field phase is the impairment of the p16/CDK/pRb pathway [22]. In case of smoking this pathway is disrupted by inactivating CDKN2A, the gene encoding p16, by mutation, chromosomal loss or promoter hypermethylation. Alternatively, this pathway can be impaired by E7, a protein that is produced by HPV. Next, clonal divergence leads to the development of one or more tumors within the contiguous field of preneoplastic cells. Considering the two etiological factors, the common and differential chromosomal events possibly involved in the progression from field to carcinoma are shown. At this moment it is difficult to point out the pathways that are involved in this stage of carcinogenesis. CCND1 (Cyclin D1) is likely involved in smoking-related HNSCC, since the 11q13 region is found to be highly significantly amplified; cyclin D1 is part of the p16/CDK/pRb pathway, a component of the CDK/Cyclin complex that phosphorylates pRb causing S-phase entry.
MATERIAL AND METHODS

Patients and tumor specimens
We obtained tumor specimens from 143 consecutive patients who underwent surgical treatment for a tumor in the head and neck region at the VU University Medical Center. Primary tumor samples were immediately snap-frozen in liquid nitrogen and stored at −80 oC. In twelve tumors HPV16 DNA and E6/E7 mRNA expression was determined. This was considered to reflect direct viral involvement and was used as selection criterion for the case group [12]. The HPV-positive case group was compared with a HPV DNA-negative control group of 12 HNSCC, in part (10/12) similar as the group previously published [12]. Cases and controls were stratified according to tumor site (oral cavity or oropharynx), as well as degree of differentiation and further selected in such a way that a similar distribution over the groups was ensured for those clinical parameters that might in theory confound the analysis (i.e., age, sex, smoking and drinking behavior and TNM-stage). The Fisher’s exact test and chi-square test were used to assess the statistical significance of frequency distributions between case and control groups (Table 1). The study was approved by the Institutional Review Board of the VU University Medical Center, and written informed consent was obtained from all patients.

Information on patient tobacco and alcohol use was obtained from the medical files. Patients were classified as never, former (at least one year of tobacco abstinence), or current daily tobacco smokers as well as never, former, or current alcohol drinkers.

Detection of high-risk HPV DNA and E6/E7 transcripts
Frozen tumor samples were used for HPV analysis and for all tumor samples microdissection was performed to enrich for tumor tissue. DNA and RNA was extracted as described previously [22]. Detection of the presence of HPV16 DNA and expression of the oncogenes E6 and E7 was performed as previously described [12].

Micro-array Comparative Genomic Hybridization
Arrays of Bacterial Artificial Chromosomes (BACs) were prepared from the 1 Mb resolution Sanger BAC set (http://www.ensembl.org/Homo_sapiens/index.html), and the OncoBAC set (http://informa.bio.caltech.edu/Bac_onc.html), and in house clones of interest, supplemented with clones from the Children’s Hospital Oakland Research Institute (CHORI), amounting to a total of 5,659 clones with known chromosomal
positions. DNA of BAC clones was isolated according to published protocols available on the web (http://bacpac.chori.org/dnaprep.htm).

Amplification of BAC clone DNA was performed by ligation-mediated polymerase chain reaction (PCR) according to Snijders et al. [23], followed by purification with Montage PCR µ96 filter plates (Millipore BV, Amsterdam, the Netherlands). This DNA was spotted at a concentration of 10 µg/µL in 150mM sodium phosphate buffer, pH8.5) on CodeLinkTM microscopic glass slides (Amersham Bioscience, Roosendaal, the Netherlands), using a Spot Array 72 robot (Perkin Elmer Life Sciences, Zaventum, Belgium). The printed glass slides were further processed according to instructions of the manufacturer.

Test DNA, extracted as described previously [22] and reference genomic DNA isolated from randomized blood donors (300 ng of each) were labeled with a random primer elongation kit (Invitrogen, Breda, The Netherlands) in a 50 µL reaction with Cy3 dCTP and Cy5 dCTP (Perkin Elmer Life Sciences). Non-incorporated nucleotides were removed using ProbeQuant G-50 Micro Columns (Amersham Biosciences). Labeled DNA was mixed with Cot-1 DNA (Invitrogen) and precipitated with ethanol. DNA was dissolved in 190 µl hybridization mix with a final composition of 50% formamide, 10% dextran sulphate, 2x SCC, 4% SDS, and 1.3 mg yeast tRNA. The hybridization solution was heated to 73 °C for 10-15 minutes to denature the DNA, and then incubated at 37 °C for one hour to allow blocking of repetitive sequences. The array-slides were prehybridized with hybridization mix containing 650 µg salmon sperm DNA without probe at 37 °C in a hybridization station (Hybstation 12; Perkin Elmer Life Sciences) for one hour. Subsequently slides were hybridized for 48 hour at 37°C. After hybridization, slides were washed six times in 50% formamide, 2x SCC, pH 7.0, at 45 °C, and twice in PN buffer (0.1 M sodium phosphate, 0.1% nonidet P40, pH 8.0) at room temperature, twice with 0.2x SCC and twice with 0.1x SCC. Slides were scanned with Scan Array Express (Perkin Elmer Life Sciences).

Image acquisition and data analysis

Image analysis was carried out with ImageneTM software (BioDiscovery, El Segundo, California, USA). Spots with non-homogeneous fluorescence were automatically flagged and excluded from further analysis. Mean log2 ratios of the triplicate Cy3 and Cy5 signals of each spot were calculated in a spreadsheet (Excel 2000, Microsoft Corp, Amsterdam, The Netherlands) after subtraction of the mean local background. BAC clones with a standard deviation above 0.2 over the triplicate
values were excluded from further analysis. Normalization per array was carried out using the modus of the log2 ratios of all unflagged clones. The clones were ordered by position in the genome according to the UCSC draft genome sequence (“freeze” July 2003, http://www.sanger.ac.uk). With a “smoothing” algorithm the gains (>0) and losses (<0) were defined over chromosome 1 to 22 on basis of the log2 values [29].

The smoothing algorithm “aCGH-Smooth” uses a heuristic algorithm, based on the assumption that the experimental noise in the data is generated by a Gaussian process. It identifies breakpoints and smooths the observed array CGH values between consecutive breakpoints to a suitable common value. All log2 ratios exceeding 1.0 were considered to indicate amplifications. Chromosome X-clones were discarded from further analysis, since all tumor samples were hybridized to reference DNA of the opposite gender. Only data of BACs were included in the analysis when at least 20/24 tumors showed a value. After the exclusion procedures a total number of 3,969 clones could eventually be analyzed per each array.

To perform this analysis, the “smoothed”-log2 ratios were converted to categorized data of `+1` for chromosomal gains, `-1` for chromosomal losses and `0` if no change was determined. To identify the differential chromosomal loci between the two tumor groups these smoothed data were used for the Wilcoxon rank sum test with ties, which was implemented in two web-based applications: “CGHMultiArray” and “CGHMultiArrayRegion” test (www.win.tue.nl/~markvdw/CGHMultiArray.html), especially developed for differential genetic analyses using CGH micro-array data [30]. Both tests take the discrete nature of the smoothed data into account. The first, “CGHMultiArray”, computes P-values per clone, and the second, “CGHMultiArrayRegion” test defines regions of adjacent and identical alterations which are then considered as one testing entity. This reduces the multiplicity problem severely and allows adaptation of the usual Benjamini-Hochberg FDR type corrections. For more information we refer to the supplementary information at http://webmathematica.win.tue.nl/mark/cghregion/index.html.

We noted that the analysis could be disturbed by opposite events that occurred only in a few tumors. These events can be considered accidental and therefore we also performed the statistical analysis by comparing two groups by adjusting either all gains or all losses to ‘no change’ (from ‘+1’ to ‘0’ or from ‘-1’ to ‘0’, respectively).
ACKNOWLEDGMENTS

We thank the mapping core and map finishing groups of the Wellcome Trust Sanger Institute for initial BAC clone supply and verification. We are indebted to M Tijssen, PP Eijk, E Hopmans and P van den IJssel for printing slides and technical assistance.
REFERENCES


Genetic classification of oral and oropharyngeal carcinomas identifies subgroups with a different prognosis

Cellular Oncology (2009); in press

Serge J. Smeets*
Ruud H. Brakenhoff*
Bauke Ylstra
Wessel N. van Wieringen
Mark A. van de Wiel
C. René Leemans
Boudewijn J. M. Braakhuis
*These authors contributed equally
ABSTRACT
The common risk factors for oral and oropharyngeal cancer are tobacco smoking and alcohol consumption, and recently the human papillomavirus (HPV) was shown to be involved in a subgroup. HPV-positive and -negative carcinomas can be distinguished on basis of their genetic profiles. Aim of this study was to investigate patterns of chromosomal aberrations of HPV-negative oral and oropharyngeal squamous cell carcinomas (OOSCC) in order to improve stratification of patients regarding outcome. Thirty-nine OOSCCs were classified on basis of their genetic pattern determined by array comparative genomic hybridization (aCGH). Resulting groups were related to patient and tumor characteristics using the Fisher’s exact test and in addition to survival with the Kaplan-Meier and log rank tests. Classification distinguished three groups, one characterized by hardly any chromosomal aberration (N=8), and another by a relatively high level (N=26), and one with a very high level (N=5) of chromosomal aberrations. This classification was significantly (p=0.003) associated with survival, with the best survival in the genetically ‘silent’ group and the worst survival in the most aberrant group. The silent profile was significantly (p<0.05) associated with wild-type TP53, an absence of alcohol consumption and a female gender. These carcinomas were negative for microsatellite instability. This classification of OOSCC was confirmed in an independent set of 89 oral carcinomas. In conclusion: The discovery of these new classes of oral and oropharyngeal cancer with unique genetic and clinical characteristics has important consequences for future basic and clinical studies.
INTRODUCTION
Oral and oropharyngeal squamous cell carcinomas (OOSCC) develop in the mucosal linings of the oral cavity and oropharynx and constitute with hypopharyngeal and laryngeal carcinomas the group of head and neck squamous cell carcinomas. Despite significant advances in loco-regional control, long-term survival of OOSCC patients has only moderately improved during the last 20 years. The identification of biological markers will be essential to make headway in detecting this malignancy at an early stage and developing novel therapies.

The major and classical risk factors for OOSCC are exposure to tobacco and alcohol, but it has recently become clear that also an infection with human papillomavirus 16 (HPV16) plays an important role in a subgroup of these tumors. There is convincing evidence that HPV–infected tumors form a completely distinct group. These tumors differ from non-infected tumors with respect to risk factors [13], absence of a TP53 mutations [38,42], a low level of allelic losses [5] and mRNA expression profiles [30]. This distinct pathological entity may be the reason for a better response rate and survival [11,43].

Genetic analysis of cancers with array comparative genomic hybridization (aCGH) makes it nowadays possible to analyze DNA copy number variations with high resolution [12]. With aCGH it could be shown that HPV-infected tumors are also distinct from non-infected OOSCC regarding DNA copy number changes [31].

We hypothesized that other subgroups of non-HPV-involved carcinomas exist that differ regarding their pattern of genetic alterations and that this might have consequences for prognosis. Previously, it was shown that breast cancers and leukemia could be divided into prognostic relevant subgroups using expression array analysis [18,40]. These classifications are based on algorithms using the T-statistics as expression data are normally distributed. Only recently, cluster algorithms like WECCA, tailor-made for ordered aCGH data sets, became available [7,41].

The present study aims to identify distinct groups of non-HPV OOSCC by means of aCGH analysis and unsupervised cluster analysis. Different genetic groups were identified and could be linked to clinical parameters.
MATERIAL AND METHODS

Patients and tumor specimens
We obtained tumor specimens from 39 patients who underwent surgical treatment for a carcinoma in the oral cavity or oropharynx at the VU University Medical Center. This group was randomly selected from our tissue collection that contains snap-frozen carcinomas, gathered during the period from 1997 to 2001. To exclude the presence of HPV, detection of viral DNA and expression analysis of the oncogenes E6 and E7 was performed as previously described [31,38]. The mutational status of TP53 of these carcinomas was determined in the evolutionarily conserved regions, exons 5 to 9 [39]. When no mutations were found in these exons, the remaining coding exons 2, 3, 4, 10 and 11 were sequenced in addition. Mutations were also classified in disruptive or non-disruptive mutations, according to the criteria used by Poeta et al. [27]. In brief, these criteria are based on the location of the mutation and the type of predicted amino acid alteration. For all studies microdissection was performed to enrich for carcinoma tissue. The study had been approved by the Institutional Review Board of the VU University Medical Center, and written informed consent was obtained from all patients. Staging was performed according to the classification of the International Union Against Cancer (UICC) [33] and information on patient tobacco and alcohol use was obtained from the medical files. Patients were classified as never, current or former tobacco smokers. Pack-years were taken as a measure of cumulative tobacco consumption. It is calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked. Patients were classified as never, current or former alcohol drinkers. Unit-years were taken as a measure of cumulative alcohol consumption and were calculated as the number of years drinking multiplied by the number of units per day. A unit is defined as one alcoholic beverage (equivalent to approximately 15 mL of alcohol).

Array Comparative Genomic Hybridization
Test DNA (300 ng), extracted as described previously [39] and reference genomic DNA (300 ng) isolated from a random panel of blood donors were labeled with a random primer elongation kit (Invitrogen, Breda, The Netherlands). Arrays of Bacterial Artificial Chromosomes (BACs) were prepared and hybridized as described previously [31].
Image acquisition and data analysis

Image analysis, the exclusion of bad spots and the calculation of the mean log2 ratios of triplicate Cy3 and Cy5 signals were performed as previously described [31]. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO), and are accessible through GEO series accession number GSE11929.

Chromosome X-clones were discarded from further analysis, since all tumor samples were hybridized to reference DNA of the opposite gender. Data of a BAC was only included in the analysis when at least 80% of the tumors showed a value for that particular BAC. A total number of 4,062 clones could be analyzed for all samples.

The log2 ratios are segmented by means of ‘DNAcopy’ [25], and called by means of ‘CGHcall’ [36]. The calling results in the so-called call probabilities, a probability vector for each feature on the array: a probability of a loss, a probability of 'no-aberration', and a probability of a gain. The maximum probability indicates the most likely aberration. The use of call probabilities in down-stream analyses prevents the loss of information associated with the use of calls. Finally, genomically adjacent features not separated by a breakpoint have the same call probability signature over the samples. Therefore, the data are reduced to the unique, non-spurious call probability signatures by means of ‘CGHregions’ [37].

The samples are clustered by means of a modified version of WECCA [41]. WECCA is a hierarchical clustering method tailor-made for called aCGH data. The modified version accommodates the use of call probabilities instead of calls. The use of the call probabilities in the unsupervised analysis will give a more subtle picture of the similarities and differences between the samples. The modified version of WECCA defines the distance between two features as the absolute difference between the cumulative call probability distributions. The distance between the call probability profiles of two samples is then defined as the average of these differences over all features. In the construction of the dendrogram we used Ward's linkage as it yields compact and well-separated clusters. Chromosomal breakpoints were calculated from the segmented data.
Differences between groups as for frequencies of patient and tumor characteristics were performed with the 2 x 2 and 2 x 3 Fisher’s exact test. Continuous variables were compared with the Mann-Whitney (two groups) or the Kruskal-Wallis (three groups) test. Survival characteristics were determined according to the Kaplan-Meier method and survival curves were compared with the log-rank test. All tests were two-sided and differences were considered to be significant if the P-value was below 0.05.

**Micro Satellite Instability (MSI)**

MSI was detected using a MSI analysis system (Promega corporation, Madison, USA), according to instructions of the manufacturer. In short, this system uses fluorescently labeled primers (marker panel) for co-amplification of seven markers for analysis of the MSI-high (MSI-H) phenotype, including five nearly monomorphic mononucleotide repeat markers (BAT-25, BAT-26, MON0-27, NR-21 and NR-24) and two highly polymorphic pentanucleotide repeat markers (Penta C and Penta D). These markers are recommended by the National Cancer Institute (NCI) workshop for identifying MSI-type colorectal cancer [35]. Amplified fragments were detected by a sequence analyzer (model 3100 from Applied Biosystems BV, Nieuwerkerk a/d IJssel, Netherlands).
Table 1. Patient and tumor characteristics in relation to genetic classification

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>1 vs. 2</th>
<th>1 vs. 2a vs. 2b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>8</td>
<td>26</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
<td></td>
<td>0.772</td>
<td>0.930</td>
</tr>
<tr>
<td>Average</td>
<td>58.6</td>
<td>59.0</td>
<td>61.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>43-81</td>
<td>38-76</td>
<td>42-79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>0.235</td>
<td>0.021</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>14</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2</td>
<td>12</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor Site</td>
<td></td>
<td></td>
<td></td>
<td>0.652</td>
<td>0.554</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral cavity</td>
<td>7</td>
<td>19</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathological tumor Stage</td>
<td></td>
<td></td>
<td></td>
<td>0.426</td>
<td>0.567</td>
</tr>
<tr>
<td>I +II</td>
<td>6</td>
<td>13</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III + IV</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathological nodal stage</td>
<td></td>
<td></td>
<td></td>
<td>0.683</td>
<td>0.781</td>
</tr>
<tr>
<td>N-</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N+</td>
<td>6</td>
<td>15</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking§</td>
<td></td>
<td></td>
<td></td>
<td>0.302</td>
<td>0.524</td>
</tr>
<tr>
<td>Current + former</td>
<td>5</td>
<td>22</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pack-years</td>
<td></td>
<td></td>
<td></td>
<td>0.644</td>
<td>0.680</td>
</tr>
<tr>
<td>Average</td>
<td>37</td>
<td>31</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-84</td>
<td>0-65</td>
<td>23-61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol‡</td>
<td></td>
<td></td>
<td></td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>Current + former</td>
<td>2</td>
<td>20</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit-years</td>
<td></td>
<td></td>
<td></td>
<td>0.199</td>
<td>0.230</td>
</tr>
<tr>
<td>Average</td>
<td>103.57</td>
<td>136.0</td>
<td>138.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-405</td>
<td>0-450</td>
<td>120-244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP53</td>
<td></td>
<td></td>
<td></td>
<td>0.049</td>
<td>0.076</td>
</tr>
<tr>
<td>Mutated</td>
<td>1</td>
<td>15</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td>7</td>
<td>11</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disruptive mutation</td>
<td>0</td>
<td>11</td>
<td>2</td>
<td>0.035</td>
<td>0.061</td>
</tr>
<tr>
<td>Non-Disruptive/Wild-type</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of breakpoints</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>2.8</td>
<td>30.0</td>
<td>52.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-6</td>
<td>12-49</td>
<td>31-80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Characteristics of the three groups as classified with WECCA (Figure 1) are listed. Group 1 tumors show the lowest and group 2b the highest level of aberrations. P values are two-sided and for the 2 x 2 and 2 x3 frequency comparisons the Fisher's exact test was used and for the mean comparisons, the Mann-Whitney or the Kruskal-Wallis test. Significant differences (P<0.05) are shown in bold. § Patients were classified as never, current or former tobacco smokers. Pack-years were taken as a measure of cumulative tobacco consumption. It is calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked. ‡ Patients were classified as never, current or former alcohol drinkers. Unit-years were taken as a measure of cumulative alcohol consumption and were calculated as the number of years drinking multiplied by the number of units per day. A unit is defined as one alcoholic beverage (equivalent to approximately 15 mL of alcohol).
CHAPTER 5

RESULTS
To investigate whether classification on the basis of genetic aberrations is possible, we performed aCGH on 39 OOSCC and analyzed the data. Unsupervised clustering enabled the discovery of two distinct groups at the top of the tree (Figure 1). Group 1 consisted of 8 tumors that showed a significantly lower level of chromosomal aberrations, when compared to the other group, labeled 2 (average number of chromosomal breakpoints 2.8 and 33.7 respectively: Table 1). The few changes that were observed in group 1 were never shared by two or more tumors of that group. Group 2 was characterized by a higher level of chromosomal aberrations. Aberrations that were present in this group in more than 50% of the tumors, were gains at chromosome 3q, 7q, 8q and 11q13 and losses at chromosome 3p, 5q, 11q23 and 18q (Figure 1). Regions that showed numerical changes in even more than 80% of carcinomas; loss at 3p (total p-arm), gain at 3q26.2-qter (29 Mb) and gain at 8q24.13-q24.21 (3 Mb). Typical examples of a CGH profile of tumors of groups 1 and 2 are shown in Figure 2.
Looking further down the tree it appeared that group 2 consists of two further subgroups; one (2a) of 26 tumors with a relatively high level of aberrations (average number breakpoints of 30.0) and another (2b) of five tumors with an even higher level of aberrations (average number of breakpoints 52.8).
Comparison of the clinical parameters between the identified groups revealed that group 1 (with infrequent aberrations), significantly differed from groups 2a and 2b, as it was characterized by the absence of alcohol consumption, female gender and wild type TP53 in the carcinoma (Table 1). Regarding the mutational status in TP53, the differences between group 1 and 2 were more pronounced when the mutations were classified as disruptive vs. non-disruptive and wild-type (Table 1). A remarkable feature of group 2b was the gender dysbalance, as only males were observed in this group.
Figure 1. WECCA heatmap of 39 OOSCCs without HPV involvement. At the top of the tree two tumor groups can be distinguished: a group consisting of 8 OOSCC with in general genetically silent profiles and another group of 31 tumor with many genetic aberrations. In group 2 downwards, two subgroups (a and b) can be distinguished with 2b as the group with the most genetic aberrations. The x-axis represents tumor numbers and the y-axis chromosome numbers.
Figure 2. Typical examples of an aCGH profile of each group. Group 1) Represents a tumor with a silent profile and wild-type TP53, showing just a few aberrations. Group 2) Shows a typical profile of a tumor of group 2 showing many aberrations, e.g. with gains at 3q, chromosome 8, 11q13, 17q, chromosome 20, and losses at 3p, 11q23 and 13q.

Analysis of overall survival between the groups revealed a significant difference between the groups (log rank test with a p-value of 0.003), with the best survival in group 1 and the worst survival in group 2b (figure 3). We next investigated classification after including the dataset of 12 HPV-positive OOSCC from a previous study [31]. Seven out of twelve HPV-positive OOSCC appeared to cluster as a separate group within group 2a. Five HPV-positive tumors belonged to group 1 in which hardly any numerical change was observed (see supplementary Figure 1).
It was tried to confirm the findings by analysis of an independent and external aCGH dataset of 89 oral squamous cell carcinomas described by Snijders et al. [32]. Also in this material two main groups were identified (supplementary Figure 2). One group of oral carcinomas (N=33) was characterized by a low level of chromosomal aberrations (average breakpoints: 11.5 ranging from 0 to 27) and a group of oral carcinomas (N=56) with a high level of chromosomal aberrations (average breakpoints 28.1 ranging from 7 to 53). Similar to our dataset, also in this external set a subgroup of tumors (n=18) with a much higher level of chromosomal aberrations (average breakpoints 34.7 ranging from 18 to 53) could be recognized. Unfortunately, no clinical parameters were available of these patients. The TP53 mutation status of most of the tumors was described by Snijders et al. [32]. Similar to our data we found a significant correlation between mutation status and classification as 2 out of 25 tumors in group 1, and 15 out of 34 tumors of group 2 contained mutated TP53 (p=0.02). It was noteworthy that all aberrations in our dataset were present in the dataset of Snijders et al. at high frequency as well.

Tumors with hardly any chromosomal aberrations are also known in colorectal cancer. In these tumors microsatellite instability (MSI) is mostly the driving force for progression [10]. We evaluated this possibility on this subgroup and found no indications for MSI in all evaluated markers (data not shown).
DISCUSSION
The present analysis of a large number of HPV-negative OOSCC by high resolution CGH revealed a large number of DNA copy number changes. This complex genetic pattern is a well known characteristic of OOSCC and confirms other CGH-studies [2,3,12,16,24,26,32,34,45]. Copy number changes were relatively frequent at chromosomes 3, 5p, 8, and 11q, involving over 50% of the tumors in group 2. Aberrations at these chromosomal locations have also been described in other studies, but at lower frequencies [2,3,12,16,24,26,32,34,45]. Three chromosomal regions were more pronounced and found to be altered in more than 80 percent of these tumors; loss at 3p, gain at 3q26 and gain at 8q24. The gain of 8q24.21 comprising 3 MB was identified by defining the smallest region of overlap. Gain at 8q, in particular 8q24 has also been reported by others in an unselected group OOSCC [3,44]. In this region c-Myc is located, a forceful oncogene that is commonly found to be overexpressed by genomic amplification in OOSCC [19,14,23] and that has the capability to transform primary oral epithelial cells to cancer cells in vitro [15].

By virtue of WECCA, a new classification tool for aCGH data, we were able to identify separate tumor groups. A group with a remarkably low and a group with a relatively high level of chromosomal aberrations could be discriminated; a subgroup within that latter group showed a very high level of chromosomal alterations. To find confirmation of the present findings, the data from a previously published study [32] was investigated in more detail with the WECCA cluster algorithm. Analysis of the data revealed again the presence of two major tumor groups divided on the basis of the number of aberrations. The average number of aberrations in the genetically silent group was somewhat higher than in our tumor panel and this group was somewhat larger (37.1% vs. 20.5% in our material). Nevertheless the high similarity between the two datasets, including the identification of the subgroup with the rather high number of aberrations, is noteworthy. Unfortunately the survival data were not available from that group. It has to be realized that there are geographical difference regarding the patient populations that were compared. In addition, the fact that only oral carcinomas had been included and that some cases could be HPV-positive [32], may explain some of the differences.

OOSCC harboring a genome with few chromosomal aberrations were earlier recognized in cytogenetic studies as tumors with simple karyotypic changes [20]. In fact, our sensitive aCGH method confirms in a way and extends in much greater detail of what was found with this 'older' crude method. The tumors with such a simple karyotype comprised approximately 36% of the study population, but were
never discussed as a separate tumor group and this characteristic was not correlated with clinical parameters like prognosis.

When comparing the presently classified tumor groups regarding clinical and biological parameters it was found that survival, alcohol drinking history, gender and TP53 mutation status were significantly different, which can be interpreted as independent evidence to support the relevance of the findings. It can be appreciated that the tumors showing a silent genetic profile were all TP53 wild type. This notion was also supported by the results of the external aCGH dataset of 89 OOSCC [32]. The question remains whether this is a non-random association or that the TP53 status plays a causative role on the genetic profile. An argument against the latter is that some tumors of group 2 with a high level of chromosomal changes also showed a wild type TP53 gene while the gene was sequenced from exons 2-11. Notwithstanding, exon deletions could have been missed and we cannot exclude that the p53 pathway is impaired in an alternative way. Genes modifying p53 function might be mutated, abrogating p53 function indirectly [4,17].

Six of the eight patients with a silent tumor profile, never had consumed alcohol. In addition, an excessive alcohol consumption appeared to be correlated with the presence of a TP53 mutation and thereby an increased genomic instability. Although, a causal relation between both tobacco and alcohol abuse and OOSCC is well established [9,22], there is controversy about the relationship between levels of tobacco and alcohol exposure vs. the number and type of genetic aberrations in the tumor. Koch et al. [21] found a distinct clinical and molecular entity of head and neck squamous cell carcinoma in non-smokers, while Singh et al. did not [29]. In addition, Brennan et al. found an association between excessive alcohol/ tobacco consumption and the presence of mutated TP53 [6]. It has recently become clear that alcohol can be a carcinogen on its own [28], but its role in generating a mutation in TP53 needs further attention.

We found evidence that the level of DNA copy number changes is associated with patient outcome. The lowest level of aberrations was associated with the best and the highest level with the worst survival. This finding is in line with what could be expected: patients with carcinomas without a TP53 mutation have a better survival [13] and a relatively low level of genetic instability has been reported to be related to a better outcome, for head and neck [1] and breast cancer [8]. Although the difference in survival between the groups was significant, it has to be added that the numbers were small. A larger study on OOSCC is needed to definitively proof the clinical value of this classification.
The genetically silent group (number 1) is for a large part characterized by tumors that are TP53 wild type and contain hardly any chromosomal aberrations. It is tempting to speculate on the molecular mechanism that might drive carcinogenesis in these tumors. This subgroup may be the result of a defect known as microsatellite instability (MSI), related to a deficiency in the mismatch repair machinery. A proportion of colorectal carcinomas show this phenomenon as well that is reflected in relatively few chromosomal aberrations [10]. We evaluated this possibility on this subgroup, but found that all were MSI negative. It has to be added that there is a possibility that MSI was missed; a set of markers was tested that may not be ideal for OOSCC, as it is normally used for colon carcinoma. Second, other forms of DNA alterations, in particular loss of heterozygosity (LOH) without numerical changes and epigenetic alterations such as promoter hypermethylation, might play a role in these tumors and might even abrogate the p53 pathway. Finally, tumor heterogeneity at the cellular level may explain the absence of specific aberrations in these OOSCC. Gains in some cells could be counterbalanced by losses in other. However it seems unlikely that this results in normal genetic profiles, but without more detailed investigations this remains unclear.

Taken together, our main finding is the identification by array CGH of a previously unrecognised group of OOSCC with a remarkably almost normal genome, regarding DNA copy numbers. Data of the current study indicate that this genomic subgroup of OOSCC shows a different outcome and patient profile; these findings and the unraveling of the possible mechanisms driving its carcinogenesis warrant additional studies in a larger OOSCC cohort. At this moment, at least three different OOSCC groups can be discriminated: 1) a HPV-positive group that was discussed in a previous study [31], 2) a ‘normal DNA’ group without TP53 mutation that we now reported on, and 3) a group with a high to very high level of chromosomal aberrations, characterized by loss of 3p, 5p, 11q23 and gains at 3q, chromosome 8 and 11q13. These findings underline the clinical importance of biological classification of OOSCC. Stratification in groups is critical when analyzing the role of potential cancer genes, early detection of tumor markers, and selection of treatments.
Supplementary Figure 1. WECCA heat map of 51 OOSCC, 39 without (present study) and 12 with HPV16-infection. Results of the HPV-infected tumors are derived from a previous publication [31]. Seven out of 12 HPV-positive tumors formed a separate group and five of them rearranged among the group of silent tumors (indicated with black bars near the x-axis). The x-axis represents tumor numbers and the y-axis chromosome numbers.
Supplementary Figure 2. WECCA heatmap of 89 OOSCCs described by Snijders et al. [32]. General classification is similar as in the present study; a group consisting of 33 OOSCC with in general a low level of chromosomal aberrations and another group of 56 tumors with many chromosomal aberrations. In group 2 downwards, two subgroups (a and b) can be distinguished with 2b as the group with the most chromosomal aberrations. The y-axis represents chromosome numbers.

ACKNOWLEDGEMENTS
This work was supported by the EU-sixth framework project: DISMAL, Contract No.: LSHC-CT-2005-018911.
REFERENCES


Genomic profiling identifies common HPV-associated chromosomal alterations in squamous cell carcinomas of cervix and head and neck

Submitted for publication

Saskia M. Wilting & Serge J. Smeets
Peter J.F. Snijders
Wessel N. van Wieringen
Mark A. van de Wiel
Gerrit A. Meijer
Bauke Ylstra
C. René Leemans
Chris J.L.M Meijer
Ruud H. Brakenhoff
Boudewijn J.M. Braakhuis &
Renske D.M. Steenbergen
CHAPTER 6

ABSTRACT
It is well known that a persistent infection with high-risk human papillomavirus (hrHPV) is causally involved in the development of squamous cell carcinomas of the uterine cervix (CxSCCs) and a subset of squamous cell carcinomas of the head and neck (HNSCCs). The latter differ from hrHPV-negative HNSCCs at the clinical and molecular level.

To determine whether hrHPV-associated SCCs arising from different organs have specific chromosomal alterations in common, we compared genome-wide chromosomal profiles of 10 CxSCCs (all hrHPV-positive) with 12 hrHPV-positive HNSCCs and 30 hrHPV-negative HNSCCs. Potential organ-specific alterations and alterations shared by SCCs in general were investigated as well.

Unsupervised hierarchical clustering resulted in one mainly hrHPV-positive and one mainly hrHPV-negative cluster. Interestingly, loss at 13q and gain at 20q were frequent in HPV-positive carcinomas of both origins, but uncommon in hrHPV-negative HNSCCs, indicating that these alterations are associated with hrHPV-mediated carcinogenesis. Within the group of hrHPV-positive carcinomas, HNSCCs more frequently showed gains of multiple regions at 8q whereas CxSCCs more often showed loss at 17p. Finally, gains at 3q24-29 and losses at 11q22.3-25 were frequent (>50%) in all sample groups.

In this study hrHPV-specific, organ-specific, and pan-SCC chromosomal alterations were identified. The existence of hrHPV-specific alterations in SCCs of different anatomical origin, suggests that these alterations are crucial for hrHPV-mediated carcinogenesis.
INTRODUCTION
In the pathogenesis of uterine cervical cancer the necessary and causal involvement of high-risk types of the human papillomavirus (hrHPV) is widely accepted and supported by strong epidemiological and molecular evidence [1]. HrHPV is present in virtually all cervical carcinomas and the viral oncogenes E6 and E7 are consistently expressed in cervical cancers and precancers. Deregulated expression of these oncogenes in the basal, dividing cells of the epithelium interferes with cell cycle control due to their ability to induce degradation of the tumour suppressor proteins p53 and pRb, respectively. This results in uncontrolled cell proliferation and accumulation of specific (epi)genetic changes in the host cell genome, driving progression to a malignant phenotype [2,3]. In a previous study we have used array-based comparative genomic hybridisation (array CGH) to determine frequent chromosomal alterations in cervical cancer, which included gains at 1q, 3q and 20q and losses at 8q, 10q, 11q, and 13q [4]. The necessity of these and other additional (epi)genetic alterations in the carcinogenic process is illustrated by the fact that their frequency increases with increasing severity of cervical disease.

On the other hand, head and neck squamous cell carcinoma (HNSCC) is known to be mainly caused by well-established life-style related habits, such as tobacco and excessive alcohol consumption. However, besides the influence of these life-style carcinogens, high risk human papillomavirus (hrHPV) is present in 15-35% of HNSCCs and has been suggested to be a separate aetiological factor in head-and-neck carcinogenesis [5-8]. Several studies have shown that hrHPV-positive HNSCCs are associated with a better clinical outcome [7-9]. Moreover, molecular differences were found between hrHPV-positive and hrHPV-negative HNSCCs, supporting the idea of two separate carcinogenic pathways to HNSCC, one determined by life-style carcinogens and the other by hrHPV [5,6,10-13].

In a previous study, using array CGH, we identified a number of chromosomal alterations specific for hrHPV-negative HNSCCs that were absent in hrHPV-positive HNSCCs, including loss at 3p, 5q, and 9p, and amplifications at 11q [13]. The hrHPV-positive HNSCCs were characterised by a lower level of chromosomal alterations, none of which were hrHPV-specific. To investigate potential organ-independent, hrHPV-associated chromosomal alterations, genomic profiles of cervical SCCs (CxSCCs), hrHPV-positive and hrHPV-negative HNSCCs were compared using sophisticated clustering and statistical approaches [4,13]. In addition, we also investigated the presence of organ-specific alterations and alterations shared by all SCCs included in this study.
METHODS

CGH microarrays
We used chromosomal profiles of 30 hrHPV-negative HNSCCs, 12 hrHPV-positive HNSCCs and 10 CxSCCs all of which were previously described (Table 1) [4,13] (Smeets et al., in press Cell Oncol 2009). To avoid possible confounding of our results, we excluded a subset of hrHPV-negative HNSCCs described in Smeets et al., which showed little to no alterations and should therefore be considered a separate group (Smeets et al., in press, Cell Oncol 2009).

CGH BAC microarrays produced at the Microarray facility of the VU Medical Center were used. These arrays included the 1Mb resolution Sanger BAC clone set and a subset of clones from the Children’s Hospital Oakland Research Institute (CHORI). Spots were quantified using ImaGene 5.6.1 software (BioDiscovery Ltd, Marina del Rey, CA, USA) with default settings for the flagging of bad quality spots.

The entire dataset described here is available from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/projects/geo/) through series accession numbers GSE6473 (CxSCC) and GSE12020 (HNSCC).

This study followed the ethical guidelines of the Institutional Review Board of the VU University Medical Center and informed consent was obtained from all patients included.

Array CGH analysis
Calling of gains and losses
BAC clones were positioned along the genome according to the May 2004 freeze. After exclusion of clones with one or more flagged spots, the average of the triplicate spots was calculated for each BAC clone. Log2 ratios were normalised per spotted sub-array by subtraction of the median value of all BAC clones spotted within that sub-array. Segmentation and subsequent calling of gained, amplified and lost regions was done using CGHCall, an automated calling algorithm. Segments with a probability score of ≥0.5 were considered gained, amplified or lost [14].
Table 1. Summary of clinical data of carcinomas included in this study.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>origin</th>
<th>hrHPV (+/-)</th>
<th>age (yrs)</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>CxSCC2</td>
<td>cervix</td>
<td>+</td>
<td>39</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC4</td>
<td>cervix</td>
<td>+</td>
<td>62</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC12</td>
<td>cervix</td>
<td>+</td>
<td>44</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC15</td>
<td>cervix</td>
<td>+</td>
<td>47</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC27</td>
<td>cervix</td>
<td>+</td>
<td>49</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC28</td>
<td>cervix</td>
<td>+</td>
<td>48</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC32</td>
<td>cervix</td>
<td>+</td>
<td>37</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC36</td>
<td>cervix</td>
<td>+</td>
<td>72</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC38</td>
<td>cervix</td>
<td>+</td>
<td>51</td>
<td>female</td>
</tr>
<tr>
<td>CxSCC39</td>
<td>cervix</td>
<td>+</td>
<td>40</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9741</td>
<td>oropharynx</td>
<td>-</td>
<td>53</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9762</td>
<td>oral cavity</td>
<td>-</td>
<td>76</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9773</td>
<td>oral cavity</td>
<td>-</td>
<td>49</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9830</td>
<td>oral cavity</td>
<td>-</td>
<td>42</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9848</td>
<td>oropharynx</td>
<td>-</td>
<td>59</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9892</td>
<td>oral cavity</td>
<td>-</td>
<td>67</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9897</td>
<td>oral cavity</td>
<td>-</td>
<td>72</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9902</td>
<td>oral cavity</td>
<td>-</td>
<td>53</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9942</td>
<td>oral cavity</td>
<td>-</td>
<td>55</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9952</td>
<td>oral cavity</td>
<td>-</td>
<td>61</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9956</td>
<td>oral cavity</td>
<td>-</td>
<td>38</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC2014</td>
<td>oral cavity</td>
<td>-</td>
<td>68</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC2034</td>
<td>oral cavity</td>
<td>-</td>
<td>45</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9738</td>
<td>oral cavity</td>
<td>-</td>
<td>60</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9745</td>
<td>oropharynx</td>
<td>-</td>
<td>52</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9750</td>
<td>oral cavity</td>
<td>-</td>
<td>65</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9812</td>
<td>oral cavity</td>
<td>-</td>
<td>65</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9827</td>
<td>oropharynx</td>
<td>-</td>
<td>57</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9829</td>
<td>oral cavity</td>
<td>-</td>
<td>74</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9831</td>
<td>oral cavity</td>
<td>-</td>
<td>55</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9841</td>
<td>oropharynx</td>
<td>-</td>
<td>55</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9847</td>
<td>oropharynx</td>
<td>-</td>
<td>57</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9880</td>
<td>oral cavity</td>
<td>-</td>
<td>71</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9907</td>
<td>oral cavity</td>
<td>-</td>
<td>46</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9914</td>
<td>oropharynx</td>
<td>-</td>
<td>78</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9926</td>
<td>oropharynx</td>
<td>-</td>
<td>60</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9957</td>
<td>oral cavity</td>
<td>-</td>
<td>51</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9970</td>
<td>oral cavity</td>
<td>-</td>
<td>49</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9981</td>
<td>oropharynx</td>
<td>-</td>
<td>55</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9987</td>
<td>oral cavity</td>
<td>-</td>
<td>79</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9881</td>
<td>oral cavity</td>
<td>+</td>
<td>51</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9729</td>
<td>oropharynx</td>
<td>+</td>
<td>52</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9838</td>
<td>oropharynx</td>
<td>+</td>
<td>60</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9860</td>
<td>oropharynx</td>
<td>+</td>
<td>67</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9901</td>
<td>oropharynx</td>
<td>+</td>
<td>57</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9948</td>
<td>oropharynx</td>
<td>+</td>
<td>70</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9951</td>
<td>oropharynx</td>
<td>+</td>
<td>46</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9808</td>
<td>oral cavity</td>
<td>+</td>
<td>46</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9859</td>
<td>oral cavity</td>
<td>+</td>
<td>40</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC9924</td>
<td>oral cavity</td>
<td>+</td>
<td>41</td>
<td>female</td>
</tr>
<tr>
<td>HNSCC9947</td>
<td>oral cavity</td>
<td>+</td>
<td>72</td>
<td>male</td>
</tr>
<tr>
<td>HNSCC2015</td>
<td>oral cavity</td>
<td>+</td>
<td>65</td>
<td>male</td>
</tr>
</tbody>
</table>
Reduction of dataset into chromosomal regions

We used the CGHregions algorithm to reduce our dataset to chromosomal regions, accepting maximally 0.1% information loss (Threshold=0.001). It was shown by Van de Wiel et al. that the use of regions instead of single BAC clones improved the effectiveness of subsequent statistical analyses and facilitated interpretation of the results [15].

Clustering analysis

The samples were clustered by means of a modified version of WECCA [16]. WECCA is a hierarchical clustering method tailor-made for called aCGH data. The modified version accommodates the use of call probabilities instead of calls. The use of the call probabilities in the unsupervised analysis will give a more subtle picture of the similarities and differences between the samples. The modified version of WECCA defines the distance between two features as the symmetric Kullback-Leibler divergence. The distance between the call probability profiles of two samples is then defined as the average of these divergences over all features. In the construction of the dendrogram we used Ward's linkage as it yields compact and well-separated clusters.

Statistical analysis

The association between clustering results and HPV status was determined by chi-square testing. The average total number of altered regions was compared between HPV-positive and HPV-negative carcinomas using the non-parametric Mann Whitney test. Two-sided p-values below 0.05 were considered statistically significant. Alteration patterns between HPV-positive (HNSCCs and CxSCCs) and HPV-negative tumors as well as between HNSCCs and CxSCCs were compared using a binomial differential proportion test. The test procedure includes a permutation-based false discovery rate (FDR) correction for multiple testing, needed to discriminate real differences from chance effects [17]. An FDR below 0.10 was considered statistically significant.
Gene ontology analysis

To interpret the biological significance of the genes that are located at altered chromosomal regions of interest, a gene ontology analysis was performed using Ingenuity Pathways Analysis (Ingenuity Systems®, Redwood City, USA). Biological/molecular functions were considered to be significantly overrepresented if they contained more than 1 gene and the Benjamini-Hochberg corrected p-value was p<0.10 [18].

Figure 1. Result from the unsupervised hierarchical clustering analysis. Cluster 1 contains in majority hrHPV-positive carcinomas, as is indicated by the black boxes in the legend underneath the heatmap (p<0.0001).
RESULTS

hrHPV-positive carcinomas cluster together
To obtain an overview of the similarities between samples, unsupervised hierarchical clustering was performed. This method enabled us to determine in an unbiased manner whether chromosomal profiles of hrHPV-positive HNSCCs were more closely related to hrHPV-negative HNSCCs or hrHPV-positive CxSCCs. As is shown in Figure 1, two clusters emerged. Cluster 1 contained 24 samples, 18 of which were hrHPV-positive (75%) and 6 were hrHPV-negative. Cluster 2 contained 28 samples of which 4 samples were hrHPV-positive and 24 were hrHPV-negative (86%). This association between cluster assignment and hrHPV status was statistically significant \((p<0.0001)\). The hrHPV-positive cluster 1 included both samples of cervical \((n=8)\) and head and neck origin \((n=10)\), indicating similarities between the chromosomal profiles of hrHPV-induced carcinomas of different anatomical origins. Within cluster 1, however, 7 out of 8 cervical samples formed a separate sub-cluster. This suggests that organ-specific alterations exist as well in hrHPV-positive HNSCCs and CxSCCs.

hrHPV-associated loss at chromosome 13q and gain at 20q
To assess the differences between all three sample groups, the frequency of gains (including amplifications) and losses was analysed for all chromosomal regions (Figure 2). In general, hrHPV-negative carcinomas showed significantly more altered regions than hrHPV-positive carcinomas \((p=0.022)\).
Figure 2. Frequency plots for all 3 sample groups. The frequency of gains (positive axis) and losses (negative axis) are shown for A. hrHPV-negative HNSCCs, B. hrHPV-positive HNSCCs and C. CxSCCs for chromosome 1-22.
Table 2. Significantly different chromosomal alterations between hrHPV+ and hrHPV- carcinomas.

<table>
<thead>
<tr>
<th>chromosome</th>
<th>start position (bp)</th>
<th>end position (bp)</th>
<th>Cytoband</th>
<th>FDR</th>
<th>% loss</th>
<th>% gain</th>
<th>% amp</th>
<th>% loss</th>
<th>% gain</th>
<th>% amp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1082137.5</td>
<td>5868019.5</td>
<td>1p36.33-p32.2</td>
<td>0.08</td>
<td>0.0</td>
<td>26.7</td>
<td>15.6</td>
<td>40.0</td>
<td>0.0</td>
<td>14.6</td>
</tr>
<tr>
<td>1</td>
<td>10974674.5</td>
<td>11726378.5</td>
<td>1p13.3-p13.1</td>
<td>0.04</td>
<td>0.0</td>
<td>13.6</td>
<td>16.7</td>
<td>10.0</td>
<td>0.0</td>
<td>17.8</td>
</tr>
<tr>
<td>2</td>
<td>96169500</td>
<td>97679803</td>
<td>2q11.2</td>
<td>0.08</td>
<td>4.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>165891891</td>
<td>180528807</td>
<td>2q24.3-q31.3</td>
<td>0.06</td>
<td>9.1</td>
<td>8.3</td>
<td>10.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>282248</td>
<td>86178494</td>
<td>3p26.3-p12.1</td>
<td>0.00</td>
<td>17.8</td>
<td>20.5</td>
<td>14.6</td>
<td>9.4</td>
<td>3.8</td>
<td>16.2</td>
</tr>
<tr>
<td>5</td>
<td>51158140.5</td>
<td>180569503</td>
<td>5q11.2-q35.3</td>
<td>0.00</td>
<td>8.6</td>
<td>7.5</td>
<td>10.0</td>
<td>9.1</td>
<td>8.3</td>
<td>10.0</td>
</tr>
<tr>
<td>7</td>
<td>9952497</td>
<td>25665820.5</td>
<td>7p21.3-p15.2</td>
<td>0.05</td>
<td>13.6</td>
<td>16.7</td>
<td>10.0</td>
<td>0.0</td>
<td>0.0</td>
<td>3.3</td>
</tr>
<tr>
<td>7</td>
<td>65521728.5</td>
<td>11645099</td>
<td>7q11.21-q31.2</td>
<td>0.03</td>
<td>4.5</td>
<td>0.0</td>
<td>10.0</td>
<td>0.3</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>384747.5</td>
<td>34510486</td>
<td>8p23.3-p12</td>
<td>0.02</td>
<td>6.8</td>
<td>12.5</td>
<td>0.0</td>
<td>17.3</td>
<td>31.7</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>75648193.5</td>
<td>9392184</td>
<td>8q21.11-q22.1</td>
<td>0.06</td>
<td>0.0</td>
<td>22.7</td>
<td>41.7</td>
<td>0.0</td>
<td>3.3</td>
<td>57.8</td>
</tr>
<tr>
<td>8</td>
<td>11471693.5</td>
<td>145681302</td>
<td>8q23.3-q24.3</td>
<td>0.05</td>
<td>0.3</td>
<td>0.0</td>
<td>0.7</td>
<td>34.4</td>
<td>55.4</td>
<td>9.3</td>
</tr>
<tr>
<td>11</td>
<td>69184373.5</td>
<td>70313976.5</td>
<td>11q13.3-q13.4</td>
<td>0.04</td>
<td>11.4</td>
<td>12.5</td>
<td>10.0</td>
<td>13.6</td>
<td>0.0</td>
<td>4.5</td>
</tr>
<tr>
<td>13</td>
<td>55718989</td>
<td>71489428.5</td>
<td>13q21.1-q21.33</td>
<td>0.06</td>
<td>66.7</td>
<td>64.8</td>
<td>68.9</td>
<td>0.0</td>
<td>0.0</td>
<td>33.0</td>
</tr>
<tr>
<td>14</td>
<td>47082923</td>
<td>61966540</td>
<td>14q21.3-q23.2</td>
<td>0.08</td>
<td>11.4</td>
<td>12.5</td>
<td>10.0</td>
<td>2.3</td>
<td>0.0</td>
<td>5.0</td>
</tr>
<tr>
<td>14</td>
<td>79430708.5</td>
<td>86910217</td>
<td>14q31.1-q31.3</td>
<td>0.09</td>
<td>18.2</td>
<td>25.0</td>
<td>10.0</td>
<td>9.1</td>
<td>0.0</td>
<td>20.0</td>
</tr>
<tr>
<td>18</td>
<td>28336092</td>
<td>75619995</td>
<td>18q12.1-q23</td>
<td>0.03</td>
<td>10.0</td>
<td>8.3</td>
<td>12.0</td>
<td>18.2</td>
<td>25.0</td>
<td>10.0</td>
</tr>
<tr>
<td>20</td>
<td>1557745.5</td>
<td>60281352</td>
<td>20p12.1-q13.33</td>
<td>0.06</td>
<td>1.5</td>
<td>2.8</td>
<td>0.0</td>
<td>58.5</td>
<td>41.7</td>
<td>78.7</td>
</tr>
<tr>
<td>21</td>
<td>14683331</td>
<td>46846395.5</td>
<td>21q11.2-q22.3</td>
<td>0.00</td>
<td>1.1</td>
<td>0.0</td>
<td>2.5</td>
<td>25.0</td>
<td>14.6</td>
<td>37.5</td>
</tr>
</tbody>
</table>

hrHPV-specific alterations are printed in bold. Alterations specific for hrHPV-negative HNSCCs are italicised.

Bp; base pair, FDR; False Discovery Rate, amp; amplification
Table 3. Significantly different chromosomal alterations between hrHPV-positive HNSCCs and CxSCCs.

<table>
<thead>
<tr>
<th>Chromosome Start Position</th>
<th>Chr Band</th>
<th>HNSCC+</th>
<th>CxSCC</th>
<th>HNSCC+ Chromosome Gain</th>
<th>CxSCC Chromosome Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3q13.2-q29</td>
<td>0.0</td>
<td>95.6</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>8q12.1-q21.11</td>
<td>0.0</td>
<td>9.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>8q22.1-q22.2</td>
<td>0.0</td>
<td>9.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>8q24.3</td>
<td>0.0</td>
<td>9.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>11</td>
<td>11q22.1-q22.3</td>
<td>0.0</td>
<td>95.6</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td>17</td>
<td>17p13.3</td>
<td>0.0</td>
<td>9.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>17</td>
<td>17p11.2</td>
<td>0.0</td>
<td>95.6</td>
<td>2.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Alterations specific for HNSCCs are printed in bold, alterations specific for CxSCCs are italicised.

Bp: base pair, FDR: False Discovery Rate, amp: amplification
To determine hrHPV-specific chromosomal alterations, the frequency of alterations was compared between hrHPV-positive (12 HNSCCs and 10 CxSCCs) and hrHPV-negative tumors (30 HNSCCs) for all chromosomal regions. Regions showing a significant difference (FDR<0.10) in this comparison are shown in Table 2. Interestingly, loss of 13q21.1-21.33 and gain of 20p12.1-q13.33 were significantly more frequent in HPV-positive SCCs compared to hrHPV-negative HNSCCs (66.7% and 58.5% in hrHPV-positive SCCs compared to 33% and 24% in hrHPV-negative SCCs, respectively). As was also shown in our previous study, loss of regions at 3p and 5q, and gains/amplifications of a small region at 11q (CCND1 locus) were specific for hrHPV-negative HNSCCs [13]. A region at 8p showed loss in hrHPV-negative HNSCCs, gain in hrHPV-positive HNSCCs and no alteration in CxSCCs. Gain at chromosome 8q was more frequent in hrHPV-negative HNSCCs, which was due to absence of this alteration in CxSCCs.

Our unsupervised classification results indicated that, within our hrHPV-positive cluster, CxSCCs formed a separate group. To identify potential organ-specific alterations, we therefore also compared the frequency of alterations for all regions between hrHPV-positive HNSCCs (n=12) and CxSCCs (n=10) (Table 3). Gains at 3q and losses at 17p were significantly more frequent in CxSCCs than HNSCCs (FDR<0.10). On the other hand, HNSCCs showed frequent gains at chromosome 8q and losses at 11q. It is important to note that even though gain at 3q and loss at 11q were significantly different between hrHPV-positive HNSCCs and CxSCCs, these alterations were frequent in all sample groups (>50%). Significant differences for these regions can mainly be explained by the fact that the size of the exact altered region differed between groups. Therefore, the smallest regions of overlap between all samples, namely 3q24-29 and 11q22.3-25, may represent general alterations in carcinomas derived from squamous epithelium.

All results described above are summarised in a Venn diagram, showing a general overview of the frequently altered chromosomal arms specific to or common between the (sub) groups (Figure 3).
Figure 3. Summary of common and specific chromosomal alterations in the different sample groups. Chromosome arms showing frequent alterations (>50%) in one or more sample groups (hrHPV-positive HNSCCs; hrHPV-negative HNSCCs; CxSCCs) are shown in a Venn diagram. ↑ indicates gain; ↓ indicates loss.

Biological functions of genes located within hrHPV-specific chromosomal alterations
As described above gain at 20p12.1-q13.33 and loss at 13q21.1-21.33 were significantly more frequent in hrHPV-positive SCCs compared to hrHPV-negative ones. In fact, a 4.5 megabase (Mb) region on chromosome 20 (20q11.21-q11.23), and a 2Mb region on chromosome 13 (13q21.1) formed the smallest regions of overlap (SRO) at these respective loci when all hrHPV-positive carcinomas were considered (Figure 4A and B). All genes located within the SROs found at chromosome 20q (78 genes) and 13q (6 genes) are listed in Supplementary Table 1. Within the SRO at chromosome 20, two genes reside, i.e. NCOA6 and RBM39, which showed elevated expression in hrHPV16 E7 expressing cells in vitro [19]. Other cancer-related genes located within this SRO include PIGU, E2F1, and DNMT3B. The SRO on chromosome 13 encompasses the PCDH17 gene and a cluster of five identical loci all of which are predicted to encode proline-rich proteins that contain several dopamine D4 receptor signatures.

Subsequent gene ontology analysis of all genes located in these SROs identified a number of significantly overrepresented GO biological/molecular functions, including cell cycle, cell-to-cell signalling and interaction, cellular growth and proliferation, and a number of cellular maintenance functions (i.e. DNA replication, recombination and repair, gene expression, cellular function and maintenance, cellular assembly and organisation, cellular compromise, cell death, cell morphology, cellular development, nucleic acid metabolism, and lipid metabolism).
Figure 4. Genomic coordinates of A. losses at chromosome 13 and B. gains at chromosome 20 are shown for all hrHPV-positive carcinomas. Chromosomal alterations in CxSCCs are shown by dashed lines and alterations in hrHPV-positive HNSCCs by solid lines. In C. the smallest regions of overlap (SROs) between hrHPV-positive carcinomas at chromosome 13 and 20 are summarised.
DISCUSSION

In the present study we compared genome-wide chromosomal profiles of hrHPV-negative HNSCCs, hrHPV-positive HNSCCs and hrHPV-positive CxSCCs, to determine whether hrHPV-associated carcinomas of different origins have similar chromosomal signatures. In addition, potential organ-specific alterations were determined within the total group of hrHPV-positive SCCs.

Unsupervised hierarchical clustering resulted in a separate hrHPV-positive cluster, indicating similarities in the chromosomal profiles of hrHPV-induced carcinomas. Subsequent supervised statistical analysis identified a number of hrHPV-associated chromosomal alterations, including gains at 20p13-q13.33 and losses at 13q21.1-21.33, which were frequent (>50%) in hrHPV-positive carcinomas, but were only rarely observed in hrHPV-negative carcinomas. In contrast, lost regions at 3p and 5q as well as amplifications at 11q13.3 (CCND1 locus) were frequent in hrHPV-negative carcinomas, but not in hrHPV-positive carcinomas. In our previous study comparing only hrHPV-positive and hrHPV-negative HNSCCs, the same regions were identified as well as being specific for hrHPV-negative HNSCCs [13].

In the present study we found that gains at 20q and losses at 13q are specific for hrHPV-positive carcinomas of different anatomical origins. Our previous study, including only hrHPV-positive and hrHPV-negative HNSCCs, did not identify these alterations as specific for hrHPV-positive HNSCCs. However, in the present study a larger number of hrHPV-negative hrHNSCCs and HPV-positive SCCs was included. Furthermore, a sophisticated, objective calling method was presently used to determine gained and lost regions [14]. Interestingly, in a number of studies describing hrHPV E6 and/or E7 mediated immortalisation of keratinocytes of multiple anatomical origins, gains of chromosome 20q and losses of chromosome 13q were consistently observed [19-22]. Low-level gains of chromosome 20q are suggested to be caused by E7 expression and consequent inactivation of the pRb pathway in epithelial cells [19,20,23]. Deletion of part of the long arm of chromosome 13 is frequently found in a number of hrHPV-associated tumors, including cervical, anal and head and neck carcinomas [4,13,24]. Sabbir et al. showed that loss of 13q in HNSCCs was associated with the presence of hrHPV, which is in agreement with our findings [25].

In a recent study Pyeon et al. showed that the gene expression patterns of hrHPV-positive HNSCCs and CxSCCs differed yet shared many changes compared to hrHPV-negative HNSCCs [10]. Our study shows that the same holds true on a chromosomal level. Interestingly, 28% of the genes Pyeon et al. found to be
differentially expressed between hrHPV-positive and hrHPV-negative carcinomas is located within the chromosomal regions identified in this study and showed expression changes concordant with the chromosomal alterations [10]. Of these genes, 39% was located at chromosome 1p, 25% at 5q, and 14% at 3p. The other genes were located at 11q, 18q and 21q. Only one gene, SYCP2, was located within the hrHPV-specific chromosomal alterations found in this study (20q), but was not located within our SRO. To the best of our knowledge none of the genes are known to directly interact with hrHPV E6 and/or E7 [26,27]. Pathway analysis of all genes overlapping with our findings identified cell cycle/proliferation as most overrepresented biological function, which is in concordance with the observations made by Pyeon et al. [10].

Pathway analysis of all genes located within the hrHPV-associated SROs at chromosome 20q and 13q found in this study, again underlined the importance of cell cycle (replication and proliferation) related genes in hrHPV-mediated carcinogenesis. This may be related to the continuous E7-regulated E2F1 activation and is accompanied by changes in overall cellular maintenance systems, such as nucleic acid metabolism, as was also found in HPV16 E7 expressing epithelial cells in vitro [19]. Two genes, NCOA6 and RBM39, overlapped between this in vitro study and our results, warranting further investigation of their role in hrHPV-mediated transformation. NCOA6 encodes a transcriptional coactivator interacting with basal transcription factors, histone acetyltransferases, and methyltransferases. RBM39 encodes an RNA binding protein and possible splicing factor and may act as a transcriptional coactivator for the AP-1 transcription activator complex and estrogen receptors. Other cancer-related genes at 20q include E2F1, which is specifically targeted by hrHPV-mediated degradation of pRb, PIGU, which may play a role in cell cycle control and was identified as an oncogene in bladder cancer [28], and DNMT3B, a de novo DNA methyl transferase. We previously showed that DNMT3B is amplified in the cervical cancer cell line SiHa and found a correlation between increased DNMT3B gene copy numbers and elevated mRNA expression in 78% of CxSCCs [4]. The increased DNMT3B levels are most likely related to the high frequency of tumour suppressor gene promoter hypermethylation events during hrHPV-mediated carcinogenesis [29-31]. The SRO at chromosome 13q includes one known gene, PCDH17, a member of the protocadherin family, which is a subfamily of the cadherin superfamily. At present no reports are available describing interactions between hrHPV and protocadherins. However hrHPV presence has been related to decreased E-cadherin and subsequent impaired immune response [32-34].
Next to the hrHPV-related common events, organ-specific alterations for CxSCCs and HNSCCs were identified as well. CxSCCs more frequently showed loss at 17p, while HNSCCs were characterised by frequent gains at chromosome 8q, harbouring the oncogene c-Myc [35].

Gains at 3q and losses at 11q were found frequently in all SCCs included in this study, suggesting their involvement in carcinogenesis of squamous epithelial cells in general. Interestingly, we previously showed that gains of 3q were highly frequent in CxSCCs but not in adenocarcinomas of the same organ, further emphasising that this alteration may be specific for squamous epithelium [4].

CONCLUSIONS
Together with results from previous studies, our findings support a causal role for hrHPV in the development of a subset of HNSCCs. Consequently, hrHPV-positive and hrHPV-negative HNSCCs should be regarded as different disease entities requiring different diagnostic and therapeutic approaches. The fact that hrHPV-associated SCCs of different organs have chromosomal alterations in common, suggests that these alterations are crucial for hrHPV-induced carcinogenesis. Diagnostic and/or therapeutic targets based on these alterations may therefore be relevant to hrHPV-associated SCCs of all anatomical origins.

ACKNOWLEDGEMENTS
We would like to thank the Mapping Core and Map Finishing groups of the Wellcome Trust Sanger Institute for initial clone supply and verification. This work was partly supported by the Centre for Medical Systems Biology (CMSB) in the framework of the Netherlands Genomic Initiative (NGI).
### Supplementary Table 1. BAC clones and genes included in the SROs of the hrHPV-specific chromosomal alterations at chromosomes 13q and 20q.

<table>
<thead>
<tr>
<th>Cytoband</th>
<th>Start position (bp)</th>
<th>End position (bp)</th>
<th>BAC clones</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>20q11.21-q11.23</td>
<td>29779091.5</td>
<td>34257710.5</td>
<td>RP5-857M17; RP1-310O13; RP5-854E16; RP11-410N8; RP5-1018D12; RP5-1093G12; RP5-1085F17; RP4-733O23; RP5-1137F22; RP5-1184F4; RP4-553F4; RP1-18C9; RP11-234K24</td>
<td>TPX2; MYLK2; FOXS1; DUSP15; TTL9; PDRG1; XKR7; C20orf160; HCK; TM9SF4; TSPYL3; PLAGL2; POFUT1; KIF3B; ASXL1; C20orf112; LOC284805; COMMD7; DNMT3B; MAPRE1; EFCAB8; SPAG4L; BPI1; BPI3; C20orf185; C20orf186; C20orf70; BASE; C20orf71; PLUNC; C20orf114; CDK5RAP1; SNTA1; CBFA2T2; NECAB3; C20orf144; C20orf134; E2F1; PXMP4; ZNF341; CHMP4B; RALY; EIF2S2; ASIP; AHCL; ITCH; DYNLRB1; MAP1LC3A; PIGU; TP53INP2; NCOA6; HMGB3L1; GG7; ACSS2; GSS; MYH7B; TRPC4AP; EDEM2; PROCR; MMP24; EIF6; FAM83C; UQCC; GDF5; CEP250; C20orf173; ERGIC3; FER1L4; SPAG4; CPNE1; RBM12; NFS1; ROMO1; RBM39; PHF20; SCAND1; C20orf152; EPB41L1</td>
</tr>
<tr>
<td>13q21.1</td>
<td>55718989</td>
<td>57862597.5</td>
<td>RP11-640E11; RP1-205J24; RP11-516G5; RP11-204N9; RP11-435P18; RP11-98F3; RP11-111C7; RP11-522F22; RP11-334O13; RP11-538C21; RP11-168J5</td>
<td>LOC729233; LOC729240; LOC729246; LOC729250; PRR20; PCDH17</td>
</tr>
</tbody>
</table>
REFERENCES


CHAPTER 7

Functional and molecular consequences of p53 and pRb pathway inactivation in oral keratinocytes

Submitted for publication

Serge J. Smeets
Tieneke B.M. Schaaij-Visser
Eva A.M. van Veen
Johan van Meerloo
Boudewijn J.M. Braakhuis
Renske D. Steenbergen
Ruud H. Brakenhoff
ABSTRACT
Abrogation of the p53 and pRb pathways is critical in head and neck squamous cell
carcinogenesis. TP53 is mostly inactivated by mutation and the pRb pathway either
by downregulation of p16 (methylation or gene loss) and/or overexpression of cyclin
D1. In head and neck tumors that arise by infection with the human papillomavirus
(HPV) these pathways are abrogated by the viral oncogenes E6 and E7, respectively. Aim of our study was to establish the functional and molecular
consequences of p53 and pRb pathway inactivation irrespective whether this might
occur by HPV infection or a genetic event. We applied a model with conditionally
transformed oral squamous cells that allows studying the effect of gene manipulation
on immortalization and down-stream gene expression. We show that abrogation of
the p53 pathway by either shRNA mediated knock-down of the TP53 gene
(mimicking a nonsense mutation), overexpression of dominant-negative missense
mutant p53R(175)H or overexpression of the HPV16 oncoprotein E6, led to an
extended life span. Despite a seemingly identical phenotype, gene expression
profiles considerably varied in relation to the type of p53 inactivation. The dominant-
negative mutant p53R(175)H showed highly significant different molecular changes
associated with an activation of the hypoxia signalling (HIF1) and WNT signaling
pathways. Knock-down of p16, overexpression of cyclin D1 or overexpression of
HPV16 E7 had no direct effect on the lifespan, but all led in combination with p53
pathway abrogation to an immortal phenotype. Our data show the critical role of p53
and pRb pathway abrogation in early squamous carcinogenesis by either HPV
infection or genetic events, and indicate that these are the earliest changes in the
carcinogenic process. The way by which the p53 pathway is inactivated has
functional consequences at the molecular level.
INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) comprises about 5% of all newly diagnosed cancer cases in the Northern and Western European countries and the United States, and is the fifth most common cancer worldwide [1]. Despite advances in local tumor control, the five-year survival rates have only moderately improved during the last 20 years. Identification of the cancer genes or pathways causally involved in carcinogenesis seems crucial to enable development of new diagnostic tools and targeted therapies.

Based on epidemiological data and in vitro transformation experiments, it has been estimated that four to six genetic events are required to transform a normal human cell into a malignant cell [2;3]. In various tumor types these events entail the impairment of the p53 and the pRb pathways, which are involved in cell cycle regulation and apoptosis [4]. In some tumor types such as colorectal cancer, p53 abrogation is considered a late event, while in other types this seems an early change [5;6]. Also in HNSCC carcinogenesis, the p53 pathway appears to be abrogated in more than 60% by mutations of the TP53 gene, often accompanied by allelic loss at the 17p13 region [5]. Evidence has been provided suggesting that these genetic changes belong to the earliest events [5;7]. TP53 inactivating mutations are either nonsense mutations or deletions and insertions that lead to a truncated protein. In addition, missense mutations are frequently found resulting into a protein with impaired DNA binding. Many missense mutants exert a dominant-negative effect in a heterozygous cell by forming mixed tetramers with wild type p53; this effective reduction of the cellular levels of wild-type homotetramers contributes to p53 inactivation. Recently, it was shown that the type of TP53 mutation seem to have clinical consequences in HNSCC, and that tumors with so called disruptive mutations show a less favorable outcome [8].

Besides TP53 mutations, chromosomal loss of 9p21 and amplification of 11q13 are found in 60-80% of the HNSCC, also in precursor stages, suggesting early impairment of the pRb pathway by p16 down- and cyclin D1 upregulation. Although there is compelling evidence for p16, the supportive role of cyclin D1 is less clear [9;10]. Moreover, the precise timing of the different events has not been elucidated and is still debated.

The hypothesis that abrogation of the p53 and pRb pathways are key events in HNSCC carcinogenesis is supported by the etiological role of the human papillomavirus (HPV). A subgroup of HNSCC, particularly those in the oropharynx, are caused by HPV infection [11-13], the virus known to be involved in 100% of
cervical cancers. The virus produces two oncoproteins encoded by the E6 and E7 genes, which inactivate the p53 and pRb proteins, respectively. These two viral oncoproteins are crucial in the carcinogenic process, but are not sufficient to induce malignant transformation into an invasive carcinoma as in HPV-positive head and neck tumors specific additional genetic changes are found in the cancer cell genome [11;14].

It is at present unknown whether the inactivation of the p53 and pRb pathways by both HPV E6 and E7 is functionally equivalent to mutational inactivation of p53 or changes in p16 and/or cyclin D1. A compelling approach to identify and characterize cancer genes involved in squamous carcinogenesis is to assess the transforming ability of these genes in normal mucosal keratinocytes in vitro, but experimental transformation of primary human oral keratinocytes is very difficult due to the rapid premature senescence of keratinocytes in culture [15;16]. In a few passages the cells stop proliferation and die, a biological phenomenon seriously hampering functional experiments. Here we present a conditionally immortalized model of oral squamous cells of a normal healthy individual. With this model we performed a detailed investigation of p53 and pRb pathway abrogation in oral carcinogenesis, and the functional consequences at the molecular level.

RESULTS

Establishment of the in vitro model of conditionally transformed oral keratinocytes.

We cultured keratinocytes isolated from the oral mucosa and transduced the cells at their first passage with amphotropic retroviral vectors carrying a temperature-sensitive mutant of the SV40 large T-antigen (tsLT) and the telomerase subunit hTERT. Before we generated these cultures, we checked that primary oral keratinocytes grow well at both 32 °C and 39 °C. In total 17 different clones developed; the remaining cells stopped proliferation and died. The clones were individually expanded to cell lines. Three of these SV40 transformed cell lines have been cultured for more than 150 population doublings at 32°C confirming their immortal state at the temperature at which tsLT is active. The 17 lines were further characterized for their sensitivity to calcium, a strong differentiation inducer of primary keratinocytes, and their ability to grow on soft agar, a phenotype associated with invasive growth, at the permissive temperature of 32 °C. These are typical characteristics of established malignant squamous carcinoma cell lines. None of the tsLT-hTERT-keratinocyte lines showed growth in soft-agar or calcium containing
medium (data not shown), which suggested that these cells are still in an early state of carcinogenesis. All cell lines were shifted to 39°C to test the effect of temperature induced inactivation of tsLT. After the temperature shift cells immediately went into proliferation arrest (less than one additional doubling), and died during the following one to three weeks. To confirm that these lines express telomerase (hTERT), a TRAP assay measuring telomerase activity was performed on the three cell lines that were chosen for further study. All tested SV40-LT-hTERT keratinocyte cell lines, but not primary oral keratinocytes, were positive in the TRAP assay (Table I).

<table>
<thead>
<tr>
<th>Table I. Results of the TRAP-assay measuring telomerase activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
</tr>
<tr>
<td>SiHa (+)</td>
</tr>
<tr>
<td>tsLT-OK</td>
</tr>
<tr>
<td>OK (-)</td>
</tr>
</tbody>
</table>

RTA = Relative telomerase activity. A lysate of the conditionally immortalized oral keratinocytes (tsLT-OK) transduced by p53R(175)H and cultured at 39 °C for 2 weeks was analyzed for telomerase activity. These cells show an enhanced telomerase activity as compared to primary keratinocytes. SiHa cells were used as positive control.

Abrogation of p53 leads to an extended lifespan
We determined the effect of p53 abrogation on the proliferation arrest that occurred after temperature shift to 39°C. We inhibited p53 expression by using a p53 shRNA construct [17], mimicking a nonsense mutation. In addition, we expressed the dominant-negative missense mutant p53R(175)H. Finally, we introduced the HPV16 oncogene E6. Knockdown of TP53 gene expression by introduction of shRNA as well as blocking the wild type p53 protein by either overexpression of p53R(175)H or HPV16 viral oncogene E6, resulted all in an extended lifespan of approximately 15 cell doublings in a 2-month period (Figure 1A). After this extended lifespan all cells simultaneously entered a proliferation arrest within a week. These experiments were repeated with three different lines in duplicate. In one of these experiments we analyzed the p53 protein level in the cultures four weeks after gene manipulation by Western blot. In Figure 1B the effect of the various gene manipulations on the level of p53 expression is shown, as well as the effect on the level of the downstream protein p21. For p53R(175)H enhanced expression of the mutant protein is detected and for the other two manipulations, the relative levels of p53 protein were much lower. The protein level of p21, a well-known protein of which the expression is regulated by p53, was low in all manipulated cells lines, a molecular confirmation that p53 is not active. Both p53 and p21 were not downregulated at 32 °C when the SV40 large T
antigen is active. Although this seems somewhat unexpected, it confirms previous observations that the exact working mechanism of SV40 large T remains elusive [18].

**Figure 1.** Lifespan of SV40-LT-hTERT keratinocytes at 39°C. Values are averaged over three cell lines. A: Effect of p53 pathway abrogation. Number of population doublings (PDs) are depicted after introduction of empty vector ((-) negative control), TP53 shRNA (17.3±2.3 PDs), p53R(175)H cDNA (20.1±0.5 PDs) or HPV16 E6 cDNA (17.0±2.4 PDs). B: Downregulation of p53 and p21 protein levels after genetic manipulation assayed by Western blot. Cell lysates of tsLT oral keratinocytes cultured at 32°C were used as control. Manipulated cells were cultured at 39 °C for four weeks before harvest. For p53R(175)H the (mutant) protein is detected but the low expression of p21 indicates that the detected p53 protein is not active. Inactivation of p53 by HPV16-E6 and p53 shRNA causes a low level of p53 and downstream p21 proteins. C: Effect of pRB pathway abrogation: Number of PDs after introduction of empty vector (negative control), HPV16 E6 cDNA (positive control) (12±0.5 PDs), cyclin D1 cDNA, p16 shRNA, a combination of both, or E7 cDNA. None of the cells with a pRb pathway abrogation showed an extended lifespan.

*Abrogation of the pRb pathway has no effect on life span*

To investigate pRb pathway abrogation, we manipulated the pathway in four different ways: by 1) overexpression of cyclin D1, 2) downregulation of p16, 3) a combination of both, or 4) overexpression of HPV-E7. None of these pRb pathway abrogations caused a change in the lifespan of SV40-LT-hTERT keratinocytes, compared to cells transduced with an empty vector. All cells died within one to three weeks after the temperature shift (Figure 1C), which also hampered analysis of the effect of the gene manipulations at the molecular level.
Combination of p53 and pRb pathway abrogation causes immortalization

Based on all existing descriptive data collected from the genetic analysis of HNSCC and its precursor lesions as well as the fact that HPV contains two oncogenes that block both pathways simultaneously we raised the hypothesis that pRb pathway abrogation only has effect in combination with functional inactivation of p53. Therefore, we inactivated p53 by overexpression of the viral oncogene HPV-E6, followed by perturbation of the pRb pathway, using the same manipulations as indicated above. This resulted in four different SV40-LT-hTERT keratinocyte variants: E6/p16 shRNA; E6/cyclin D1; E6/p16 shRNA and cyclin D1, as well as E6/E7. All combinations of p53 and pRb pathway perturbation led to an immortal phenotype as the cells could be grown for more than 100 Population Doublings (PDs) at 39°C (Figure 2A) with a similar proliferation rate (Figure 2B).

To proof that HPV16-E6 was successfully transcribed, the expression of E6 was determined by RT-PCR, and an Enzyme Immuno Assay (EIA). The cells that were transduced by the E6 containing vector were all highly positive in the EIA (Figure 2C). The effect of overexpression of HPV16-E6 on p53 protein level was investigated by Western blot, and a clear downregulation was observed (Figure 2C). Cyclin D1 was investigated at the RNA level using qRT-PCR and an increased mRNA expression level was detected (data not shown). P16 protein levels were also investigated by Western blot. All cells harboring the p16-specific shRNA displayed a lowered expression of p16 protein when compared to the control cells (Figure 2D).
Figure 2. A: P53 inactivation in combination with pRb pathway abrogation: Number of Population Doublings after introduction of respectively empty vector (-), E6 cDNA, E6 cDNA in combination with p16 shRNA, cyclin D1 cDNA (CD1), a combination of both, or E7 cDNA. Values are averaged over three cell lines. In contrast to the negative control (-) and E6 cDNA, all manipulations abrogating the pRb pathway in the context of p53 downregulation by HPV16 E6 led to immortalization (>100 PD’s). B: Proliferation rates between the various conditions described in A. C: Levels of p53 protein under the conditions indicated in A analyzed by Western blotting using monoclonal antibody DO-7. SV40-tsLT keratinocytes grown at 39°C were used as control. ß-actin was included as a loading control. Presence of HPV16 E6 transcripts was measured by Enzyme Immuno Assay (EIA); (+) = EIA OD value that reached at least 3x the background, (-) = EIA OD value is below 3x background. D: Analysis of p16 expression levels by Western Blotting. Cell lysates of tsLT-keratinocytes with and without E6 overexression and p16 downregulation, cultured at 39°C were analyzed. Cell lines SiHa and UM-SCC-22A were included as positive and negative controls, respectively. Equal amounts of protein were separated by 12 % SDS/PAGE, blotted and p16 detected with a p16-specific monoclonal antibody. ß-actin was included as a loading control.
Expression profiling of cells with different ways of p53-pathway abrogation

We observed that the various methods of p53 pathway abrogation showed a comparable extended lifespan without apparent differences in growth rate. This was unexpected as several lines of evidence suggest that dominant-negative TP53 mutations like R(175)H, can lead to a gain-of-function tumorigenic activity in vivo [19-21]. Also HPV16 E6 interacts with a variety of cellular proteins [22], of which the targeting for proteasome-mediated degradation of the p53 protein is best characterized [23]. To investigate potential differences between the three methods of TP53 inactivation at the molecular level, we used microarray expression profiling. Our hypothesis was that introduction of the dominant-negative mutant p53R(175)H or HPV16-E6 might have additional molecular effects when compared to the knockdown of TP53 by short hairpin RNA. Therefore we used cells manipulated by shRNA transfection as reference condition. P53 inactivation by the dominant-negative mutant R(175)H or HPV16-E6 resulted in respectively 2467 or 2399 significantly (p<0.05, corrected for the false discovery rate (FDR) differentially expressed genes compared to the shRNA reference (Figure 3A; Supplementary Table 1). Some of these genes were shared (n=784), but a larger number seemed specific for either p53R(175)H or HPV16-E6. To interpret the biological significance of the lists of differentially expressed genes, a gene ontology analysis was performed by Ingenuity Pathways Analysis. As the expression of TP53 was manipulated in the reference condition, we removed this gene from the dataset prior to the analysis. Cells with a p53 pathway abrogation by HPV16 oncoprotein E6 did not reveal any significant additional pathway changes as compared to the shRNA reference condition. However, abrogation of p53 function by overexpression of the R(175)H mutant showed significant specific involvement of 2 canonical pathways, i.e. ‘Hypoxia signaling (HIF1)’ and ‘WNT/ beta-catenin signaling’ (p<0.05, FDR corrected) (Figure 3B; Supplementary Figure 1). Significant involvement of only these two pathways were confirmed using Panther Gene Ontology [24].
Figure 3. Genes that showed differential expression in keratinocytes with an extended lifespan induced by either overexpression of HPV16-E6 or p53R(175)H. Knockdown of TP53 by shRNA was used as reference condition. A: Venn diagram showing the differentially expressed genes (p<0.05, FDR corrected [39]) specific for and shared by HPV16-E6 and p53R(175)H. B: Significant canonical pathways discovered by Ingenuity Pathway Analysis (IPA) using the differentially expressed genes indicated under A. Expression of cells with HPV16-E6 (black) or mutant p53R(175)H (grey) were compared to p53 knockdown by shRNA as reference. The threshold line represents the p-value 0.05 (corrected for multiple testing according to Benjamini-Hochberg [39]).
DISCUSSION
The functional consequences of various p53 and pRb pathway abrogations found in HNSCC by either HPV infection or specific genetic events, has not been investigated before. To allow these studies we have established an in vitro model of early carcinogenesis using conditionally transformed oral keratinocytes. All different approaches used for p53 pathway inactivation; overexpression of dominant-negative mutant p53R(175)H, overexpression of HPV16-E6, or knockdown of the TP53 gene by shRNA, caused an extended lifespan of about 15 cell doublings. Abrogation of the pRb pathway by cyclin D1 overexpression, p16 knockdown or HPV16-E7 overexpression had no effect, but in combination with p53 inactivation led to an immortal phenotype, albeit in the context of telomerase expression. These results implicate in our view that p53 and pRb pathway abrogation are early genetic hits in HNSCC carcinogenesis as had already been hypothesized by Braakhuis et al. in their progression model [5]. Our data are in line with those from previous studies showing that p53 and pRb pathway manipulations are necessary to transform primary oral keratinocytes [25;26]. It should be stressed, however, that although the large majority of HNSCCs show a clear manifestation of p53 pathway abrogation, there are tumors that are TP53-wild type and do not contain HPV [11;27]. It is unclear whether the p53 pathway is functionally impaired in these tumors by other mechanisms, or that there are p53 independent routes of HNSCC carcinogenesis. Our current data strongly support the role of p53 (and pRb) pathway abrogation in the malignant transformation of oral keratinocytes, but do not exclude alternative routes.

With respect to the proliferation arrest, it seems of no importance how the p53- or pRb pathways are abrogated. Every manipulation tested for p53 abrogation led to the same extended lifespan, and every manipulation of the pRb pathway in the context of E6 overexpression led to immortalization. In this respect it is remarkable that loss of 9p21 and amplification of 11q13, chromosomal regions containing the p16 and cyclin D1 genes, respectively, are frequently found together in the same tumor [9]. Detailed analysis of the various key players in this same model might help to elucidate whether cyclin D1 is indeed always the driving gene at the 11q13 locus [28], or that other candidate genes within this amplicon like FADD could also be involved [29].

It has previously been shown that different p53 mutations lead to different cancer-associated phenotypes. In transgenic mouse models it was shown that overexpression of the mutant p53R(175)H (in mouse R(172)H) shows a cancer phenotype different from that in p53-/- mice [20;21]. Evidence was provided that...
deregulation of the p63/p73 proteins by the R(175)H mutant might explain at least part of the different cancer phenotypes observed [20]. We investigated the downstream effects of the various ways of p53 inactivation in our model by microarray expression profiling using p53 shRNA as reference condition. When we compared the molecular effects of the dominant-negative missense mutation R(175)H to TP53 gene knock-down we noted that two pathways were specifically identified: the hypoxia signalling (HIF1) and the WNT signaling pathway. The reliability of this finding is supported by the fact that all the determinant genes indicating these pathways are in the non-overlapping differential gene set of 2,467 genes in Figure 3A. The apparent link between p53 and HIF1 has been reported previously. The triad of p53, Myc and HIF1 appears to be the key regulator of the glycolysis, as in many cancers with changes in the activity of these genes the glucose metabolism is deregulated. This phenomenon is known as the “Warburg” effect and recently even called the “Seventh hallmark of cancer” [30]. In this context the function of p53 is considered to be downregulated and those of Myc and HIF1 upregulated. Metabolic profiling of cells with different functional abrogations of p53 might reveal whether this is indeed associated with the type of mutation as suggested by our data.

The second pathway that was significantly associated with overexpression of p53R(175)H, was the WNT signaling. Activation of the WNT signaling pathway is the initial step in the development of colorectal cancer, and seems to play an important role in the epithelial stem cell homeostasis of the colon. There is no compelling evidence at present that the WNT signaling pathway is involved in head and neck carcinogenesis. However, there is strong evidence for a role of the WNT signaling pathway in HPV-mediated cellular transformation [31]. Introduction of HPV in primary foreskin keratinocytes leads to immortalization, but without transformation to a malignant phenotype. Only by subsequent activation of the WNT signaling pathway a transformed phenotype is obtained, an observation that was confirmed in clinical material. This indicated that WNT activation might be an important second hit in HPV-mediated squamous cancer progression [31]. Our data suggest that this might be true for tumors with a TP53 nonsense mutation as well. Tumors with a TP53 missense mutation comparable to p53R(175)H might already have a change in the WNT signaling pathway caused by the oncogenic mutation. Whether this effect on gene expression profiles leads to other cancer-associated phenotypes or might be specific for this particular TP53 mutation awaits further investigation.
In summary, our study provides evidence for TP53 downregulation as an early hit in HNSCC carcinogenesis, and leads to immortalization of mucosal keratinocytes when the p16-cyclin D1-pRb pathway becomes abrogated as well, albeit in the context of telomerase expression. Based on these data we postulate that HNSCC carcinogenesis starts with TP53 mutation and pRb pathway abrogation, an observation that fits with the genetic changes found in precursor lesions [5;7]. The apparent different molecular and clinical effects of different p53 mutant proteins inactivation might be exploited to select patients for other treatment regimens and for the development of specific targeted therapies.
MATERIALS AND METHODS

Cell culture and vector constructs

Normal human oral keratinocytes were isolated from an excised uvula of a non-smoking individual who never had consumed alcoholic beverages. Isolation and culturing of the primary oral keratinocytes was performed as previously described [32], except that the cells were cultured at 32°C and that 0.1% of bovine serum albumin (BSA) was added to the medium to inactivate trypsin after passaging the cells.

Amphotropic retroviral supernatants were produced by transient transfection of Phoenix packaging cells with retroviral vector constructs [33] using Fugene 6 (Roche Diagnostics, Woerden, The Netherlands). Phoenix cells were cultured in DMEM supplemented with 5% heat-inactivated fetal calf serum (LONZA, Basel, Switzerland) at 37°C. Viral supernatants were filtered through a 0.45 mm filter and infections were carried out at three following days for 4 hours in the presence of 3 µg/ml polybrene (Sigma-Aldrich, Zwijndrecht, The Netherlands). After infection the viral supernatant was discarded, and cells were washed thoroughly with phosphate buffered saline (PBS) to remove the Ca²⁺ residues from the viral supernatant to avoid keratinocyte differentiation and proliferation arrest.

Vector constructs used were pBabe-hTERT-hygro and pMESVts-neo [17;34], containing the telomerase catalytic subunit and a temperature-sensitive allele of the SV40-Large T antigen (tsLT), respectively. Drug selection was performed with 50 µg/ml Neomycin (Invitrogen, Breda, The Netherlands).

Control transfections were performed with retroviral vectors without insert. LZRS-GFP, a retroviral vector marked with the Green Fluorescence Protein (GFP) gene, was used to check the transfection efficiency by UV-microscopy.

Culturing of the tsLT-keratinocytes was performed according to the same procedure as primary keratinocytes at incubation temperatures of 32°C or 39°C. Media were refreshed twice a week and cells were passaged when 70% confluence was reached.

Calcium sensitivity and soft agar growth

The tsLT-keratinocytes were monitored for two hallmarks of transformation; their resistance to calcium-induced differentiation and their capability to grow in soft agar.

Calcium sensitivity was analyzed by adding 2.5 mM Ca²⁺ to the culture medium during a culturing period of two weeks at 32°C. Cells were considered sensitive to
Ca\(^{2+}\) when the cells differentiated, stopped proliferating and reached premature senescence.

For the analysis of anchorage independent soft agar growth, 5,000 cells were mixed in 0.35% TOP agarose in Keratinocyte Growth Medium (KGM) (Invitrogen, Breda, The Netherlands) [32] and plated onto a bottom of 0.6% agarose in KGM. Cells were cultured for 20 days before colonies were counted. As a positive and negative control UM-SCC-22A cells and primary oral keratinocytes were taken, respectively.

**P53 and pRb pathway abrogation: vector-constructs and lifespan analysis**

Vector constructs used for p53 abrogation were pWZL-p53DN-hygro (encoding the human p53 dominant-negative mutant Arg(175)His or R(175)H), pRS-p53shRNA-puro containing a short hairpin RNA specific for TP53 as described in [17] and LZRS-E6-GFP (containing the HPV16-E6 oncogene: kindly provided by Dr. E. Hooijberg, Amsterdam, Netherlands).

Vector constructs used for pRb-pathway abrogation were pRS-p16shRNA-puro containing a p16 specific short hairpin sequence [17], pBabe-cyclinD1 and pBabe HPV16-E7-puro (kindly provided by Dr. K. Münger, Bethesda, United States). Cells were counted by microscope and population doublings (PD) calculated with the formula: $PD = 2 \log (n_c/n_0)$; $n_c$ stands for the number of counted cells and $n_0$ for the initial number of cells seeded. The total number of PDs was taken as lifespan.
Molecular analysis of gene manipulations

RNA and DNA isolation
RNA and DNA isolation from cell pellets was performed using RNAbee and DNAsstat as previously described [35].

Detection of cyclin D1 expression by real-time quantitative RT-PCR
Real time quantitative RT-PCR was performed using the ABI Prism 7700 Sequence Detection System (Applied Biosystems, Nieuwerkerk AD IJssel, The Netherlands) as was previously described [35]. Intron-spanning primers were chosen to avoid amplification of contaminating genomic DNA. Primers and probes specific for cyclin D1 were selected using PrimerExpress software (Applied Biosystems, Nieuwerkerk AD IJssel, The Netherlands). Selected sequences of primers and probes (Isogen, IJsselstein, The Netherlands) are listed in Table II. RNA input was determined by using the housekeeping gene BGUS as a reference.

Detection of HPV16 E6 mRNA.
Primers to detect HPV E6 are listed in Table II. PCR products were detected on agarose gel and confirmed using an enzyme immunoassay (EIA) as described previously [36;37] with a HPV16-E6 specific probe (Table II). The EIA for E6 was only considered positive when the EIA OD value reached at least 3x the background value.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amplicon size (bp)</th>
<th>Primer/probe 5'-Sequence-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV16-E6</td>
<td>86</td>
<td>Forward TACTCGCACGTGGAGGTGTA Reverse GGAATCTTTGCTTTTTGTC Probe AAGCCACTGCTTCTGAAGAAAGCAA</td>
</tr>
<tr>
<td>CCND1</td>
<td>81</td>
<td>Forward TACTACCCCTCACACGCTTC Reverse TTCGATCTGCTCCTGGCAG Probe ATCGAGTGAGCCAGAGACTGCTCCG</td>
</tr>
</tbody>
</table>

TRAP assay
For quantitative determination of telomerase activity, the TeloTAGGG Telomerase PCR ELISA plus kit (Roche Diagnostics, Mannheim, Germany) was used. The procedure was performed according to the manufacturer’s protocol, with an input of 250 ng protein. All samples were analysed in duplicate. The tumor cell line SiHa was included as a positive control. Primary oral keratinocytes were included as a negative control.
Western Blotting

Cells were lysed using a buffer of 10 mM Tris, 0.65% NP40 and 150 mM NaCl (pH 8.0) and protein concentrations were measured using a nanodrop spectrophotometer (Isogen). Western Blot was performed as previously described [38]. Cell lines SiHa and UM-SCC-22A were used as controls. Proteins were detected with antibodies specific for p16\(^{INK4A}\) (clone DCS50.1; Pharmigen 1:500) and p53 (clone DO7; DAKO, 1:800). beta-actin (clone 150 1R; Chemicon, 1:3,000) was used as a loading control.

Gene expression profiling of cells with different p53-pathway abrogation

In two different tsLT keratinocyte lines, p53 was inactivated in triplicate by either TP53 shRNA, HPV16-E6 or p53R(175)H, allowing six replicates per manipulated condition. Cells were transduced at 32°C as described above and checked for transduction efficiency by LZRS-GFP control. Cells were harvested synchronously after three weeks of culturing at 39°C. Microarray hybridization, using the Agilent Low RNA Input Fluorescent Linear Amplification Kit and 4x44K Whole Human Genome Arrays, was carried out according to the manufacturer (Agilent Technologies, Amstelveen, The Netherlands). We assumed that overexpression of dominant-negative mutant p53R(175)H and viral oncogene HPV16-E6 would have increased activity as compared to TP53 shRNA inactivation, and therefore analyzed the data using the shRNA manipulated cells as reference. The gene expression data have been deposited in the NCBI’s Gene Expression Omnibus (GEO), with series accession number GSE12553.

For interpretation of the biological processes and canonical pathways, we used Ingenuity Pathways Analysis 6.0 (Ingenuity Systems®, Redwood City, USA). Gene lists containing the significant genes (FDR corrected p-value \(p<0.05\)), of the two conditions used for p53 abrogation (HPV16-E6 and p53R(175)H) versus TP53 shRNA as reference condition, were uploaded into the application. The significance of a canonical pathway is controlled by a Benjamini-Hochberg corrected p-value [39], considering \(p<0.05\) as significant. The same analysis was repeated using Panther Gene Ontology analysis [24].
ACKNOWLEDGMENTS

We thank K. Berns and R. Bernards for providing vector constructs and helpful discussions, A.T. Hesselink for the detection of the HPV-E6 transcript, D. Claassen-Kramer for performing the TRAP assay, N. Nassipour for detailed phenotypic characterization of the tsLT-keratinocyte cell lines, P.P. Eijk as well as M.A. van de Wiel for support with the microarray analysis, G. van Liempt, M. Stigter-van Walsum and A. Huizenga for experimental assistance.

Supplementary Figure 1. Ingenuity pathway analysis schemes of the hypoxia signalling (HIF1) and the Wnt/β-catenin signalling pathway, respectively. Increased RNA levels are indicated in red, decreased levels in green and unaffected levels in white. The color intensity correlates to the fold change. When multiple ‘family-members’ of one particular gene shows both increased and decreased expression, a mixture of green and red is used. As we manipulated the TP53 transcript levels by shRNA knockdown, we removed the gene from the dataset.
REFERENCES


CHAPTER 8

Figures in color
Figure 2. HNSCC progression model. Braakhuis et al. suggested an adapted progression model for cancer in the head and neck. A stem cell acquires one or more genetic alterations and forms a patch, a small cluster of less than 200 cells in diameter that can be detected by p53 immunostaining. Eventually the stem cell escapes normal growth control. A precursor lesion arises and laterally displaces the normal mucosa. Additional genetic hits give rise to different subclones within a field and eventually a subclone evolves into invasive cancer [61].
Figure 3. Principle of array CGH. This figure shows the steps in BAC array CGH. (A) BAC clones are selected from a physical map of the genome. (B) DNA samples are extracted from selected BAC clones and their identity is confirmed by DNA fingerprinting or sequence analysis. (C) A multi-step amplification process generates sufficient material from each clone for array spotting. Each clone is spotted in replicate onto a solid support. (D) Reference DNA and test DNA are differentially labeled with cyanine 3 and cyanine 5 respectively. (E) The two labeled products are combined and hybridized onto the spotted slide. (F) Images from hybridized slides are obtained by scanning in two channels. Signal intensity ratios from individual spots can be displayed as a simple plot (G) or by using more complex software such as Imagene, which can display copy number alterations throughout the whole genome [88].
Colored figure on backside of this page
Colored figure on backside of this page
Figure 2. Typical examples of HPV16 FISH on sections of paraffin embedded HNSCC specimens. All three tumors were positive for HPV DNA and E6/E7 expression. The images A and B show punctuate nuclear FISH signals indicating HPV16 DNA integrated into the host genome. Image C shows an area with diffuse nuclear FISH staining indicative for episomal HPV16 DNA.

Figure 3. Representative examples of p16 immunostaining on three tumors of the HPV D+/R+ group (A-C) and one of the HPV-negative group (D). A = Tumor 1 (table 2) with 100% of the cells stained with high intensity in both nucleus and cytoplasm; B = Tumor 5 (table 2) with approximately 50% of the cells positively stained with high intensity; C = Tumor 11 (table 2) with 100% of the cells stained, but with a lower intensity; D = Tumor 26 (table 2) negative for p16 immunostaining.
Figure 2. A genetic progression model of multi-step head and neck carcinogenesis is proposed. The two etiological factors, smoking and HPV16 are incorporated into the "patch-field-cancer" model that in essence has previously been published [9]. In this model the development of a field with genetically altered cells play a central role. In the initial phase a "patch" develops, a clonal unit in which the stem cell and its daughter cells acquire a genetic alteration. Some evidence point to an alteration in p53/MDM2 pathway as a likely first event. This pathway is disrupted by a mutation of TP53 in case of smoking as a causative factor, or alternatively, by the HPV effect of E6, that results in a degradation of p53. The conversion of a patch into a field is the next step in this progression model, and this field expands at the expense of normal epithelium. An important event in the field phase is the impairment of the p16/CDK/pRb pathway [22]. In case of smoking this pathway is disrupted by inactivating CDKN2A, the gene encoding p16, by mutation, chromosomal loss or promoter hypermethylation. Alternatively, this pathway can be impaired by E7, a protein that is produced by HPV. Next, clonal divergence leads to the development of one or more tumors within the contiguous field of preneoplastic cells. Considering the two etiological factors, the common and differential chromosomal events possibly involved in the progression from field to carcinoma are shown. At this moment it is difficult to point out the pathways that are involved in this stage of carcinogenesis. CCND1 (Cyclin D1) is likely involved in smoking-related HNSCC, since the 11q13 region is found to be highly significantly amplified; cyclin D1 is part of the p16/CDK/pRb pathway, a component of the CDK/Cyclin complex that phosphorylates pRb causing S-phase entry.
Colored figure on backside of this page
Colored figure on backside of this page
Figure 1. WECCA heatmap of 39 OOSCCs without HPV involvement. At the top of the tree two tumor groups can be distinguished: a group consisting of 8 OOSCC with in general genetically silent profiles and another group of 31 tumor with many genetic aberrations. In group 2 downwards, two subgroups (a and b) can be distinguished with 2b as the group with the most genetic aberrations. The x-axis represents tumor numbers and the y-axis chromosome numbers.
Supplementary Figure 1. WECCA heatmap of 51 OOSCC, 39 without (present study) and 12 with HPV16-infection. Results of the HPV-infected tumors are derived from a previous publication [31]. seven out of 12 HPV-positive tumors formed a separate group and five of them rearranged among the group of silent tumors (indicated with black bars near the x-axis). The x-axis represents tumor numbers and the y-axis chromosome numbers.
Colored figure on backside of this page
Supplementary Figure 2. WECCA heatmap of 89 OOSCCs described by Snijders et al. [32]. General classification is similar as in the present study; a group consisting of 33 OOSCC with in general a low level of chromosomal aberrations and another group of 56 tumors with many chromosomal aberrations. In group 2 downwards, two subgroups (a and b) can be distinguished with 2b as the group with the most chromosomal aberrations. The y-axis represents chromosome numbers.
Figure 1. Result from the unsupervised hierarchical clustering analysis. Cluster 1 contains in majority hrHPV-positive carcinomas, as is indicated by the black boxes in the legend underneath the heatmap (p<0.0001).
Colored figure on backside of this page
Colored figure on backside of this page
Figure 3. Genes that showed differential expression in keratinocytes with an extended lifespan induced by either overexpression of HPV16-E6 or p53R(175)H. Knockdown of TP53 by shRNA was used as reference condition. **A:** Venn diagram showing the differentially expressed genes (p<0.05, FDR corrected [39]) specific for and shared by HPV16-E6 and p53R(175)H. **B:** Significant canonical pathways discovered by Ingenuity Pathway Analysis (IPA) using the differentially expressed genes indicated under **A.** Expression of cells with HPV16 E6 (purple) or mutant p53R(175)H (blue) were compared to p53 knockdown by shRNA as reference. The threshold line represents the p-value 0.05 (corrected for multiple testing according to Benjamini-Hochberg [39]).

Supplementary Figure 1. Ingenuity pathway analysis schemes of the hypoxia signalling (HIF1) and the Wnt/β-catenin signalling pathway, respectively. Increased RNA levels are indicated in red, decreased levels in green and unaffected levels in white. The color intensity correlates to the fold change. When multiple ‘family-members’ of one particular gene shows both increased and decreased expression, a mixture of green and red is used. As we manipulated the TP53 transcript levels by shRNA knockdown, we removed the gene from the dataset.
CHAPTER 9

Summary and Future perspectives
SUMMARY AND FUTURE PERSPECTIVES

Despite major improvements in surgical techniques and radiotherapy in the clinical management of squamous cell carcinoma in the head and neck (HNSCC), the long term survival of HNSCC patients has only moderately improved during the last 20 years [1]. Significant improvements may be reached when the mechanisms underlying HNSCC carcinogenesis would be better characterized and the causative genes identified. The main risk factors for the development of HNSCC are tobacco smoking and excessive alcohol consumption [2]. Recently, also the involvement of the human papillomavirus (HPV) in a subset of HNSCC cases was firmly established [3-8]. The reported frequencies of HNSCC with high-risk HPV involvement vary tremendously, ranging from 0 to 100% [9;10]. Part of the variation in HPV prevalence could be explained by differences in the location of the tumor, i.e. the prevalence is relatively low in the oral cavity and high in the tonsil [5;6;11]. More importantly, variations in the type of tissue material studied and the HPV detection method used, may have a major impact on the discrepancy in the reported prevalence rates [5;9;12;13]. In Chapter 2, an explanation is provided for the discrepancy in the literature about HPV prevalence in HNSCC. We evaluated commonly used HPV detection methods and used that experience to find a combination of techniques to reach the most reliable HPV detection. This HPV detection algorithm had to be applicable for archival formalin-fixed paraffin-embedded (FFPE) material, since the preparation of FFPE specimens is routine in histopathological diagnostics worldwide. We compared quantitative reverse transcriptase polymerase chain reaction (RT-PCR), fluorescence in-situ hybridization (FISH), general primer PCR (GP-PCR), HPV protein detection in sera, and p16-immunostaining. As a golden standard we used the presence of viral transcripts (oncogenes E6/E7) in frozen tissue samples of the same tumors, a method that reflects active virus involvement. None of the tested methods was optimal with respect to sensitivity and specificity, but an algorithm based on a combination of two frequently used diagnostic techniques reached 100% for both parameters. All HNSCC tumors should be tested first for p16 overexpression by immunostaining, followed by a GP5+/6+ DNA PCR on the p16-positive cases.

In this thesis we also focused on a genetic fingerprint of HNSCC in relation to clinical and molecular variables. For this purpose, a micro-array CGH platform with a genomic resolution of 1 mega base (Mb) was established. This platform allows mapping of the genome-wide numerical genetic alterations in a high throughput manner. When genomic data is generated in a high-throughput manner good bioinformatics becomes necessary. In Chapter 3 a 'bioinformatics'-tool is described
which enables calculation of significance for genetic alterations that differ between tumor groups [14]. Specific statistical approaches need to be applied for ordinal data as produced by array CGH. Furthermore, a correction for multiple testing was incorporated by defining a false discovery rate (FDR) according to Benjamini and Yekutieli [15].

The established array CGH platform and developed bioinformatics were applied to genetically characterize HNSCC that are positive or negative for transcriptionally active HPV16 (Chapter 4). The main question was if genetic differences between HPV-positive and HPV-negative HNSCC could be detected, which would be indicative for two different routes of carcinogenesis. Furthermore, this study also defined the relevant genetic regions that differ and are shared between these two groups of tumors. A gain of 18q was shown to be specific for HPV-positive HNSCC. HPV-negative tumors showed more specific genetic alterations such as loss of 9p21, amplification of 11q13.1 and TP53 mutations. These changes all relate to the abrogation of the p53 and pRb-pathways, and are consistent with and provide further evidence for the hypothesis that HNSCCs develop by two different etiologies [16-18]; one driven by exposure to environmental carcinogens (i.e. tobacco and alcohol) without HPV involvement, and one involving infection with transcriptionally active HPV16. On the other hand, it seems that the two carcinogenic routes also partly overlap. The alterations in common may be necessary events in HNSCC irrespective of the etiological factor, and a comprehensive genetic progression model of multi-step head and neck carcinogenesis was proposed.

In Chapter 5, we studied the hypothesis if additional subgroups within HPV-negative oral and oropharyngeal squamous cell carcinoma (OOSCC) could be distinguished on basis of the pattern of chromosomal aberrations, which might be relevant for further stratification. Thirty-nine OOSCCs were classified on basis of their genetic pattern determined by array CGH using a recently developed cluster algorithm. The tumors clustered in three groups, one (n=8) characterized by a low chromosomal instability index (low CIN), another by a relatively high chromosomal instability index (n=26), and one with a very high chromosomal instability index (n=5). This classification was significantly (p=0.003) associated with survival, with the best survival in the low CIN group and the worst survival in the group with very high CIN. Low CIN was also significantly (p<0.05) associated with the presence and type of TP53 mutation, absence of alcohol consumption and a female gender. In addition, we excluded that these tumors showed microsatellite instability, and they were shown to be diploid. We confirmed this classification of OOSCC on basis of low CIN
and high CIN with an independent set of 89 oral carcinomas [19]. To elucidate the biological basis of the low CIN genotype, these tumors were tested for microsatellite instability, but this could not be detected. The discovery of these new classes of OOSCC with unique genetic and clinical characteristics might have important consequences for future biological and clinical studies. The most important conclusion is, taking also HPV-positive HNSCC in consideration, that at least three and possibly four genetically distinct subgroups of HNSCC exist which might each have their own different route of carcinogenesis. It is too early to speculate on the role of this classification for clinical management. There is a consensus in the field that at least HPV-positive HNSCC should be considered separately in clinical trials, but even for this specific group adjustment of clinical management is not yet implemented. An important problem at present is that there is no consensus on the most reliable assay to assess HPV, which has consequences for the claims on the more favorable prognosis of HPV-positive tumors. Our algorithm might help to assess the involvement of HPV in large retrospective series.

Currently, the role of HPV in a subset of HNSCC is under intensive research. In contrast to the well established causal role of transforming hrHPV infections in carcinomas of the uterine cervix [20], less is known about the precise role of hrHPV infections in HNSCC. We hypothesized that the genomic patterns of HPV-positive HNSCC might show similarity to that of cervical cancers, irrespective the difference in anatomical site. In Chapter 6 genomic profiles of HNSCC, positive and negative for transcriptionally active HPV16 (described in Chapters 4 and 5) were compared to those of cervical carcinomas (all HPV-positive). This comparison showed that a number of alterations common to HPV-negative head and neck carcinomas were rarely detected in HPV-positive carcinomas either from the cervix or the head and neck, which included losses at 3p, 5q, and 8p as well as amplifications at 11q13.3 (CCND1 locus). On the other hand gains of chromosome 20 and losses at 13q seemed specific for HPV-induced tumors, irrespective the anatomical site. Frequently shared changes between all tumors encompassed 3q gains and 11q losses, suggesting critical cancer genes at these loci. Finally, a number of organ-specific alterations were found, including a gain at 8q in HNSCCs and losses at 17p in cervical carcinomas. The high frequencies in which these alterations are detected in these tumor types suggest that they are crucial in carcinogenesis.

HNSCC carcinogenesis is driven by a few genetic events that entail the impairment of the p53 and the pRb pathways, which are involved in cell cycle regulation and apoptosis [21]. Based on epidemiological data and in vitro transformation
experiments, it has been estimated that four to six genetic events are required in humans to transform a normal cell into a malignant cell [22;23]. The current high-throughput methods that allow genome wide analysis of genetic changes with high resolution such as array CGH described in this thesis, but also the increasing body of sequencing data, have tremendously accelerated the identification of candidate cancer genes and chromosomal regions harboring candidate cancer genes [24]. Of many of these genes, the role and importance in the pathogenesis is unclear. Tumors are intrinsically genetically unstable and many DNA changes or aberrantly expressed genes should be considered a consequence and not as a cause of carcinogenesis. It is therefore of utmost importance to discriminate the carcinogenic ‘driving’ from the ‘passenger’ events.

At present suitable transgenic mouse models are absent for HNSCC carcinogenesis, and functional characterization of candidate cancer genes in established HNSCC cell lines only allows analysis of phenotypes in the highly deregulated cellular environment of an established invasive carcinoma cell line. We therefore generated an *in vitro* model of conditionally immortalized primary oral squamous keratinocytes. Establishment of this model system and its application to investigate the phenotypes associated with p53 and pRb pathway abrogation in HNSCC are described in Chapter 7. The abrogation of the p53 pathway, either by knock-down of the p53 gene by overexpression of dominant-negative p53, or by the human papillomavirus oncoprotein E6, led to an extended life span. Despite the seemingly identical phenotypes, dominant-negative mutant p53 showed most severe molecular changes including upregulation of the hypoxia signaling (HIF1) and WNT pathways. Knock-down of p16, overexpression of cyclin D1 or HPV E7 alone did not seem to have effect on the life span. In combination with abrogation of the p53 pathway this led to an immortal phenotype. Our data suggest a critical order of genetic hits in squamous carcinogenesis with abrogation of pRb and p53 pathways as essential and likely primary hits. This *in vitro* model of conditionally immortalized primary oral squamous keratinocytes further allows the functional analysis of other candidate cancer genes, as well as high-throughput functional genetic screens.

**FUTURE PERSPECTIVES**

The Human Genome Project has provided an enormous amount of knowledge about the human DNA sequence. This allowed the global profiling of copy number imbalances in tumors and the precise determination of the breakpoints of regions that are gained and/or lost. During the last few years, micro-array and sequencing
technology rapidly became the main techniques for studying chromosomal aberrations. Array CGH has proven valuable for a better understanding of the biology of human cancer [25;26] and has the potential to improve diagnosis, outcome prediction and treatment response. Diagnostic genomic signatures, as are already present for gene expression [27], will soon be at hand [28] and will improve tailored therapy in the near future. Also for HNSCC we have shown by array CGH analysis the existence of subgroups of tumors with distinct genomic profiles. These subgroups with different genomic entities may also have specific underlying mechanisms of progression and patient prognosis. For instance, HPV-positive HNSCC have distinct genomic alterations similar to those found in cervical cancers, in which HPV has a confident causal role, and should therefore be investigated as a separate tumor group. Use of reliable detection methods as RT-PCR, FISH or the algorithm provided in this thesis, are necessary to select the true HPV-positive patients and to estimate the world-wide prevalence for HPV in HNSCC. The question remains if HPV-induced HNSCC can be prevented by vaccination. Currently prophylactic HPV vaccination is carried out for cervical carcinoma prevention and in theory could be implemented for the prevention of HPV-induced HNSCC as well. To date, there is only evidence that the vaccine prevents infection and it will take years before beneficial effects on the cervical carcinoma incidence will become clear. Nevertheless, no serious side effects have been reported so far, supporting the view to start vaccination studies in the near future in both girls and boys.

Within the group of HPV-negative HNSCC additional subgroups can be distinguished on basis of the genetic profile. This classification seems associated with p53 mutation status, the number of genomic alterations, etiological risk factors, and prognosis. Investigation of larger cohorts should substantiate this initial observation. Prognostic separation of patients in different subgroups based on different aetiology, genomic classification and other parameters clearly has major clinical implications. If unequivocally proven, these subgroups of patients need to be analyzed as separate groups in future studies for several reasons. First, the effect of targeted drugs could be very different and this needs to be considered in future clinical trials. Second, the current clinical management could be adapted either for the tumors that show a clear favorable prognosis (treatment less intense) or for the tumors that show with current treatment regimen a more unfavorable prognosis (more intense: shift to primary surgical treatment with adjuvant chemo- or bioradiation). Thirdly, novel therapies could be developed directed to the cell signaling pathways involved in the different subgroups of tumors (e.g. directed to HPV as causative factor).
Our array CGH studies were of value for tumor classification, but did not improve our understanding of the underlying mechanisms driving cancer because of its descriptive nature. The genomic resolution of the array CGH platform we established and used had a limited resolution of 5 Mbs. The identified genomic aberrations were therefore too large to pinpoint candidate driver genes. Nevertheless, the differential analysis showed the genomic regions that we have to focus on in the future, mainly 3q26, 8q24 and 11q. At present, the resolution and accuracy of array CGH platforms are improving enormously and will soon enable us to detect genomic aberrations at the gene level. Furthermore, sophisticated data analysis tools facilitate the integration of array CGH with expression, single nucleotide polymorphism (SNP), methylation and proteomics data. This and other upcoming high resolution applications like massive parallel sequencing will lead to an excess of candidate cancer genes [24;29;30]. Functional characterization of all these genes and their relevance to HNSCC development necessitates in vitro models. The in vitro model we established will be of importance in understanding the underlying gene interactions and pathways. Large high-throughput genetic screens using cDNA, short hairpin RNA or microRNA libraries are at hand to identify novel candidate cancer genes. The elucidation of the driving cancer genes and signaling pathways causing head and neck cancer, taking the HNSCC subgroups into account, will be a major step to the identification of novel drug targets and to the improvement of the clinical outcome of this disease in the future.
REFERENCES


Samenvatting en Toekomstperspectief
SAMENVATTING EN TOEKOMSTPERSPECTIEF

De laatste 20 jaar is er veel vooruitgang geboekt in de behandeling van plaveiselcelcarcinomen in het hoofd-halsgebied (HHPCC). Desalniettemin is de kans op overleving op de lange termijn slechts beperkt verbeterd. Grote verbeteringen kunnen alleen worden bereikt als de onderliggende biologie van het ontstaan van deze tumoren wordt opgehelderd. Hiervoor moeten de verantwoordelijke kankergenen en signaalpaden waarin ze actief zijn worden geïdentificeerd. De belangrijkste risicofactoren voor het ontwikkelen van HHPCC zijn het roken van tabak en overmatig alcoholgebruik. Recentelijk is ook beschreven dat het humaan papillomavirus (HPV) HHPCC kan veroorzaken. Het exacte percentage van tumoren waarbij HPV een rol speelt is onduidelijk. Dit komt doordat de gerapporteerde prevalenties van HPV-positieve HHPCCs sterk uiteenlopen. Deze variatie kan gedeeltelijk verklaard worden doordat verschillende tumorlocaties onderzocht zijn; het virus wordt namelijk vaker gevonden in tumoren van de keelholte dan in die van de mondholte. Een andere belangrijke oorzaak van de variatie in prevalentie is het soort monster dat wordt geanalyseerd (ingevroren of formaline-gefixeerd paraffine-ingebed weefselmateriaal) en de methode waarmee HPV wordt gedetecteerd. Dit laatste wordt onderzocht in hoofdstuk 2. Hier worden de meest gebruikte HPV detectie methoden met elkaar vergeleken en zijn de verkregen resultaten gecombineerd om de meest gevoelige en specifieke HPV bepaling te verkrijgen die eenvoudig werkt op formaline-gefixeerd paraffine-ingebed weefselmateriaal. Dit laatste is van belang omdat weefsel wereldwijd zo opgeslagen wordt en zal worden. Wij hebben in deze studie kwantitatieve “reverse transcriptase-polymerase chain reaction” (RT-PCR), fluorescentie in situ hybridisatie (FISH), General Primer PCR (GP-PCR), p16-immunokleuring en detectie van HPV specifieke eiwitten in bloed met elkaar vergeleken. De aanwezigheid van viraal mRNA (expressie van de oncogenen E6/E7 bepaald met RT-PCR) in vers ingevroren materiaal werd als afspiegeling van de aanwezigheid van actief HPV beschouwd, de gouden standaard voor deze vergelijking. Geen enkele test bereikte een specificiteit en gevoeligheid van 100%. Dit kon alleen worden bereikt door een combinatie van twee standaard diagnostische technieken; alle HHPCCs worden dan eerst getest op p16 overexpressie, gevolgd door een GP5+/6+ DNA-PCR op de p16-positieve tumoren. Kanker ontstaat door een verzameling van genetische veranderingen in het DNA van de cel. De genetische veranderingen beïnvloeden de activiteit van genen die betrokken zijn bij onder andere celgroei, invasie, uitzaaiing en celdood. Genen die tumorgroei bevorderen zijn oncogenen en genen die een remmende werking hebben op de
tumorgroeit noemt men tumorsuppressorgenen. Dit proefschrift richt zich op het identifieren van de genetische veranderingen in HHPCC in relatie tot klinische en moleculaire variabelen. Voor dit doel werd array CGH met een genomische resolutie van \(~1\) Megabase (Mb) opgezet. Met array CGH is het mogelijk om op een snelle wijze de numerieke genetische veranderingen over het gehele genoom van een tumor in kaart te brengen. Met behulp van array CGH wordt in een korte tijd een enorme hoeveelheid data geproduceerd die goede bioinformatica noodzakelijk maakt voor de analyses. In hoofdstuk 3 wordt een analyse-methode beschreven waarmee kan worden berekend welke genetische veranderingen tussen twee tumorgroepen significant verschillend zijn. Deze analyse is specifiek voor geordende data, zoals die door array CGH worden gegenereerd. Bovendien is een statistische correctie voor meervoudige testen, zoals beschreven door Benjamini en Yekutieli, in het programma verwerkt. Deze methode is vervolgens toegepast op array CGH data van een groep HHPCCs die wel of geen HPV16 bevatten (hoofdstuk 4). HPV-negatieve tumoren bleken zeer specifieke chromosomale veranderingen te hebben, zoals het verlies van 9p21, amplificatie van 11q13.1 en de aanwezigheid van \(TP53\) mutaties. Deze DNA veranderingen zijn geassocieerd met een blokkering van de p53 en pRb signaalpaden en zijn in overeenstemming met onze hypothese dat HHPCC zich ontwikkelt langs twee verschillende routes; één gedreven door de blootstelling aan kankerverwekkende stoffen (bijvoorbeeld tabak en alcohol) zonder HPV, en één gedreven door infectie met HPV16. Tevens ontdekten we dat de twee routes naar kanker genetisch ook gedeeltelijk overlappen. Deze gemeenschappelijke genetische veranderingen kunnen genen bevatten die betrokken zijn bij het ontstaan van HHPCC, ongeacht de etiologische factor. Deze resultaten zijn tevens schematisch weergegeven in een genetisch progressie model voor ‘hoofd-hals carcinogenese’. In hoofdstuk 5 hebben we de hypothese onderzocht of er nog extra subgroepen binnen de HPV-negatieve carcinomen kunnen worden onderscheiden op basis van chromosomale afwijkingen. Van negenendertig HHPCCs werden de chromosomale veranderingen in kaart gebracht met behulp van array CGH en vervolgens in groepen ingedeeld met behulp van een recentelijk ontwikkeld cluster algoritme. Er konden drie groepen tumoren onderscheiden worden: (I) één gekarakteriseerd door zeer weinig chromosomale veranderingen (n = 8), (II) een ander door een relatieve hoge hoeveelheid chromosomale veranderingen (N = 26) en (III) een met een zeer hoge hoeveelheid chromosomale veranderingen (N = 5). Deze indeling correleerde significant met een aantal klinische en moleculaire variabelen. De patiënten met tumoren met zeer weinig chromosomale veranderingen hadden de beste
overlevingskans (p = 0,003). Verder bleek dat afwezigheid van chromosomale veranderingen in het tumorgenoom gerelateerd is aan het vrouwelijke geslacht, het ontbreken van een *TP53* mutatie en het ontbreken van alcoholgebruik in de voorgeschiedenis. Tevens werd uitgesloten dat microsatelliet instabiliteit een factor van belang is in deze tumoren en ze waren ook allemaal diploid. Het bestaan van groepen HHPCC met zulke grote verschillen in aantal chromosomale veranderingen hebben we bevestigd met behulp van classificatie analyse van een eerder beschreven onafhankelijke set van 89 orale carcinomen. De ontdekking van deze nieuwe indeling van HHPCC met unieke genetische en klinische kenmerken kan belangrijke gevolgen hebben voor toekomstige biologische en klinische studies. De belangrijkste conclusie is, met inachtneming van de groep van HPV-positieve HHPCC, dat minstens drie genetisch verschillende subgroepen van HHPCC bestaan met mogelijk hun eigen specifieke route naar kanker. Het is nog te vroeg om te speculeren over het precieze belang van deze classificatie voor het klinisch handelen. Er is in ieder geval consensus over het feit dat HPV-positieve HHPCC als verschillend moeten worden beschouwd en voor deze specifieke groep is de discussie omtrent een specifieke behandeling in volle gang. Er is in ieder geval consensus over de meest betrouwbare HPV-detectie methode. Hierna kan namelijk pas de juiste prevalentie en de veronderstelde gunstigere prognose van patiënten met HPV positieve tumoren met zekerheid bepaald worden. Onderzoek op grote retrospectieve series HHPCC met behulp van ons detectie algoritme zou kunnen helpen om de rol van HPV en de klinische gevolgen hiervan nauwkeurig te bepalen.

In tegenstelling tot tumoren van de cervix uteri (baarmoederhals) waar de causale rol van hrHPV duidelijk is vastgesteld, is de precieze rol van HPV in HHPCC nog onduidelijk. Indien de werking van het virus in HHPCC vergelijkbaar is aan die in baarmoederhals tumoren, dan zouden de chromosomale veranderingen van HPV positieve HHPCC veel gelijkenis met die van baarmoederhals tumoren moeten vertonen. Om deze hypothese te toetsen werd in hoofdstuk 6 de array CGH data van HPV-positieve en -negatieve HHPCC (beschreven in hoofdstuk 4 en 5) vergeleken met die van baarmoederhalscarcinomen (allen HPV positief). Uit deze vergelijking werd bevestigd dat een aantal chromosomale veranderingen gevonden in HPV-negatieve HHPCC zelden werden aangetroffen in HPV-positieve carcinomen (baarmoederhalsPCC en HHPCC), waaronder verlies van de chromosomale regio’s 3p, 5q, en 8p evenals toename van 11q13.3 (CCND1 locus). Verlies van chromosoom 20 en 13q bleek specifiek voor HPV-positieve tumoren ongeacht de anatomische herkomst. Tenslotte is binnen de groep van HPV-positieve carcinomen,
ook een aantal weefsel-specifieke chromosomale veranderingen gevonden, zoals vermeerdering van 8q in HHPCCs en verlies van 17p in baarmoederhalstumoren. De hoge frequentie waarin deze HPV-specifieke chromosomale veranderingen voor komen, suggereren dat zij belangrijk zijn naast de kanker initiërende rol van HPV zelf. Het identificeren van deze chromosomale veranderingen zal nuttig blijken om de verantwoordelijke genen op te sporen.

HHPCC carcinogenese wordt gedreven door enkele genetische veranderingen die leiden tot blokkade van de p53 en de pRb signaalpaden. Deze signaalpaden zijn betrokken bij de regulatie van o.a. de celcyclus en geprogrammeerde cel dood (apoptose). Op basis van epidemiologische gegevens en onderzoek naar celtransformatie in in vitro modelsystemen is bepaald dat vier tot zes genetische gebeurtenissen nodig zijn om een normale humane cel in een kwaadaardige kankercel te veranderen. De huidige snelle methoden, waarmee zeer precies minimale chromosomale veranderingen en mutaties in kaart worden gebracht, hebben de identificatie van kandidaat kankergenen enorm versneld. Voor veel van deze genen is de rol en betekenis in het ontstaan van kanker nog onduidelijk.

Tumoren zijn genetisch instabiel en vele DNA veranderingen moeten worden beschouwd als een gevolg van het carcinogenese proces en niet als een oorzaak. Het is daarom van het allergrootste belang om de 'bestuurder' van de 'passagier' genetische veranderingen functioneel te onderscheiden. Op dit moment zijn er geen geschikte transgene muismodellen beschikbaar voor HHPCC om kandidaat kankergenen te onderzoeken. De functionele karakterisering van kandidaat genen wordt daarom vooral gedaan in HHPCC celllijnen die alleen een vergevorderd stadium van kanker nabootten. Om die reden hebben wij een celkweekmodel van conditioneel onsterfelijke primaire orale plaveiselcellen ontwikkeld. De ontwikkeling van dit model en het onderzoek naar het effect van p53 en pRb signaalpad blokkade in HHPCC is beschreven in hoofdstuk 7. Blokkade van de p53-pathway, hetzij door; (I) specifiek short hairpin RNA (shRNA) van het p53-gen, (II) overexpressie van een dominant-negatieve p53 mutant, of (III) overexpressie van het HPV oncoprotein E6, leidde alle tot een langere levensduur van de cellen. Ondanks de ogenschijnlijk identieke effecten op de levensduur van de cellen, toonde cellen met de dominant-negatieve p53 mutant grotere moleculaire veranderingen, waaronder activatie van de hypoxia (HIF1) en WNT signaalpaden. Blokkade van het pRb signaalpad door vermindering van p16 expressie, vermeerdering van cyclin D1 of HPV E7 expressie bleken allen geen effect te hebben op de levensduur van de cellen. Echter, in combinatie met blokkade van het p53-pad leidde dit tot immortalisatie, een van de
belangrijke kanker-geassocieerde kenmerken. Onze resultaten suggereren eveneens dat de volgorde van genetische veranderingen in de HHPCC carcinogenese kritisch is, met blokkade van de pRB en p53 pathways als essentiële en waarschijnlijk eerste veranderingen. Dit celkweek model van conditioneel geïmmortaliseerde primaire orale plaveiselcellen vereenvoudigt de functionele analyse van andere kandidaatgenen, en maakt ook het uitvoeren van grootschalige functionele genetische screens mogelijk.

TOEKOMSTPERSPECTIEF
Het Human Genome Project heeft enorm veel kennis over het menselijk DNA opgeleverd. Hierdoor werd het mogelijk de basenvolgorde van kankergenomen te bepalen en methoden te ontwikkelen voor grootschalige analyse van chromosomale veranderingen en genexpressie. Micro-array technologie en grootschalige basenvolgorde bepaling zijn daar nu de belangrijkste technieken voor. Deze methodieken hebben hun toegang tot de kliniek al gevonden, enerzijds in het beter voorspellen van de prognose, en anderzijds om gerichter te kiezen voor een behandeling waar de tumor gevoelig voor is. Verder heeft deze methodologie er toe geleid dat verschillende subgroepen van tumoren kunnen worden onderscheiden op basis van moleculaire verschillen, hetgeen belangrijke consequenties heeft voor het onderzoek naar de oorzakelijke kankergenomen. Ook voor HHPCC hebben we door middel van array CGH analyse laten zien dat er subgroepen van tumoren bestaan. De identificatie van deze subgroepen en de gevonden associaties met specifieke klinische karakteristieken zijn een belangrijke waarneming in dit proefschrift. HHPCC die geïnfecteerd zijn met HPV hebben specifieke chromosomale veranderingen en specifieke genexpressie profielen. Dit inzicht in combinatie met de literatuurgegevens dat deze tumoren ook klinisch een gunstiger beloop lijken te hebben, maken duidelijk dat deze tumoren als een afzonderlijke groep dienen te worden beschouwd. Gebruik van betrouwbare detectiemethoden zoals RT-PCR, of het detectie algoritme uit dit proefschrift, zijn nodig om de HPV-positieve patiënten te identificeren. De vraag is of HPV-geïnduceerde HHPCC voorkomen zou kunnen worden door vaccinatie. Preventie door profylactische HPV vaccins wordt al toegepast om baarmoederhalskanker te voorkomen en zou in theorie ook toegepast kunnen worden om HPV-positieve HHPCC te voorkomen in de gehele populatie. De introductie van grootschalige HPV vaccinatie bij meisjes in Nederland verliep echter bepaald niet zonder slag of stoot en de vraag is of op dit moment het draagvlak bestaat om ook jongens te gaan immuniseren. Het zal ook nog jaren gaan duren...
voordat een effect van vaccinatie op de baarmoederhalskanker incidentie kan worden aangetoond. Tot op heden is alleen bewezen dat infectie wordt voorkomen. Omdat er op dit moment geen bijwerkingen van het vaccin worden gerapporteerd lijkt het toch in de rede te liggen om niet langer te wachten.

Ook binnen de HPV-negatieve HHPCC blijken er nog additionele subgroepen met specifieke patronen van chromosomale veranderingen te bestaan. Verder onderzoek met grotere aantallen tumoren is echter nodig om deze eerste resultaten te bevestigen. Het onderscheiden van de patiënten in verschillende subgroepen naar etiologische factoren, specifieke chromosomale veranderingen en de relatie die deze hebben met de prognose heeft een belangrijke klinische implicatie. Ten eerste kan het effect van een bepaald geneesmiddel verschillen per subgroep. Ten tweede zouden de tumoren met een duidelijk gunstige prognose minder intensief behandeld kunnen worden in verhouding tot tumoren met een ongunstige prognose. Een op het eerste gezicht logische consequentie die echter niet zo snel ingang zal vinden om mogelijke onderbehandeling te voorkomen. Eerder zal men geneigd zijn zwaardere behandelingen zoals postoperatieve chemoradiatie te beperken tot de patiëntengroep met de meest ongunstige prognose. Ten derde kunnen nieuwe therapieën worden ontwikkeld gericht op de afwijkende signaalpaden die specifiek zijn voor een bepaalde subgroep. Verder zullen in klinisch vervolgonderzoek deze variabelen moeten worden bepaald omdat ze net als de TNM stadiëring van invloed kunnen zijn op het beloop en dus moeten worden meegewogen in de analyse.

In het onderzoek beschreven in dit proefschrift zijn nog geen nieuwe kandidaat kankergenen geïdentificeerd. Een van de oorzaken is de beperkte resolutie van het toegepaste array CGH platform, dat slechts grote chromosomale afwijkingen detecteert en dus niet zomaar tot de ontdekking van specifieke kandidaat kankergenen leidt. Niettemin geven de differentiële analyses wel aan op welke chromosomale gebieden we ons voorlopig moeten concentreren, in het bijzonder op 3q26, 8q24 en 11q. Momenteel is de resolutie en de nauwkeurigheid van de array CGH enorm verbeterd en is het detecteren van zeer kleine chromosomale veranderingen mogelijk. Bovendien maakt geavanceerdere data analyse de integratie van array CGH data met expressie profielen, single nucleotide polymorfismen (SNPs), methylatie patronen en proteomics data mogelijk. Deze en andere aankomende toepassingen met hoge resolutie, zoals massale parallelle sequencing (MPS), zullen leiden tot de identificatie van kandidaat kankergenen. Functionele karakterisering van al deze genen en hun relevantie voor HHPCC vereist goede functionele modellen. Het door ons opgezette celkweekmodel van
conditioneel geïmmortaliseerde plaveiselcellen uit de mondholte is hiervoor zeer geschikt. Grote functionele genetische screens met behulp van cDNA, shRNA of microRNA, zijn hiermee eveneens mogelijk om nieuwe kandidaat genen te identificeren. De ontregeling van de signaalpaden in een tumor zijn tegelijkertijd ook hun achilleshiel, en we zullen derhalve moeten proberen de relevante signaalpaden voor hoofdhalskanker te ontrafelen. Daarmee zullen we beschikken over de meest betrouwbare biomarkers voor diagnostische toepassingen en een ingang voor nieuwe behandelingen.
CURRICULUM VITAE

The author of this thesis was born on the 8th of July 1971 in Amsterdam. In 1994 he received his bachelor degree in ‘Technical Microbiology’ at the Fontys University of Applied Sciences in Venlo (NL). The research project for his graduation was carried out at the Dept. of Pathology at the VU University Medical Center (VUmc) Amsterdam. He continued this project instead of military service. From 1995 on, he worked as research technician on the project ‘Determinants of retinoid sensitivity in the prevention and treatment of head and neck squamous cell carcinomas’ at the Tumor Biology section of the Dept. of Otolaryngology/Head-Neck Surgery at the VUmc. In 1999, he became Senior Research technician on the research subject ‘The human Ly-6 antigens: suitable targets for diagnosis and (immuno)therapy of squamous cell carcinoma’ and managed the Integral Molecular Biology Lab. (IMBL) that housed four research groups. In 2003 he started as PhD-student at the same department on the research-project ‘Genetic and functional analysis of head and neck carcinogenesis’, which resulted in this thesis. As a PhD-student, he presented his results at several international conferences (the Cancer Congress, Dusseldorf; American Association for Cancer Research (AACR), Anaheim and San Diego; the International Academy of Oral Oncology (IAOO), Amsterdam; MC-GARD, Amsterdam). In July 2007 he started at a Post Doctoral position at the Micro Array Core Facility within the Cancer Center Amsterdam at VUmc and is currently involved in various (inter-)national genome projects.
PUBLICATIONS:


*authors contributed equally.
Dankwoord
DANKWOORD

Pfff, ik heb heel wat ‘opbeurende’ spreekwoorden langs horen komen en aan ‘De laatste loodjes wegen het zwaarst’ hecht ik nu veruit de meeste waarde. Maar nu is het af en ben ik hartstikke trots op alles wat hierin staat. En daarvoor wil ik vele mensen bedanken, want anders was dit nooit gelukt. Helaas kan ik onmogelijk iedereen aanspreken, maar ik wil graag toch de belangrijkste spelers noemen.

Laat ik beginnen met mijn beide promotoren. Prof. Dr. R.H. Brakenhoff. Bestie Ruud, hierbij wil ik je allereerst bedanken voor het geloof in mijn kunnen. Je gaf me op het Integraal Moleculair Biologisch Laboratorium (IMBL) alle vrijheid om nieuwe technieken op te zetten en vervolgens een belangrijk deel van de organisatie op me te nemen. Later als promovendus veranderde dit in het opzetten van een model systeem in samenwerking met Dr. Katrien Berns, Prof. Dr. René Bernards en Dr. Reuven Agami op het NKI, waar ik hen zeer erkentelijk voor ben. Ook het uitvoeren van kostbare micro arrays werd onderdeel van mijn takenpakket. Iets waar ik tot de dag van vandaag nog altijd de vruchten van pluk. Je bevlogenheid en kritische blik werden me wel eens te veel, maar je enorme enthousiasme en ‘drive’ weet je als geen ander over te brengen. Je hebt je ooit in het zonnetje mogen zetten met een sketch op jouw oratiefeest. Ik hoop dat jij mij ook zal sparen op mijn feestje… Prof. Dr. C.R. Leemans, beste René, ook jij bedankt voor je interesses in dit onderzoek en de mogelijkheden die mij geboden zijn.

Dan natuurlijk mijn beide co-promotoren. Dr. B.J.M. Braakhuis, beste Boudewijn, ik leerde je pas echt kennen toen ik als promovendus ‘uitverkorene’ werd om met jou de kamer te gaan delen. Sinds die tijd ging mijn “know-how” met sprongen vooruit. Jouw enorme kennis van de literatuur, jouw relativering van wetenschappelijke frustraties en functie als rustgevende buffer van Ruud heb ik altijd enorm gewaardeerd. Dr. B. Ylstra, beste Bauke. Eerst samen aan de BAC arrays en nu aan de ultra high resolution arrays! Lange tijd heb ik de afronding van dit onderzoek en proefschrift in de avond-uren moeten combineren met het mijn werk op de micro-array faciliteit. Iets wat regelmatig leidde tot begrijpelijke wederzijdse frustraties. Ik wil je danken voor de kritische blik en waardevolle discussies in de laatste fase van dit onderzoek en hoop op een goede voortzetting in de toekomst.
De leden van de promotie commissie, Prof. Dr. Frank Speleman, Prof. Dr. René Bernards, Dr. Katrien Berns, Prof. Dr. Rob J Baatenburg de Jong, Prof. Dr. Henk MW Verheul, Prof. Dr. Elisabeth Bloemena en Prof. Dr. Peter JF Snijders, wil ik danken voor de tijd en aandacht die ze aan mijn werk hebben besteed.

Al mijn collega AIO’s bij de KNO, met in het bijzonder mijn paranimfen Hester en Lars. Hester, bedankt voor al je meedenken en meeleven met alle keratinocyten in voor en tegenspoed. Jij begreep als geen ander dat kera’s alleen meewerken als je ze vertroetelt en met ze praat. Lars, ook al heb je een totaal andere onderzoeksrichting, meepriaten over micro arrays ging al snel goed door je lieve vriendin en zeer aanstaande vrouw Esther. We hebben ook buiten het lab een hoop lol gehad en ik hoop dat we binnenkort in rustiger tijden het sporten weer op kunnen pakken. Uiteraard zullen we met het hele ploegje blijven schaatsen, zodat we in de Skeeve skaets weer een ‘werkbespreking’ kunnen houden. We zouden die biertjes eigenlijk moeten declareren. Ook alle andere collega KNO-AIO’s: Pontus, Asaf, Jantien, Tienieke, Bernard, Peggy, Lisa en Maria. Bedankt voor alles en ik wens jullie allen heel veel succes met de afronding van jullie eigen promotie.

Ook dank aan al mijn hard werkende labmaatjes en toenmalige collega’s Arjen, Henrique, Thijs, Michiel, Marianne, met in het bijzonder Janny voor alle p16 kleuringen en Aafke en Marijke voor het afronden van mijn ‘never-ending’ kweekproeven en het behalen van de 100 celdelingen!

Al mijn studenten: Saman, Geert, Nadia, Johan en Eva. Bedankt voor jullie onnoemlijke inzet en het zetten van jullie tanden in het model-systeem. Ik heb met veel plezier met jullie gewerkt en hopelijk houden we contact. De meeste van jullie zijn in het onderzoek gebleven en ik hoop jullie later nog als groot onderzoeker te treffen.

Mijn collega’s bij de Pathologie, Renske, Peter, Saskia, Bart en Debbie. Jullie zijn de beste HPV-speurders ter wereld en gelukkig hebben jullie veel van die HPV “know-how” met mij willen delen. Ik ben blij dat onze samenwerking tot meerdere hoofdstukken heeft geleid. Hopelijk in de toekomst nog meer. Renske, ik betreur het nog steeds dat je bent verhuisd naar Aalsmeer. Ondanks dat we vaak stevig moesten doortrappen om op tijd bij de kinderopvang te zijn, hadden we toch nog adem genoeg voor goede gesprekken.

Dan mijn vroegere en huidige collega’s bij de Micro-array Core Faciliteit (MACF) en Tumor Profiling Group (TP). In den beginne met Bauke, Paul Eijk, Paul van den IJssel, Sandra, Beatriz, Cindy en Marianne. Bedankt voor jullie hulp met het opzetten van de 5K BAC array en de eerste array experimenten. Tegenwoordig zijn Danielle,

Ook nog mijn collegae bij de Hematologie, met in het bijzonder Pauline, Jacqueline en mijn goede vriend en vervolg BVF-er Fedor. Het was leuk om met jullie de labs te delen. Ik wil jullie danken voor alles en nog wat, zoals het doornemen van SOP’s, het ontwerpen en inrichten van de CCA laboratoria en nog veel meer.

Vrienden en kennissen vragen zich wel eens af. Leeft die jongen nog? Al was ik sowieso nooit goed in het onderhouden van contacten, nu heb ik me de laatste jaren toch echt verstopt. Aan iedereen die ik de laatste jaren verwaarloosd heb, bij dezen mijn excuses. Ik hoop het nu weer helemaal goed te gaan maken.


Iedereen weet dat kinderen krijgen en een promotieonderzoek doen een onmogelijke combinatie is. Of je promotie legt het loodje of je relatie krijgt een dreun. Lieve Nicolle, wat ben ik blij met jou en wat heb ik me vaak schuldig gevoeld, omdat ik je weer moest teleurstellen. Of ik kwam laat thuis of ik zat tijdens de kerst mijn andere kinderen (de keratinocyten) te vertroetelen. Ik ben ongelofelijk blij dat je in mij bent blijven geloven, ondanks dat je altijd tegen een labtop moest praten of me verwenste als de kinderen midden in de nacht wakker werden van mijn geratel op het toetsenbord. Gelukkig, aan alles komt een eind, al duurt het soms iets langer dan gehoopt. Zonder jou had ik dit nooit kunnen doen en je bent de basis van ons gezinnetje. Lieve Ghislaine en Philine. Wat zijn jullie mooie meiden! Mijn trots is werkelijk onbeschrijfelijk. Ghislaine, je hebt vanaf jouw eerste dag een rol gespeeld in