การแสดงออกของยีน ในมะเร็งชนิดสความัสเซลล์คาร์ซิโนมา บริเวณศีรษะและลำคอ

นางชิดชนก ลีธนะกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีววิทยาช่องปาก
คณะทันตแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2543

ISBN 974-347-047-6

PATTERN OF GENES EXPRESSION IN SQUAMOUS CELL CARCINOMA OF HEAD AND NECK

Chidchanok Leethanakul

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Oral Biology

Faculty of Dentistry

Chulalongkorn University

Academic year 2000

ISBN 974-347-047-6

Thesis Title	PATTERN OF GENE EXPRESSION IN SQUAMOUS CELL CARCINOMA OF HEAD AND NECK
By	Chidchanok Leethanakul
Program	Oral Biology
Thesis Advisor	Asst. Prof. Dr. Visaka Limwongse
Thesis Co-advisor	Assoc. Prof. Dr. Sittichai Koontongkaew
	Dr. Silvio Gutkind
	Dr. Vyomesh Patel

Accepted by Faculty of Dentistry, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

…….Dean of Faculty of Dentistry
(Assoc. Prof. Surasith Kiatpongsan)

THESIS COMTTEE

…….Chairman
(Assoc. Prof. Dr. Ratana Sereenirach)

…….Thesis Advisor
(Asst. Prof. Dr. Visaka Limwongse)

…….Thesis Co-advisor
(Assoc. Prof. Dr. Sittichai Koontongkaew)

…….Thesis Co-advisor
(Dr. J. Silvio Gutkind)

…….Thesis Co-advisor
(Dr. Vyomesh Patel)

…….Member
(Assoc. Prof. Woranat Weerapradist)

…….Member
(Dr. Suconta Chareonvit)

ชิดชนก ลีธนะกุล: การแสดงออกของยีนชนิดสความัสเซลล์คาร์ซิโนมาบริเวณ-ศีรษะและลำคอ อ.ที่ปรึกษา : ผศ. ดร. วิสาขะ ลิ่มวงศ์ อ.ที่ปรึกษาร่วม : รศ. ดร. สิทธิชัย ขุนทองแก้ว, Dr. J. Silvio Gutkind, Dr. Vyomesh Patel, 139 หน้า. ISBN 974-347-047-6

มะเร็งชนิดสความัสเซลล์บริเวณศีรษะและลำคอ พบได้บ่อยเป็นลำดับที่ 6 ในกลุ่มประเทศพัฒนาแล้ว โดยส่วนใหญ่บริเวณช่องปาก ริมฝีปาก กล่องเสียง และคอหอย จุดประสงค์ของการศึกษานี้ เพื่อที่จะอธิบายถึงการเปลี่ยนแปลงทางด้านพันธุกรรม ซึ่งมีผลต่อการเกิดมะเร็งชนิดนี้ และเพื่อใช้ความรู้นี้ในการค้นหา molecular marker ของการเป็นมะเร็ง ความคาดหวังจากการศึกษานี้ คือ นำผลที่ได้มาศึกษาต่อ เพื่อเอื้ออำนวยในการวินิจฉัยระยะเริ่มแรกของการเกิดมะเร็งชนิดนี้ และยังอาจจะเป็นประโยชน์ต่อการรักษาโรคนี้โดยการใช้ยา เพื่อการฆ่าทำลายเนื้อเยื่อมะเร็ง

ความเป็นไปได้ในการที่จะศึกษาค้นคว้ารูปแบบการแสดงออกของยีน ในแต่ละระยะของโรคมะเร็งบริเวณศีรษะและคอมักจะพบข้อจำกัดจากการที่เซลล์มะเร็งมีการปนเปื้อนกับเนื้อเยื่อปกติ แต่ในปัจจุบัน พัฒนาการของเทคโนโลยีสมัยใหม่ เช่น กล้องจุลทรรศน์ชนิดเลเซอร์ (Laser Capture Microdissection, LCM) ซึ่งมีคุณสมบัติในการแยกเซลล์เฉพาะออกจากเนื้อเยื่อปนเปื้อนอื่น ๆ และนำเซลล์เฉพาะกลุ่มนั้นมาศึกษาการแสดงออกในระดับยีนหรือโปรตีนได้ ดังนั้นการศึกษานี้จึงได้นำกล้องจุลทรรศน์ชนิดเลเซอร์นี้มาทำการตัดแยกกลุ่มเซลล์สความัสเซลล์คาร์ซิโนมา ทั้งที่เป็นเซลล์มะเร็งและเซลล์ปกติจากเนื้อเยื่อที่ได้รับจากผู้ป่วยด้วยโรคมะเร็งชนิดสความัสเซลล์คาร์ซิโนมาบริเวณศีรษะและลำคอ โดยพบว่าการใช้เลเซอร์ในการผ่าตัดแยกเซลล์ประมาณ 5,000 เซลล์นั้น เพียงพอในการสกัด RNA (14.7 -18.6 นาโนกรัม) เพื่อใช้ในการสังเคราะห์ cDNA probes และนำไป hybridize กับยีนที่เกี่ยวกับมะเร็งในคน ซึ่งเคลือบอยู่บนเยื่อแผ่น (membrane) จากนั้นนำผลที่ได้ไปเปรียบเทียบกับกลุ่มเซลล์ปกติ ซึ่งการศึกษานี้ พบว่ากลุ่มเซลล์มะเร็งมีการแสดงออกของยีนบางกลุ่ม เช่น cytokeratins ลดลง และมีการเพิ่มขึ้นของการแสดงออกของยีนในกลุ่มของ signal transducting และ cell cycle regulatory molecular growth และ angiogenic factors และ matrix degrading proteases นอกจากนั้นยังพบการแสดงออกของยีนในกลุ่ม Wnt และ notch ในกลุ่มเซลล์มะเร็ง ซึ่งอาจจะมีผลต่อการเกิดมะเร็งชนิดสความัสเซลล์คาร์ซิโนมา การทดลองนี้เป็นการนำกล้องจุลทรรศน์ชนิดเลเซอร์มาใช้ในการเริ่มต้นดำเนินการค้นหายีนโดยการสร้าง cDNA Library จากกลุ่มเซลล์มะเร็งศีรษะและลำคอและเซลล์ปกติ ที่ได้รับจากเนื้อเยื่อที่ได้รับจากผู้ป่วยคนเดียวกัน RNA ที่สกัดได้จะถูกสังเคราะห์เป็น Blunt-ended, double strand cDNA โดยใช้ Oligo (dT) Reverse transcriptions และเพิ่มตัวเชื่อม (Linker) เพื่อทำให้ง่ายต่อการ Transform DNA เข้าสู่แบคทีเรียชนิด pAMP 10 แล้วทำการประเมินคุณภาพของ cDNA Library โดยการตรวจลำดับเบส (Sequencing) clones จำนวน 96 clones ที่ได้มาจากการสุ่มตัวอย่าง จากทั้ง 6 cDNA Library โดยพบว่า 76-96% ของ Sequencing ที่ได้นั้นเป็น Anonymous ESTs (25-48%), ยีนที่ทราบแล้ว (9-29%) และ ยีนใหม่ (27-51%) โดยมีการซ้ำกันของยีนและ Ribosomal RNA น้อยมาก ซึ่งเป็นตัวบ่งชี้ว่า Library เหล่านี้ มีคุณภาพดี และมีการค้นพบยีนตัวใหม่จำนวนมาก ประโยชน์ที่ได้จากการศึกษานี้คือการประยุกต์ความรู้ไปค้นคว้าระบบกลไกการทำให้เกิดมะเร็งชนิดสความัสเซลล์คาร์ซิโนมาบริเวณศีรษะและลำคอ ดังนั้น cDNA library ทั้ง 6 Library นี้จึงถูกนำมาวิเคราะห์โดยการใช้ Bio-Informatic Tools และพบว่า cDNA Library 6 ชนิด คือ HN7 (normal) และ HN8 (well differentiated invaive carcinomas), HN9 (normal) และ HN10 (carcinoma *in situ*) HN11 (normal) และ HN12 (moderate to poorly differentiated invasive carcinomas) ประกอบไปด้วยยีนใหม่ที่ไม่เคยรายงาน จำนวน 138 ชนิด ผลที่ได้จากการวิเคราะห์ครั้งนี้ เป็นจุดเริ่มต้นในการที่จะศึกษาถึงรูปแบบเฉพาะของการแสดงออกของยีนในมะเร็งชนิดนี้ในระยะต่าง ๆ ได้ และได้มีการวิเคราะห์เพื่อใช้ในการศึกษาต่อไปโดยการใช้เทคโนโลยีของกล้องจุลทรรศน์ชนิดเลเซอร์ และ high throughput gene arrays ซึ่งมีผลต่อข้อมูลที่มีคุณภาพ และเกี่ยวข้องกับการที่จะทำให้เกิด มะเร็งชนิดสความัสเซลล์คาร์ซิโนมาบริเวณศีรษะและลำคอ นอกจากนี้ ผลจากการสร้าง cDNA Library นี้ ทำให้มีการค้นพบยีนใหม่อย่างน้อย 189 ชนิด ซึ่งอาจมีผลต่อพยาธิสภาพของมะเร็งชนิดสความัสเซลล์คาร์ซิโนมาบริเวณศีรษะและลำคอ และอาจจะเป็นโมเลกุลตัวใหม่ที่สำคัญ สำหรับการวินิจฉัยมะเร็งในระยะเริ่มแรก หรืออาจเกี่ยวข้องต่อการรักษามะเร็งโดยเคมีบำบัด โดยข้อมูลและ DNA clones ที่ได้ทั้งหมดจากการศึกษานี้ ได้มีการการเผยแพร่ต่อสาธารณชน เพื่อใช้ในการวิจัยต่อไป

สาขาวิชา : ชีววิทยาช่องปาก
ปีการศึกษา : 2543

ลายมือชื่อนิสิต.....
ลายมือชื่ออาจารย์ที่ปรึกษา.....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

397 04458 32; MAJOR: ORAL BIOLOGY
KEY WORD : HNSCC / GENE EXPRESSION / LCM / cDNA LIBRARY /TUMOR PROGRESSION / TUMOR MARKER / HUMAN TISSUES

CHIDCHANOK LEETHANAKUL: PATTERN OF GENE EXPRESSION IN SQUAMOUS CELL CARCINOMA OF HEAD AND NECK. THESIS ADVISOR: ASSIST PROF.VISAKA LIMWONGSE, D.D.S., Ph.D. THESIS COADVISOR: ASSOC PROF. SITTICHAI KOONTONGKAEW, D.D.S., Ph.D., Dr. J. SILVIO GUTKIND Ph.D., Dr. VYOMESH PATEL, Ph.D 139 pp. ISBN 974-347-047-6

Squamous cell carcinoma of the head and neck (HNSCC) is the 6th most common cancer in the developed world. The vast majority of these malignancies involve neoplastic lesions in the oral cavity, lip, larynx, and pharynx. The goal of this study is to help elucidate the genetic changes contributing to HNSCC, and to use this knowledge to develop molecular markers heralding malignancy. We expect these efforts will facilitate the early detection of oral cancer lesion, as well as the discovery of novel potential targets for pharmacoligical intervention in this devastating disease. The ability to investigate gene expression profiles at different stages of tumor progression in HNSCC is usually limited by the remarkable heterogeneity of these neoplastic lesions. New technological breakthroughs, such as the development of laser capture microdissection (LCM), have now provided a unique platform for gene and protein expression analysis in specific cell populations. We have recently developed LCM-based techniques to procure neoplastic and phenotypically normal cells from representative sets of HNSCCs and their matching normal tissues. Indeed, we found that the laser assisted microdissection of 5,000 cells was sufficient to extract total RNA (14.7-18.6 ng) of high integrity for the synthesis of labeled amplified cDNA probes which could then be hybridized to membranes arrayed with known human cancer-related cDNAs. By this approach, HNSCCs were compared to normal tissues, and found that cancer cells exhibit a consistent decrease in expression of differentiation markers such as cytokeratins, and an increase in the expression of a number of signal transducing and cell cycle regulatory molecules, growth and angiogenic factors, and matrix degrading proteases. Unexpectedly, most of the HNSCCs overexpress members of the *wnt* and *notch* growth and differentiation regulatory systems, thus suggesting that they may contribute to squamous cell carcinogenesis. Furthermore, we took advantage of the LCM technology for the launching of a gene discovery effort, which involved the generation of cDNA libraries from microdissected HNSCC tissues. HNSCC tissue sets comprised oral squamous cell carcinomas and matching normal tissues. Isolated RNAs were used for the synthesis of blunt-ended, double strand cDNAs by oligo (dT)-mediated reverse transcription, followed by addition of linkers. Primers specific for these linkers with UDG-compatible ends were used to amplify these cDNAs by PCR and the product was subcloned into the pAMP10 cloning vector. For our initial analysis, ninety-six clones from each of the 6 libraries were randomly sequenced. Results indicated that 76-96% of the inserts represented either anonymous ESTs (25-48%), known genes (9-29%) or novel sequences (27-51%), respectively, with very little redundancy. These findings indicated that high quality, representative cDNA libraries can be generated from microdissected tissues, and led to the identification of a number of novel HNSCC specific genes. In an effort to begin addressing the molecular basis of HNSCC, these 6 microdissection libraries were further analyzed using powerful bioinformatic tools. Using newly developed search engines, we found that these libraries [HN7 (normal) and HN8 (well differentiate invasive carcinoma); HN9 (normal) and HN10 (carcinoma *in situ*); HN11 (normal) and HN12 (moderate to poorly differentiated invasive carcinoma)] include 138 unknown unique genes, and a large number of unknown non-unique genes. From the available information on known genes, we have also begun to appreciate the unique pattern of gene expression in this tumor type. This was further analyzed using the LCM platform and high-throughput gene array technologies, which has provided a wealth of information on genes that are likely responsible for the establishment and growth of squamous carcinoma cells. Furthermore, these efforts led to the discovery of at least 189 novel genes, which may have a role in the pathogenesis of HNSCC, and thus may represent novel markers for early detection as well as targets for pharmacological intervention in this disease. Finally, all data and DNA clones have been deposited in the public domain, and are available for further investigation by our scientific community.

Field of study : Oral Biology

Academic year : 2000

Student's signature…….

Advisor's signature…….

Co-advisor's signature……..

ACKNOWLEDGEMENT

This dissertation has by no means been a solitary undertaking. I am deeply grateful for the financial support I received partly from the Ministry of the University Affairs of Thailand during the course of these studies and I would like to express my sincere gratitute for the guidance and support provided by my colleagues, family and friends.

I wish to express my deepest gratitude to Dr. Narong Suksu-art, Asst. Prof. Dr. Kanok Sorathesn, and Assoc Prof. Dr. Sittichai Koontongkaew who encouraged me to pursue my doctorate degree and have been a constant source of inspiration.

I am deeply grateful to my advisor, Asst. Prof. Dr. Visaka Limwongse, for her valuable advice and continuing support throughout my study in this program. I am particularly grateful to Assoc Prof. Dr. Sittichai Koonthongkaew, my co-advisor at PSU, whose support, enchoragement, sense of direction, and timely advise enabled me to finish this program. I am also in debt to Dr. J. Silvio Gutkind, my Branch Chief and co-advisor at Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health (OPCB/NIDCR/NIH), with whom I worked closely, for his excellent guidance, encouragement, and invaluable suggestions that helped me carry out my dissertation work at NIH. In addition to his help with my dissertation, he provided me with financial support and chances to grow scientifically. To me, he has been a model scientist and wonderful mentor. I also owe special thanks to Dr.Vyomesh Patel, my co-advisor at OPCB/NIDCR/NIH, for being a patient mentor who gave me so much of his precious time, energy, and advise which helped me learn not only about sciences but also American culture and language, and who has been taking care of me as a member of his extended family for the past three years in the US.

In addition to my advisors, I sincerely thank the rest of my dissertation committee: Assoc Prof. Dr. Ratana Sereenirach, Assoc Prof. Woranat Weerapradist, and Dr. Sukonta Chareonvit. for their insightful comments and personal efforts in reviewing my work.

A great many people encouraged me and stood behind me as I attempted to complete this dissertation. I am thankful to all co-workers at NIH including Drs. Vladimir Knezevic, David B. Krizman, Micheal R. Emmert-Buck, and John Gillespie; to all OPCB members; and to others, who were involved in one way or another, for their excellent help, hospitality, advise, comments, and willingness to cooperate with my work. Their contribution will always be a treasured memory. In particular, I wish to express my gratitude to Asst. Prof. Dr. Prasit Pavasant for the skillful advise he provided in solving all problems occurring along the duration of this study.

Special thanks go to all my good friends, the members of the Preventive Department at Prince of Songkla University and the members of the Orthodontic Department at Chulalongkorn University for their consistent and warmest support that became an invaluable source of strength and confidence.

Finally, I am blessed to have a wonderful family and a great and beloved husband, Weraphan Leethanakul, to whom I wish to whole-heartedly thank for their endless love and support that have always been there when they were needed.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Background and Significance

Cancer arises as a consequence of multiple molecular changes that are initiated by carcinogenic induced genetic damage in certain cells and, if undetected by DNA repair mechanisms, develop a selective growth advantage and form a tumor. In all recognized tissues the fine and controlled balance between cell proliferation and cell death is controlled by genes whose expression is often unregulated in cancer. Hence, the alteration of gene regulation mechanism occurs during cancer development. By using cellular and molecular biological techniques, previous studies have demonstrated that many cancers have a strange genetic basis. Oncogenes are defined as being able to induce or maintain cellular transformation, thus giving a cell a growth advantage (Gelehrter and Collin F.S., 1990; Miller and Dmitrovsky, 1991; Pillai, 1992) whereas proto-oncogenes have critical functions in growth or differentiation of normal cellular genes. The latter are usually strictly regulated by other genes that either promote or inhibit their transcription. When the regulation of proto-oncogenes escapes these controls, they become oncogenic, and the balance stimulation of the cell by their protein products is unimpeded (Fishleder, 1990). Recently, the knowledge on the normal and aberrant function of these genes provides unique opportunities to understand and, ultimately, to control the processes leading to human cancer. Mutation of two classes of genes, oncogenes and tumor suppressor genes, plays major roles in triggering cancer. These genes control cell growth by

integrating the information generated by extra-cellular stimuli and intracellular mediators. Oncogenes promote cell proliferation, while tumor suppressor genes inhibit cell growth.

Cancer is a complex cellular disease, the defining characteristic of which is the ability of cells to divide under abnormal conditions, resulting ultimately in their loss negative growth controlling mechanisms. This uncontrolled proliferation can be explained, in part by the gain or loss of protein function that constitute the cell cycle regulatory machanism (Clurman and Roberts, 1995). For instance, following DNA damage, loss of function of p53, a protein that acts as a transcription factor will not stop cell cycle progression in the G1 phase by induction of the *p21* (el-Deiry *et al*., 1993). Head and neck squamous cell carcinoma (HNSCC) is considered the 6^{th} most common cancer among men in the developed world**.** In spite of the recent advances in our understanding, prevention and treatment of other types of cancers, the five-year survival rate after diagnosis for HNSCC is still very low, approximately 53%, which is considerably lower than that for other neoplasias, such as those of colorectal, cervix and breast origin (Landis *et al*., 1999; Parkin *et al*., 1999)**.** This high morbidity rate can be attributed to many factors, which include failure to respond to available chemotherapy, late presentation of the lesions and the lack of suitable markers for early detection. The poor prognosis of HNSCC patients may be a reflection of the fact that while many of the risk factors involved in HNSCC pathogenesis, such as alcohol and tobacco, are well recognized (van Oijen *et al*., 1998; van Oijen *et al*., 1999), by contrast very little is known about the molecular mechanisms responsible for this type of cancer. The primary sites have been varied in frequency as follows: 40% oral cavity, 25% larynx, 15% pharynx and 20% salivary gland. Moreover 50% of the

limited survival rate has been related to cervical lymph node metastasis (van Oijen *et al*., 1998; van Oijen *et al*., 1999)

Many risk factors are associated with HNSCC such as the use of tobacco (cigarette smoking and betel quid chewing), alcohol and genetic variation. About 90% of HNSCC in Western countries occur in smokers and majority is moderate to heavy alcohol drinkers. In India and South East Asian countries, however, betel quid chewing may appear to represent the main cause of HNSCC (Greenblatt *et al*., 1994; Kaur *et al*., 1994). While in China, smoked tobacco and areca nut appeared to be the main causal factors. In support, a recent study suggested that areca nut and components of the pan/betel quid can be carcinogens (Gupta *et al*., 1992: Zheng, 1990 #186). In some western countries, such as, France and Denmark, 90% smoked tobacco and this effect appear to be synergistic with alcohol (Franceschi *et al*., 1990). Other potential causes of HNSCC are viral infection (Snijders *et al*., 1992) and occupational factor (Maier *et al*., 1991). In addition although, little is known about the role of hereditary factors, a genetic predisposition is an important consideration for the development of HNSCC. However, many individuals are exposed to tobacco and/or alcohol, only few develop HNSCC. In fact, this may be explained that genetically determined susceptibility to external carcinogens may be important in the etiology of HNSCC. In deed, this notion is supported by cases reported about the familial clustering of oral and laryngeal cancer, which involved the simultaneous occurrence of a similar type of squamous cell carcinoma in 2-3 members of a single family (Tashiro *et al*., 1986). The role of dental factors in HNSCC study remains controversial but however associations between tooth loss and increased risk for HNSCC development for example in China have been described (Zheng *et al*., 1990),

but these may reflect nutritional and socioeconomic differences. The risk for oropharyngeal cancer among industrial constructive workers is increased 2-3 folds after exposure to such items as dust or coal tar product. In addition recent studies suggested that air pollution might be relevant (Dietz *et al*., 1995). In spite of the limited evidence on carcinogenic food components with respect to HNSCC, two recent studies have shown significantly increased risks for heavy and long-term consumption of nitrite and nitrosamine containing foods, such as smoked or salted products. Similarly, very spicy food may also be an etiological factor (De Stefani *et al*., 1994).

Several Human papilloma virus (HPV), particularly viral proto-oncogene 16 and 18, can immortalize human keratinocytes cultured through the E5, E6, and E7 oncogenes. E6 and E7 viral proteins bind to p53 and pRb tumor suppressor proteins respectively and prevent their ability to negatively regulated cell growth (Sugerman *et al*., 1995). In HPV integration may be at chromosomal site in close proximity to oncogenes, such as *c-myc* and *ras*, thus resulting in a growth potential. The prevalence of high-oncogenic types of HPV in oral and other HNSCC is thus of great interest. However, recent data are conflicting, possibly reflecting technical variation. Interestingly, HPV have been described in non-malignant tissues of HNSCC (Sugerman *et al*., 1995). Concerning oral cancer development, it seems that HPV may be a cofactor, as all of the patients who developed oral cancer within 4-12 years were all positive for HPV, one being positive for HPV 16 (Nielsen *et al*., 1996). Epstein-Barr virus (EBV) is oncogenic for human B-lymphocytes and has long been thought to as a causative factor for nasopharyngeal carcinoma (Niedobitek and Young, 1994), a distinctively different disease from the smoking-alcohol related squamous cell

carcinoma of the oral cavity and pharynx. However, the role of EBV in HNSCC development still remains unclear.

Normal cellular functions, for instance those controlling excitatory and inhibitory pathways, include the division, differentiation, survival, and death of cell. These regulatory pathways are composed of extra-cellular ligands binding to cell surface receptors, to generate an intracellular signaling cascade. These signals either directly alter cell function or stimulate the transcription of genes, whose products may be clarified (Bishop, 1991). Carcinogenesis is a complex and multi-step process in which genetic events within normal signal transduction pathways are quantitatively and qualitatively altered. The result of these alterations may occur at any level of the pathways. As the cells accumulate these alterations or mutations, the normal cells can be transformed into malignant ones (Vogelstein *et al*., 1988). Malignancy enhances the ability of proliferation, stimulates neo-vascularization, and grows by locally invading or metastasizing to distant sites (Weiner and Cance, 1994).

Alteration of genes in the dominant excitatory pathways (proto-oncogenes) and mutation of those in the inhibitory pathways (tumor suppressor genes) result in either the gain or loss of function, respectively (Bishop, 1991). The goal of studies of HNSCC is to achieve the early detection, prevention and treatment of patients at high risk of developing this lesion. To attain this goal, potential biomarkers must be identified with a high level of reliability and consistency of expression, and designated as an indicator of early malignant transformation. Therefore, genes conferring either gain or loss of vital growth regulatory function pathways must be investigated. Mutations or altered expression of some of these cellular genes are associated with cellular transformation, and many of them are involved in the

regulation of cell growth and differentiation (Todd *et al*., 1997). The tumor suppressor genes play a role in the normal cell growth regulation, and their alteration impair the important control steps of the cell cycle, such as cell cycle arrest or apoptosis. In normal cells, tumor suppressor genes are generally related to growth regulation (*p53* and *pRb*) (Montenarh, 1992; Weinberg, 1995), cell adhesion (E-cadherin) (Schipper *et al*., 1991), or those effecting G_1 phase progression (cdk inhibitors) (el-Deiry *et al*., 1993). Any alterations of their genetic structure, either constituted or induced by external factors, may have deleterious effects on the normal functions of cell. Although oncogenes alone are not sufficient to transform normal oral keratinocytes, they appear to be the important initiators of this process (Sidransky, 1995). Currently, several candidate biomarkers have been described in oral carcinoma or HNSCC such as *c- myc, H-ras, erbB2, bcl2, hst1, p53, p16, p21, p27* etc (Scully *et al*., 2000).

Literature review

Overview

Each year, approximately 13,000 deaths are attributed in the US to cancers of the oral cavity, salivary glands, larynx and pharynx (Boring *et al*., 1993; Landis *et al*., 1999). As more than 90% of these neoplastic lesions are of squamous cell origin, they are usually referred to, collectively, as squamous cell carcinomas of the head and neck (HNSCC). This cancer is considered the 6th most common cancer among men in the developed world (Landis *et al*., 1999; Parkin *et al*., 1999)**.** In spite of the recent advances in our understanding, the five-year survival rate after diagnosis for HNSCC is still very low, approximately 53%, which is considerably lower than that for other

neoplasias, such as those of colorectal, cervix and breast origin (Landis *et al*., 1999; Parkin *et al*., 1999)**.** This high morbidity rate can be attributed to many factors, which include failure to respond to available chemotherapy, late presentation of the lesions and the lack of suitable markers for early detection. The poor prognosis of HNSCC patients may be the reflection of the fact that many of the risk factors involved in HNSCC pathogenesis.

Recent discoveries have dramatically increased our understanding of the most basic mechanisms controlling normal cell growth, and have also greatly enhanced our ability to investigate the nature of the biological processes that lead to cancer. We now know that the majority of cells in a tumor are derived from the clonal expansion of a single ancestral cell that has acquired an aberrant program of cell growth. Whereas normal cells proliferate only when needed, as a result of a delicate balance between growth promoting and growth inhibiting factors and under the influence of biochemical cues provided by neighboring cells, cancer cells override these controlling mechanisms and follow their own internal program for timing their reproduction. These cells usually grow in an unrestricted manner, and over time they can acquire the ability to migrate from their original site, invade nearby tissues and metastasize at distant anatomical sites.

The progressive changes in cellular behavior, from slightly deregulated proliferation to full malignancy, are a result of the accumulation of mutations in a limited set of genes. Among them, two classes of genes, oncogenes and tumor suppressor genes, play major roles in triggering and promoting cancerous growth (Chi *et al*., 1999; Hoffman and Liebermann, 1998). Whereas activated oncogenes promote cell proliferation, tumor suppressor genes inhibit cell growth and contribute to the

carcinogenic process when inactivated by mutations. An emerging concept is that several activating and inactivating events must occur in oncogenes and tumor suppressor genes for the initiation and progression of many types of cancer (Urbain, 1999). These genetic changes occur in a multistep process (Partridge *et al*., 1997). Thus, if molecular markers representing early and late events could be isolated, it would be then possible to identify persons at high risk of HNSCC, namely, those whose lesions are progressing through the pre-malignant state. Furthermore, the availability of biochemical markers heralding malignancy would be key for monitoring cancer recurrence as well as for the evaluation of the efficacy of novel chemopreventing agents. Clearly, the full elucidation of the genetic changes leading to the development of HNSCC will lead to improved molecular assays with important implications for the early diagnosis, therapy and prognosis of HNSCC patients.

Genetic Alterations in HNSCC

Cancer evolves in a series of distinct steps, each characterized by the sequential accumulation of additional genetic defects followed by clonal expansion (Dewanji *et al*., 1999). Consequently, it is possible to define a genetic progression model of this disease. In this regard, the pioneer work characterizing the genetic alterations in colorectal cancer has now become a paradigm for other human neoplasias (Fearon and Vogelstein, 1990). This model describes a stepwise involvement of important cancer genes during the various stages of tumor progression. For example, the loss of function of key tumor suppressor genes (*APC, DCC, DPC4* and *MCC*) appears to be critical for the conversion of normal epithelium to adenoma. (Gryfe *et al*., 1997; Laken *et al*., 1999; Saito *et al*., 1999; Takaku *et al*.,

1998). While additional mutations in the *K-ras* gene are likely to be responsible for the conversion to malignancy (Chiang, 1998; Giaretti *et al*., 1998). It is now believed that HNSCC follows a similar pattern in its development. And thus, it may be preceded by pre-malignant lesions, such as leukoplakia, erythroplakia, sub-mucous fibrosis, chronic hyperplastic candidosis and erosive lichen planus (Crosthwaite *et al*., 1996; Lo Muzio *et al*., 1998; Qin *et al*., 1999; Trivedy *et al*., 1999; Wright, 1998). The precise nature of the genetic alterations occurring at each step is still unclear, but a recent report has described a preliminary HNSCC molecular progression model (Califano *et al*., 1996). From the scheme illustrated in fig 1.1, it is now regarded that in this tumor type, these genetic lesions occur in a distinct order of events. For instance, loss of chromosomal material is thought to result in changes leading to dysplasia (9p21, 3p21, 17p13), carcinoma *in situ* (11q13, 13q21, 14q31) and invasive tumors (4q26-28, 6p, 8p, 8q). Although the available information is still limited, it nevertheless provides a framework for the understanding of the molecular pathogenesis of this cancer type.

Although cytogenetic studies have been fundamental in detecting chromosomal rearrangements in HNSCC, the actual identity of genes at these sites often remain elusive. Other techniques, such as comparative genomic hybridization (CGH) and fluorescence *in situ* hybridization, are also frequently utilized to detect pattern of chromosomal imbalances and gross chromosomal regions involved in structural rearrangements (Bockmuhl *et al*., 1998), but these, too, fail to identify genes at these sites. On the other hand, the use of polymorphic microsatellite markers in HNSCC has recently helped identify a number of areas of loss of heterozygosity

(LOH), thus suggesting the contribution of known and novel putative tumor suppressor genes. Areas of chromosome loss frequently described include; 3p, 4q,

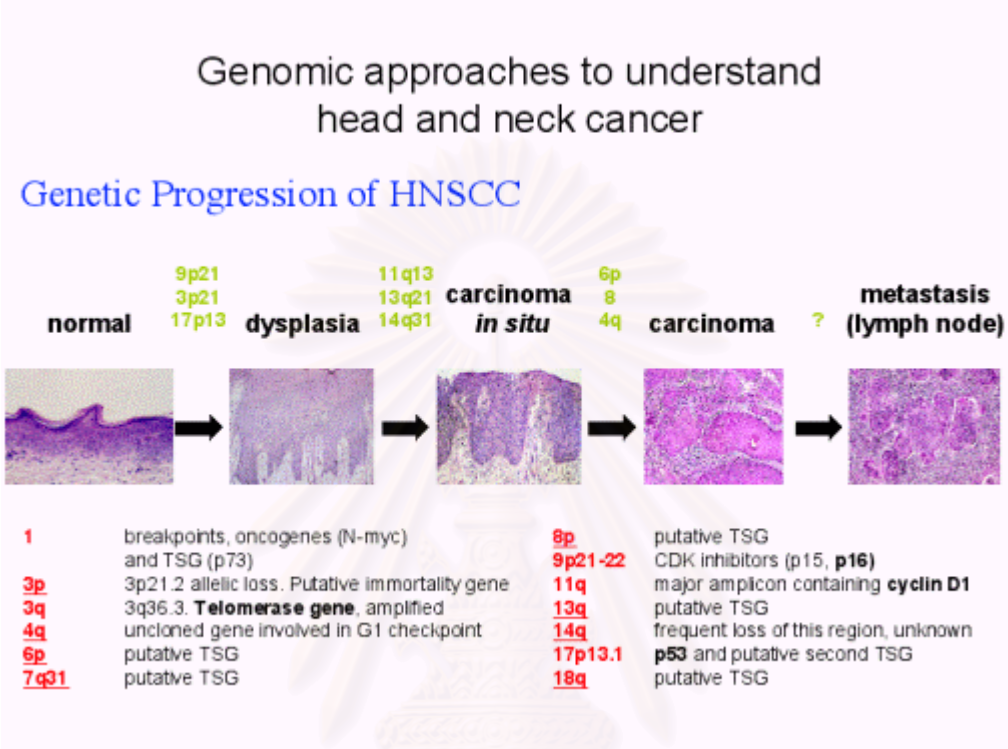


Figure 1.1: *Genetic alterations characterizing HNSCC*

An adapted scheme (Califano *et al*., 1996) illustrates those chromosomal alterations that may contribute to the pathogenesis from normal tissue (A; x200), dysplasia (B; x400), carcinoma *in situ* (C; x200), overt carcinoma (D; x200) and metastasis (E; x200). Non-random loss of material on chromosome 9p21, 3p21, 17p13 and, 11q13, 13q21, 14q31 and, 4q, 6p, 8p, 8q are thought to result in changes leading to dysplasia, carcinoma *in situ* and invasive tumors, respectively. Additional defects required for conversion to the invasive and metastatic phenotypes are currently unclear. Areas of chromosome loss suggest the involvement of important growth regulatory genes, whose identity is still largely unknown.

The mapping of a candidate gene to regions of chromosome loss does not necessarily implicate the identified gene in tumorigenesis, as the genetic instability that characterizes cancer cells may lead to frequent alterations in some susceptible areas which may have no obvious biological relevance (Bryce *et al*., 1999). Nonetheless, these studies suggest a role in HNSCC for p16 (9p21), APC (5q21-22) and p53 (17p13), which are known tumor suppressor genes. Indeed, mutations affecting the expression or function of these genes have been reported to involve nearly 70%, 50% and 40% of all head and neck cancers, respectively (Sartor *et al*., 1999; Taylor *et al*., 1999; Uzawa *et al*., 1994). The identity of additional putative tumor suppressor genes affected in HNSCC remains under intense investigation.

The complete sequence of the human genome should be known shortly, and with it, the chromosomal location of all genes, including those that may be involved in many types of cancers. It follows that if these genes can be identified as targets for non-random chromosomal alterations, it will suggest their putative contribution in malignancy, thus advancing tremendously our understanding of cancer etiology. In this regard, over 26,000 recurring and non-random chromosomal alterations have been thus far reported for all cancers (Mitelman *et al*., 1997). Obviously, this explosion of information on cancer-specific chromosomal aberrations can only be of use to the scientific community if compiled in easily accessible databases. In response to this need, the Cancer Chromosome Aberration Project (cCAP), an initiative supported by the National Cancer Institute and the National Center for Biotechnology Information, has been set up as a public domain database, to catalogue all distinct chromosomal alterations associated with malignant transformation (http://www.ncbi.nlm.nih.gov/CCAP). This, in part, could have important clinical implications, as

diagnosis of cancers based on their distinct genetic profiles may be one day used to determine treatment strategies. On the other hand, as large areas of the human genome are now known, it is possible to fluorescent label large inserts of human DNA, thus revealing genes and facilitating the identification of regions involved in cancer-specific chromosomal aberrations.

All HNSCC-specific chromosomal alterations reported thus far are summarized in the scheme depicted in fig 1.2. The location of all non-random and recurring break points, together with some of the genes known to be associated at these sites are illustrated. Collectively, this information indicates that the pathogenesis of HNSCC involves the gross re-arrangements of many loci. Unfortunately**,** only a few genes located at these sites are currently known to be involved in tumor progression (Sartor *et al*., 1999; Taylor *et al*., 1999; Uzawa *et al*., 1994), thus reflecting our still limited understanding of the main genetic lesions involved in the molecular progression of HNSCC.

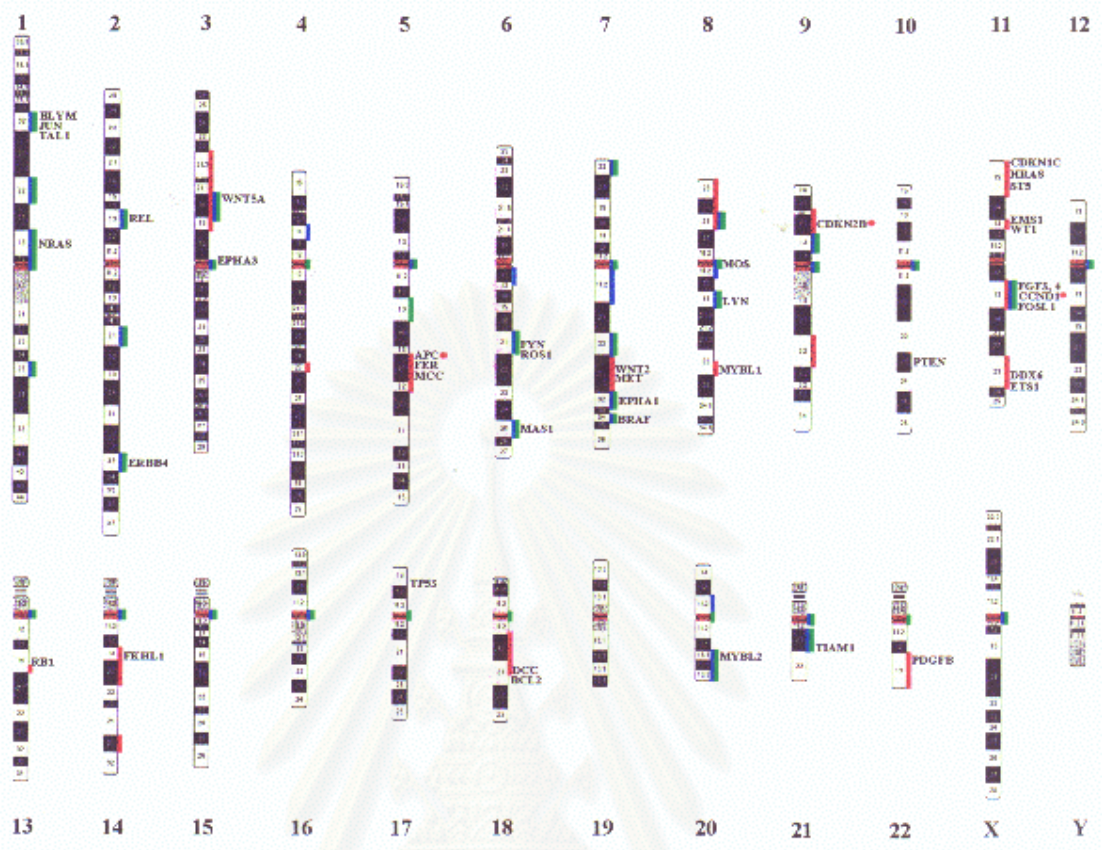


Figure 1.2:*Genome-wide map of chromosomal breakpoints in HNSCC*

Breakpoint Map of all recurrent neoplasia-associated chromosomal aberrations (balanced and unbalanced) reported thus far for HNSCC are schematically represented. The information on these non-random events are a compilation from those alterations reported in the literature (▮), Catalog of Chromosome Aberrations in Cancer (Mitelman *et al*., 1997) (▮) and from the cCAP project (▮). The loci of some of the known genes that map to these sites are indicated, and include those that are now thought to have a role in the pathogenesis of HNSCC (●). Further details regarding these genes can be obtained from the tumor suppressor and oncogene directory, available through CGAP (http://www.ncbi.nlm.nih.gov/ncicgap).

High throughput Analysis of Differential Gene Expression

Methods for identification of differentially expressed genes have a wide variety of applications in modern biology, for example, the discovery of genes that are up-or down regulated in malignant cells and tumor (Byrne *et al*., 1995; Frigerio *et al*., 1995; Gress *et al*., 1997; Ji *et al*., 1997; Watson and Fleming, 1994). In addition to contributing to the understanding of the molecular mechanism underlying carcinogenesis, such discoveries might, in the long term, serve as a basis for the development of gene therapy strategies. A number of methods have been developed over the past seven years for identification of differentially expressed genes and they all have proven successful, although to different extents. In recent years, a variety of techniques have been developed to analyze differential gene expression, including comparative expressed sequence tag (EST) sequencing, differential display (DD), representational difference analysis (RDA), cDNA microarrays, and serial analysis of gene expression (SAGE).

i. Expressed Sequence Tag (EST) sequencing

The concept of EST sequencing first came into public view in 1991 (Adams *et al*., 1991). The basic idea is simple: create cDNA libraries from tissues of interest, pick clones randomly from these libraries, and then perform a single sequencing reaction from a large number of clones. Each sequencing reaction generates 300 base pairs or so of sequence that represents a unique sequence tag for a particular transcript. An EST sequencing project is technically simple to execute, since it requires only a cDNA library, automated DNA sequencing capabilities, and standard

bioinformatics protocols. To generate meaningful amounts of data, however, high throughput template preparation, sequencing, and analysis protocols must be used.

EST sequencing can be accomplished using normalized or non-normalized cDNA libraries. A normalized cDNA library is one in which each transcript is represented in more or less equal numbers (Patanjali *et al*., 1991; Soares *et al*., 1994). The advantage of using normalized cDNA libraries is that redundant sequencing of highly expressed genes is minimized, and the potential for identification of rare transcripts is maximized (Bonaldo *et al*., 1996). An advantage of non-normalized library is that the transcript abundance of the original cell or tissue is accurately reflected in the frequency of clones in the libraries. Non-normalized libraries can be used for an EST project to identify highly expressed unknown genes and to compare the expression of highly expressed genes in different cell or tissue samples (Ji *et al*., 1997).

ii. Differential display (DD)

Differential display (Liang and Pardee, 1992) and RNA fingerprinting by arbitrarily primed PCR (Welsh *et al*., 1992) are probably the most widely used of these methods, and they have promoted the isolation of differentially expressed genes in cancers (Liang and Pardee, 1992; Welsh *et al*., 1992). The general strategy is to amplify partial cDNA sequences from subsets of mRNAs by reverse transcription and PCR, and then to display the short cDNA fragments on a sequencing gel. Pairs of primers are selected so that each will amplify DNA from 50-100 mRNAs at a time, a number that is optimal for display on the gel. The main difference between differential message display (Liang and Pardee, 1992) and RNA fingerprinting by arbitrarily primed PCR (Welsh *et al*., 1992) is in the sequence of the oligonucleotides

used to prime reverse transcription. As a result, different regions of the mRNAs are fingerprinted by these two procedures. Differential message display uses a primer based on oligo(dT) but with an “anchor” of two bases at the 3′ end. After reverse transcription and denaturation, arbitrary priming is performed on the resulting first strand cDNA. PCR is then carried out to generate a series of products derived from the 3′ end of the mRNAs. RNA fingerprinting, on the other hand, uses an arbitrary primer for reverse transcription, thus selecting those regions internal to the RNA that have 6-8 base matches with the 3′ end of the primer. This is followed by arbitrary priming of the resulting first strand cDNA with the same or a different arbitrary primer, and then PCR amplification.

iii. Representational difference analysis (RDA)

Representational difference analysis (RDA) is a PCR–based subtractive hybridization procedure that was originally developed for cloning the differences between genomes (Lisitsyn and Wigler, 1993), and then it was later adapted for cloning differentially expressed genes (Hubank and Schatz, 1994). Genomic RDA relies on the generation, by restriction enzyme digestion and PCR amplification, of simplified versions of the genomes under investigation, known as “representation”. If an amplifiable restriction fragment (the target) exists in one representation (the tester), and is absent from another (the driver), a kinetic enrichment of the target can be achieved by subtractive hybridization of the tester in the presence of excess driver. By virtue of the fact that tester and driver fragments are ligated to different primer adaptors to enable selective PCR amplification, tester sequences with homologues in the driver population are rendered unamplifiable, while the target hybridizes only to itself, and retains the ability to be amplified by PCR. Successive iterations of the

subtraction/PCR process results in increasingly higher enrichment. Critical in genomic RDA is the creation of simplified representations of the genome, called “amplicon” by restriction digestion. A high proportion of digested fragments do not fall into the amplifiable range, thus reducing the complexity of the amplicon so that the final representation contains only ~2-10% of the total genome (Lisitsyn and Wigler, 1993). In contrast, however, cellular mRNA populations are significantly less complex than genomic DNA, thus warranting the success of cDNA RDA without the need to reduce the complexity of the original cDNA populations.

iv. cDNA microarrays

The feasibility of the method based on hybridization to cDNA microarrays immobilized on glass slides was first demonstrated for 45 *Arabidopsis* genes (Lisitsyn and Wigler, 1993), and later applied to surveying human mRNAs (DeRisi *et al*., 1996; Schena *et al*., 1996). An attractive feature of this method is that it enables differential expression measurements to be made by means of simultaneous, two-color fluorescence hybridization. Although this method was shown to have enough sensitivity in the *Arabidopsis* system, it remains to be determined whether the same will hold true for more complex mRNA populations with repetitive elements occuring at a wide range of frequencies. Furthermore, it should be emphasized that although these hybridization-based approaches seem likely to become the methods of choice for large scale survey of gene expression in humans and in mice, they cannot be readily applied to other organisms because they rely on the availability of a complete set of cDNA sequences for the organism of interest. Therefore, PCR-based and sequence-based methods will remain invaluable for many years to come.

A significant advantage of cDNA microarray analysis is the ability to analyze the same set of genes under a range of experimental condition. For example, Heller et al. created an array of 96 known genes to be involved in inflammatory processes (Heller *et al*., 1997). To identify genes specifically involved in rheumatoid arthritis, they probed the array with RNA from cultured macrophages, chondrocytes, and synoviocyte, as well as arthritic tissue samples. These experiments demonstrated for the first time the involvement of several genes, including interleukin 3 and Groα, in rheumatoid arthritis. In another example, An oligonucleotide array to monitor the response of virtually all of the genes in the yeast genome to variety protein kinase inhibitors (Gray *et al*., 1998). They were able to identify genes that responded uniquely to a specific compound, as well as genes that responded similarly to a range of compounds. The ability to perform multiple assays on the same arrays provides a powerful approach to streamlining the search for gene(s) with specific characteristics of interest.

v. Serial Analysis of Gene Expression (SAGE)

Serial Analysis Gene Expression (SAGE), on the other hand, is a more recently developed sequence-based strategy (Velculescu *et al*., 1995) that allows the simultaneous analysis of a large number of transcripts. SAGE is based on two principles: firstly, a short nucleotide sequence tag of 9-10 base pairs (bp) containing enough information to unequivocally identify a transcript; and secondly, concantenation of these short sequence tags allows the efficient analysis of transcripts serially by sequencing of multiple tags within a single clone. Briefly, double-stranded cDNA is digested with a frequent-cutting restriction enzyme (four-base recognition sequence termed “anchoring enzyme restriction site”. SAGE has a tremendous

potential as a method to register the occurrence of large numbers of transcripts in particular mRNA or cDNA clone populations. For example a subtracted cDNA library could be converted into a SAGE library for rapid identification of tissue-specific transcripts. Furthermore, significant improvements in sequence technology are likely to result from ongoing efforts to obtain the complete sequence of the human genome, which could make sequence-based approaches, such as SAGE, progressively more efficient and cost effective. It should be emphasized, however, that identification of SAGE tags is only possible if matching cDNA sequences exist in public database.

The variety of methods for high throughput analysis of differential gene expression has been developed over the past several years. If these methods are used properly, they offer the opportunity to understand biological processes at a level of molecular detail that was not possible even a few years ago. However, the high througput nature of these experiments is a double-edged word: if an experiment is poorly designed, or if the biological materials are compromised, the result is a large body of data that is difficult and time-consuming to analyze. In addition, some approaches are better suited than others for addressing specific biological question. A set of guidelines to choose the method is shown in table 2.1 (Carulli *et al*., 1998).

Regardless of the strategy applied, however, it is crucial that the ongoing gene discovery efforts be completed so that a final master collection of human cDNAs and a unigene set of ESTs comprising all human genes become available for general expression studies. Accordingly, the recently implemented Cancer Genome Anatomy Project (Strausberg *et al*., 1997) (http://www.ncbi.nlm.nih.gov/ncigap) represents a remarkable initiative towards the achievement of this ultimate goal.

	RNA requirements	Throughput	Sequencing requirements	Cloning requirements	Bioinformatics requirements
EST sequencing	1.0-5.0 μg Poly A.RNA	Low	High	Full-length cloning may be required for novel genes of interest	Target databases:standard Search protocols:Standard Volumn:high
Microarray	1.0 μg or more Poly A RNA	High	Low	Full-length cloning may be required for novel genes of interest	Image analysis required Target databases:standard Search protocols : standard; Volumn:low
RDA	10-100 ng Poly A RNA	Medium	Low	Full-length cloning required for novel genes of interest	Target databases:standard Search protocols:Standard Volumn:low
DD	10-100 ng Poly A RNA	High	Medium	Full-length cloning required for novel genes of interest	Target databases:standard Search protocols:Standard Volumn:low
SAGE	1.0-5.0 μg Poly A RNA	High	High	Full-length cloning required for novel genes of interest	Target databases: specialized Search protocols: specialized Volumn:high

Table 2.1 ***Attributes of five different methods for high throughput analysis of differential gene expression***

Array Technologies

Advances in array technology now provide a high throughput approach to monitor the hybridization of RNA or DNA-derived samples to cDNA arrays, for mRNA expression and genomic polymorphisms, respectively. This technology has the potential for identifying patterns of gene expression and polymorphisms that may be indicative of certain diseases, including cancer (Khan *et al*., 1999). However, to begin to explain how some of these genes may be involved in tumor progression, a detailed comparison of the pattern of gene expression in normal tissue is strictly required (Alon *et al*., 1999).

Array technology that enables this type of analysis is currently available in two formats. These include DNA fragments that have been arrayed onto either *nylon filters* or *glass slides*, and high density oligonucleotide microarrays (Gerhold *et al*., 1999). In both cases, they are used very similarly to Northern and Southern blots, but have the added advantage of size reduction to allow hybridization of complex probes to appropriate cDNAs, thus allowing the analysis of thousands of genes simultaneously.

Commercially available filter membranes are usually affordable and require no specialized equipment. For example, some of them are specifically arrayed, in duplicate, with recognized human cancer genes of known function, the majority of which are believed to play a major role in the regulation of normal cell growth, differentiation, survival, and/or carcinogenesis. These cancer genes are usually PCR-amplified, sequence-verified fragments of 200-500 bp, which were specifically chosen for their optimized hybridization characteristics. The uses of filters thus allow an approach to monitor the expression and function of normal and aberrant genes and

understand how they may be pertinent to HNSCC. However, this does not allow for the discovery of novel genes specific to HNSCC, but it is expected to help expedite the identification of the functional role for these genes in the pathogenesis of HNSCC.

Laser Capture Microdissection

A major scientific challenge in HNSCC is the understanding of the molecular events that drive tumor progression *in vivo*. Using currently available molecular biology techniques, recent studies have identified genes that may be involved in the pathogenesis of HNSCC (Bockmuhl *et al*., 1998; Jin *et al*., 1998; Shao *et al*., 1998). However, the possibility exists that the heterogeneity of the tissues used for these studies may limit the value of this body of information. Indeed, only a fraction of the total tissue volume (<5%) is suitable for this type of analysis, thus the use of bulk tissue or tissue areas of interest manually microdissected may include abundant contamination, such as cells of lymphatic or stromal origin.

The Laser Capture Microdissection (LCM) apparatus as shown in figure 1.3, recently developed at the Laboratory of Pathology (National Cancer Institute), allows a precise and a more accurate assessment of molecular alterations in cancers. This procedure enables the procurement of pure cell populations from frozen or archival human tissue sections in one step under direct visualization (Dean-Clower *et al*., 1997; Emmert-Buck *et al*., 1996; Pappalardo *et al*., 1998; Simone *et al*., 1998). And thus, provide a platform for current efforts in defining some of the molecular basis of neoplasias as they exist *in vivo*. In this regard, the accurate procurement of specific cell types for RNA isolation by using LCM remains a critical step in the analysis of genes expressed in HNSCC and the potential of addressing their possible contribution to neoplasia, if comparison can be made with normal tissue.

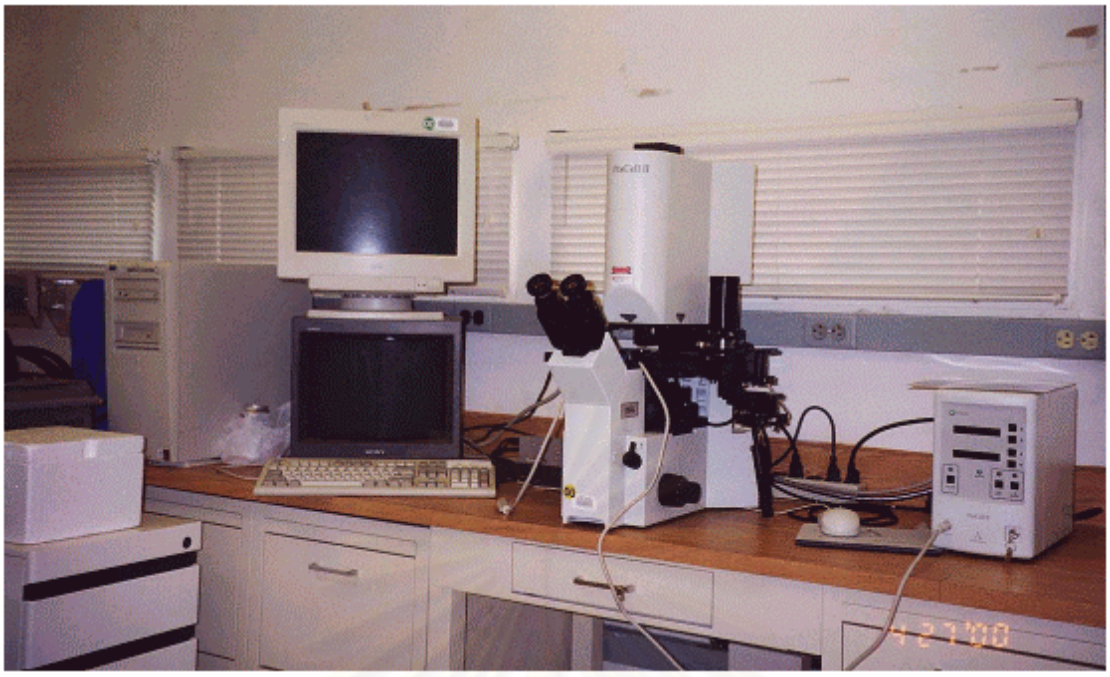


***Figure 1.3**: Laser Capture Microdissection apparatus*

Laser Capture Microdissection (LCM) was developed recently between the Laboratory of Pathology National Cancer Institute, National Institute of Health, Bethesda, MD 20892 and Arcturus Engineering Inc. Mountain View CA 94043.

The method of laser capture microdissection (LCM) illustrated in Chapter II (fig 2.2) primarily involves visualizing the tissue with the microscope and using a pulse of laser beam to procure those cells of interest onto a plastic cap coated with ethylene vinyl acetate (EVA). These caps are transferred to microcentrifuge tubes (Eppendorf, Westbury, New York, USA, Cat # 2236430-8) that contain appropriate lysis buffer, for extraction of RNA. In addition, a video camera on top of the microscope enables the user to monitor the progress of the microdissection, thus providing a convenient reference point for subsequent serial sections. The sequence of events of procuring a pure population of tumor cells whereby an area of HNSCC tissue of interest is visualized and targeted for capture onto caps with a laser beam.

The homogeneity of the captured cells is then confirmed under a light microscope prior to proceeding with RNA extraction. With this method, a pure population of cells of interest (>95% purity) can be rapidly concentrated, and with maximum preservation of RNA. The isolated RNA from these cells can then be used with the currently available array technology for the detailed analysis of gene expression.

With consideration to the laser energy and beam diameter, each laser pulse is able to capture approximately 1-7 cells. In preliminary experiments, it has been determined that approximately 5,000 cells (1,000 shots) are required for gene expression analysis. Therefore, an advantage with this approach is that only a small amount of starting material (5,000 microdissected cells) is required to extract a sufficient quantity of total RNA. Furthermore, the quality and integrity of the RNA makes this approach suitable for use with available array technology, thus affording the possibility of defining a pattern of gene expression in a tumor progression model of HNSCC.

It is important to consider that many of the previously reported genetic alterations may be affected by the heterogeneity of tumor tissue. In this regard, the procedure of LCM can be extended for use in the accurate analysis of DNA and this can be advantageous even if only archival tissue samples are available. Using this approach, extracted DNA from microdissected cells can now be successfully used, for example, to identify loss of heterozygosity (LOH), mutational changes associated with known tumor suppressor genes (Park *et al*., 1999) and, more applicable for high throughput use, single nucleotide polymorphisms (SNPs). LCM technology has been successfully applied to DNA and RNA analysis, from frozen and paraffin embedded

tissues, to understand some of the underlying molecular basis of cancer, namely genomic alterations and expression pattern of critical genes. However, if this body of information could be complemented with proteomics, this may then expedite a more thorough understanding of molecular pathways that ultimately result in cancer. Information that can be acquired from protein based analysis include translational efficiency, post-translational modifications, protein-protein interactions, and the expression and activity of transcription factors. Therefore, any information that can be obtained regarding protein products of mutated or dysfunctional genes will provide an insight into tumor progression and will help to identify potential markers for early detection and drug development.

This type of analysis was previously limited by the inability to extract, quantify and characterize functional proteins expressed by a small population of cell of interest in tumor tissues. Furthermore, tissues available were usually heterogeneous and manually dissected, and the protein identification involved micro sequencing and mass spectrometry. In addition, these approaches did not allow a high throughput application for the analysis of many samples simultaneously. However, it is possible to gain important information regarding critical steps of tumor progression, as described recently for bladder cancer (Celis *et al*., 1999), using the available methods for proteomics, such as 2-D gel electrophoresis, western blot analysis and ELISA.

A new protein analysis system, based on the SELDI (Surface-Enhanced Laser Desorption/Ionization) technology has been recently applied to the high throughput separation, detection, and analysis of proteins in very small amounts (~10 ng) of microdissected cancer tissue. This system enables protein capture, purification,

analysis, and processing from complex biological mixtures directly on Protein Chip Array surfaces and the detection of the purified proteins is performed by laser desorption /ionization flight time mass analysis. In this regard, the potential exists to characterize proteins associated with specific forms of cancer, from benign to pre-invasive carcinoma to invasive carcinoma, if comparison with normal tissue is made. Indeed, this approach has been used successfully to protein "fingerprint" microdissected cell extract from different cancers (Ornstein *et al*., 2000) and may be used as markers of tumor progression.

Cancer Genome Anatomy Project (CGAP)

There are estimated to be approximately 100,000 genes in the human genome and of these 4,000 may be related to disease, including cancer (Trent *et al*., 1997). Altered expression of some of these genes is now thought to be the basis of most cancers and it follows that the identity of these genes will not only enhance our understanding of the molecular basis of the disease, but will also provide means of early detection and subsequent treatment. In response, the Cancer Genome Anatomy Project (CGAP), supported by the National Cancer Institute (NCI), was established with the goal of achieving a comprehensive molecular characterization of normal, precancerous, and malignant cells, in order to create a complete information infrastructure of genes expressed during cancer development (http://www.ncbi.nlm.nih.gov/ncicgap). In only a few years, this project has become the leading effort in gene discovery, and has united the newest technologies, along with those both cost-effective and capable of high-throughput, to identify all the genes responsible for the establishment and growth of cancer. The CGAP initiative involves the generation of cDNA libraries from cancer cells, and after random sequencing, expressed genes are

then catalogued and compared with those from normal corresponding tissues. Key to the success of this effort has been the development of robust databases and easily accessible Web-based analytical tools.

The Head and Neck CGAP was established as a cooperative effort between the National Institute of Dental and Craniofacial Research (NIDCR) and the NCI CGAP. Initially, six high quality cDNA libraries from representative cancer cell lines of head and neck (Patel *et al*., 1997) and normal and immortalized gingival keratinocytes were submitted. Overall, from the 1,160 clones sequenced, 38 of these were identified as novel genes. Furthermore, a distinct pattern of gene expression can already be observed when comparing normal to cancerous cells using analytical tools available at the CGAP site. An early report on this genome-wide analysis of oral cancer are available (Shillitoe *et al*., 2000).

Problems and hypothesis

Many studies have examined the differences in the expression of one or a few genes in HNSCC, but no comprehensive and systematic study of gene expression profiles in these neoplastic cells has yet been undertaken. Therefore, available molecular models are not able to explain fully the complexity of the genetic changes that occur during the development and progression of HNSCC. The recent development of several high-throughput hybridization-based methods utilizing cDNAs arrayed on nylon membrane and glass slides now allows the detailed analysis of thousands of genes simultaneously. This provides a unique opportunity to identify the nature of those genes that are differentially expressed in normal and tumor squamous cells, particularly when combined with the recently developed technology

of laser capture microdissection to procure specific cell populations. It is now recognized that all cancers arise from the accumulation of genetic changes within a limited number of cells. Thus, the ability to define the gene expression profile of normal as well as tumor cells at each stage of cancer development is key to advance our still limited knowledge of this devastating disease. This information is invaluable in our fight against cancer, as it is expected to facilitate cancer prevention, its early detection and diagnosis, and the selection of optimal treatment modalities. We hypothesized that the pattern of genes expressed at each stage of tumor progression is different and related to the pathogenesis of HNSCC, and possibly distinct from those of other better-understood tumor types. We hypothesized that their unique gene expression profile can be revealed by procuring specific cell populations upon microdissection, and by the construction of stage-specific cDNA libraries and the use of recently developed high density cDNA array technologies.

Specific Aims

1: To identify genes differentially expressed between normal and tumor tissues from the same patient by using LCM and human cancer cDNA arrays. This specific aim was achieved (Chapter II) and published (Leethanakul *et al*., 2000) (Appendix A).

2: To establish cDNA libraries from HNSCC patient tissues. This specific aim was accomplished (Chapter III). Six HNSCC cDNA libraries had been constructed and nucleotide information made available in the public domain at the web site: http://www.ncbi.nlm.nih.gov/ncicgap. This research was published (Leethanakul *et al*., 2000) (Appendix B).

3: To elucidate the nature of those genes that may be involved in the pathogenesis of HNSCC. This investigation was conducted using analytical tools that are available at the web site: http://cgap.nci.nih.gov/Tools. Results were prepared as a manuscript for Oral Oncology (Chapter IV).

Benefits

The results obtained from this study will provide a better understanding of gene expression patterns in different stages of HNSCC. The construction of cDNA libraries will also provide the basis for the creation of a publicly available catalog of genes involved in squamous carcinogenesis. Information provided by this investigation will be beneficial for the diagnosis, prevention and treatment of head and neck cancers.

CHAPTER II

Distinct Pattern of Expression of Differentiation and Growth-Related Genes in Squamous Cell Carcinomas of the Head and Neck Revealed by the Use of Laser Capture Microdissection and cDNA arrays

Introduction

Squamous cell carcinoma of the head and neck (HNSCC) is the 6th most common cancer in developed countries, and of the 44,000 annual cases reported in the United States, approximately 11,000 will result in an unfavorable outcome (Landis *et al*., 1999; Parkin *et al*., 1999). In spite of its high incidence, the molecular mechanisms of this disease remain poorly understood. However, the recently gained knowledge of normal and aberrant function of oncogenes and tumor suppressor genes has provided unique opportunities to understand, and ultimately to control, the processes leading to malignancy. Thus, the identification of the molecular and genetic events involved in each step of tumor progression may be central to understand HNSCC, and for the development of diagnostic markers and novel treatment strategies.

Although HNSCC is thought to result from the progressive accumulation of genetic lesions leading to malignancy (Mao *et al*., 1998) the precise nature of the affected molecules is still largely unknown. The recent development of several high throughput, hybridization-based methods utilizing cDNAs arrayed on nylon membranes and glass slides allows the analysis of hundreds of genes simultaneously, and thus provides a unique opportunity to identify genes expressed in normal and

tumor tissues, as well as to analyze gene expression profiles in tumor progression. However, an accurate procurement of specific cell types for RNA isolation is a critical step influencing the validity of this analysis. In this regard, a novel technique of Laser Capture Microdissection (LCM) developed at the Laboratory of Pathology (National Cancer Institute), enables the procurement of pure cell populations from frozen human tissue sections (Emmert-Buck *et al*., 1996; Simone *et al*., 1998), a key consideration as many tumors, including HNSCC, are heterogeneous, and include areas of connective tissues, blood vessels and even inflammatory cells that infiltrate into the tumor mass. Most importantly, microdissection allows analysis of gene expression in specific cell populations as it exists *in situ*. In this study, we have used LCM to procure specific cell populations from a representative set of tumors and their matching normal tissues to explore the feasibility of establishing a pattern of expression of cancer-related genes for HNSCC.

Materials and Methods

Tissue samples

Tissue samples from HNSCC patients undergoing surgery were immediately snap frozen in liquid nitrogen and H&E sections from these samples were subsequently analyzed and confirmed by a board certified pathologist, as either normal, hyperplasia or malignant. The tissue samples were then processed using a standard embedding protocol, currently available at the following web site: http://dir.nichd.nih.gov/lcm. Briefly, empty cryomolds were placed on dry ice for 1 min prior to immobilizing the frozen tissue samples in a thin layer of OCT embedding medium (Sakura Finetek Torrance, CA). Tissue in the cryomold was then covered

with OCT and allowed to harden at -70°C. Eight micron thick sections were cut on to RNAase free glass slides, using a cryostat and stored at -70°C until use.

Tissue Staining of Tissue Sections

Frozen tissue section slides were stained prior to Laser capture microdissection (LCM) as described on the web site: http://dir.nichd.nih.gov/lcm. Briefly, slides were fixed in 70% ethanol, washed in purified ddH_20, stained in Mayer's hematoxylin and placed in blueing reagent. After a brief rinse in 70% ethanol, the slides were counter stained with eosin and dehydrated with 95-100% ethanol and xylene. All chemical reagents were from Sigma.

Laser Capture Microdissection (LCM)

The use of LCM (Arcturus Engineering, Mountain View, CA) was essentially as described (Emmert-Buck *et al*., 1996; Simone *et al*., 1998). Briefly, an RNAase free plastic cap with a transparent ethylene vinyl acetate (EVA) thermoplastic film was maneuvered with the transport arm, precisely over an area of tissue of interest, which had been previously selected with the microscope. The laser parameters on the microscope were set at 20 mW, 50 msec and 30 μm for power, pulse duration and beam diameter respectively, in order to facilitate capture of cells onto the EVA thermoplastic film on the caps. With this procedure, one thousand laser shots procured approximately 5000 cells. The caps containing the procured cells were subsequently transferred to a 0.5 ml RNAase free microcentrifuge tube containing lysis buffer for total RNA extraction. Caps coated with EVA for use with LCM were supplied by Arcturus Engineering.

RNA Extraction

Captured cells contained within the cap and microcentrifuge tube were digested with 200 μl guanidium isothiocyanate (GITC) buffer and 1.6 μl β-mercaptoethanol for 30 sec. After centrifugation (14000 rpm at 4°C, for 30 min), the supernatant was transferred to a 1.5 ml RNAase free microcentrifuge tube and after the addition of 20 μl of 2M sodium acetate, 220 μl saturated phenol and 60 μl chloroform-isoamyl alcohol, the samples were vortexed and placed on ice for 15 min. All reagents described above were from Stratagene, La Jolla, CA. The resulting aqueous layer after centrifugation (14000 rpm at 4°C for 30min) was transferred to a fresh RNAase free microcentrifuge tube and the RNA precipitated with 200 μl cold isopropanol (Stratagene) and 2 μl of glycogen (10 μg/μl; GenHunter, Nashville, TN) at –80°C for 1 h. After centrifugation (14000 rpm at 4°C for 30min) the RNA pellet was treated with 2 μl DNAase I (10 units/μl**;** GenHunter), 1 μl RNAase inhibitor (20 units/μl**;** Perkin-Elmer, Branchburg, NJ), 2 μl 10x reaction buffer (GenHunter) and incubated at 37°C for 2 h. The RNA was reprecipitated as described, then resuspended in 3.5 μl deionized water and 1 μl RNAase inhibitor (20 Units/μl) and used for cDNA synthesis.

RT-PCR and RNA integrity

Total amount of RNA from each sample was reverse transcribed using the SMART PCR cDNA synthesis kit and according to instruction provided (CLONETECH Laboratories, Inc., Palo Alto, CA). Briefly, for annealing, 5 μl RNA samples containing 1 μl (10 μM) cDNA synthesis (CDS) primers, 0.5 μl (10 μM)

SMART II oligonucleotides were heated for 2 min at 72°C and cooled to 42°C. RNA was subsequently reverse transcribed with the addition of a master mix (2 μl 5x first strand buffer, 1 μl 20 mM DTT, 1 μl 10mM dNTP and 1 μl 200 U/μl Superscript RT; (Life Technologies, Rockville, MD) and reaction left to incubate for 1 h at 42°C. The integrity of each RNA sample was assessed by the ability to amplify *GAPDH* by RT-PCR. One microliter of RNA from each sample was used as template. The reaction was performed using the GeneAmp RNA PCR kit (Perkin Elmer) except *GAPDH* primers (Stratagene). Conditions for PCR were 95°C for 105 sec followed by 30 cycles at 95°C and 60°C for 15 sec and 30 sec respectively followed by 7 min at 72°C and the PCR product subsequently analyzed on a 1.2% agarose/EtBr gel.

cDNA Probe preparation

Radioactive probes were synthesized using 2 μl of cDNA from each tissue set (normal and tumor), 2 μl of dNTP mix (2.5 mM of dGTP, dATP, dTTP and 0.5 mM of dCTP; Perkin Elmer), 75 μl ddH_20, 10 μl 10x PCR reaction buffer and 4 μl 10 mM PCR primers (CLONTECH Smart kit), 5 μl 10 mCi/ml α-^{33}P dCTP (NEN Life Science Products Inc., Boston MA) and 2 μl Taq polymerase (Perkin-Elmer). The samples were mixed and used to synthesis complex cDNA probes by PCR (1min at 95°C followed by 30 cycles of 15 sec at 95°C, 5 sec at 65 °C and 5 min at 68°C, and then cooled to 4°C). The specific activity of each cDNA probe, after purification (PCR SELECT-II columns; 5Prime-3Prime, Inc., Boulder, CO), was assessed by scintillation.

Hybridization of cDNA Arrays and Analysis of Gene Expression

Synthesized cDNA radioactive probes from normal and tumor samples from the same tissue set, were used to simultaneously hybridize human cancer cDNA expression Arrays (CLONTECH) and the conditions used were essentially as described in the provided protocol. Briefly, membranes were prehybridized in ExpressHyp (CLONTECH) at 68°C for 30 min and subsequently hybridized overnight at 68°C with the appropriate cDNA probe. After washes as described, the membranes were analyzed by PhosphorImaging (Molecular Dynamics) and auto-radiography. The hybridization of the cDNA's on the arrays for each of the samples (normal and tumor) was documented and the identity of the genes was determined from the relative position and list provided by the manufactures. The intensity of the hybridization signal of genes from both sets of membrane (normal and tumor) was quantified by PhosphorImaging and the degree of expression of genes between the two samples was compared with those of the house keeping genes.

Results

Clinical features characterizing HNSCC

Clinical characteristics of the human biopsies from HNSCC patients that had undergone surgery and were chosen for the study are indicated in table 2.1. The anatomical sites of these lesions are representative of the most frequent HNSCC sites, and include the tongue (WSU 1, 62 and 63), larynx (WSU 51), and pharynx (WSU 58). All lesions, except for WSU 1, were confirmed carcinomas, which were either poorly (WSU 51, 63) or moderate to well differentiated (WSU 58, 62) and all were invasive**.** Tissue WSU 1 was confirmed as hyperplasia. Corresponding normal tissue,

from the same anatomical site and patient as the lesion, was part of the tissue set and consisted of normal epithelium. Thus, we considered these samples to be representative of HNSCC and suitable for gene expression analysis.

Histological Features of HNSCC Tissue Sets

We initially chose five tissue sets, as indicated above, to assess the feasibility of using LCM for the detailed analysis of gene expression in HNSCC. Before proceeding, we confirmed the provided histopathology of these samples. Microscopic visualization of representative frozen tissue sections (8 μm thickness), stained with hematoxylin-eosin is shown in fig 2.1. Normal oral squamous epithelium (A) shows an orderly architecture that ranges from immature small round cells in the basal layer to those of mature flattened cells with abundant cytoplasm and small nucleus at the superficial layer. Furthermore, cell mitoses in this layer are rare and generally confined to cells in the basal layer. The basement membrane is observed to be intact. Oral hyperplasia (B) is seen to be characterized by an increased cellularity of the lower third of the epithelium and normal maturation of the remaining upper two third. However, the integrity of the basement membrane is similar to that observed for the normal epithelium. Tissue sections from squamous cell carcinomas (C) shows infiltration of the underlying stroma, loss of normal architecture, occasional keratin pearls. There is mild nuclear pleomorphism and increased numbers of mitotic figures. Invasive carcinoma (D) are characterized by cells that have breached the basement membrane and have invaded into the subjacent soft tissue. Furthermore, this tumor type was observed to vary from being well-differentiated squamous cell carcinomas with evidence of keratin formation (keratin pearls) and few mitoses, to that of a

poorly differentiated type, exhibiting no recognizable squamous features. The data therefore illustrates the integrity and the squamous origin of the tissues and thus suitable for detailed molecular analysis of HNSCC.

Case	Origin	Lesion	Pathology
WSU 1	Tongue	Hyperplasia	Severe hyperplasia
WSU 51	Laryngeal	Carcinoma	High clinical grade Invasive and poorly differentiated
WSU 58	Pharyngeal	Carcinoma	Mild to moderate dysplasia, moderate to well differentiated and invasive in places.
WSU 62	Tongue	Carcinoma	Invasive, moderate to well differentiated
WSU 63	Tongue	Carcinoma	Moderate to poorly differentiated and invasive

Table 2.1 *Clinical Characteristics of HNSCC tissues*

HNSCC lesions upon biopsy were analyzed for clinical classification. The five tissue sets (WSU 1, 51, 58, 62, 63) were biopsies from patients previously confirmed to have neoplastic lesions of the head and neck. Anatomical site and severity of the lesions are indicated.

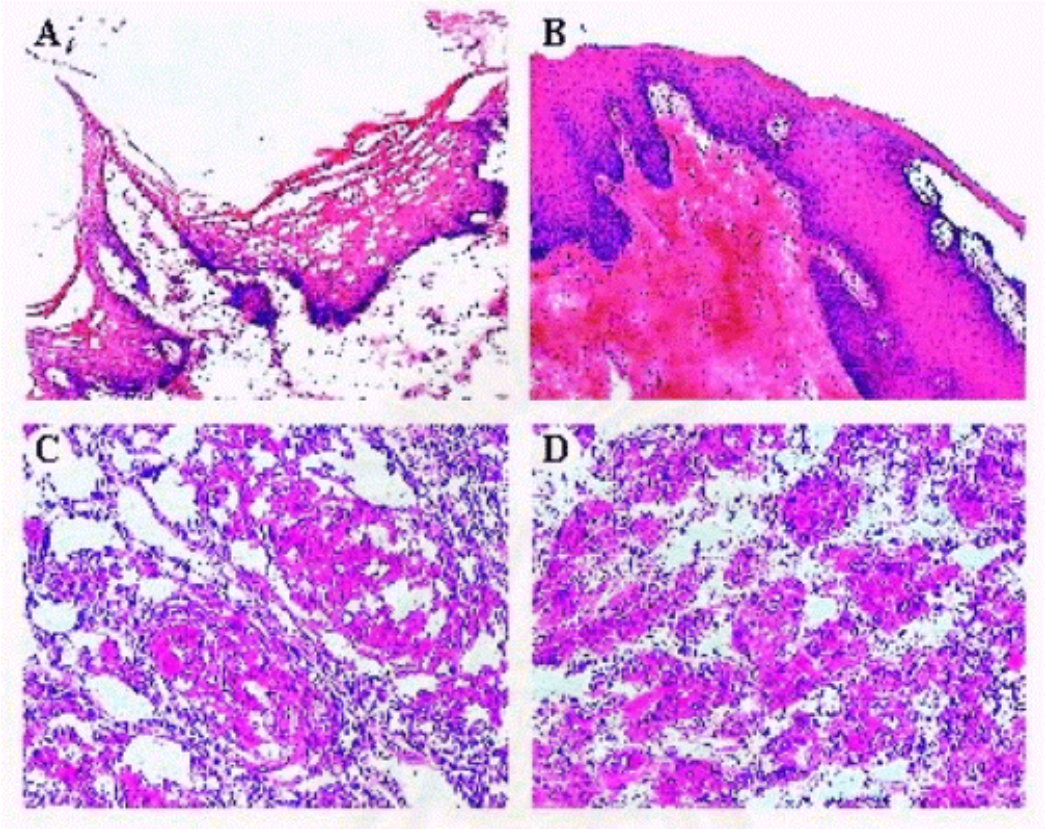


Figure 2.1 *Histopathological Features of HNSCC*

Tissue sets, comprising of both normal and tumor from the same HNSCC patient, were snap frozen and 8 μm sections were stained with H&E. Histopathological features of progression of HNSCC from normal to carcinoma are illustrated. Representative normal squamous epithelium (A) from head and neck region shows an orderly maturation from deep to superficial cell layers marked by progressive flattening of the cells and nuclei (magnification X200). Hyperplasia (B) is characterized by an increased layers of epithelial cells throughout the lower third of the epithelium, normal maturation, and the retention of the basement membrane (magnification X200). Well-differentiated invasive squamous cell carcinoma (C) shows infiltration of the underlying stroma, loss of normal architecture, occasional

keratin pearls. There is mild nuclear pleomorphism and increased numbers of mitotic figures (magnification X500). Moderate to poorly-differentiated invasive carcinoma (D) shows infiltration of the underlying stroma with sheets of cancer cells showing marked nuclear pleomorhism and hyperchromasia. There is notable absence of keratin pearls (magnification X500).

Laser Capture Microdissection (LCM)

The method of laser capture microdissection (LCM) as shown in fig 2.2 (Upper panel**)**, was developed by the Laboratory of pathology, NCI, to provide a reliable method for the procurement of a pure population of cells. Using the joystick, the platform holding the slide can be easily manipulated until an area of interest is determined with the microscope. An additional advantage with LCM is that with the video camera, microscopic images can be viewed and the progress of the microdissection monitored. Similarly, these images can be captured, thus providing a convenient reference point for subsequent serial sections. A pulse of laser beam is then used to capture areas of tissues of interest onto EVA coated caps. Using this method, frozen tissue sections can be microdissected immediately after staining and thus, a pure population of cells of interest (>95% purity) can be rapidly concentrated with maximum preservation of RNA.

Procurement of cells from HNSCC tissue

Under direct microscopic visualization, squamous tumor cells are captured with individual laser shots (fig 2.2 lower panel). Each shot activates the EVA on the cap surface, which then attaches to the cells of interest. With consideration to the laser

energy and beam diameter, each laser pulse is able to capture approximately 3-7 cells. In preliminary experiments, we had predetermined that approximately 5,000 cells (1,000 shots) are sufficient for our analysis. Lifting the cap separates the captured cells from the remaining heterogenous tissue, which is retained onto the glass slide (C). This ensures that the correct cell population is obtained, by visualization under light microscopy (D). The same procedure was followed for procuring cells from normal tissue (data not shown).

RNA Extraction and Analysis by RT-PCR

Prior to analysis of microdissected cells, the integrity of the nucleic acids in these cells was assessed. Quality assessment of the RNA from each tissue was performed by RT-PCR of GAPDH. As demonstrated in fig 2.2 (E), RNA from two representative tissue sets (WSU 62, 63) was considered to be of sufficient quality to reverse transcribe and amplify GAPDH. Using specific primers for GAPDH, a 600 bp product was observed**.** Similar results were obtained with the remaining three tissue sets (data not shown). The data demonstrates that LCM preserves the integrity of the RNA extracted from procured cells and this can be reverse transcribed directly using anchored modified 5′-end enhancer oligonucleotides, thereby facilitating the detailed analysis of gene expression in HNSCC.

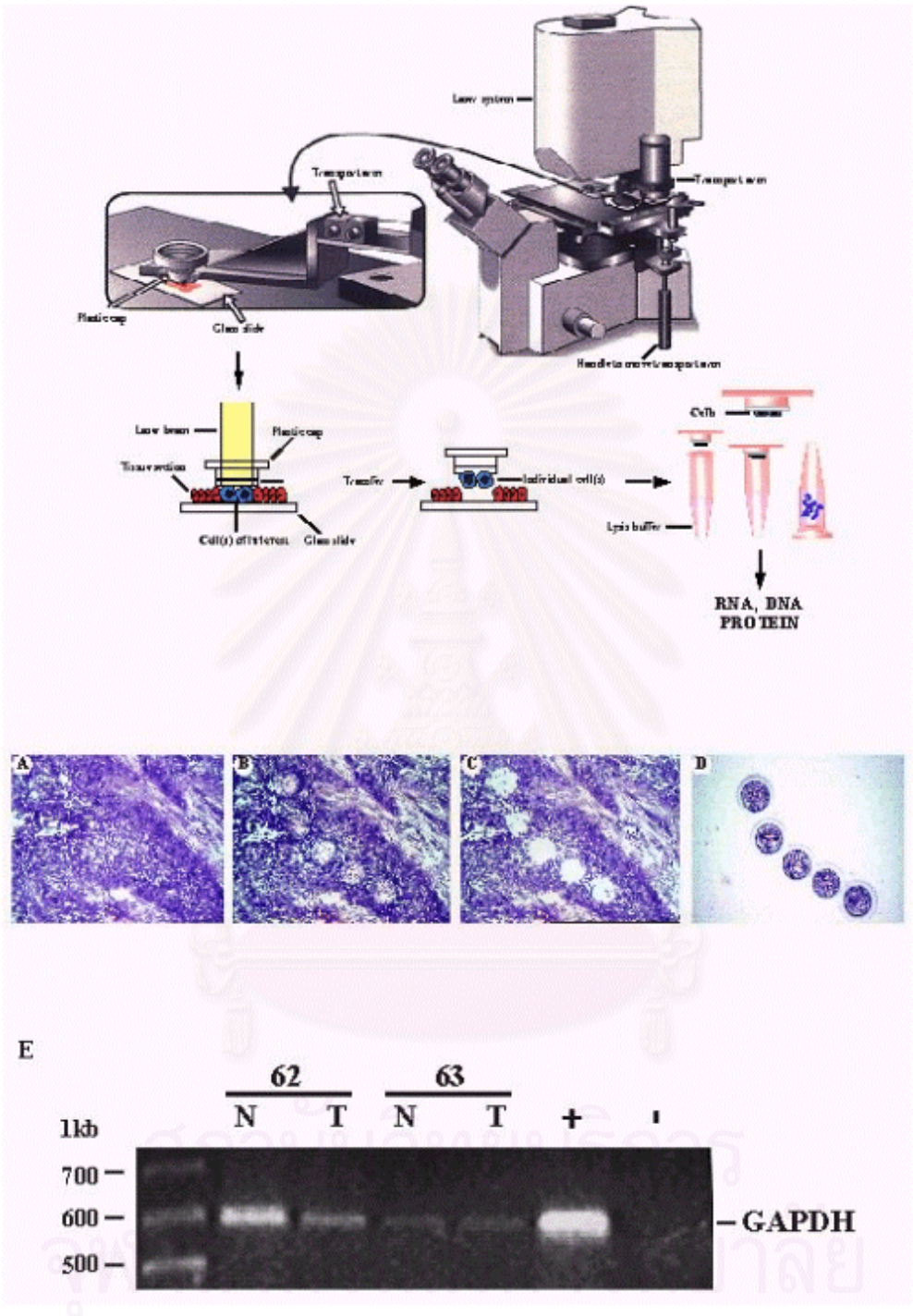


Figure 2.2 *Laser Capture Microdissection (LCM) and its Use in Procuring Pure Populations of Cells from HNSCC and Assessment of RNA Quality by RT-PCR*:

The use of the LCM apparatus is illustrated (upper panel). A pure population of cells from H&E stained tissue sections are microdissected and captured onto EVA coated caps with a laser beam. The caps containing the cells are then transferred to microcentrifuge tubes for RNA extraction and processing as described in Materials and Methods. For procurement (lower panel), an area of tumor, containing cells of interest are visualized (A) and targeted for capture with a 30 μm diameter laser beam (B). The caps containing the captured cells are lifted off the tissue section (C), and the homogeneity of these cells is confirmed under a light microscope (D) prior to processing for RNA extraction. Each laser beam procures 3-7 cells. (magnification X500). Total RNA was extracted from microdissected tissues (approximately 5,000 cells) as described in Materials and Methods. The integrity of the RNA was assessed amplifying *GAPDH* by RT-PCR. Using specific primers for *GAPDH*, a 600 bp fragment is amplified, as observed for both normal (N) and tumor (T) tissues from HNSCC patients (WSU62 and 63). Appropriate positive and negative controls are indicated (E).

Hybridization and Gene Expression

Prior to synthesis of probe by PCR, cDNAs synthesized for each sample were assessed for integrity. In all cases, the average size of the reverse-transcribed messages was demonstrated to be approximately 500 bp (range 300-800 bp, data not shown), which is similar to that achieved for the construction of cDNA libraries (Peterson *et al.*, 1998). cDNAs corresponding to normal and tumor tissue from the same set were simultaneously amplified and labeled with α-^{33}P dCTP. For each sample, 3-4 independent PCR reactions were carried out and these were subsequently

combined and used to hybridize commercially available membranes containing 200-500 bp DNA fragments, in duplicates, for 588 known human cancer and 9 housekeeping genes (http://www.clontech.com/atlas/genelist). Membranes hybridized with complex cDNA probes from a representative tissue set are shown in fig 2.3 (left panel) and illustrates the comparative differences in expression of genes belonging to different functional groups in both normal and tumor tissue from the same patient.

Genes Differentially Expressed

Examples of genes differentially expressed are shown in fig 2.3 (right panel) and include the cytokeratins (A) and those genes belonging to the MAPK (mitogen-activated protein kinase) (B) and *wnt* (wingless) (C) signaling pathways. The amount of radiolabeled probe hybridized to each arrayed DNA was quantified using PhosphorImaging, and normalized by that hybridized to the housekeeping genes. For each DNA, the ratio between the expression in the cancer and normal tissue was then estimated for each tissue set. To simplify the analysis, the main functional groups of genes assessed to be differentially expressed by at least two fold in three or more of the five tissue sets were and considered of likely biological significance are listed in the table 2.2. They include genes involved in the control of cell growth and differentiation, angiogenesis, apoptosis, cell cycle, and signaling, most of which have not been previously implicated in HNSCC when using other analytical approaches (see below). These data indicate that complex cDNA probes labeled with α-^{33}P dCTP can be successfully synthesized from small amounts of total RNA for use in comparative hybridization studies of cancer genes expressed in HNSCC. Furthermore,

our findings have helped identify a number of new candidate genes, which might play an unexpected role in squamous carcinogenesis.

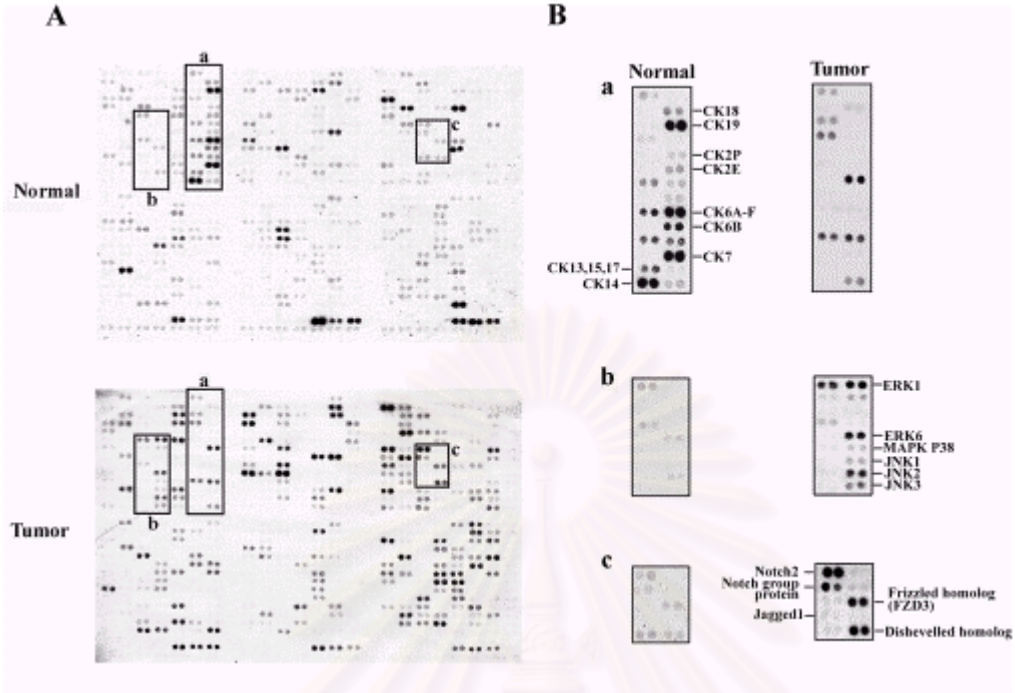


Figure 2.3 *Analysis of Gene Expression in HNSCC Using cDNA Arrays and Genes differentially expressed in HNSCC*:

For each HNSCC tissue set, cDNA probes were prepared and used simultaneously for the hybridization of nylon membranes arrayed in duplicate with human cancer and housekeeping genes, as described in Materials and Methods. The comparative level of expression for each gene was assessed by PhosphoImaging, and expressed relative to that of the house keeping genes. Pattern of gene expression for a representative tissue set from the same HNSCC patient is shown (left panel). Expression levels of genes were assessed by PhosphoImaging and compared with those of housekeeping genes**.** Differentially expressed genes in three or more HNSCC tissue sets and considered of likely biological significance are indicated (right panel).

Cell cycle		**Signaling Molecules**	
L33264	cdc-2 Related kinase	X60811	ERK1
X59768	Cyclin D1	X79483	ERK6
U11791	Cyclin H	U 82532	GDI Dissociation protein
U40345	p19INK4D	L35253	MAP kinase p38
U18422	DP-2	L26318	JNK1
		L31951	JNK2
Angiogenesis		U34819	JNK3
		U39657	MKK6
L12350	Thrombospondin2 precursor	U78876	MEKK3
U43142	VEGF-C	M31470	Ras-like protein (TC10)
X07819	MMP-7	L25080	Rho A
X07820	MMP-10		
D50477	MMP-16	**Apoptosis**	
Z30183	TIMP-3		
U76456	TIMP-4	U45878	Inhibitor of Apoptosis Protein
		U37448	Caspase 7 precursor
Growth Factors and Oncogenes		U60520	Caspase 8 precursor
		U28014	Caspase 5 precursor
M11730	HER2	M77198	AKT2
M37722	bFGFR 1	U59747	BCL-W
M37722	bFGFR 1	U78798	TRAF2
Z12020	MDM2	S83171	BAG-1
M31213	c-RET	L22474	BAX
D17517	SKY		
X87838	β-Catenin	**WNT/Notch signaling system**	
X14445	FGF 3		
M37825	FGF 5	M73980	Notch 1
X63454	FGF 6	U77493	Notch 2
M65062	IGFBP5	M99437	Notch group protein
M60718	HGF precursor	AF028593	Jagged 1
K03222	TGF-α	L37882	Frizzled
X02812	TGF-β	U82169	Frizzled homolog (FZD3)
M96956	*EGF Cripto protein CR1 and 2*	*U46461*	*Dishevelled* homolog
X06374	PDGF-a	U43148	Patched homolog
X02811	c-SIS	U94352	Manic fringe
M74088	APC	U94354	Lunatic fringe

Table 2.2 Genes Differentially Expressed in HNSCC:

Gene expression pattern for each HNSCC tissue set (WSU 1, 51, 58, 62, 63) was analyzed by PhosphorImaging and the relative amount of expression was compared with those of housekeeping genes. Those genes judged to be differentially expressed at >2 folds in at least 3 of the cancer tissue sets were considered of likely biological significance, and are listed in their corresponding functional groups. The GeneBank access number for each gene is included.

Discussion

Using currently available molecular biology techniques, recent studies have identified genes that may be involved in the pathogenesis of HNSCC (Bockmuhl *et al*., 1998; Jin *et al*., 1998; Shao *et al*., 1998). However, the heterogeneity of the tissues used for those studies may have limited the value of this body of information. Indeed, only a fraction of the total tissue volume (<5%) is suitable for this type of analysis; thus the use of bulk tissue or tissue areas of interest manually microdissected, may include abundant contamination, such as cells of lymphatic or stromal origin. In contrast, the recently developed technique of laser captured microdissection (LCM) (Emmert-Buck *et al*., 1996) allows the isolation of pure populations of cells from neoplastic lesions, thus facilitating their subsequent analysis. Furthermore, cells from their corresponding normal matched controls can be procured and subsequently used for the accurate analysis of differential gene expression in many types of cancers. This technique allows the analysis of gene expression in the cells of interest at the tissue level, thus enabling a comprehensive molecular characterization of normal, premalignant and malignant cells. Thus, further analysis of appropriate tissue samples using this recently available technology will

make it possible to define a pattern of gene expression in a tumor progression model of HNSCC. In this study, we have explored the feasibility of such a comprehensive analysis, utilizing a representative set of HNSCC tissues, comprising of tumors and the corresponding normal tissues from the same patient. The patho-physiological characteristics of tumors used in this study ranged from overt carcinomas to a lesion that was assessed histologically to be hyperplasia, thus providing means of ascertaining the pattern of gene expression in late and intermediate stages of tumor progression.

In this study, we demonstrate the successful procurement of a pure population of both tumor and normal cells from five HNSCC tissue sets. We have estimated the purity of these cells histologically to be approximately 95% and from the approximately 5000 procured cells from each tissue, we were able to isolate approximately 15 ng of total RNA of good integrity**.** High molecular weight cDNA (300-800 bp), considered to be representational of HNSCC were synthesized from these very small amounts of total RNA and are therefore suitable for use in the comparative analysis of gene expression between cells from normal and tumor tissues. However, the mRNA recovery is usually < 10 % of the total RNA and the possibility exists for that genes of moderate or low abundance, may be undetected. One solution would be to increase the amount of starting material but this is usually limited by both time and effort involved in dissecting > 5,000 cells. Furthermore, the amount of tissue available is usually small, thus limiting LCM. Sensitive techniques utilizing amplification of low amounts of mRNA have been successfully used and reported (Eberwine, 1996). These offer the advantage of analysis of intermediate stages of neoplastic processes, which usually consist of a relatively small number of cells.

However, as with any amplification-based strategies, target bias is still a concern, and is currently being addressed (Warnecke *et al*., 1997).

Here, we were able to label directly the first strand cDNAs, and using these as targets, perform hybridization analysis on membranes arrayed with known cancer genes. The single most important consideration for the selection of these commercially available membranes is that they are arrayed in duplicate with known genes of known function, the majority of which are believed to play a role in the regulation of normal cell growth, differentiation, survival, and/or carcinogenesis. Furthermore, these are all PCR-amplified, sequence-verified fragments of 200-500 bp, which were chosen for their optimized hybridization characteristics. Thus, although this approach does not allow for the discovery of novel genes specific to HNSCC, it is expected to help expedite the identification of the functional role for these genes in HNSCC pathogenesis.

In this study, we were able to demonstrate the existence of numerous changes in gene expression when comparing HNSCC to their matching normal epithelial cells. One of the most remarkable was the general decrease in the expression of cytokeratins (2E, 2P, 6A-F, 7, 13, 14, 15, 17, 18, 19), which were readily detected in normal tissue but nearly absent in the cancer cell population (1-3 fold). This reduction was expected, and most likely reflects loss of differentiation in tumor cells. On the other hand, we observed a clear increase in the levels of *cyclin D1* (2-3 fold), as previously reported (Bartkova *et al*., 1995), and those of metaloproteases (*MMP-7, MMP-10, MMP-14*), which reflects the highly invasive behavior of this tumor type. Furthermore, we observed a remarkable increase in the levels of many growth and angiogenic factors including *TGFα, TGFβ, EGF Cripto*

protein, *PDGF A chain* and *B chain* (*c-Sis*), different *FGF* isoforms, *HGF*, and *VEGF-C*. This supports the conclusion that this tumor type secretes factors that are likely to induce epithelial cell growth in an autocrine fashion in addition to promoting the growth of stromal cells and the process of neovascularization (Desai and Libutti, 1999). Furthermore, these tumors also overexpress several proapoptotic molecules, including caspase precursors, *Bcl-W*, *Bax*, and *Bag-1*, but might survive apoptotic signals through the overexpression of apoptosis inhibiting molecules, including *Akt2* and *IAP* (inhibitor of apoptosis) (Granville *et al*., 1998). Also shown in the figure 2.3 overexpressed signaling molecules participating in the MAP kinase pathway, including *ERK1*, all isoforms of *JNK* (1-3), two *p38* related MAPKs, *p38* and *ERK6*, and their upstream activators, *MEKK3* and *MKK6*, which is likely to contribute to the enhanced growth stimulation in these cells. Thus, collectively these results demonstrate that HNSCC exhibit a distinct pattern of gene expression, which might help explain many of the cellular abnormalities described in this tumor type.

Unexpectedly, genes involved in the *wnt* and *notch* signaling pathway, were found to be highly represented in tumor tissues (1-3 fold). High expression levels of some of these genes have been demonstrated in many neoplasias and may have an implication in maintaining an undifferentiated epithelium (Liu *et al*., 1996; Shelly *et al*., 1999). For *wnt*, two *wnt* receptors, *frizzled* and *FDZ3*, and their downstream targets, *dishevelled* and *β-catenin* (Wodarz and Nusse, 1998) were highly expressed. In the case of *notch*, the detection of both receptor and ligand (*notch* and *jagged*, respectively) also suggests strongly their constitutive activation (Artavanis-Tsakonas *et al*., 1999). Furthermore, two of the *fringe* genes, *Manic* and *Lunatic*, which encode pioneer secretory proteins that modulate Notch-ligand interactions (Panin *et al*., 1997)

were similarly highly represented. Thus, together these findings support an unexpected role for the *notch* signaling system in squamous cell carcinogenesis (Cooper and Bray, 1999; Thelu *et al*., 1998). However, their precise role in the pathogenesis of HNSCC is currently unknown, and warrants further investigation. Taken together, we can conclude that the use of LCM and DNA arrays has allowed the detailed analysis of gene expression in HNSCC. Although data obtained involved a limited set of tissue samples, a general trend is already observed, implicating cell cycle regulating and signaling molecules, growth and angiogenic factors, matrix degrading proteases, and survival and apoptotic molecules. Furthermore, we obtained evidence implicating, for the first time, the *notch* and *wnt* pathways in squamous cell carcinogenesis. We are currently assessing the validity of these observations in a panel of HNSCC tissue sets using available antisera and cDNA probes. With the imminent completion of the Human Genome Project and recent advances in array technology (micro arrays and filter arrays), it is now possible to study and better understand the expression pattern of those genes that may be causal in neoplasia as they exist *in vivo*. With this strategy, using samples representing each stage during tumor progression, it would be possible to ascribe a function in HNSCC to genes whose role in carcinogenesis might have not been suspected. Furthermore, this experimental approach will facilitate the identification of earlier markers heralding malignancy, thus providing valuable tools of diagnostic and prognostic value to study premalignant lesions.

CHAPTER III

Gene Expression Profiles in Squamous Cell Carcinomas of the Oral Cavity: Use of Laser Capture Microdissection for the Construction and Analysis of Stage-specific cDNA Libraries

Introduction

From the estimated 100,000 genes in the human genome, 4,000 of these may be directly related to disease, including cancer (Trent *et al*., 1997). Indeed, altered expression of some of these genes is now thought to be the basis of most neoplasias, either because they are expressed at abnormally high or low levels, or due to their ability to encode aberrant proteins upon mutations in their coding sequence (Urbain, 1999). In this regard, the availability of a catalog of genes expressed in tumor cells may provide a fingerprint of their genetic make up, and comparison with that of their matching cells exhibiting a normal phenotype can help identify genes that either by their presence or absence, can be causal in cancer. It follows that knowing the identity of these genes will not only enhance our understanding of the molecular basis of this disease and its progression, but it will also provide novel means for its early detection and subsequent treatment.

In response to our limited knowledge of the molecular mechanisms of many neoplasias, the Cancer Genome Anatomy Project (CGAP) supported by the National Cancer Institute (NCI), was established with the goal of creating a complete information infrastructure of genes expressed during tumor progression, which is also expected to yield early markers of cancer, thus providing an opportunity to improve

our ability to match patients with appropriate treatment strategies. The CGAP initiative involves the generation of cDNA libraries from cancer cells, and after random sequencing, expressed genes are then catalogued and compared with those from the corresponding normal tissues. In doing so, CGAP has also become the leading effort in gene discovery. Further success of this approach has been the development of robust databases and easily accessible Web-based analytical tools for comparative use (Strausberg *et al*., 2000; Strausberg *et al*., 1997).

Squamous cell carcinoma of the head and neck (HNSCC) are neoplastic lesions found predominantly in the oral cavity, including the salivary glands, larynx and pharynx (Boring *et al*., 1993). Despite recent advances in our understanding, prevention, and treatment of other types of neoplasias, HNSCC still remains the 6th most common cancer among men in the developed world (Parkin *et al*., 1999) and in the United States alone approximately 13,000 deaths occur yearly as a result of this disease (Landis *et al*., 1999). The high morbidity rate for this malignancy can be attributed to many factors, which include lack of suitable markers for early detection, late presentation, insensitivity to available treatment, and our limited understanding of the molecular mechanisms responsible for this disease (Califano *et al*., 1996). In this regard, the identity of those genes that may have a role in the progression of HNSCC has yet to be fully elucidated. Therefore, in an attempt to begin addressing the molecular basis of this cancer, the Head and Neck CGAP (HNCGAP) was established as a cooperative effort between the National Institute of Dental and Craniofacial Research (NIDCR) and the National Cancer Institutes CGAP initiative.

A major scientific challenge in HNSCC is our understanding of the molecular events that drive tumor progression *in vivo* (Bockmuhl *et al*., 1998). This

problem is further compounded by the heterogeneity of this tumor type. Thus, gene expression analysis using bulk tissue or tissue areas of interest manually microdissected, might not be representational and of limited value when using this body of information for assessing gene expression profiles in HNSCC. Of interest, the use of Laser Capture Microdissection (LCM) (Bohm *et al*., 1997; Emmert-Buck *et al*., 1996; Simone *et al*., 1998) allows the procurement of pure cell populations for RNA isolation, thus providing an appropriate platform for current efforts in defining the nature of those genes expressed in HNSCC, and their potential contribution to neoplasia (Krizman *et al*., 1996; Zhang *et al*., 1997).

In this study we have used HNSCC and their matching normal tissues from patients with oral cancer lesions. We demonstrate the successful use of LCM to procure specific cell populations. Furthermore, we show that 5,000 cells are sufficient to extract RNA of high integrity for the synthesis of high quality representational cDNAs libraries. Furthermore, sequence analysis of randomly selected clones from each library indicates that 76-96% of the inserts represented anonymous ESTs (25-48%), known genes (9-29%) or novel sequences (27-51%), respectively, and with very little redundancy among libraries. Emerging sequence information suggests the existence of many novel genes, whose function in tumor development can now begin to be evaluated.

Materials and Methods

Tissue samples and Laser Capture Microdissection

Biopsies from patients confirmed to have carcinomas of the oral cavity were immediately fixed in 70% ethanol and subsequently embedded in OTC as described

(http://dir.nichd.nih.gov/lcm). Using a cryostat, 8 μm thick tissue sections were cut onto RNAase free glass slides, and prior to laser capture microdissection (LCM), H&E stained sections were analyzed and confirmed by a board certified pathologist. The use of LCM (Arcturus Engineering, Mountain View, CA) was essentially as described (Simone *et al*., 1998). This procedure enables the enrichment of a pure cell population onto plastic caps coated with a transparent ethylene vinyl acetate (EVA) thermoplastic film. Microdissected cells were assessed microscopically and the caps containing the procured cells were subsequently transferred to a 0.5 ml RNAase free microcentrifuge tube containing appropriate lysis buffer for extraction of total RNA.

Extraction of RNA and Assessment of RNA Integrity

The procedure of extracting RNA from microdissected cells is described in detail elsewhere (http://dir.nichd.nih.gov/lcm). Briefly, cells were digested in lysis buffer containing guanidium isothiocyanate and β-mercaptoethanol and total RNA extracted from the supernatant with 2M sodium acetate, saturated phenol and chloroform-isoamyl alcohol. The RNA was subsequently precipitated from the resulting aqueous layer with ice-cold isopropanol and glycogen (10 μg/μl). The resulting RNA pellet was treated with DNAase I (10 units/μl) and RNAase inhibitor (20 units/μl), reprecipitated and resuspended in 3.5 μl deionized water and 1 μl of 20 U/μl RNAase inhibitor. The integrity of each RNA sample was assessed as described. Briefly, 1 μl of RNA was reverse transcribed using the GeneAmp RNA PCR kit (Perkin Elmer, Branchburg, NJ) following the supplied protocol, and 2 μl of the resulting cDNA was used to amplify *GAPDH* with specific primers using PCR

conditions according to the manufacturer's recommendations (Perkin Elmer).

Quantification of Total RNA Extracted from Microdissected cells.

Total RNA extracted from microdissected cells was quantified using the VersaFluor^{TM} Fluorometer system following the manufacturer's recommendations (Bio-Rad Laboratories, Hercules, CA). Briefly, 1 ml of TE (1X) containing 1 μl of total RNA was mixed with an equal volume of diluted Ribogreen probe (1:3000 in 1X TE). After incubation (5 min), the samples were measured for RNA using excitation and emission wavelength of 495 and 525 nm, respectively. The amount of total RNA was quantified against standards (1-50 ng), which were prepared according to the manufacturer's recommendations.

Synthesis of cDNAs from total RNA and library Construction

Synthesis of double-strand cDNAs from total RNA extracted from microdissected cells was as previously described (Krizman *et al*., 1996). Briefly, using the Superscript Choice System (Life Technologies Inc. Gaithersburg MD) and following the supplied protocol, RNA samples from microdissected tissues (5,000 cells) were used as templates to reverse transcribe first strand cDNAs. The second-strand replacement and *Eco*RI linker addition reactions were followed as described (Krizman *et al*., 1996). Briefly, RT reactions containing 5X second strand buffer, 10 mM dNTPs, *E. Coli* DNA ligase, *E. Coli* DNA Pol I, and *E. Coli* Rnase H were incubated at 16 $^{\circ}$C for 2 h and after the addition of 2 μl of T4 DNA polymerase, the reactions were incubated for a further 10 min. After extracting and precipitating with phenol/chloroform and ethanol, respectively, the cDNA pellets were resuspended in

ddH_2O and used to ligate *Eco*RI linkers. The linker-ligated cDNAs were purified in 1% low-melting agarose and cDNA products (0.4-2 Kb) were excised and the agarose was digested overnight at 37 °C with β-agarase (New England Biolabs, Beverly, MA). After extracting with phenol/chloroform and precipitating with ethanol and glycogen (20 μg /ml) respectively, the cDNA pellets were washed with 70% ethanol, air dried and resuspended in 20 μl ddH_2O. 2 μl of linker-ligated cDNAs was used to assess the quality on a 1.2 % agarose gel.

PCR Amplification of double strand cDNA

Amplification of double strand cDNAs by PCR was essentially as described (Krizman *et al*., 1996). Briefly, 18 μl cDNA of template, ddH_2O, 10X PCR buffer, 10mM dNTPs, 1 μM linker-specific primers and *Taq* polymerase, was used for a single PCR reaction. Conditions for PCR were 3 min at 94 °C followed by 15 cycles at 94 °C for 15 sec, 65 °C for 15 sec, 72 °C for 3 min and a final extension at 72 °C for 5 min. The PCR products were purified, isopropanol precipitated and washed with 70% ethanol prior to resuspending in 10 μl ddH_2O.

Subcloning and Assessment of cDNA libraries

One μl of each of the amplified cDNAs was non-directionally cloned into the UDG (uracil deglycosylase) cloning vector pAMP10 (Life Technologies) and the cloning reactions were subsequently used for transformation. Resulting clones were randomly picked for assessing insert size and sequencing. Universal primers (M13f and M13r) for pAMP10 were used to amplify inserts with 30 cycles of standard PCR

reaction and these reactions subsequently analyzed for diversity by assessing the insert size on a 1.2 % agarose gel.

Sequencing of cDNA Libraries and Data Analysis

Recombinant clones from each library were randomly picked, expanded in 96 well plates, and sequenced using M13 forward primers. Individual clone sequences were subsequently analyzed by searching available data bases (GenBank, dbEST) by the BLAST program accessed through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov), for the search of homology to known genes or ESTs.

Results

Scheme of Experimental Procedure

The experimental strategy for this study is illustrated in fig 3.1. Normal and pathological oral squamous epithelium was visualized under the microscope and appropriate cells were microdissected with individual laser shots. Caps containing approximately 5,000 cells were processed for RNA and subsequently assessed for quality. The mRNA served as a template for library production. After transformation, clones from each cDNA library were sequence analyzed and prepared for CGAP submission. The sequencing data were analyzed using the BLAST software to assess whether individual clones matched previously identified genes or anonymous ESTs, or represented novel sequences.

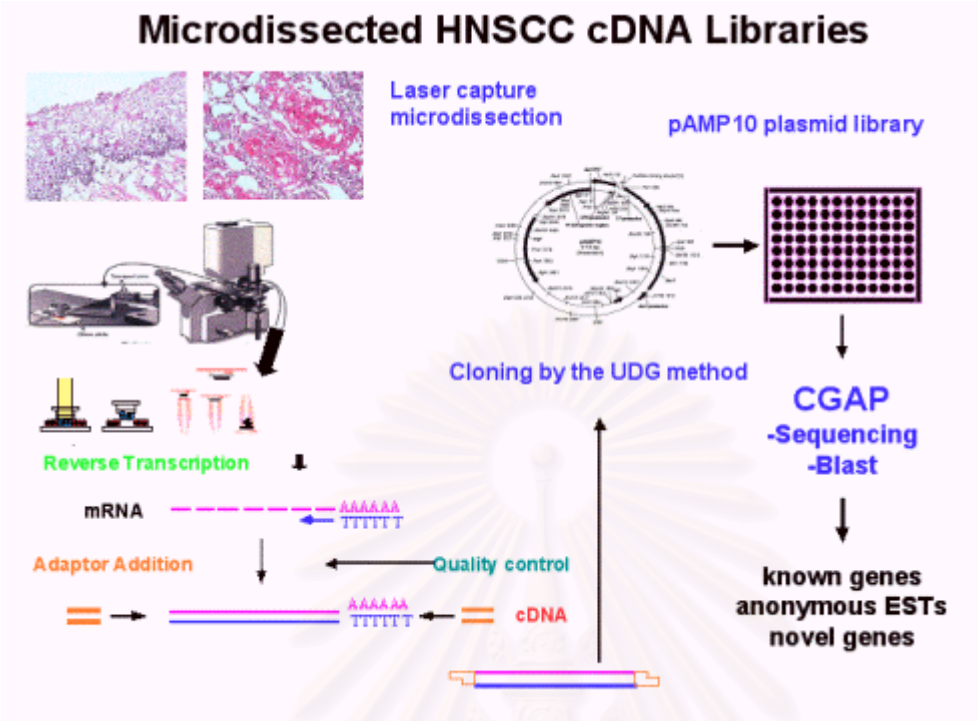


Figure 3.1. *Scheme Illustrating the Experimental Procedure used in the Study*: Normal and malignant oral keratinocytes are procured by laser capture microdissection and the RNA extracted. After assessment the quality, the RNA samples are reverse transcribed and with the resulting cDNAs, adaptors are ligated to enable cloning. Clones from each cDNA library are sequence analyzed and prepared for CGAP. The available data are further analyzed using the BLAST program to determine whether any of the information matches to anonymous ESTs, known genes or represent novel genes.

Clinical and Histological Feature of HNSCC tissues samples

Clinical samples were obtained from HNSCC patients that had undergone surgery. Samples used for this study were from male patients and the age ranged from 46-71 years upon presentation of the lesions. The main clinical features of these lesions as diagnosed by a certified board pathologist are summarized in table 3.1. However, before proceeding, we confirmed the pathophysiology of these tissue samples. Their anatomical sites are representative of that of the most frequently detected neoplastic lesions, and include the retromolar trigone region (S1), floor of the mouth (S2) and tongue (WSU 62). The corresponding lesions were carcinoma *in situ* (CIS), well-differentiated invasive carcinoma and moderated to poorly differentiated carcinoma, respectively. For each tumor sample, adjacent normal tissue comprising squamous epithelium was also provided and in all cases the histopathology was confirmed. Frozen tissue sections (8 μm) stained with H&E are illustrated in fig 3.2. Normal squamous epithelium S1 (A), S2 (B) and WSU 62 (C) that match to the tumor samples, showed an orderly architecture that ranges from immature small round cells in the basal layer to those of mature flattened cells with abundant cytoplasm and small nuclei at the superficial layer (X500). In all cases, the basement membrane was observed to be intact. In contrast, tumor sample S1 (D), a carcinoma *in situ*, showed a variation in size and shape of cells, nuclear pleomorphism, hyperchromasia, and loss of normal cellular maturation (X500). Tumor sample S2 (E), a well-differentiated squamous cell carcinoma, showed infiltration of the underlying stroma with sheets and islands of cancer cells displaying nuclear pleomorphism and hyperchromasia. Furthermore, occasional keratin pearls were observed (X500). Tumor sample WSU62 (E), a moderate to poorly differentiated invasive carcinoma, showed infiltration of

underlying stroma as sheets and islands of cancer cells showing marked nuclear pleomorphism and hyperchromasia. In this sample, there is distinct absence of keratin pearls (X500). The data illustrates that anatomical and histological differences in tumor samples make them suitable for their use in identifying those genes that may be critical in tumor development.

Sample	Age	Sex	Origin	Pathology
S1	65	M	Retromolar trigone	Carcinoma *in situ*
S2	71	M	Floor of mouth	Well differentiated invasive
WSU 62	46	M	Tongue	Moderate to poorly Differentiated invasive

Table 3.1 *Clinical characteristics of lesions from patients with HNSCC*:

Tissues, including adjacent normal mucosa, were surgically removed from patients with oral carcinomas for clinical classification. Anatomical site and severity of the carcinomas, S1, S2 and WSU62 are indicated. In all cases, corresponding normal tissues were confirmed as squamous epithelium.

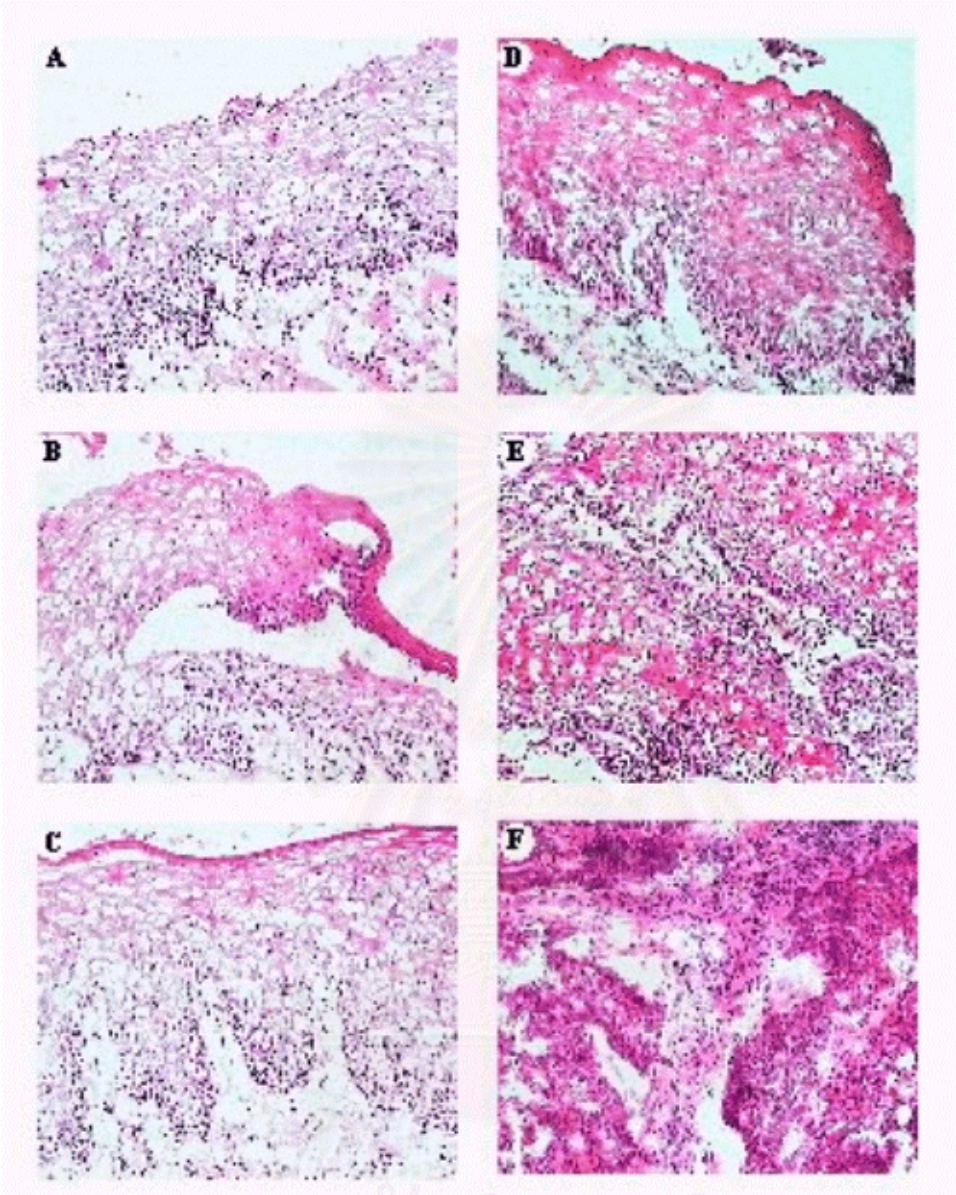


Figure 3.2. *Histopathological Features of HNSCC*:

Tissue sets, comprising of both normal and pathological lesion from the same HNSCC patient, were snap frozen and 8 μm thick sections were stained with H&E. Histopathological features of progression of HNSCC from normal to carcinoma are demonstrated. Representative normal squamous epithelium (A, B, C) from the oral cavity shows an orderly maturation with progressive flattening of the cells going from

deep to superficial cell layers, uniform small nuclei, and low nuclear:cytoplasmic ratio (magnification X500). Carcinoma *in situ* (D) shows a variation in size and shape of cells, nuclear pleomorphism, hyperchromasia and a loss of normal cellular maturation is observed. There is no breach of the basement membrane (magnification X500). The well-differentiated squamous cell carcinoma (E) shows infiltration of the underlying stroma with sheets and islands of cancer cells showing nuclear pleomorphism and hyperchromasia. Occasional keratin pearls are also observed (magnification X500). Moderate to poorly differentiated invasive carcinoma (F) shows infiltration of the underlying stroma as sheets of cancer cells showing marked nuclear pleomorphism and hyperchromasia. However, a notable absence of keratin pearls is noted (magnification X500).

Total RNA extractable from Microdissected tissues

Before proceeding with the construction of representational cDNAs from the tissue sets, we sought to determine the amount and integrity of total RNA that was extractable from approximately 5,000 microdissected cells. As shown in table 3.2, the average amount of total RNA that was quantifiable using the fluorometric system, was 16.7 ng and this was demonstrated to be similar (14.7-18.6 ng) among samples. In addition, when 2,000 and 10,000 cells were microdissected from each tissue sample, the average amount of total RNA extracted was 7.9 ng and 21.4 ng respectively (data not shown).

Tissue	S1		S2		WSU 62		
	N	T	N	T	N	T	Average
RNA (ng)	18.6	15.6	15.9	17.3	17.8	14.7	16.7

Table 3.2. *Quantification of Extractable RNA from Microdissected Tissues*:
RNA extracted from 5,000 cells microdissected from the three tissue sets (S1, S2, WSU 62) was labeled with a fluorescent dye and quantified using appropriate standards and using a VersaFluor^{TM} Fluorometer. The amount of total RNA extractable for normal (N) and tumor (T) epithelium, using the experimental conditions described in Materials and Methods and the average for all six samples is indicated. The data are representative of three separate experiments.

The integrity of the RNA in all cases was assessed by RT-PCR of GAPDH. As demonstrated in fig 3.3, RNA extracted from all tissue sets (S1, S2, WSU 62) was considered to be of sufficient quality to be reverse transcribed and amplified using specific primers, which yielded a 600 bp PCR product of GAPDH. The data demonstrate that small amounts (ng) of total RNA extracted from microdissected cells are of sufficient integrity to synthesize and construct representational cDNA libraries from oral carcinomas and matching normal epithelium.

Quality Assessment of Amplified double strand cDNAs

As the integrity of total RNA was maintained, we next determined whether this would be reflected in the synthesis of high quality cDNA libraries. Double strand cDNA product of each tissue sample was amplified by PCR and the quality determined on an agarose gel. A homogenous smear (200-1,500 bp) was observed for all samples, indicating sufficient complexity of transcripts for library construction (data not shown). Amplified cDNAs were purified, subcloned, and randomly picked clones (twelve) from each library were analyzed to determine the average insert size. In all cases, the average size was observed to be around 600 bp, with a range of 300-1,500 bp (data not shown). The data indicate a complexity of cDNAs that may be representational of the genes expressed in oral epithelium.

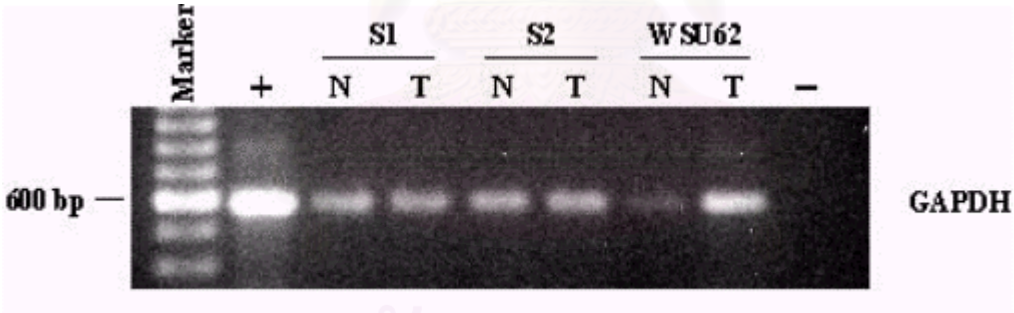


Figure 3.3. *Assessment of RNA Integrity by RT-PCR*:
Total RNA was extracted from microdissected tissues (approximately 5,000 cells) as described in Materials and Methods. RNA integrity was assessed by RT-PCR of the *GAPDH* gene. Using specific primers for *GAPDH*, a 600 bp fragment is amplified, as observed for all the tissue samples. Appropriate positive and negative controls are indicated. Data shown is representative of four independent experiments.

Sequence Analysis of cDNA libraries

After quality control, 96 clones were randomly picked from each of the six cDNA libraries, expanded, sequenced and their nucleotide sequences analyzed using the BLAST program. The cumulative information is summarized in table 3.3. Analysis of the nucleotide information obtained from the successful sequencing of 76-96% of the inserts indicated that these DNAs represent either anonymous ESTs (25-48%), known genes (9-29%) or novel transcripts (27-51%), respectively. Both *Alu* repeats and ribosomal RNA were detected at low frequencies in all six libraries (2-9%) and no sequences of mitochondrial or bacterial origin were recorded. Approximately 7% of the total number of clones analyzed were either without insert or contained short sequences that were uninformative. The data indicate the successful cloning of a large number of novel transcripts that are likely to be representative of those expressed in normal and malignant oral epithelium.

CDNA library	S1		S2		WSU62	
	N	T	N	T	N	T
Clones Analyzed	92	92	73	84	82	90
Known genes	8	27	22	11	9	14
Anonymous ESTs	44	35	18	25	42	41
Novel sequences	37	24	30	43	22	30
Expressed *Alu* repeats	3	4	3	2	7	4
Ribosomal RNA	0	2	0	3	2	1

Table 3.3*. Cumulative sequence analysis of HNSCC-specific cDNA clones*:
Random clones (96) from libraries constructed from normal (N) and tumor (T) epithelium were sequenced and this information was used to search available databases, using the BLAST program. Clones from the six libraries were subsequently categorized as having sequences similar to known genes, those that matched to ESTs and those that showed no significant homology. Frequency of *Alu* repeats and Ribosomal RNA sequences are indicated.

Identification of known genes in cDNA libraries from Oral Epithelium

We next determined, by using the BLAST program, whether any of these sequences matched known genes. Those that were identified as being expressed in any of the six libraries are listed in table 3.4. Of interest, known genes identified more than once in a library include calgranulin A, kappa casein precursor, elafin precursor,

monocyte chemotactic protein 3 precursor, cyclin I, keratin 13, keratin 56 kD, cornifin B and the interleukin-1 receptor antagonist protein precursor. Collectively, of the known genes identified, six were represented in two or more of the libraries and are listed in table 3.5. Kappa casein precursor was common in all the libraries and monocyte chemotactic protein 3 precursor was represented in libraries made from all tumor tissues and one normal (62N). The keratin 4 was detected in two normal libraries (S1, S2) whereas the keratin 13, calgranulin A (calcium binding protein A8) and cornifin B (small proline rich protein, SPRR 1B) were identified in 2 of the libraries.

14-3-3 protein σ
Acyl-CoA-binding protein
Annexin I
Antigen peptide transporter2 (Transporter, ATP-binding cassette)
Antileukoproteinase 1
ATP synthase A chain
Breast basic conserved protein 1 (Breast basic conserved gene 1)
Calgranulin A (calcium binding protein A8)
Calgranulin B (calcium binding protein A9)
Calpain 2 large
CD9 antigen
Cds 1 human SPR2-1 gene for small proline rich protein
Cornifin B (small proline rich protein, SPRR 1B)
Cyclin I
Cystatin A
Cytochrome C oxidase polypeptide via liver, 2, 3
Dolichyl-phosphate β-glucosyltransferase
E25B protein mRNA
Elafin precursor
Erythrocyte adducin α subunit
Erythrocyte adducin β subunit
Ferritin heavy chain
GAP junction β-2 protein (Connexin 26)
Growth relating protein BB1
Histocompatibility antigen, M alpha chain
Histone H3.3
Interferon γ
Interleukin-1 receptor antagonist protein precursor
Kappa casein precursor
Keratin, type I cytoskeletal 13
Keratin, type II cytoskeletal 4, 5, 7, 56kD
L-3 phosphoserine-phosphatase homologue
Leukocyte elastase inhibitor
Mitotic kinesin-like protein-1
Monocyte chemotactic protein 3 precursor
Myosin like chain Alkali smooth muscle Isoform
NADH-Ubiquinone oxidoreductase chain 1 and 4
Protein phosphatase PP1-γ
Ribosomal protein S11, S20, S18, S24, L3, L8, L37, L38
Signal recognition particle receptor β subunit
Transcriptional coactivator

Table 3.4 *Known genes identified from six HNSCC-specific libraries:*

Sequence analysis of random clones from cDNA libraries constructed from normal (N) and tumor epithelium (T) identified many as having homology to known genes. Some of these known genes were identified more than once in one or more of the six libraries. The 55 known genes that were identified are listed by name alphabetically.

Kappa casein Precursor	S1 (N, T), S2 (N, T), WSU 62 (N, T)
Monocyte chemotactic protein 3 precursor	ST1, ST2, WSU 62 (N, T)
Keratin, type II cytoskeletal 4	SN1, SN2
Keratin, type I cytoskeletal 13	SN2, ST1
Calgranulin A	S2 (N, T)
Cornifin B	SN2, ST1

Table 3.5. *Known genes identified in two or more HNSCC-specific cDNA libraries*: Six known genes were identified in two or more HNSCC-specific cDNA libraries (N; normal, T; tumor) and listed. Only one known gene was present in all 6 libraries analyzed.

Discussion

Approximately 10 % of the total number of genes are suspected to be expressed in a given cell type (Lisitsyn and Wigler, 1993). Determining their identity is an important first step towards understanding the patterns of gene expression that mediate normal cellular physiology and disease process. In this study, we report the construction of six high quality cDNA libraries from tissues of oral origin, including normal and malignant epithelium. Previous studies have reported on genes that are expressed in tissue specimens containing squamous epithelium (van der Velden *et al*., 1999). However, these studies utilized heterogenous tissues as starting material, thus the expression data may not reflect the genes that are specifically active in the epithelium. In contrast, for the present investigation, we procured pure populations of cells (approximately 95% purity) from normal and malignant epithelium using laser capture microdissection (LCM).

A molecular fingerprint of every expressed gene in each cell type is its mRNA, but this usually constitutes approximately 5% of the total RNA (Eberwine, 1996). Therefore, the quantity and integrity of RNA extracted from the starting material are important considerations to ensure that the gene expression analysis is representational. In addition, procedures involved in tissue isolation using LCM may themselves limit this type of analysis by facilitating RNA degradation. However, from approximately 5,000 cells procured from each tissue sample by LCM, we were able to demonstrate that the RNA extracted was quantifiable and of good integrity as assessed by the ability to amplify GAPDH.

cDNA libraries were prepared from three patient tissue sets, comprised of normal and tumor oral epithelium. Our primary goal from this study was to identify

those genes whose expression may be solely confined to the malignant oral epithelium and these may be causal in the development of HNSCC. In addition, we wanted to use these high quality cDNA libraries to contribute to the HNCGAP, in their gene discovery efforts and those of systematically cataloguing cancer genes from different neoplasias. We and others have previously contributed to the HNCGAP initiative (http://www.ncbi.nlm.nih.gov). However, cDNA libraries for these efforts were either constructed from bulk tissues or from representative HNSCC cell lines and from normal and immortalized gingival keratinocytes (Shillitoe *et al*., 2000). To date, this report is the first describing the construction of cDNA libraries from microdissected oral epithelium that may represent more accurately those genes expressed *in vivo*.

From the first round of analysis, sequence information was obtained from approximately 76-96% of clones analyzed, which included anonymous ESTs and known genes. Furthermore, these six libraries collectively demonstrated a remarkable sequence novelty, with most transcripts likely representing unknown genes. However, it is still possible that these novel sequences may be found in one or more of the libraries. An important parameter indicating the quality of cDNA libraries is the presence of contaminating sequences. In this regard, no bacterial or vector sequences were detected, while a very low frequency (0-7.3%) of *Alu* repeats and rRNA sequences were present. Thus, the high percentage of quality sequencing suggests these libraries are representative with a concomitant maintenance of complexity.

The high incidence of detection of novel sequences in only the first round of analysis is in itself, remarkable and surpasses the figure (4% per library) set out by the gene discovery efforts of CGAP. Additional sequence analysis of these libraries is likely to identify many new transcripts. These yet to be identified transcripts may

indeed be uniquely expressed in oral epithelium. It therefore follows that the expression pattern and the function of these genes may help to identify gene products involved in the transformed and/or the metastatic phenotype, as well as additional molecules that, without playing an obvious role in the neoplastic process, can nevertheless be used as clinically useful markers of tumor development.

In this regard, the 55 known genes identified from this first round of sequence analysis is in its self interesting, but whether they may play a role in HNSCC tumor development is not documented. Of particular interest was the sequence match to the monocyte chemotactic protein 3 precursor gene (MCP3), which was readily detected in all three cDNA libraries from tumor tissues (ST1, ST2, 62T). MCP3 is a chemokine known to induce the production of gelatinase B and chemotaxis of monocytes. In addition, MCP3 is also known to be produced by tumor cells and its expression in HNSCC may play a role in tumor progression (Mantovani *et al*., 1992; Opdenakker *et al*., 1994; Opdenakker *et al*., 1993; Opdenakker and Van Damme, 1992). Additional known genes, while not causal to tumor development, may be useful makers. These include cornifin B, also known as small proline rich protein 1 (SPRR1). Its gene product has been reported expressed in the sublingual and tongue epithelium and in malignant oral epithelium. While its precise function is unclear it may be involved in the terminal differentiation status of keratinocytes (Fischer *et al*., 1998; Fujimoto *et al*., 1997; Hohl *et al*., 1995). Its use as a marker of tumor progression has been highlighted in a recent report (Nacht *et al*., 1999). Also included is calgranulin A, which has a functional role in epithelium differentiation and whose expression has been reported in a restricted subset of normal stratified squamous epithelium of the tongue and buccal mucosa (Wilkinson *et al*., 1988). In this regard,

the expression of cytokeratins (4, 5 and 13) can be used as makers of epithelial differentiation and altered expression of these proteins have been reported in oral malignancies (Vaidya *et al*., 1996). Sequence match to kappa casein precursor and many ribosomal proteins (S11, S20, S24, L3, L8, L38, L37) were readily detected but whether they have any role in tumor development is currently unclear.

Although data presented in this report involves a limited set of tissue samples, a high frequency of novel transcripts have been already identified from the first round of sequence analysis, thus suggesting the presence of many unknown genes that may play a critical role in the biology of oral epithelium. Indeed, by subsequent sequencing of these cDNA libraries HNCGAP has already identified the additional 57 novel transcripts (http://www.ncbi.nlm.nih.gov/ncicgap). A detailed analysis of the genes identified by the HNCGAP effort in these and additional cDNA libraries constructed from tissues of oral origin, including different sites and stages of malignancies, is already in progress and will be reported shortly. Thus, we can conclude that the construction of representational cDNA libraries form normal and neoplastic oral tissues has resulted in the availability of cDNAs for many novel genes which are expressed in normal oral epithelium and in HNSCCs. These efforts, together with the use of gene array technologies and LCM will soon make it possible to define a pattern of gene expression in a tumor progression model of HNSCC. This experimental approach is also expected to facilitate the identification of earlier markers heralding malignancy, thus providing valuable tools of diagnostic and prognostic value to study premalignant lesions.

discovery of genes that are directly involved in tumor development, thus potentially providing the means for early detection.

A large amount of gene expression data, including ESTs (expressed sequence tags) and partial and full-length sequences has been recently generated by a number of public and private organizations by the high throughput sequencing of cDNA libraries and genomic fragments. The Cancer Genome Anatomy Project (CGAP), whose primary goal has been to compile a complete catalogue of all genes expressed in different tumor cells, can typify this approach. This effort is reliant on the analysis of RNA isolated from a small number of isolated normal and tumor cells, and the development of efficient methods to generate cDNA libraries that truly reflect their expression pattern, as they exist *in vivo* (Krizman *et al*., 1996; Krizman *et al*., 1999; Peterson *et al*., 1998). The available sequence information derived from these representational cDNAs libraries, is aimed primarily to provide an annotated tumor index that can lead to the identification of tumor specific genes, specific molecular targets for each step of tumor progression, and for their use in prevention and treatment strategies (Strausberg *et al*., 2000; Strausberg *et al*., 1997). Key to the success of this effort has been the development of robust databases and easily accessible Web-based analytical tools (http://www.ncbi.nlm.nih.gov/cgap), which has generated a wealth of information freely available to the scientific community.

The CGAP effort has been primarily focused on the five most prominent tumor types: breast, colon, lung, prostate, and ovarian. Other cancers, such as squamous cell carcinoma of the head and neck (HNSCC), are also characterized by high morbidity rates, affecting more than 500,000 new patients each year worldwide (Landis *et al*., 1999). The 5 year survival rate of this cancer type, ~ 50% (Parkin *et*

al., 1999), is relatively low, and its poor prognosis likely reflects disease heterogeneity, the absence of any notable precursor lesion, and our limited understanding about the molecular basis for this malignancy. To address these deficits, the Head and Neck CGAP (HNCGAP) was established as a joint venture between the National Institute of Dental and Craniofacial Research (NIDCR) and the National Cancer Institutes CGAP initiative. The rational was to generate sequence information initially from representational cDNA libraries derived from HNSCC cell lines, primary oral keratinocytes and E6/E7 of HPV immortalized oral keratinocytes (Shillitoe *et al*., 2000), to gain information on the nature of those genes expressed in this particular cell type. However, cells in culture often exhibit genetic mutations similar to those of the tumor of origin, but the molecules expressed are likely to be quite different from those present in their natural setting, and highly dependent on the particular culturing conditions. Thus, we next sought to obtain sequence information from high complexity cDNA libraries derived from laser capture microdissected (LCM) HNSCC tissue (normal and tumor tissue resected from HNSCC patient), to begin defining more accurately, those genes that are expressed in this cancer type and their molecular changes as they occur during tumor progression *in vivo*. Six high quality LCM derived cDNA libraries has been thus far generated. Their initial sequence analysis revealed that their gene discover rate is relatively high, 4-7%, thus suggesting that many transcripts represent unknown genes which may indeed be uniquely expressed in oral epithelium. Furthermore, it is expected that expression pattern and function of these genes may help to identify gene products that may have a role in HNSCC pathogenesis (Leethanakul *et al*., 2000).

In this study we report the generation of extensive sequence information from cDNA libraries derived from LCM-procured normal and cancerous squamous epithelium. Furthermore, with the use of the extensive CGAP database and bioinformatics tools, we surveyed this sequence information comparing the data generated from libraries representing normal oral epithelium and different stages of HNSCC. The list of genes compiled from this work provides a first glimpse at the pattern of gene expressed in HNSCC.

Materials and Methods

Construction of Representational non-normalized cDNA Libraries from HNSCC Cells

The construction of cDNA libraries from HNSCC cell lines, primary oral keratinocytes, and E6/E7 of HPV immortalized oral keratinocytes and their inclusion in the CGAP has already been reported (Shillitoe *et al*., 2000). In order to gain information on those novel as well as known genes expressed *in vivo,* we chose to use laser-assisted microdissection as the platform for the construction of representational cDNA libraries. Briefly, RNA isolate from LCM derived cells from normal and tumor epithelium resected from HNSCC patients, was subsequently used for the construction of non-normalized cDNA libraries as outlined in the Krizman's protocol 1 (http://www.ncbi.nlm.nih.gov/cgap). Total RNA (14.7-18.6 ng) was then used as template to reverse transcribe the first strand cDNAs using oligo-dT priming, followed by second-strand replacement and *Eco*RI linker addition. The resulting linker-ligated cDNAs were assessed for quality on a 1.2 % agarose gel before amplification by PCR using LINK-CUA primers designed specifically for the *Eco*RI linkers (5′ CUACUACUACUAAATTCGCGGCCGCGTCGAC 3′) and cloned non-

directionally into the UDG (uracil deglycosylase) cloning vector pAMP10 (BRL). The resulting HNSCC cDNA libraries were initially assessed for quality assurance prior to submission to CGAP for nucleotide sequencing and bioinformatics. Briefly, 1 μl of each of the library was used for transformation and recombinant clones from each library were randomly picked, expanded in 96 well plates, and sequenced using universal primers (M13f and M13r). Sequence information was used to assess the overall diversity of the clones by searching available databases (GenBank, dbEST) using BLAST (Basic Local Alignment Search Tool) program accessed through the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST), for homology to known genes. For additional confirmation of diversity, inserts were amplified for 30 cycles by standard PCR reactions and assessed for size on an agarose gel.

Sequence Analysis of HNSCC cDNA Libraries.

After in house quality assurance, the cDNA libraries were transferred to the Lawrence Livermore National Laboratories (LLNL, CA United States), for preparation for sequencing and bioinformatic analysis prior to submission to the CGAP database. All procedures at LLNL were performed using automated working station, and cDNA libraries upon receipt were transformed and colonies subsequently arrayed in 384 well plates. Replicate plates were made and sent to the Washington University Human Genome Center (St Louis, MO), for nucleotide sequencing.

The human Tumor Gene Index and Data Analysis

Nucleotide sequence information generated from the six high quality cDNA libraries from normal and tumor tissue samples is publicly available by accessing the CGAP database (http://www.ncbi.nlm.nih.gov/CGAP/hTGI) and currently forms the basis of a comprehensive human tumor gene index (hTGI). The goal of hTGI is primarily to identify genes expressed during development of human tumors to discover new human genes. Furthermore, specialized web-based analytical tools, such as XProfiler, can be interfaced with the CGAP database following the available instructions, to simplify analysis of various data sets (http://cgap.nci.nih.gov/Tools). These sites are expected to provide all the necessary information online, enabling investigators anywhere in the world to perform analysis that may help identify those genes differentially expressed during development of human tumors and additionally aid in the discovery of cancer genes not previously described.

For the analysis of HNSCC cDNA libraries, the tool browser, XProfiler (http://cgap.nci.nih.gov/Tools) was used. The libraries were chosen on the bases of tissue (normal and tumor), different stages of tumor development and sample preparation (bulk and LCM derived RNA) and XProfiler then provided a list of genes that from statistical analysis, may be differentially expressed in the tissue of origin. This list includes genes that may be either unique or non-unique, known or unknown genes, in either of the two libraries under analysis, for instance normal and tumor. Criteria used to determine whether a sequence (gene/EST) is unique or not are based on whether it has been previously detected in other CGAP cDNA libraries. Genes were considered as “unknown” if no match was found in available databases (GenBank). Furthermore, the researcher is also able to obtain more information of

each cDNA clone from each library, for instance: the average insert size, type of vector, and protocol, including tissue of origin.

Results

LCM Derived cDNA Libraries from HNSCC Tissues.

For the construction of cDNA libraries we chose biopsies from HNSCC patients that had undergone surgery, which were received following appropriate procedures. Where possible the inclusion of normal epithelium whether adjacent or distant was included. Clinical features of these samples include the anatomical sites, which ranged from the retromolar trigone region, floor of the mouth and the tongue, therefore, considered representative of that of the most frequently detected HNSCC neoplastic lesions. The actual lesions were carcinoma *in situ* (CIS), well-differentiated invasive carcinoma and moderated to poorly differentiated carcinoma. The matching normal epithelium was assessed to be of squamous origin. CGAP annotation of cDNA libaries include HN7, HN9 and HN11 which correspond to the normal epithelium matching to HN 8, HN 10 and HN12 are those form the lesions of the same set.

Analysis of LCM derived cDNA libraries

All HN cDNA library information is available and can be accessed at http://www.ncbi.nlm.nih.gov/cgap/hTGI and http://cgap.nci.nih.gov. By searching the CGAP database information on 14 cDNA libraries related to head and neck can be retrieved. These include those derived from appropriate cell lines, which have been reported previously (Shillitoe *et al*., 2000). For this study, we chose to analyze only

those that were derived from LCM procured HNSCC tissue (HN7-12). The database provides a very detailed description of each library and is summarized in table 4.1. The information includes the total number of sequences analyzed, unclustered sequences, known genes and unique genes. Furthermore, we chose to elaborate this analysis by using the library tool browser XProfiler, to determine differences in this gene classification when comparing between cDNA libraries from normal squamous epithelium and those from squamous cell carcinoma. The gene list obtained is shown in table 4.2.

Genes of interest

The online information includes a library ID link, which provides detailed information on, for example, library discovery, diversity, statistics and list of genes that may be of interest. In our analysis, shown in table 4.3, the list of genes identified as of potential interest represented 1.1% - 34.5% of their individual libraries.

Common genes and EST

As shown in table 4.4, the tool browser XProfiler identified 57 known genes and ESTs which were present in one or more of the libraries, and of these only 4 genes, *calgranulin A, eukaryotic translation elongation factor 1 alpha, ATPase, Na+/K+ transporting alpha3 polypeptide*, and EST (Hs. 153423) were expressed in all 6 cDNA libraries

Differential expressed genes and ESTs in cDNA libraries derived from normal and tumor squamous epithelium.

From our analysis using XProfiler, a total of 347 known genes were electronically detected in only the 3 cDNA libraries derived from normal squamous epithelium but not in tumors. Those genes present in 2 or more of the libraries are indicated in table 4.5. However, when the same analysis was applied to cDNA libraries from tumor tissue, a total of 295 known genes were found to be differentially expressed, and 10 genes of these are shown in table 4.6 as being present in two or more libraries.

Emerging gene expression profile in oral squamous cell carcinomas

The 3 libraries derived from tumor tissues are representative of different clinical grade (table 4.1). Therefore by using XProfiler, analysis can be made to obtain a list of genes differentially expressed in these samples. Gene list consisting of known genes and ESTs for tumor samples that reflect carcinoma *in situ* (HN10), well differentaited invasive carcinoma (HN8) and moderate to poorly differentiated invasive carcinoma (HN12) are provided in tables 4.7- 4.9.

Discussion

Because cancer is a genetic disease, it is important to determine the nature of the genetic alterations characterizing particular stages of tumorigenesis, which could potentially lead to improve diagnosis and therapeutic strategies. For example, to achieve a comprehensive gene expression profile of HNSCC, we need to catalog genes from cDNAs that correspond to mRNAs expressed in normal, precancerous,

and carcinoma cells. Towards this end, 6 cDNA libraries were prepared from three patient tissue sets, comprised of normal oral squamous epithelium and carcinomas, after procuring pure populations of cells from normal and malignant tissues using laser capture microdissection (Leethanakul *et al*., 2000). As there is very limited information on gene expressed in normal oral epithelium and HNSCC, we chose to use non-normalized cDNA libraries. The advantage of non-normalized, non-amplified libraries is that the transcript abundance of the original cell or tissue is accurately reflected in the frequency of clones in the libraries (Ji *et al*., 1997), thus they can be used for both gene discovery and to compare the expression of highly expressed genes in different cells or tissue samples (Ji *et al*., 1997). These libraries were plated, and a number of clones (121-882) from each library were randomly chosen for 3′ end sequencing to generate approximately 300 bp or more, which represents a unique sequence tag for a particular transcript. The BLAST search algorithm was then used to identify the homology of the sequences to those already available in databases. Using this approach, we classified the DNAs as unclustered sequences, known genes and unique genes. Their corresponding number can change over time, as more sequence information is deposited in GenBank. For example, at the time of the first round of sequencing the total number of unique genes was 189 genes, and it was reduced to 138 genes at the time of analysis for this report.

The total number of the genes discovered for each library type is shown in table 4.2. We analyzed them by grouping into normal cDNA libraries (HN7, 9, 11) and carcinoma cDNA libraries (HN8, 10, 12) using the Library XProfiler (http://cgap.nci.nih.gov/Tools). Of interest, there were a very high number of unknown unique and non-unique genes discovered in these libraries (~280 genes),

which is probably because these represent the first group of cDNA libraries from microdissected HNSCC constructed thus far. In particular, there were 189 genes, 87 genes, and 4 genes discovered from the normal squamous epithelium, carcinoma, or both, respectively. Thus, 67.5% of the unknown genes were discovered from the normal cDNA libraries, which might be cell type specific for normal squamous epithelium.

The CGAP hTGI cDNA xProfiler (http://www.ncbi.nlm.nih.gov/cgap/hTGI) allows for intergroup comparisons. For example, using the function *logical and operator* we observed that 57 genes can be found in both normal and carcinoma libraries, and 4 known genes and ESTs are found in all 6 libraries, which are listed in table 4.4. Performing the Library XProfiler intergroup comparison using the function *logical not operator*, we were able to identify genes expressed differentially in normal and carcinoma libraries. In this case, there are a total of 348 genes expressed in HN7, 9, 11 but not in HN8, 10, 12, and 298 genes cancer derived libraries but not in their phenotypically normal counterparts. Those genes expressed at least in two libraries for each group are likely to be highly represented, and are shown in table 4.5 and 4.6. As HN8, HN10 and HN12 are carcinoma *in situ*, well-differentiated invasive carcinomas and moderate to poorly differentiated invasive carcinomas, respectively, we used the Library XProfiler to classify the known genes expressed in each of these libraries that were not expressed in any of the normal libraries (HN7, HN9, and HN11). For example, HN 10, representative of carcinoma *in situ,* expressed 150 known genes differentially identified from normal squamous epithelium and well to poorly differentiated invasive carcinomas. Those that were expressed at least twice (30 genes) are shown in table 4.7. HN8, representative of a well-differentiated

invasive carcinoma, expressed 15 known genes differentially, as shown in table 4.8. In addition, 37 known genes were differentially expressed in moderate to poorly invasive carcinoma, and are shown in table 4.9.

This report is aimed to analyze the available sequence information using analytical tools to begin establish expression profiles corresponding to each HNSCC tissue grade, including carcinoma in situ, well differentiated invasive to poorly differentiated invasive carcinoma. However, there are multiple steps in HNSCC tumor progression that may need to be analyzed in more detail by engineering additional representative libraries. Such an effort, including the construction of additional cDNA libraries from HNSCCs and from putative precancerous lesions such as hyperplasia and leukoplakia, is already underway, and sequence information is expected to be available shortly.

On the other hand, these emerging gene expression profiles should be re-examined and confirmed in an extensive collection of clinical samples, for example, by using DNA fragments as probes in *in situ* hybridization techniques, or by the use of specific antibodies, when available, for the immuno-detection of their gene products. Moreover, many new genes were identified in these cDNA libraries from normal and cancerous squamous epithelial tissues. Their pattern of expression can now begin to be investigated using high throughput technologies such as cDNA microarrays and analytical software. This may lead to the identification of genes correlating with the acquisition of a malignant phenotype, which may play a direct role in squamous carcinogenesis or represent suitable clinical markers of disease progression. Furthermore, the cloning of their coding region by currently available molecular biology techniques or by the prediction of their upstream coding exons in

the human genome may provide soon an exciting opportunity for the discovery of novel proteins expressed in normal and cancerous oral epithelium, whose physiological or pathological role can then begin to be addressed.

Table 4.1: *The general information from CGAP*

Library ID.	Library title	Library type	Library protocol	Origin	Histology	Sequences	Unclustered	Known genes	Unique genes
Lib 1370	HN 7	LCM	Non-normalized	Floor of mouth	Normal epithelium	249	36	111	12
Lib 1362	HN 8	LCM	Non-normalized	Floor of mouth	Well differentiated invasive	121	39	54	9
Lib 1363	HN 9	LCM	Non-normalized	Retromolar trigone	Normal epithelium	879	402	177	55
Lib 1364	HN 10	LCM	Non-normalized	Retromolar trigone	Carcinoma *in situ*	582	122	276	20
Lib 1371	HN 11	LCM	Non-normalized	Tongue	Normal epithelium	494	158	198	35
Lib 1372	HN 12	LCM	Non-normalized	Tongue	Moderate to poorly differentiated invasive	133	51	70	7

Table 4.2: *The number of genes identified as unique genes, non-unique genes including know and unknown genes in these 6 libraries.*

Libraries	Unique genes		Non-Unique genes	
	Known	Unknown	Known	Unknown
NCI_CGAP_HN7, 9, 11	0	105	212	91
NCI_CGAP_HN8, 10, 12	0	33	264	60
NCI_CGAP_HN7, 9, 11 and HN 8, 10, 12	0	138	423	147
NCI_CGAP_HN7, 9, 11 or HN 8, 10, 12	0	0	53	4
NCI_CGAP_HN7, 9, 11 minus HN8, 10, 12	0	105	159	87
NCI_CGAP_HN8, 10, 12 minus HN7, 9, 11	0	33	211	56

Table 4.3: *Genes of interest*

UniGene id	Symbol	Description	Library
Hs.186571	ATP1A3	ATPase, Na+/K+ transporting alpha3 polypeptide	HN7,9,11
Hs.695	CSTB	cystatin B (stefin B)	HN7
Hs.135084	CST3	cystatinC	HN11
Hs.240137		DKFZP586J1923 protein	HN8
Hs.79368	EMP1	epithelium membrane protein 1	HN7
Hs.250591		ESTs	HN9
Hs.256262		ESTs	HN9
Hs.253854		ESTs	HN9
Hs.250866		ESTs	HN9
Hs.250591		ESTs	HN10
Hs.253901		ESTs	HN10
Hs.113943		ESTs	HN11
Hs.254825		ESTs	HN11
Hs.254353		ESTs	HN11
Hs.254824		ESTs	HN11
Hs.254821		ESTs	HN11
Hs.254563		ESTs	HN12
Hs.254828		ESTs	HN12
Hs.254566		ESTs	HN12
Hs.254565		ESTs	HN12
Hs.254564		ESTs	HN12
Hs.254350		ESTs	HN12
Hs.254349		ESTs	HN12
Hs.254348		ESTs	HN12
Hs.254347		ESTs	HN12
Hs.250597		ESTs	HN12
Hs.254325		ESTs	HN7
Hs.254083		ESTs	HN8
Hs.254082		ESTs	HN8
Hs.253905		ESTs	HN8
Hs.253904		ESTs	HN8
Hs.253848		ESTs	HN8
Hs.253847		ESTs	HN8
Hs.253846		ESTs	HN8
Hs.253845		ESTs	HN8
Hs.153423		ESTs	HN10,11
Hs.56105		ESTs, weakly similar to WDNM1 protein precursor	HN7
Hs.181165	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	HN7,9,11
Hs.81075	ECM1	extracellular matrix protein 1	HN7
Hs.111334	FTL	ferritin, light polypeptide	HN10
Hs.172153	GPX	glutathione peroxidase 3 (plasma)	HN9
Hs.79022	GEM	GTP-binding protein overexpressed in skeletal muscle	HN9
Hs.3235	KRT4	Keratin 4	HN7,9
Hs.74070	KRT13	keratin13	HN7,10,11
Hs.111758	KRT6B	keratin6B	HN10
Hs.3297	RPS27A	Ribosomal protein S27a	HN9

Hs.100000	S100A8	S100 calcium-binding protein A8 (CalgranulinA)	HN7,10,11
Hs.5476	VAKTI	serine proteinase inhibitor	HN10
Hs.139322	SPRR3	small proline-rich protein 3	HN10

Table 4.4: *Common genes and EST*

UniGene	Symbol	Description	Library
Hs.100000	S100A8	S100 calcium-binding protein A8 (CalgranulinA)	HN7-12
Hs.181165	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	HN7-12
Hs.186571	ATP1A3	ATPase, Na+/K+ transporting alpha3 polypeptide	HN7-12
Hs.153423	-	ESTs	HN7-12

Table 4.5: *Common genes in normal squamous epithelium*

Unigene id	Symbol	Description	Library
Hs.1526	ATP2A2	ATPase, Ca ++transporting, cardiac muscle, slow twitch2	HN9, 11
Hs.74316	DSP	Desmoplakin (DPI, DPII)	HN7, 11
Hs.253872	-	ESTs	HN9, 11
Hs.46423	H4FG	H4 histone family, member G	HN7, 9
Hs.179943	RPL11	Ribosomal protein L11	HN7, 11
Hs.182979	RPL12	Ribosomal protein L12	HN7, 9, 11
Hs.539	RPS 29	Ribosomal protein S29	HN7, 11

Table 4.6*: *Common genes in HNSCC

UniGene id	Symbol	Description	Library
Hs.174050	EDF1	othelial differentiation-related factor 1	HN10, 12
Hs.278546	-	ESTs	HN10, 12
Hs.181307	H3F3A	H3 histone, family 3A	HN8, 10
Hs.189509	HMG1	High-mobility group protein 1	HN8, 10
Hs.228785	-	Homo sapiens cDNA FLJ10168 fis	HN10, 12
Hs.75545	IL4R	Interleukin 4 receptor	HN8, 10
Hs.178391	L44L	L44-like ribosomal protein	HN8, 10
Hs.184108	RPL21	Ribosomal protein L21	HN10, 12
Hs.151604	RPS8	Ribosomal protein S8	HN10, 12
Hs.112408	S100A7	S100 calcium-binding protein A7 (psoriasin1)	HN8, 10

***Table 4.7**: Differential genes expressed* in *HN10 (Carcinoma in situ)*

UniGene id	Symbol	Description	Repeat
Hs.79274	ANXA5	annexin A5	3
Hs.178452	CTNNA1	catenin (cadherin-associated protein), alpha 1 (102kD)	3
Hs.21490	LOC51009	CGI-101 protein	3
Hs.75887	COPA	coatomer protein complex, subunit alpha	2
Hs.83834	CYB5	cytochrome b-5	2
Hs.92323	FXYD3	FXYD domain-containing ion transport regulator 3	3
Hs.226795	GSTP1	glutathione S-transferase pi	3
Hs.5662	GNB2L1	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	2
Hs.75258	H2AFY	H2A histone family, member Y	2
Hs.118625	HK1	hexokinase 1	2
Hs.279921	LOC51669	HSPC035 protein	2
Hs.250911	IL13RA1	interleukin 13 receptor, alpha 1	2
Hs.180446	KPNB1	karyopherin (importin) beta 1	2
Hs.117729	KRT14	keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)	2
Hs.69559	KIAA1096	KIAA1096 protein	3
Hs.181357	LAMR1	laminin receptor 1 (67kD, ribosomal protein SA)	3
Hs.3709	QP-C	low molecular mass ubiquinone-binding protein (9.5kD)	2
Hs.80395	MAL	mal, T-cell differentiation protein	2
Hs.75789	NDRG1	N-myc downstream regulated	2
Hs.155396	NFE2L2	nuclear factor (erythroid-derived 2)-like 2	2
Hs.184582	RPL24	ribosomal protein L24	2
Hs.119598	RPL3	ribosomal protein L3	2
Hs.169793	RPL32	ribosomal protein L32	2
Hs.184109	RPL37A	ribosomal protein L37a	2
Hs.174131	RPL6	ribosomal protein L6	3
Hs.241507	RPS6	ribosomal protein S6	3
Hs.73742	RPLP0	ribosomal protein, large, P0	3
Hs.1076	SPRR1B	small proline-rich protein 1B (cornifin)	2
Hs.171581	UCH37	ubiquitin C-terminal hydrolase UCH37	2

Table 4.8: *Differential genes expressed* in *HN8 (well differentiated invasive carcinoma)*

UniGene id	Symbol	Description	Repeat
Hs.82396	OAS1	2',5'-oligoadenylate synthetase 1	1
Hs.90336	ATP6J	ATPase, H+ transporting, lysosomal (vacuolar proton pump), member J	1
Hs.10029	CTSC	cathepsin C	1
Hs.65134	CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen)	1
Hs.8986	C1QB	complement component 1, q subcomponent, beta polypeptide	1
Hs.147916	DDX3	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3	1
Hs.1742	IQGAP1	IQ motif containing GTPase activating protein 1	1
Hs.69423	KLK10	kallikrein 10	1
Hs.165296	KLK13	kallikrein 13	1
Hs.771	PYGL	phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)	1
Hs.148027	POLR2B	polymerase (RNA) II (DNA directed) polypeptide B (140kD)	1
Hs.75721	PFN1	profilin 1	1
Hs.112341	PI3	protease inhibitor 3, skin-derived (SKALP)	1
Hs.101850	RBP1	retinol-binding protein 1, cellular	1
Hs.111065	USP6	ubiquitin specific protease 6 (Tre-2 oncogene)	1

Table 4.9: *Differential genes expressed* in *HN12 moderate to poorly differentiated invasive carcinomas*

UniGene id	Symbol	Description	Repeat
Hs.82425	ARPC5	actin related protein 2/3 complex, subunit 5 (16 kD)	1
Hs.7957	ADAR	adenosine deaminase, RNA-specific	1
Hs.80986	ATP5G1	ATP synthase, H+ transporting, mitochondrial F0 complex,	1
Hs.101025	BTF3	basic transcription factor 3	1
Hs.8383	BAZ2B	bromodomain adjacent to zinc finger domain, 2B	1
Hs.98658	BUB1	budding uninhibited by benzimidazoles 1 (yeast homolog)	1
Hs.155560	CANX	Calnexin	1
Hs.140452	TIP47	cargo selection protein	1
Hs.26584	DIAPH1	diaphanous (Drosophila, homolog) 1	1
Hs.30661	ETFDH	electron-transferring-flavoprotein dehydrogenase	1
Hs.82193	ESD	esterase D/formylglutathione hydrolase	1
Hs.106673	EIF3S6	eukaryotic translation initiation factor3, subunit 6(48kD)	1
Hs.6527	GPR56	G protein-coupled receptor 56	1
Hs.119537	SAM68	GAP-associated tyrosine phosphoprotein p62 (Sam68)	1
Hs.278589	GTF2I	general transcription factor II, i	2
Hs.87889	KIAA0928	helicase-moi	1
Hs.40154	JMJ	jumonji (mouse) homolog	2
Hs.84087	KIAA0143	KIAA0143 protein	1
Hs.83419	KIAA0252	KIAA0252 protein	1
Hs.183006	KIAA0836	KIAA0836 protein	1
Hs.4014	KIAA0946	KIAA0946 protein;Huntingtin interacting protein H	1
Hs.75061	MLP	MARCKS-like protein	1
Hs.83916	NDUFA5	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	2
Hs.256526	NARF	nuclear prelamin A recognition factor	1
Hs.91747	PFN2	profilin 2	1
Hs.99858	RPL7A	ribosomal protein L7a	1
Hs.182426	RPS2	ribosomal protein S2	1
Hs.1948	RPS21	ribosomal protein S21	1
Hs.256290	S100A11	S100 calcium-binding protein A11 (calgizzarin)	1
Hs.77496	SNRPG	small nuclear ribonucleoprotein polypeptide G	1
Hs.44450	SP3	Sp3 transcription factor	1
Hs.184510	SFN	Stratifin	1
Hs.155188	TAF2F	TATA box binding protein-associated factor, RNA polymerase II, F, 55kD	1
Hs.77356	TFRC	transferrin receptor (p90, CD71)	1
Hs.180248	ZNF124	zinc finger protein 124 (HZF-16)	2
Hs.2110	ZNF9	zinc finger protein 9 (a cellular retroviral nucleic acid binding protein)	1

CHAPTER V

Discussion and Conclusions

DISCUSSION

Until recently, the molecular biology of cancer was studied by concentrating all efforts on one or a few genes at a time. However, recent advances in DNA sequencing and microarray technologies, among others, have provided a unique opportunity to monitor, simultaneously, thousands of genes as a high throughput approach. Use of these technologies therefore, allows the scanning of gene expression patterns and the search for those correlating with a disease stage. Furthermore, gene expression profiles can now be investigated within a histologically defined, homogeneous population of cells, thus affording the possibility of applying these newly available techniques to investigate expression patterns in normal as well as in neoplastic tissue. These revolutionary approaches are likely to have many advantages in cancer biology, and particularly in the search for the still unknown mechanisms involved in head and neck squamous cell carcinogenesis. However, this will also require the development of effective methods to validate the biological relevance of the newly identified candidate genes. Indeed, the use of tissue culture systems and animal models to recapitulate this complex disease will be a central component of these efforts. Furthermore, the effective use of DNA and RNA labeling techniques and the development of immunological tools would be expected to allow the direct examination of expression profiles in clinical specimens, including potential premalignant lesions. Thus, exciting opportunities are ahead of us, which will certainly contribute to advance our understanding of the molecular basis of HNSCC.

However, it is becoming increasingly clear that it will take a concerted effort from the entire scientific and health professional community to battle the ravaging consequences of this disease.

The cancerous tissues often contain little relevant target tissue. Specimens are often measured in millimeters and are inadequately preserved. The epithelium is often "contaminated" by surrounding connective tissue, blood vessels and inflammatory infiltrates. In addition, the epithelium may display different histopathologic features of HNSCC carcinogenesis: normal, hyperplasia, dysplasia, carcinoma *in situ* and invasive carcinomas. LCM allows the precise identification, dissection and harvesting of pure cell populations that are more reflective of the disease process *in vivo* than cell culture, which could be distorted by conditions/selection pressures. Presently, LCM is being used on a wide variety of human cancers. RNA from LCM-isolated HNSCC tissue has been successful isolated and used to generate target samples for hybridization to cDNA filter microarrays. This approach was used to isolate and amplify approximately 16.7 ng of high quality total RNA from approximately 5,000 epithelial cells procured by LCM using RT-PCR. These data indicate that complex AcDNA probes labeled with α-^{33}P dCTP can be successfully synthesized even from small amounts of total RNA for their use in comparative hybridization studies of cancer genes expressed in HNSCC. This study has helped identify a number of new candidate genes, which might play an unexpected role in this carcinogenesis. Certainly, the expression of these genes needs to be verified using other standard method such as: *in situ* hybridization, immunocytochemistry and real time RT-PCR. Nonetheless, we can conclude that

LCM generated target sample may provide the most accurate representation of *in vivo* gene expression available.

Although data obtained involved a limited set of tissue samples, a general trend is already observed, implicating cell cycle regulating and signaling molecules, growth and angiogenic factors, matrix degrading proteases, and survival and apoptotic molecules. Furthermore, we obtained evidence implicating, for the first time, the *notch* and *wnt* pathways in squamous cell carcinogenesis. Further analysis of a more extensive sample collection using conventional and these recently available technologies will make it possible to define a pattern of gene expression in a tumor progression model of HNSCC. This experimental approach is also expected to facilitate the identification of candidate markers potentially correlated with malignancy, thus providing valuable tools of diagnostic and prognostic value to study premalignant lesions. From these experiments, it can be concluded that the use of LCM and cDNA arrays has allowed the detailed analysis of gene expression *in vivo* in HNSCC, and provided evidence for the feasibility of performing a detailed molecular characterization in normal, premalignant, and malignant HNSCC cells.

Although cDNA arrays are likely to identify known genes with known function, they may not provide a suitable platform for gene discovery. Instead, the high throughput analysis of gene expression by nucleotide sequencing of clones derived from HNSCC specific cDNA libraries was considered as an approach. The effectiveness of sequencing ESTs for gene discovery was illustrated by the high frequency of novel genes in non-normalized representational cDNA libraries constructed from LCM derived tissues. These studies were successful in the construction 6 cDNA libraries from LCM procured cells from normal and tumor

tissues derived from HNSCC patients. The results form the quality assessment demonstrated that these cDNA libraries were of a high diversity and the average size was 500 bp. Collectively, our results indicated that high quality cDNA libraries can be generated from microdissected HNSCC tissues. Furthermore, sequencing information revealed the existence of at least 132 novel genes that may be preferentially expressed in HNSCC. It is likely that these genes may play a role in the pathogenesis of HNSCC, and may represent novel markers for early detection as well as targets for pharmacological intervention in this disease. Sequence information from CGAP identified additional novel genes. To date, approximately 138 genes can be identified as uniquely expressed in HNSCC. However, this number is likely to decrease after their discovery in other cDNA libraries. In response to the limited understanding of the molecular mechanisms responsible for HNSCC, the identity of these novel genes may help to elucidate their potential role in tumor development. This effort thus paves the way for a detailed, systematic analysis of these clones, aimed to explore whether any of these cDNAs are involved in the pathogenesis of this cancer type.

As web-based bio-informatic tools are available for their use with CGAP databases, these cDNA libraries were analyzed in order to examine which clones were differentially expressed. The advantage of this study was the possibility of performing the analysis of genes differentially expressed between the normal and tumor tissue derived from the same patient. Moreover, we began to explore which genes are expressed at different stages of tumor development, including carcinoma *in situ*, well to moderate differentiate invasive carcinoma, and severe differentiate invasive carcinoma. Although these analytical tools are not able to analyze the differential expression of all EST in each library, some of these unique genes are clearly shown to

exhibit a distinct level of expression at different tumor stages. With regard to the limited set of tissue samples in this study, certainly additional cDNA libraries need to be constructed from tissues of head and neck origin, and from different site and stage of malignancy, including hyperplasia, dysplasia and metasasis.

As part of the future direction of this project, all novel genes identified in these libraries could be arrayed for the high throughput analysis using the combination of advanced technologies, such as LCM and probe labeling, to scan gene expression profiles and to search for those correlating with different stages of HNSCC. The most interesting candidate genes could be quickly identified by this approach. For those that may be deemed of importance, full length cDNA clones can then be obtained for detailed functional testing, particularly to confirm whether these genes play a role in tumor progression in HNSCC. We believe that this information will aid our understanding of the molecular mechanisms responsible for HNSCC, and will be helpful ultimately for the prevention, diagnosis, and treatment of this cancer type.

CONCLUSION:

1. The successful use of laser assisted microdissection for procuring specific cells, in particular normal and HNSCC populations, provided a novel approach and exciting opportunities for understanding the molecular basis of HNSCC *in vivo.*

2. The use of LCM and commercial cDNA arrays has allowed the detailed analysis of gene expression in HNSCC, and provided the first evidence for the feasibility of performing a comprehensive molecular characterization of normal, premalignant and malignant HNSCC cells.

3. The general trends observed in this study included the higher expression of cell cycle regulating and signaling molecule, growth and angiogenic factors, matrix degrading proteases and survival and apoptotic genes in cancer cells, which may play a role in tumor progression in HNSCC. Furthermore, this study obtained evidence implicating, for the first time, the *notch* and *wnt* pathways in squamous cell carcinogenesis.

4. The construction of representational cDNA libraries from microdissected normal and tumor tissues from head and neck has resulted in the availability of cDNAs for many novel genes expressed in normal squamous epithelium and in HNSCC.

5. Differential gene expression profile was examined using standard nucleotide sequencing to identify all ESTs expressed in these libraries, and by the use of extensive databases at the CGAP and bioinformatic tools. Novel and unique genes where identified by this approach, and their specific expression at each stage of tumor progression evaluated. Furthermore, these unique genes may represent important candidates to facilitate the identification of early markers heralding malignancy, thus providing valuable tools of diagnostic and prognosis value to study premalignant lesions.

6. The differential expression of known genes in each tumor stage was evaluated, and available evidence suggests that some of these genes may be involved in head and neck carcinogenesis.

7. All databases have been made available in the public domain at http://www.ncbi.nlm.nih.gov/ncicgap

REFERENCES

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. and et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252:** 1651-6.

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D. and Levine, A.J. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* **96:** 6745-50.

Artavanis-Tsakonas, S., Rand, M.D. and Lake, R.J. 1999. Notch signaling: cell fate control and signal integration in development. *Science* **284:** 770-6.

Bartkova, J., Lukas, J., Muller, H., Strauss, M., Gusterson, B. and Bartek, J.1995. Abnormal patterns of D-type cyclin expression and G1 regulation in human head and neck cancer. *Cancer Res* **55:** 949-56.

Bishop, J.M. 1991. Molecular themes in oncogenesis. *Cell* **64:** 235-48.

Bockmuhl, U., Wolf, G., Schmidt, S., Schwendel, A., Jahnke, V., Dietel, M. and Petersen, I. 1998. Genomic alterations associated with malignancy in head and neck cancer. *Head Neck* **20:** 145-51.

Bohm, M., Wieland, I., Schutze, K. and Rubben, H. 1997. Microbeam MOMeNT: non-contact laser microdissection of membrane-mounted native tissue. *Am J Pathol* **151:** 63-7.

Bonaldo, M.F., Lennon, G. and Soares, M.B.1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res* **6:** 791-806.

Boring, C.C., Squires, T.S. and Tong, T. 1993. Cancer statistics, 1993. *CA Cancer J Clin* **43:** 7-26.

Bryce, S.D., Forsyth, N.R., Fitzsimmons, S.A., Clark, L.J., Bertram, M.J., Cuthbert, A.P., Newbold, R.F., Pereira-Smith, O.M. and Parkinson, E.K. 1995. Genetic and functional analyses exclude mortality factor 4 (MORF4) as a keratinocyte senescence gene. *Cancer Res* **59:** 2038-40.

Byrne, J.A., Tomasetto, C., Garnier, J.M., Rouyer, N., Mattei, M.G., Bellocq, J.P., Rio, M.C. and Basset, P.1995. A screening method to identify genes commonly overexpressed in carcinomas and the identification of a novel complementary DNA sequence. *Cancer Res* **55:** 2896-903.

Califano, J., van der Riet, P., Westra, W., Nawroz, H., Clayman, G., Piantadosi, S., Corio, R., Lee, D., Greenberg, B., Koch, W. and Sidransky, D.1996. Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res* **56:** 2488-92.

Carulli, J.P., Artinger, M., Swain, P.M., Root, C.D., Chee, L., Tulig, C., Guerin, J., Osborne, M., Stein, G., Lian, J. and Lomedico, P.T. 1998. High throughput analysis of differential gene expression. *J Cell Biochem Suppl* **31:** 286-96.

Celis, J.E., Celis, P., Ostergaard, M., Basse, B., Lauridsen, J.B., Ratz, G., Rasmussen, H.H., Orntoft, T.F., Hein, B., Wolf, H. and Celis 1999. A., Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: a novel strategy for identifying metaplastic lesions. *Cancer Res* **59:** 3003-9.

Chi, S., Kitanaka, C., Noguchi, K., Mochizuki, T., Nagashima, Y., Shirouzu, M., Fujita, H., Yoshida, M., Chen, W., Asai, A., Himeno, M., Yokoyama, S. and Kuchino, Y. 1999. Oncogenic Ras triggers cell suicide through the

activation of a caspase- independent cell death program in human cancer cells. *Oncogene* **18:** 2281-90.

Chiang, J.M. 1998. Role of K-ras mutations in colorectal carcinoma. *Cancer Lett* **126:** 179-85.

Clurman, B.E. and Roberts, J.M. 1995. Cell cycle and cancer [editorial; comment]. *J Natl Cancer Inst* **87:** 1499-501.

Cooper, M.T. and Bray, S.J. 1999. Frizzled regulation of Notch signalling polarizes cell fate in the Drosophila eye. *Nature* **397:** 526-30.

Crosthwaite, N., Teale, D., Franklin, C., Foster, G.A. and Stringer, B.M. 1996. p53 protein expression in malignant, pre-malignant and non-malignant lesions of the lip. *J Clin Pathol* **49:** 648-53.

De Stefani, E., Fierro, L., Larrinaga, M.T., Balbi, J.C., Ronco, A. and Mendilaharsu, M. 1994. Smoking of hand-rolled cigarettes as a risk factor for small cell lung cancer in men: a case-control study from Uruguay. *Lung Cancer* **11:** 191-9.

Dean-Clower, E., Vortmeyer, A.O., Bonner, R.F., Emmert-Buck, M., Zhuang, Z. and Liotta, L.A. 1997. Microdissection-based genetic discovery and analysis applied to cancer progression. *Cancer J Sci Am* **3:** 259-65.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer [see comments]. *Nat Genet* **14:** 457-60.

Desai, S.B. and Libutti, S.K. 1999. Tumor angiogenesis and endothelial cell modulatory factors [see comments]. *J Immunother* **22:** 186-211.

Dewanji, A., Goddard, M.J., Krewski, D. and Moolgavkar, S.H. 1999, Two stage model for carcinogenesis: number and size distributions of premalignant clones in longitudinal studies. *Math Biosci* **155:**1-12.

Dietz, A., Senneweld, E. and Maier, H. 1995. Indoor air pollution by emissions of fossil fuel single stoves: possibly a hitherto underrated risk factor in the development of carcinomas in the head and neck. *Otolaryngol Head Neck Surg* **112:** 308-15.

Eberwine, J. 1996. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *Biotechniques* **20:** 584-91.

el-Deiry, W.S., Tokino, T., Velculescu, V.E., Levy, D.B., Parsons, R., Trent, J.M., Lin, D., Mercer, W.E., Kinzler, K.W. and Vogelstein, B. 1993. WAF1, a potential mediator of p53 tumor suppression. *Cell* **75:** 817-25.

El-Naggar, A.K., Coombes, M.M., Batsakis, J.G., Hong, W.K., Goepfert, H. and Kagan, J. 1998. Localization of chromosome 8p regions involved in early tumorigenesis of oral and laryngeal squamous carcinoma. *Oncogene* **16:** 2983-7.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. and Liotta, L.A. 1996. Laser capture microdissection [see comments]. *Science* **274:** 998-1001.

Fearon, E.R. and Vogelstein, B.1990. A genetic model for colorectal tumorigenesis. *Cell* **61:** 759-67.

Fischer, D.F., van Drunen, C.M., Winkler, G.S., van de Putte, P. and Backendorf, C.1998 Involvement of a nuclear matrix association region in the regulation of the SPRR2A keratinocyte terminal differentiation marker. *Nucleic Acids Res* **26:** 5288-94.

Fishleder, A.J.1990. Oncogenes and cancer: clinical applications. *Cleve Clin J Med* **57:** 721-6.

Franceschi, S., Talamini, R., Barra, S., Baron, A.E., Negri, E., Bidoli, E., Serraino, D. and La Vecchia, C.1990. Smoking and drinking in relation to cancers of the oral cavity, pharynx, larynx, and esophagus in northern Italy. *Cancer Res* **50:** 6502-7.

Frigerio, J.M., Berthezene, P., Garrido, P., Ortiz, E., Barthellemy, S., Vasseur, S., Sastre, B., Seleznieff, I., Dagorn, J.C. and Iovanna, J.L.1995. Analysis of 2166 clones from a human colorectal cancer cDNA library by partial sequencing. *Hum Mol Genet* **4:** 37-43.

Fujimoto, W., Nakanishi, G., Arata, J. and Jetten, A.M. 1997. Differential expression of human cornifin alpha and beta in squamous differentiating epithelial tissues and several skin lesions. *J Invest Dermatol* **108:** 200-4.

Gelehrter, T. D, Collin F.S.1990. Cancer genetics. In: Gelehrter, T. D., Collin F.S (Eds) Genetics. Baltimore. Williams and Wilkins: 229-252.

Gerhold, D., Rushmore, T. and Caskey, C.T.1999. DNA chips: promising toys have become powerful tools. *Trends Biochem Sci* **24:** 168-73.

Giaretti, W., Rapallo, A., Geido, E., Sciutto, A., Merlo, F., Risio, M. and Rossini, F.P.1998. Specific K-ras2 mutations in human sporadic colorectal adenomas are associated with DNA near-diploid aneuploidy and inhibition of proliferation. *Am J Pathol* **153:** 1201-9.

Gonzalez, M.V., Pello, M.F., Ablanedo, P., Suarez, C., Alvarez, V. and Coto, E.1998. Chromosome 3p loss of heterozygosity and mutation analysis of the FHIT and beta-cat genes in squamous cell carcinoma of the head and neck. *J Clin Pathol* **51:** 520-4.

Granville, D.J., Carthy, C.M., Hunt, D.W. and McManus, B.M.1998. Apoptosis: molecular aspects of cell death and disease. *Lab Invest* **78:** 893-913.

Gray, N.S., Wodicka, L., Thunnissen, A.M., Norman, T.C., Kwon, S., Espinoza, F.H., Morgan, D.O., Barnes, G., LeClerc, S., Meijer, L., Kim, S.H., Lockhart, D.J. and Schultz, P.G.1998. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **281:** 533-8.

Greenblatt, M.S., Bennett, W.P., Hollstein, M. and Harris, C.C. 1994. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res* **54:** 4855-78.

Gress, T.M., Wallrapp, C., Frohme, M., Muller-Pillasch, F., Lacher, U., Friess, H., Buchler, M., Adler, G. and Hoheisel, J.D. 1997. Identification of genes with specific expression in pancreatic cancer by cDNA representational difference analysis. *Genes Chromosomes Cancer* **19:** 97-103.

Gryfe, R., Swallow, C., Bapat, B., Redston, M., Gallinger, S. and Couture, J.1997. Molecular biology of colorectal cancer. *Curr Probl Cancer* **21:** 233-300.

Gupta, P.C., Mehta, F.S., Pindborg, J.J., Bhonsle, R.B., Murti, P.R., Daftary, D.K. and Aghi, M.B. 1992. Primary prevention trial of oral cancer in india: a 10-year follow-up study. *J Oral Pathol Med* **21:** 433-9.

Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis, R.W.1997. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci U S A* **94:** 2150-5.

Hoffman, B. and Liebermann, D.A. 1998. The proto-oncogene c-myc and apoptosis. *Oncogene* **17:** 3351-7.

Hohl, D., de Viragh, P.A., Amiguet-Barras, F., Gibbs, S., Backendorf, C. and Huber, M.1995. The small proline-rich proteins constitute a multigene family of differentially regulated cornified cell envelope precursor proteins. *J Invest Dermatol* **104:** 902-9.

Hubank, M. and Schatz, D.G. 1994. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res* **22:** 5640-8.

Ishwad, C.S., Shuster, M., Bockmuhl, U., Thakker, N., Shah, P., Toomes, C., Dixon, M., Ferrell, R.E. and Gollin, S.M.1994. Frequent allelic loss and homozygous deletion in chromosome band 8p23 in oral cancer. *Int J Cancer* **80:** 25-31.

Jen, K.Y. and Gewirtz, A.M. 2000. Suppression of gene expression by targeted disruption of messenger RNA: available options and current strategies. *Stem Cells* **18:** 307-19.

Ji, H., Liu, Y.E., Jia, T., Wang, M., Liu, J., Xiao, G., Joseph, B.K., Rosen, C. and Shi, Y.E.1997. Identification of a breast cancer-specific gene, BCSG1, by direct differential cDNA sequencing. *Cancer Res* **57:** 759-64.

Jin, Y., Jin, C., Wennerberg, J., Mertens, F. and Hoglund, M. 1998 Cytogenetic and fluorescence in situ hybridization characterization of chromosome 1 rearrangements in head and neck carcinomas delineate a target region for deletions within 1p11-1p13. *Cancer Res* **58:** 5859-65.

Kaur, J., Srivastava, A. and Ralhan, R. 1994. Overexpression of p53 protein in betel- and tobacco-related human oral dysplasia and squamous-cell carcinoma in India. *Int J Cancer* **58:** 340-5.

Khan, J., Bittner, M.L., Chen, Y., Meltzer, P.S. and Trent, J.M. 1999. DNA microarray technology: the anticipated impact on the study of human disease. *Biochim Biophys Acta* **1423:** M17-28.

Krizman, D.B., Chuaqui, R.F., Meltzer, P.S., Trent, J.M., Duray, P.H., Linehan, W.M., Liotta, L.A. and Emmert-Buck, M.R. 1996. Construction of a representative cDNA library from prostatic intraepithelial neoplasia. *Cancer Res* **56;** 5380-3.

Krizman, D.B., Wagner, L., Lash, A., Strausberg, R.L. and Emmert-Buck, M.R. 1999. The Cancer Genome Anatomy Project: EST sequencing and the genetics of cancer progression. *Neoplasia* **1:** 101-6.

Laken, S.J., Papadopoulos, N., Petersen, G.M., Gruber, S.B., Hamilton, S.R., Giardiello, F.M., Brensinger, J.D., Vogelstein, B. and Kinzler, K.W.1999. Analysis of masked mutations in familial adenomatous polyposis. *Proc Natl Acad Sci U S A* **96:** 2322-6.

Landis, S.H., Murray, T., Bolden, S. and Wingo, P.A., 1999. Cancer statistics, [see comments]. *CA Cancer J Clin* **49:** 8-31, 1.

Lazar, A.D., Winter, M.R., Nogueira, C.P., Larson, P.S., Finnemore, E.M., Dolan, R.W., Fuleihan, N., Chakravarti, A., Zietman, A. and Rosenberg, C.L.1998. Loss of heterozygosity at 11q23 in squamous cell carcinoma of the head and neck is associated with recurrent disease. *Clin Cancer Res* **4:** 2787-93.

Lee, D.J., Koch, W.M., Yoo, G., Lango, M., Reed, A., Califano, J., Brennan, J.A., Westra, W.H., Zahurak, M. and Sidransky, D.1997. Impact of chromosome 14q loss on survival in primary head and neck squamous cell carcinoma. *Clin Cancer Res* **3:** 501-5.

Leethanakul, C., Patel, V., Gillespie, J., Pallente, M., Ensley, J.F., Koontongkaew, S., Liotta, L.A., Emmert-Buck, M. and Gutkind, J.S. 2000. Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene* **19:** 3220-4.

Leethanakul, C., Patel, V., Gillespie, J., Shillitoe, E., Kellman, R.M., Ensley, J.F., Limwongse, V., Emmert-Buck, M.R., Krizman, D.B. and Gutkind, J.S. 2000. Gene expression profiles in squamous cell carcinomas of the oral cavity: use of laser capture microdissection for the construction and analysis of stage-specific cDNA libraries [In Process Citation]. *Oral Oncol* **36:** 474-83.

Liang, P. and Pardee, A.B. 1992 Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction [see comments]. *Science* **257:** 967-71.

Lisitsyn, N. and Wigler, M. 1993. Cloning the differences between two complex genomes. *Science*, **259**, 946-51.

Liu, Y., Dehni, G., Purcell, K.J., Sokolow, J., Carcangiu, M.L., Artavanis-Tsakonas, S. and Stifani, S. 1996. Epithelial expression and chromosomal location of human TLE genes: implications for notch signaling and neoplasia. *Genomics*, **31**, 58-64.

Lo Muzio, L., Mignogna, M.D., Favia, G., Procaccini, M., Testa, N.F. and Bucci, E. 1998. The possible association between oral lichen planus and oral squamous cell carcinoma: a clinical evaluation on 14 cases and a review of the literature. *Oral Oncol* **34:** 239-46.

Maier, H., De Vries, N. and Snow, G.B. 1991. Occupational factors in the aetiology of head and neck cancer. *Clin Otolaryngol* **16:** 406-12.

Mantovani, A., Bottazzi, B., Colotta, F., Sozzani, S. and Ruco, L. 1992. The origin and function of tumor-associated macrophages [see comments]. *Immunol Today* **13:** 265-70.

Mao, E.J., Schwartz, S.M., Daling, J.R. and Beckmann, A.M. 1998. Loss of heterozygosity at 5q21-22 (adenomatous polyposis coli gene region) in oral squamous cell carcinoma is common and correlated with advanced disease. *J Oral Pathol Med* **27:** 297-302.

Miller, W.H., Jr. and Dmitrovsky, E. 1991. Oncogenes and clinical oncology. *Curr Opin Oncol* **3:** 65-9.

Mitelman, F., Mertens, F. and Johansson, B. 1997. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia [see comments]. *Nat Genet* **15: Spec No,** 417-74.

Miyakawa, A., Wang, X.L., Nakanishi, H., Imai, F.L., Shiiba, M., Miya, T., Imai, Y. and Tanzawa, H. 1998. Allelic loss on chromosome 22 in oral cancer: possibility of the existence of a tumor suppressor gene on 22q13. *Int J Oncol* **13:** 705-9.

Montenarh, M. 1992. Biochemical properties of the growth suppressor/oncoprotein p53. *Oncogene* **7:** 1673-80.

Nacht, M., Ferguson, A.T., Zhang, W., Petroziello, J.M., Cook, B.P., Gao, Y.H., Maguire, S., Riley, D., Coppola, G., Landes, G.M., Madden, S.L. and Sukumar, S. 1999. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer [In Process Citation]. *Cancer Res* **59:** 5464-70.

Niedobitek, G. and Young, L.S. 1994. Epstein-Barr virus persistence and virus-associated tumours [see comments]. *Lancet* **343:** 333-5.

Nielsen, H., Norrild, B., Vedtofte, P., Praetorius, F., Reibel, J. and Holmstrup, P. 1996. Human papillomavirus in oral premalignant lesions. *Eur J Cancer B Oral Oncol* **32B:** 264-70.

Ogawara, K., Miyakawa, A., Shiba, M., Uzawa, K., Watanabe, T., Wang, X.L., Sato, T., Kubosawa, H., Kondo, Y. and Tanzawa, H. 1998. Allelic loss of chromosome 13q14.3 in human oral cancer: correlation with lymph node metastasis. *Int J Cancer* **79:** 312-7.

Opdenakker, G., Fiten, P., Nys, G., Froyen, G., Van Roy, N., Speleman, F., Laureys, G. and Van Damme, J. 1994. The human MCP-3 gene (SCYA7): cloning, sequence analysis, and assignment to the C-C chemokine gene cluster on chromosome 17q11.2-q12. *Genomics* **21:** 403-8.

Opdenakker, G., Froyen, G., Fiten, P., Proost, P. and Van Damme, J. 1993. Human monocyte chemotactic protein-3 (MCP-3): molecular cloning of the cDNA and comparison with other chemokines. *Biochem Biophys Res Commun* **191:** 535-42.

Opdenakker, G. and Van Damme, J. 1992. Chemotactic factors, passive invasion and metastasis of cancer cells [letter; comment]. *Immunol Today* **13:** 463-4.

Ornstein, D.K., Gillespie, J.W., Paweletz, C.P., Duray, P.H., Herring, J., Vocke, C.D., Topalian, S.L., Bostwick, D.G., Linehan, W.M., Petricoin, E.F. 2000. 3rd and Emmert-Buck, M.R., Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines [In Process Citation]. *Electrophoresis* **21:** 2235-42.

Pai, S.I., Wu, G.S., Ozoren, N., Wu, L., Jen, J., Sidransky, D. and El-Deiry, W.S. 1998. Rare loss-of-function mutation of a death receptor gene in head and neck cancer. *Cancer Res* **58:** 3513-8.

Panin, V.M., Papayannopoulos, V., Wilson, R. and Irvine, K.D. 1997. Fringe modulates Notch-ligand interactions. *Nature* **387:** 908-12.

Pappalardo, P.A., Bonner, R., Krizman, D.B., Emmert-Buck, M.R. and Liotta, L.A. 1998. Microdissection, microchip arrays, and molecular analysis of tumor cells (primary and metastases). *Semin Radiat Oncol* **8:** 217-23.

Park, J.S., Dong, S.M., Kim, H.S., Lee, J.Y., Um, S.J., Park, I.S., Kim, S.J. and Namkoong, S.E. 1999. Detection of p16 gene alteration in cervical cancer using tissue microdissection and LOH study. *Cancer Lett* **136:** 101-8.

Parkin, D.M., Pisani, P. and Ferlay, J. 1999. Global cancer statistics. *CA Cancer J Clin* **49:** 33-64, 2.

Partridge, M., Emilion, G., Pateromichelakis, S., Phillips, E. and Langdon, J. 1997. Field cancerisation of the oral cavity: comparison of the spectrum of molecular alterations in cases presenting with both dysplastic and malignant lesions. *Oral Oncol* **33:** 332-7.

Patanjali, S.R., Parimoo, S. and Weissman, S.M. 1991. Construction of a uniform-abundance (normalized) cDNA library [published erratum appears in Proc Natl Acad Sci U S A 1991 Jun 1;88(11):5066]. *Proc Natl Acad Sci U S A* **88:** 1943-7.

Patel, V., Jakus, J., Harris, C.M., Ensley, J.F., Robbins, K.C. and Yeudall, W.A. 1997. Altered expression and activity of G1/S cyclins and cyclin-dependent kinases characterize squamous cell carcinomas of the head and neck. *Int J Cancer* **73:** 551-5.

Pearlstein, R.P., Benninger, M.S., Carey, T.E., Zarbo, R.J., Torres, F.X., Rybicki, B.A. and Dyke, D.L. 1998. Loss of 18q predicts poor survival of patients

with squamous cell carcinoma of the head and neck. *Genes Chromosomes Cancer* **21:** 333-9.

Peterson, L.A., Brown, M.R., Carlisle, A.J., Kohn, E.C., Liotta, L.A., Emmert-Buck, M.R. and Krizman, D.B. 1998. An improved method for construction of directionally cloned cDNA libraries from microdissected cells. *Cancer Res* **58:** 5326-8.

Pillai, R. 1992. Oncogenes and oncoproteins as tumor markers. *Eur J Surg Oncol* **18:** 417-24.

Qin, G.Z., Park, J.Y., Chen, S.Y. and Lazarus, P. 1999. A high prevalence of p53 mutations in pre-malignant oral erythroplakia. *Int J Cancer* **80:** 345-8.

Ransom, D.T., Barnett, T.C., Bot, J., de Boer, B., Metcalf, C., Davidson, J.A. and Turbett, G.R. 1998. Loss of heterozygosity on chromosome 2q: possibly a poor prognostic factor in head and neck cancer. *Head Neck* **20:** 404-10.

Saito, M., Yamaguchi, A., Goi, T., Tsuchiyama, T., Nakagawara, G., Urano, T., Shiku, H. and Furukawa, K. 1999. Expression of DCC protein in colorectal tumors and its relationship to tumor progression and metastasis. *Oncology* **56:** 134-41.

Sartor, M., Steingrimsdottir, H., Elamin, F., Gaken, J., Warnakulasuriya, S., Partridge, M., Thakker, N., Johnson, N.W. and Tavassoli, M. 1999. Role of p16/MTS1, cyclin D1 and RB in primary oral cancer and oral cancer cell lines. *Br J Cancer* **80:** 79-86.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. and Davis, R.W. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* **93:** 10614-9.

Schipper, J.H., Frixen, U.H., Behrens, J., Unger, A., Jahnke, K. and Birchmeier, W. 1991. E-cadherin expression in squamous cell carcinomas of head and neck: inverse correlation with tumor dedifferentiation and lymph node metastasis. *Cancer Res* **51:** 6328-37.

Scully, C., Field, J.K. and Tanzawa, H. 2000. Genetic aberrations in oral or head and neck squamous cell carcinoma 3: clinico-pathological applications [In Process Citation]. *Oral Oncol* **36:** 404-13.

Shao, X., Tandon, R., Samara, G., Kanki, H., Yano, H., Close, L.G., Parsons, R. and Sato, T. 1998. Mutational analysis of the PTEN gene in head and neck squamous cell carcinoma [published erratum appears in Int J Cancer 1999 Feb 9;80(4):636]. *Int J Cancer* **77:** 684-8.

Shelly, L.L., Fuchs, C. and Miele, L. 1999. Notch-1 inhibits apoptosis in murine erythroleukemia cells and is necessary for differentiation induced by hybrid polar compounds. *J Cell Biochem* **73:** 164-75.

Shillitoe, E.J., May, M., Patel, V., Lethanakul, C., Ensley, J.F., Strausberg, R.L. and Gutkind, J.S. 2000. Genome-wide analysis of oral cancer--early results from the Cancer Genome Anatomy Project. *Oral Oncol* **36:** 8-16.

Sidransky, D. 1995. Molecular genetics of head and neck cancer. *Curr Opin Oncol* **7:** 229-33.

Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R. and Liotta, L.A. 1998. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet* **14:** 272-6.

Snijders, P.J., Cromme, F.V., van den Brule, A.J., Schrijnemakers, H.F., Snow, G.B., Meijer, C.J. and Walboomers, J.M. 1992. Prevalence and expression of

human papillomavirus in tonsillar carcinomas, indicating a possible viral etiology. *Int J Cancer* **51:** 845-50.

Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci U S A* **91:** 9228-32.

Strausberg, R.L., Buetow, K.H., Emmert-Buck, M.R. and Klausner, R.D. 2000. The Cancer Genome Anatomy Project: building an annotated gene index. *Trends Genet* **16:** 103-106.

Strausberg, R.L., Dahl, C.A. and Klausner, R.D. 1997. New opportunities for uncovering the molecular basis of cancer [news; comment] [see comments]. *Nat Genet* **15: Spec No** 415-6.

Sugerman, P.B., Joseph, B.K. and Savage, N.W. 1995. Review article: The role of oncogenes, tumour suppressor genes and growth factors in oral squamous cell carcinoma: a case of apoptosis versus proliferation. *Oral Dis* **1:** 172-88.

Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin, M.F. and Taketo, M.M. 1998. Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. *Cell* **92:** 645-56.

Tashiro, H., Abe, K. and Tanioka, H. 1986. Familial occurrence of cancer of the mouth. *J Oral Maxillofac Surg* **44:** 322-3.

Taylor, D., Koch, W.M., Zahurak, M., Shah, K., Sidransky, D. and Westra, W.H. 1999. Immunohistochemical detection of p53 protein accumulation in head and neck cancer: correlation with p53 gene alterations. *Hum Pathol* **30:** 1221-5.

Thelu, J., Viallet, J.P. and Dhouailly, D. 1998. Differential expression pattern of the three Fringe genes is associated with epidermal differentiation. *J Invest Dermatol* **111:** 903-6.

Todd, R., Donoff, R.B. and Wong, D.T. 1997. The molecular biology of oral carcinogenesis: toward a tumor progression model. *J Oral Maxillofac Surg* **55:** 613-23; discussion 623-5.

Trent, J.M., Bittner, M., Zhang, J., Wiltshire, R., Ray, M., Su, Y., Gracia, E., Meltzer, P., De Risi, J., Penland, L. and Brown, P. 1997. Use of microgenomic technology for analysis of alterations in DNA copy number and gene expression in malignant melanoma. *Clin Exp Immunol* **107:Suppl 1,** 33-40.

Trivedy, C., Warnakulasuriya, K.A., Hazarey, V.K., Tavassoli, M., Sommer, P. and Johnson, N.W. 1999. The upregulation of lysyl oxidase in oral submucous fibrosis and squamous cell carcinoma. *J Oral Pathol Med* **28:** 246-51.

Urbain, J.L. 1999. Oncogenes, cancer and imaging. *J Nucl Med* **40:** 498-504.

Uzawa, K., Yoshida, H., Suzuki, H., Tanzawa, H., Shimazaki, J., Seino, S. and Sato, K. 1994. Abnormalities of the adenomatous polyposis coli gene in human oral squamous-cell carcinoma. *Int J Cancer* **58:** 814-7.

Vaidya, M.M., Borges, A.M., Pradhan, S.A. and Bhisey, A.N. 1996. Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa. *Eur J Cancer B Oral Oncol* **32B:** 333-6.

van der Velden, L.A., Manni, J.J., Ramaekers, F.C. and Kuijpers, W. 1999. Expression of intermediate filament proteins in benign lesions of the oral mucosa. *Eur Arch Otorhinolaryngol* **256:** 514-9.

van Oijen, M.G., Rijksen, G., ten Broek, F.W. and Slootweg, P.J. 1998. Increased expression of epidermal growth factor receptor in normal epithelium adjacent to head and neck carcinomas independent of tobacco and alcohol abuse. *Oral Dis* **4:** 4-8.

van Oijen, M.G., van de Craats, J.G. and Slootweg, P.J. 1999. p53 overexpression in oral mucosa in relation to smoking. *J Pathol* **187:** 469-74.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. 1995. Serial analysis of gene expression [see comments]. *Science* **270:** 484-7.

Venugopalan, M., Wood, T.F., Wilczynski, S.P., Sen, S., Peters, J., Ma, G.C., Evans, G.A. and Srivatsan, E.S. 1998. Loss of heterozygosity in squamous cell carcinomas of the head and neck defines a tumor suppressor gene region on 11q13. *Cancer Genet Cytogenet* **104:** 124-32.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M. and Bos, J.L. 1988. Genetic alterations during colorectal-tumor development. *N Engl J Med* **319:** 525-32.

Waber, P., Dlugosz, S., Cheng, Q.C., Truelson, J. and Nisen, P.D. 1997. Genetic alterations of chromosome band 9p21-22 in head and neck cancer are not restricted to p16INK4a. *Oncogene* **15:** 1699-704.

Wang, X.L., Uzawa, K., Imai, F.L. and Tanzawa, H. 1999. Localization of a novel tumor suppressor gene associated with human oral cancer on chromosome 4q25. *Oncogene* **18:** 823-5.

Warnecke, P.M., Stirzaker, C., Melki, J.R., Millar, D.S., Paul, C.L. and Clark, S.J. 1997. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* **25:** 4422-6.

Watson, M.A. and Fleming, T.P. 1994. Isolation of differentially expressed sequence tags from human breast cancer. *Cancer Res* **54:** 4598-602.

Weinberg, R.A. 1995. The retinoblastoma protein and cell cycle control. *Cell* **81:** 323-30.

Weiner, T. and Cance, W.G. 1994. Molecular mechanisms involved in tumorigenesis and their surgical implications. *Am J Surg* **167:** 428-34.

Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D. and McClelland, M. 1992. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* **20:** 4965-70.

Wilkinson, M.M., Busuttil, A., Hayward, C., Brock, D.J., Dorin, J.R. and Van Heyningen, V. 1988. Expression pattern of two related cystic fibrosis-associated calcium- binding proteins in normal and abnormal tissues. *J Cell Sci* **91:** 221-30.

Wodarz, A. and Nusse, R. 1998. Mechanisms of Wnt signaling in development. *Annu Rev Cell Dev Biol* **14:** 59-88.

Wright, J.M. 1998. A review and update of oral precancerous lesions. *Tex Dent J* **115:** 15-9.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W. 1997. Gene expression profiles in normal and cancer cells. *Science* **276:** 1268-72.

Zheng, T.Z., Boyle, P., Hu, H.F., Duan, J., Jian, P.J., Ma, D.Q., Shui, L.P., Niu, S.R., Scully, C. and MacMahon, B. 1990. Dentition, oral hygiene, and risk of oral cancer: a case-control study in Beijing, People's Republic of China. *Cancer Causes Control* **1:** 235-41.

APPENDICES

APPENDIX A

Oncogene (2000) 19, 3220–3224

www.nature.com/onc

SHORT REPORT

Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays

Chidchanok Leethanakul[1], Vyomesh Patel[1], John Gillespie[2], Michael Pallente[1], John F Ensley[3], Sittichai Koontongkaew[4], Lance A Liotta[2], Michael Emmert-Buck[2] and J Silvio Gutkind*[,1]

[1]*Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, MD 20892, USA;* [2]*Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, MD 20892, USA;* [3]*Division of Hematology/Oncology, Wayne State University, Detroit, Michigan, MI, USA;* [4]*Department of Oral Biology and Occlusion, Faculty of Dentistry, Prince of Songkhla University, Songkhla, Thailand*

Although risk factors for squamous cell carcinomas of the head and neck (HNSCC) are well recognized, very little is known about the molecular mechanisms responsible for this malignancy. Furthermore, the ability to investigate gene expression profiles at different stages of tumor progression is usually limited by the remarkable heterogeneity of these neoplastic lesions. Here, we show the successful use of laser capture microdissection (LCM) to procure specific cell populations. The 5000 cells from representative sets of HNSCC and their matching normal tissues are sufficient to extract RNA of high integrity for the synthesis of labeled amplified cDNA probes which can then be hybridized to these membranes arrayed with known human cancer-related cDNAs. Furthermore, when compared to normal tissues, we demonstrate a consistent decrease in expression of differentiation markers such as cytokeratins, and an increase in the expression of a number of signal transducing and cell cycle regulatory molecules, as well as growth and angiogenic factors and tissue degrading proteases. Unexpectedly, we also found that most HNSCC overexpress members of the *wnt* and *notch* growth and differentiation regulatory system, thus suggesting that the *wnt* and *notch* pathways may contribute in squamous cell carcinogenesis. This experimental approach may facilitate the identification candidate markers for the early detection of preneoplastic lesions, as well as novel targets for pharmacological intervention in this disease. *Oncogene* (2000) **19**, 3220–3224.

Keywords: oral cancer; tumor progression; oncogenes; tumor suppressor genes; biomarkers

Squamous cell carcinoma of the head and neck (HNSCC) is the sixth most common cancer in developed countries, and of the 44 000 annual cases reported in the United States, approximately 11 000 will result in an unfavorable outcome (Landis *et al*., 1999; Parkin *et al*., 1999). In spite of its high incidence, the molecular mechanisms of this disease remain poorly understood. However, the recently gained knowledge of normal and aberrant function of oncogenes and tumor suppressor genes has provided unique opportunities to understand, and ultimately to control, the processes leading to malignancy. Thus, the identification of the molecular and genetic events involved in each step of tumor progression may be central to understand HNSCC, and for the development of diagnostic markers and novel treatment strategies.

Although HNSCC is thought to result from the progressive accumulation of genetic lesions leading to malignancy (Mao *et al*., 1998) the precise nature of the affected molecules is still largely unknown. The recent development of several high throughput, hybridization-based methods utilizing cDNAs arrayed on nylon membranes and glass slides allows the analysis of hundreds of genes simultaneously, and thus provides a unique opportunity to identify genes expressed in normal and tumor tissues, as well as to analyse gene expression profiles in tumor progression. However, an accurate procurement of specific cell types for RNA isolation is a critical step influencing the validity of this analysis. In this regard, a novel technique of Laser Capture Microdissection (LCM) developed at the Laboratory of Pathology (National Cancer Institute), enables the procurement of pure cell populations from frozen human tissue sections (Emmert-Buck *et al*., 1996; Simone *et al*., 1998), a key consideration as many tumors, including HNSCC, are heterogeneous, and include areas of connective tissues, blood vessels and even inflammatory cells that infiltrate into the tumor mass. Most importantly, microdissection allows analysis of gene expression in specific cell populations as it exists *in situ*. In this study, we have used LCM to procure specific cell populations from a representative set of tumors and their matching normal tissues to explore the feasibility of establishing a pattern of expression of cancer-related genes for HNSCC.

*Correspondence: JS Gutkind, Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive, Building 30, Room 212, Bethesda, MD 20892-4330, USA
Received 6 December 1999; revised 9 March 2000; accepted 25 May 2000

Extraction of RNA from HNSCC and normal epithelial cells procured by laser capture microdissection

Clinical characteristics of the human biopsies from HNSCC patients who had undergone surgery and were chosen for the study are indicated in Table 1. The

Table 1 Clinical characteristics of lesions from patients with HNSCC

Case	*Origin*	*Lesion*	*Pathology*
WSU 1	Tongue	Hyperplasia	Severe hyperplasia
WSU 51	Laryngeal	Carcinoma	High clinical grade Invasive and poorly differentiated
WSU 58	Pharyngeal	Carcinoma	Mild to moderate dysplasia, moderate to well differentiated and invasive in places
WSU 62	Tongue	Carcinoma	Invasive, moderate to well differentiated
WSU 63	Tongue	Carcinoma	Moderate to poorly differentiated and invasive

HNSCC lesions upon biopsy were analysed for clinical classification. The five tissue sets (WSU 1, 51, 58, 62, 63) were biopsies from patients previously confirmed to have neoplastic lesions of the head and neck. Anatomical site and severity of the lesions are indicated

anatomical sites of these lesions are representative of the most frequent HNSCC sites, and include the tongue (WSU 1, 62 and 63), larynx (WSU 51), and pharynx (WSU 58). The pathology that was provided with the tissues indicated that all lesions except WSU 1, were carcinomas, either poorly (WSU 51, 63) or moderate to well differentiated (WSU 58, 62). All tumors were invasive. Tissue WSU 1 was confirmed as hyperplasia. Corresponding normal tissue, from the same anatomical site and patient as the lesion, was part of the tissue set and consisted of normal epithelium. We considered these five tissue sets to be representative of HNSCC, and thus suitable for assessing the feasibility of using LCM for the detailed analysis of gene expression in these cancer lesions. Before proceeding, the histopathology was confirmed by a board certified pathologist. Microscopic visualization of representative frozen tissue sections (8 μm thickness), stained with hematoxylin-eosin is shown in Figure 1A (a–d).

As shown in Figure 1B, the use of LCM enables the procurement of a pure population of squamous cells (>95% purity) that can be used to extract RNA. Quality assessment of the RNA extracted from each tissue was performed by RT–PCR of GAPDH. As demonstrated in Figure 1C, RNA from two representative tissue sets (WSU 62, 63) was of sufficient quality to be reverse transcribed and amplified using specific primers for GAPDH, which generated a 600 bp product. Similar results were obtained with the remaining three tissue sets (data not shown), thus supporting that LCM preserves the integrity of the RNA extracted from HNSCC cells.

Hybridization and gene expression

Total RNA was reverse-transcribed using Superscript Reverse Transcriptase (Life Technologies) and cDNAs synthesized for each sample were assessed for integrity. In all cases, the average size of the reverse-transcribed messages was approximately 500 bp (range 300–800 bp, data not shown), which is similar to that achieved for the construction of highly representative cDNA libraries (Peterson *et al*., 1998). cDNAs corresponding to normal and tumor tissue from the same set were simultaneously amplified and labeled with α-^{33}P dCTP. Labeled amplified cDNAs (AcDNAs) were then used to hybridize human cancer cDNA arrays (CLONTECH) containing 200–500 bp DNA fragments, in duplicates, for 588 known human cancer and nine housekeeping genes (http://www.clontech.com/atlas/genelists/). In preliminary experiments, this procedure resulted in a highly reproducible pattern of gene expression when using the same RNA preparation and independently AcDNAs. Membranes hybridized with complex cDNA probes from a representative tissue set are shown in Figure 2A and illustrates the comparative differences in expression of genes belonging to different functional groups in both normal (upper panel) and tumor tissue (lower panel) from the same patient.

Genes differentially expressed

Examples of genes differentially expressed are shown in Figure 2B and include the cytokeratins (a) and those genes belonging to the MAPK (mitogen-activated protein kinase) (b) and *wnt* (wingless) (c) signaling pathways. The amount of radiolabeled probe hybridized to each arrayed cDNA was quantified using PhosphorImaging, and normalized by that hybridized to the housekeeping genes (GAPDH). In preliminary experiments, we found that under these experimental conditions differences of $\geqslant$ twofold were reproducible, and the radioactivity within the linear range of detection. To simplify the analysis, the main functional groups of genes assessed to be differentially expressed $\geqslant$ twofold in at least three of the four tumor sets when comparing each cancerous epithelium to their corresponding normal tissue were considered of likely biological significance, and are listed in Table 2. They include genes involved in the control of cell growth and differentiation, angiogenesis, apoptosis, cell cycle, and signaling, most of which have not been previously implicated in HNSCC when using other analytical approaches. These data indicate that complex AcDNA probes labeled with α-^{33}P dCTP can be successfully synthesized from small amounts of total RNA for their use in comparative hybridization studies of cancer genes expressed in HNSCC. Furthermore, these findings have helped identify a number of new candidate genes, which might play an unexpected role in squamous carcinogenesis.

One of the most remarkable changes was the general decrease in the expression of cytokeratins (2E, 2P, 6A-F, 7, 13, 14, 15, 17, 18, 19), which were readily detected in normal tissue but nearly absent in the cancer cell population (2–20-fold reduction), most likely reflecting the loss of differentiation in tumor cells. On the other hand, we observed a clear increase in the levels of *cyclin D1* (2–3-fold), as previously reported (Bartkova *et al*., 1995), and those of metalloproteases (*MMP-7, MMP-10, MMP-14*), which reflects the highly invasive behavior of this tumor type. Furthermore, we observed a remarkable increase in the levels of many growth and angiogenic factors including *TGFα*, *TGFβ*, *EGF Cripto protein*, *PDGF A chain* and *B chain* (*c-Sis*), different *FGF* isoforms, *HGF*, and *VEGF-C*. This supports the conclusion that this tumor type secretes factors that are likely to induce epithelial cell growth in an autocrine fashion in addition to promoting the growth of stromal cells and the process of neovascularization (Desai and Libutti, 1999). Furthermore, these tumors also overexpress several proapoptotic molecules, including

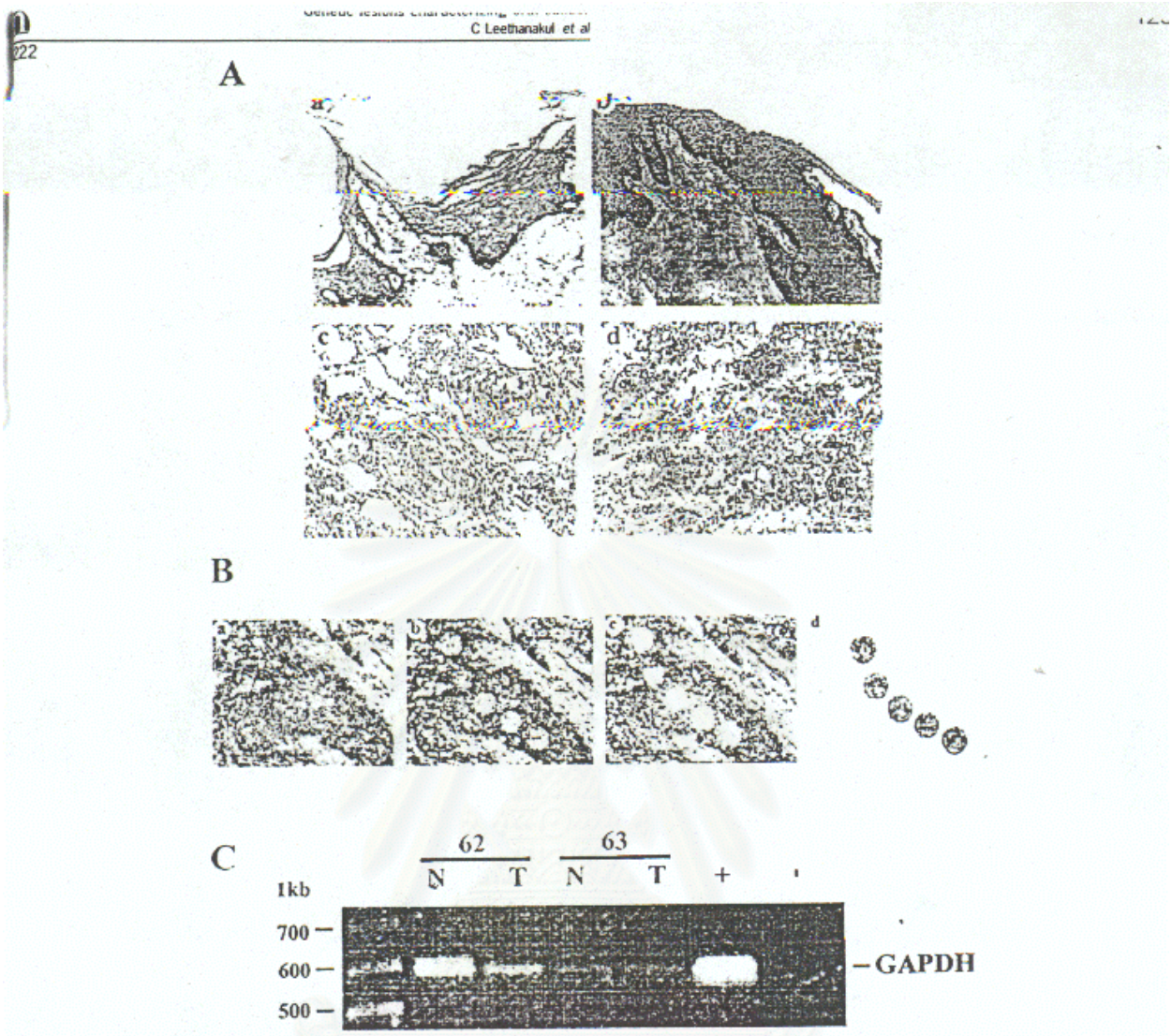


Figure 1 (A) Histopathological features of HNSCC. Tissue sets, comprising of both normal and tumor from the same HNSCC patient, were snap frozen and 8 μm sections were stained with H&E. Histopathological features of progression of HNSCC from normal to carcinoma are illustrated. Representative normal squamous epithelium (a) from head and neck region shows an orderly maturation from deep to superficial cell layers marked by progressive flattening of the cells and nuclei (magnification× 200). Hyperplasia (b) is characterized by increased layers of epithelial cells throughout the lower third of the epithelium, normal maturation, and intact basement membrane (magnification × 200). Well-differentiated invasive squamous cell carcinoma (c) shows infiltration of the underlying stroma, loss of normal architecture, occasional keratin pearls. There is mild nuclear pleomorphism and increased numbers of mitotic figures (magnification × 500). Moderate to poorly-differentiated invasive carcinoma (d) shows infiltration of the underlying stroma with sheets of cancer cells showing marked nuclear pleomorphism and hyperchromasia. There is notable absence of keratin pearls (magnification × 500). (B) For LCM procurement, an area of tumor, containing cells of interest are visualized (a) and targeted for capture with a 30 μm diameter laser beam (b). The caps containing the captured cells are lifted off the tissue section (c), and the homogeneity of these cells is confirmed under a light microscope (d) prior to processing for RNA extraction. Each laser beam procures 3–7 cells (magnification × 500). (C) Total RNA was extracted from microdissected tissues (approximately 5000 cells) using guanidium isothiocyanate (GITC) and phenol/chloroform extractions, isopropanol precipitation, and treatment with DNAse I in the presence of RNAse inhibitors. The integrity of the RNA was assessed by amplifying GAPDH by RT–PCR, using specific primers for GAPDH (sense primer 5CCACCCATGGCAAATTCCATGGCA-3, antisense primer 5-TCTAGACGGCAGGTCAGGTCCACC-3). PCR reactions were performed using GeneAmp RNA PCR kit (Perkin Elmer), and DNA products analysed on a 1.2% agarose/EtBr gel. A 600 bp fragment is amplified, as observed for both normal (N) and tumor (T) tissues from HNSCC patients (WSU62 and 63). Appropriate positive and negative controls are indicated

caspase precursors, *Bcl-W*, *Bax*, and *Bag-1*, but might survive apoptotic signals through the overexpression of apoptosis inhibiting molecules, including *Akt2* and *IAP* (inhibitor of apoptosis) (Ambrosini *et al*., 1997; Granville *et al*., 1998). Also shown in Figure 2B, cancer cells overexpressed signaling molecules participating in the MAP kinase pathway, including *ERK1*, all isoforms of *JNK* (1-3), two *p38* related MAPKs, *p38* and *ERK6*, and their upstream activators, *MEKK3* and *MKK6*, which is likely to contribute to the enhanced growth stimulation in these cells. Interestingly, the hyperplastic tissue exhibited a pattern of gene expression nearly identical to that of the adjacent normal epithelium, being only *cyclin D1* and DP_2 over-

Figure 2 Analysis of gene expression in HNSCC using cDNA arrays. For each HNSCC tissue set, AcDNA probes were prepared and used simultaneously for the hybridization of nylon membranes arrayed in duplicate with human cancer and housekeeping genes. Probes were synthesized using 2 μl of cDNA for each reaction from tissue sets (normal and tumor), 2 μl of dNTP mix (2.5 mM of dGTP, dATP, dTTP and 0.5 mM of dCTP, Perkin Elmer), 75 μl ddH_2O, 10 μl 10 × PCR reaction buffer and 4 μl 10 mM PCR primers (CLONTECH Smart kit), 5 μl 10 mCi/ml α-^{33}P dCTP and 2 μl Taq polymerase (Perkin Elmer). Each sample was mixed and used to synthesis complex cDNA probe by PCR (1 min at 95C, then 30 cycles of 15 s at 95 C, 5 s at 65 C and 5 min at 68 C, and then cooled to 4 C). For each sample, 3–4 independent PCR reactions were carried out and combined after purification (PCR SELECT-II columns; 5Prime-3Prime, Inc.). The specific activity of each cDNA probe was assessed by scintillation counting. Human cancer cDNA expression arrays (CLONTECH) and the conditions used for hybridization were essentially as described in the manufacturer's protocol. After extensive washes, the membranes were analysed by PhosphorImaging (Molecular Dynamics) and auto-radiography. The hybridization of the cDNAs on the arrays for each of the samples (normal and tumor) was documented, and the identity of the genes determined from the relative position, as provided by the manufacturer. The comparative level of expression for each gene was assessed by PhosphoImaging, and expressed relative to that of the housekeeping genes (GAPDH). Pattern of gene expression for a representative tissue set from the same HNSCC patient is shown. Differentially expressed genes in three or more HNSCC tissue sets were considered of likely biological significance, and examples of those are indicated

expressed, which might correlate with increased cell proliferation. Thus, collectively these results demonstrate that squamous cells exhibit a distinct pattern of gene expression, which might help explain many of the cellular abnormalities described in this tumor type.

Unexpectedly, genes involved in the *wnt* and *notch* signaling pathway, were found to be highly represented in tumor tissues (2–5-fold). High expression levels of some of these genes have been demonstrated in many neoplasias and may have an implication in maintaining an undifferentiated epithelium (Liu *et al*., 1996; Shelly *et al*., 1999). For *wnt*, two *wnt* receptors, *frizzled* and *FDZ3*, and their downstream targets, *dishevelled* and *β-catenin* (Wodarz and Nusse, 1998) were highly expressed. In the case of *notch*, the detection of both receptor and ligand (*notch* and *jagged*, respectively) also suggests strongly their constitutive activation (Artavanis-Tsakonas *et al*., 1999). Furthermore, two of the *fringe* genes, *Manic* and *Lunatic*, which encode pioneer secretory proteins that modulate Notch-ligand interactions (Panin *et al*., 1997) were similarly highly represented. Thus, together these findings support an unexpected role for the *notch* signaling system in squamous cell carcinogenesis. Their precise role in the pathogenesis of HNSCC is currently unknown, and warrants further investigation.

Taken together, we can conclude that the use of LCM and cDNA arrays has allowed the detailed analysis of gene expression in HNSCC, and provided the first evidence for the feasibility of performing a comprehensive molecular characterization of normal, premalignant, and malignant HNSCC cells. Although data obtained involved a limited set of tissue samples, a general trend is already observed, implicating cell cycle regulating and signaling molecules, growth and angiogenic factors, matrix degrading proteases, and survival and apoptotic molecules. Furthermore, we obtained evidence implicating, for the first time, the *notch* and *wnt* pathways in squamous cell carcinogenesis. Further analysis of a more extensive sample collection using conventional and these recently available technologies will make it possible to define a pattern of gene expression in a tumor progression model of HNSCC. This experimental approach is also expected to facilitate the identification of candidate markers potentially correlated with malignancy, thus providing valuable tools of diagnostic and prognostic value to study premalignant lesions.

Table 2 Genes highly represented in HNSCC

Cell cycle		**Signaling molecules**	
L33264	cdc-2 Related kinase	X60811	ERK1
X59768	Cyclin D1	X79483	ERK6
U11791	Cyclin H	U82532	GDI Dissociation protein
U40345	p19INK4D	L35253	MAP kinase p38
U18422	DP-2	L26318	JNK1
		L31951	JNK2
Angiogenesis		U34819	JNK3
		U39657	MKK6
L12350	Thrombospondin2 precursor	U78876	MEKK3
U43142	VEGF-C	M31470	Ras-like protein (TC10)
X07819	MMP-7	L25080	RhoA
X07820	MMP-10		
D50477	MMP-16	Apoptosis	
Z30183	TIMP-3		
U76456		U45878	Inhibitor of Apoptosis Protein
		U37448	Caspase 7 precursor
Growth Factors and Oncogenes		U60520	Caspase 8 precursor
		U28014	Caspase 5 precursor
M11730	HER2	M77198	AKT2
M37722	bFGFR 1	U59747	BCL-W
Z12020	MDM2	U78798	TRAF2
M31213	c-RET	S83171	BAG-1
D17517	SKY	L22474	BAX
X87838	β-Catenin		
X14445	FGF 3	WNT/Notch signaling system	
M37825	FGF 5		
X63454	FGF 6	M73980	Notch 1
M65062	IGFBP5	U77493	Notch 2
M60718	HGF precursor	M99437	Notch group protein (N)
K03222	TGF-α	AF028593	Jagged 1
X02812	TGF-β	L37882	Frizzled
M96956	EGF Cripto protein CR1 and 2	U82169	Frizzled homolog (FZD3)
X06374	PDGF-α	U46461	Dishevelled homolog
X02811	c-SIS	U43148	Patched homolog
M74088	APC	U94352	Manic fringe
		U94354	Lunatic fringe

Gene expression pattern for each HNSCC tissue set (WSU 1, 51, 58, 62, 63) was analysed by PhosphorImaging and the relative amount of expression was compared with those of housekeeping genes. Those genes judged to be differentially expressed at twofold in at least three of the cancer tissue sets were considered of likely biological significance, and are listed in their corresponding functional groups. The GeneBank access number for each gene is included

Acknowledgments

We thank Dr Salvatore Pece and Dr Hiroshi Miyazaki for the critical reading of the manuscript and discussions. We are grateful to Drs Kristina Cole and William D Swaim for helpful advice on cDNA labeling and histology respectively.

References

Ambrosini G, Adida C and Altieri DC. (1997). *Nat. Med.*, 3, 917–921.

Artavanis-Tsakonas S, Rand MD and Lake RJ. (1999). *Science*, **284**, 770–776.

Bartkova J, Lukas J, Muller H, Strauss M, Gusterson B and Bartek J. (1995). *Cancer Res.*, 55, 949–956.

Desai SB and Libutti SK. (1999). *J. Immunother.*, **22**, 186–211.

Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA and Liotta LA. (1996). *Science*, **274**, 998–1001.

Granville DJ, Carthy CM, Hunt DW and McManus BM. (1998). *Lab. Invest.*, **78**, 893–913.

Landis SH, Murray T, Bolden S and Wingo PA. (1999). *CA Cancer J. Clin.*, **49**, 8–31.

Liu Y, Dehni G, Purcell KJ, Sokolow J, Carcangiu ML, Artavanis-Tsakonas S and Stifani S. (1996). *Genomics*, 31, 58–64.

Mao EJ, Schwartz SM, Daling JR and Beckmann AM. (1998). *J. Oral Pathol. Med.*, 27, 297–302.

Panin VM, Papayannopoulos V, Wilson R and Irvine KD. (1997). *Nature*, **387**, 908–912.

Parkin DM, Pisani P and Ferlay J. (1999). *CA Cancer J. Clin.*, **49**, 33–64.

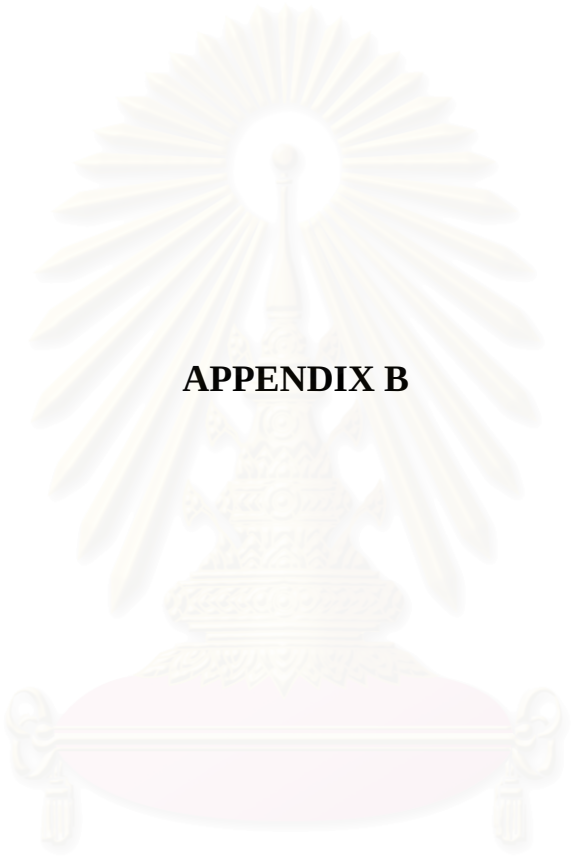
Peterson LA, Brown MR, Carlisle AJ, Kohn EC, Liotta LA, Emmert-Buck MR and Krizman DB. (1998). *Cancer Res.*, **58**, 5326–5328.

Shelly LL, Fuchs C and Miele L. (1999). *J. Cell Biochem.*, **73**, 164–175.

Simone NL, Bonner RF, Gillespie JW, Emmert-Buck MR and Liotta LA. (1998). *Trends Genet.*, **14**, 272–276.

Wodarz A and Nusse R. (1998). *Annu. Rev. Cell. Dev. Biol.*, **14**, 59–88.

APPENDIX B



PERGAMON

Oral Oncology 36 (2000) 474–483

ORAL ONCOLOGY

www.elsevier.com/locate/oraloncology

Gene expression profiles in squamous cell carcinomas of the oral cavity: use of laser capture microdissection for the construction and analysis of stage-specific cDNA libraries

C. Leethanakul [a,1], V. Patel [a,1], J. Gillespie [b], E. Shillitoe [c], R.M. Kellman [d], J.F. Ensley [e], V. Limwongse [f], M.R. Emmert-Buck [b], D.B. Krizman [g,*], J.S. Gutkind [a]

[a] *Oral & Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive, Building 30, Room 212, Bethesda, MD 20892-4330, USA*

[b] *Pathogenetics Unit, Laboratory of Pathology, Cancer Genome Anatomy Project, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-4330, USA*

[c] *Department of Microbiology and Immunology, SUNY Upstate Medical University, Syracuse, NY 13210, USA*

[d] *Department of Communication Sciences, SUNY Upstate University, Syracuse, NY 13210, USA*

[e] *Division of Hematology/Oncology, Wayne State University, Detroit, MI 48201, USA*

[f] *Faculty of Dentistry, Chulalongkorn University, Bangkok 10330, Thailand*

[g] *Cancer Genome Anatomy Project, Office of the Director, National Cancer Institute, National Institutes of Health, Advanced Technology Center, 8717 Grovemont Circle, 134F, Gaithersburg, MD 20877, USA*

Received 27 March 2000; accepted 3 April 2000

Abstract

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer among men in the developed world affecting the oral cavity, salivary glands, larynx and pharynx. Utilizing tissue from patients with HNSCC, we sought to systematically identify and catalog genes expressed in HNSCC progression. Here, we demonstrate the successful use of laser capture microdissection for procuring pure populations of cells from patient tissue sets comprised of oral squamous cell carcinomas (OSCCs) and matching normal tissue. From the estimated 5000 cells procured for each sample, we were able to extract total RNA (14.7–18.6 ng) of sufficient quality to transcribe GAPDH by reverse transcriptase-polymerase chain reaction (RT-PCR). The RNA was used for the synthesis of blunt-ended, double-strand complementary DNAs (cDNAs) by oligo (dT)-mediated reverse transcription, followed by addition of linkers. Primers specific for these linkers with uracil deglycosylase-compatible ends were used to amplify these cDNAs by PCR and the product was subcloned into the pAMP10 cloning vector. Ninety-six clones from each of six libraries were randomly sequenced and results indicated that 76–96% of the inserts represent either anonymous expressed sequence tags (ESTs) (25–48%), known genes (9–29%) or novel sequences (27–51%), respectively, with very little redundancy. These results demonstrate that high quality, representative cDNA libraries can be generated from microdissected OSCC tissue. Furthermore, these finding suggest the existence of at least 132 novel genes expressed in our cDNA libraries, which may have a role in the pathogenesis of HNSCC, and may represent novel markers for early detection as well as targets for pharmacological intervention in this disease. Published by Elsevier Science Ltd.

Keywords: Oral cancer; Gene expression; CGAP; Tumor progression

1. Introduction

From the estimated 100,000 genes in the human genome, 4000 of these may be directly related to disease, including cancer [1]. Indeed, altered expression of some of these genes is now thought to be the basis of most neoplasias, either because they are expressed at abnormally high or low levels, or due to their ability to encode aberrant proteins upon mutations in their coding sequence [2]. In this regard, the availability of a catalog of genes expressed in tumor cells may provide a fingerprint of their genetic make up, and comparison with that of their matching cells exhibiting a normal phenotype can help identify genes that either by their presence

* Corresponding author. Tel.: +1-301-402-3134; fax: +1-301-435-5155.

E-mail addresses: dkrizman@helix.nih.gov (D.B. Krizman), sg39v@nih.gov (J.S. Gutkind).

[1] Equal contribution by the authors.

1368-8375/00/$ - see front matter Published by Elsevier Science Ltd.
PII: S1368-8375(00)00039-7

or absence, can be causal in cancer. It follows that knowing the identity of these genes will not only enhance our understanding of the molecular basis of this disease and its progression, but it will also provide novel means for its early detection and subsequent treatment.

In response to our limited knowledge of the molecular mechanisms of many neoplasias, the Cancer Genome Anatomy Project (CGAP) supported by the National Cancer Institute (NCI), was established with the goal of creating a complete information infrastructure of genes expressed during tumor progression, which is also expected to yield early markers of cancer, thus providing an opportunity to improve our ability to match patients with appropriate treatment strategies. The CGAP initiative involves the generation of complementary DNA (cDNA) libraries from cancer cells, and after random sequencing, expressed genes are then cataloged and compared with those from the corresponding normal tissues. In doing so, CGAP has also become the leading effort in gene discovery. Further success of this approach has been the development of robust databases and easily accessible Web-based analytical tools for comparative use [3–5].

Squamous cell carcinoma of the head and neck (HNSCC) are neoplastic lesions found predominantly in the oral cavity, including the salivary glands, larynx and pharynx [6]. Despite recent advances in our understanding, prevention, and treatment of other types of neoplasias, HNSCC still remains the sixth most common cancer among men in the developed world [7] and in the United States alone approximately 13,000 deaths occur yearly as a result of this disease [8]. The high morbidity rate for this malignancy can be attributed to many factors, which include lack of suitable markers for early detection, late presentation, insensitivity to available treatment, and our limited understanding of the molecular mechanisms responsible for this disease [9]. In this regard, the identity of those genes that may have a role in the progression of HNSCC has yet to be fully elucidated. Therefore, in an attempt to begin addressing the molecular basis of this cancer, the Head and Neck CGAP (HNCGAP) was established as a cooperative effort between the National Institute of Dental and Craniofacial Research (NIDCR) and the National Cancer Institute's CGAP initiative.

A major scientific challenge in HNSCC is our understanding of the molecular events that drive tumor progression in vivo [10]. This problem is further compounded by the heterogeneity of this tumor type. Thus, gene expression analysis using bulk tissue or tissue areas of interest manually microdissected, might not be representational and of limited value when using this body of information for assessing gene expression profiles in HNSCC. Of interest, the use of laser capture microdissection (LCM) [11–13] allows the procurement of pure cell populations for RNA isolation, thus providing an appropriate platform for current efforts in defining the nature of those genes expressed in HNSCC, and their potential contribution to neoplasia [14,15].

In this study we have used HNSCC and their matching normal tissues from patients with oral cancer lesions. We demonstrate the successful use of LCM to procure specific cell populations. Furthermore, we show that 5000 cells are sufficient to extract RNA of high integrity for the synthesis of high-quality representational cDNAs libraries. Furthermore, sequence analysis of randomly selected clones from each library indicates that 76–96% of the inserts represented anonymous expressed sequence tags (ESTs) (25–48%), known genes (9–29%) or novel sequences (27–51%), respectively, and with very little redundancy among libraries. Emerging sequence information suggests the existence of many novel genes, whose function in tumor development can now begin to be evaluated.

2. Materials and methods

2.1. Tissue samples and LCM

Biopsies from patients confirmed to have carcinomas of the oral cavity were immediately fixed in 70% ethanol and subsequently embedded in optical cutting temperature (OTC) as described (http://dir.nichd.nih.gov/lcm). Using a cryostat, 8-μm thick tissue sections were cut onto RNAase free glass slides, and prior to LCM, hematoxylin and eosin (H&E) -stained sections were analyzed and confirmed by a board-certified pathologist. The use of LCM (Arcturus Engineering, Mountain View, CA, USA) was essentially as described [13]. This procedure enables the enrichment of a pure cell population onto plastic caps coated with a transparent ethylene vinyl acetate thermoplastic film. Microdissected cells were assessed microscopically and the caps containing the procured cells were subsequently transferred to a 0.5 ml RNAase free microfuge tube containing appropriate lysis buffer for extraction of total RNA.

2.2. Extraction of RNA and assessment of RNA integrity

The procedure of extracting RNA from microdissected cells is described in detail elsewhere (http://dir.nichd.nih.gov/lcm). Briefly, cells were digested in lysis buffer containing guanidium isothiocyanate and β-mercaptoethanol and total RNA extracted from the supernatant with 2 M sodium acetate, saturated phenol and chloroform-isoamyl alcohol. The RNA was subsequently precipitated from the resulting aqueous layer with ice-cold isopropanol and glycogen (10 μg/μl). The resulting RNA pellet was treated with DNAase I

(10 units/μl) and RNAase inhibitor (20 units/μl), reprecipitated and resuspended in 3.5 μl deionized water and 1 μl of 20 U/μl RNAase inhibitor. The integrity of each RNA sample was assessed as described. Briefly, 1 μl of RNA was reverse transcribed using the GeneAmp RNA PCR kit (Perkin Elmer, Branchburg, NJ, USA) following the supplied protocol, and 2 μl of the resulting cDNA was used to amplify *GAPDH* with specific primers using polymerase chain reaction (PCR) conditions according to the manufacturer's recommendations (Perkin Elmer).

2.3. Quantification of total RNA extracted from microdissected cells

Total RNA extracted from microdissected cells was quantified using the VersaFluor™ Fluorometer system following the manufacturer's recommendations (Bio-Rad Laboratories, Hercules, CA, USA). Briefly, 1 ml of TE (1X) containing 1 μl of total RNA was mixed with an equal volume of diluted Ribogreen probe (1:3000 in 1X TE). After incubation (5 min), the samples were measured for RNA using excitation and emission wavelength of 495 and 525 nm, respectively. The amount of total RNA was quantified against standards (1–50 ng), which were prepared according to the manufacturer's recommendations.

2.4. Synthesis of cDNAs from total RNA and library construction

Synthesis of double-strand cDNAs from total RNA extracted from microdissected cells was as previously described [14]. Briefly, using the Superscript Choice System (Life Technologies Inc., Gaithersburg, MD, USA) and following the supplied protocol, RNA samples from microdissected tissues (5000 cells) were used as templates to reverse transcribe first-strand cDNAs. The second-strand replacement and *Eco*RI linker addition reactions were followed as described [13]. Briefly, RT reactions containing 5X second strand buffer, 10 mM dNTPs, *Escherichia coli* DNA ligase, *E. coli* DNA Pol I, and *E. coli* Rnase H were incubated at 16°C for 2 h and after the addition of 2 μl of T4 DNA polymerase, the reactions were incubated for a further 10 min. After extracting and precipitating with phenol/chloroform and ethanol, respectively, the cDNA pellets were resuspended in ddH_2O and used to ligate *Eco*RI linkers. The linker-ligated cDNAs were purified in 1% low-melting agarose and cDNA products (0.4–2 Kb) were excised and the agarose was digested overnight at 37°C with β-agarase (New England Biolabs, Beverly, MA, USA). After extracting with phenol/chloroform and precipitating with ethanol and glycogen (20 μg/ml) respectively, the cDNA pellets were washed with 70% ethanol, air dried and resuspended in 20 μl ddH_2O. Two microliters of linker-ligated cDNAs was used to assess the quality on a 1.2% agarose gel.

2.5. PCR amplification of double-strand cDNA

Amplification of double-strand cDNAs by PCR was essentially as described [14]. Briefly, 18 μl cDNA of template, ddH_2O, 10X PCR buffer, 10 mM dNTPs, 1 μM linker-specific primers and *Taq* polymerase, was used for a single PCR reaction. Conditions for PCR were 3 min at 94°C followed by 15 cycles at 94°C for 15 s, 65°C for 15 s, 72°C for 3 min and a final extension at 72°C for 5 min. The PCR products were purified, isopropanol-precipitated and washed with 70% ethanol prior to resuspending in 10 μl ddH_2O.

2.6. Subcloning and assessment of cDNA libraries

One microliter of each of the amplified cDNAs was non-directionally cloned into the uracil deglycosylase (UDG) cloning vector pAMP10 (Life Technologies) and the cloning reactions were subsequently used for transformation. Resulting clones were randomly picked for assessing insert size and sequencing. Universal primers (M13f and M13r) for pAMP10 were used to amplify inserts with 30 cycles of standard PCR reaction and these reactions subsequently analyzed for diversity by assessing the insert size on a 1.2% agarose gel.

2.7. Sequencing of cDNA libraries and data analysis

Recombinant clones from each library were randomly picked, expanded in 96-well plates, and sequenced using M13 forward primers. Individual clone sequences were subsequently analyzed by searching available data bases (GenBank, dbEST) by the Basic Local Alignment Search Tool (BLAST) program accessed through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) for the search of homology to known genes or ESTs.

3. Results

3.1. Scheme of experimental procedure

The experimental strategy for this study is illustrated in Fig. 1. Normal and pathological oral squamous epithelium were visualized under the microscope and appropriate cells were microdissected with individual laser shots. Caps containing approximately 5000 cells were processed for RNA and subsequently assessed for quality. The mRNA served as a template for library production. After transformation, clones from

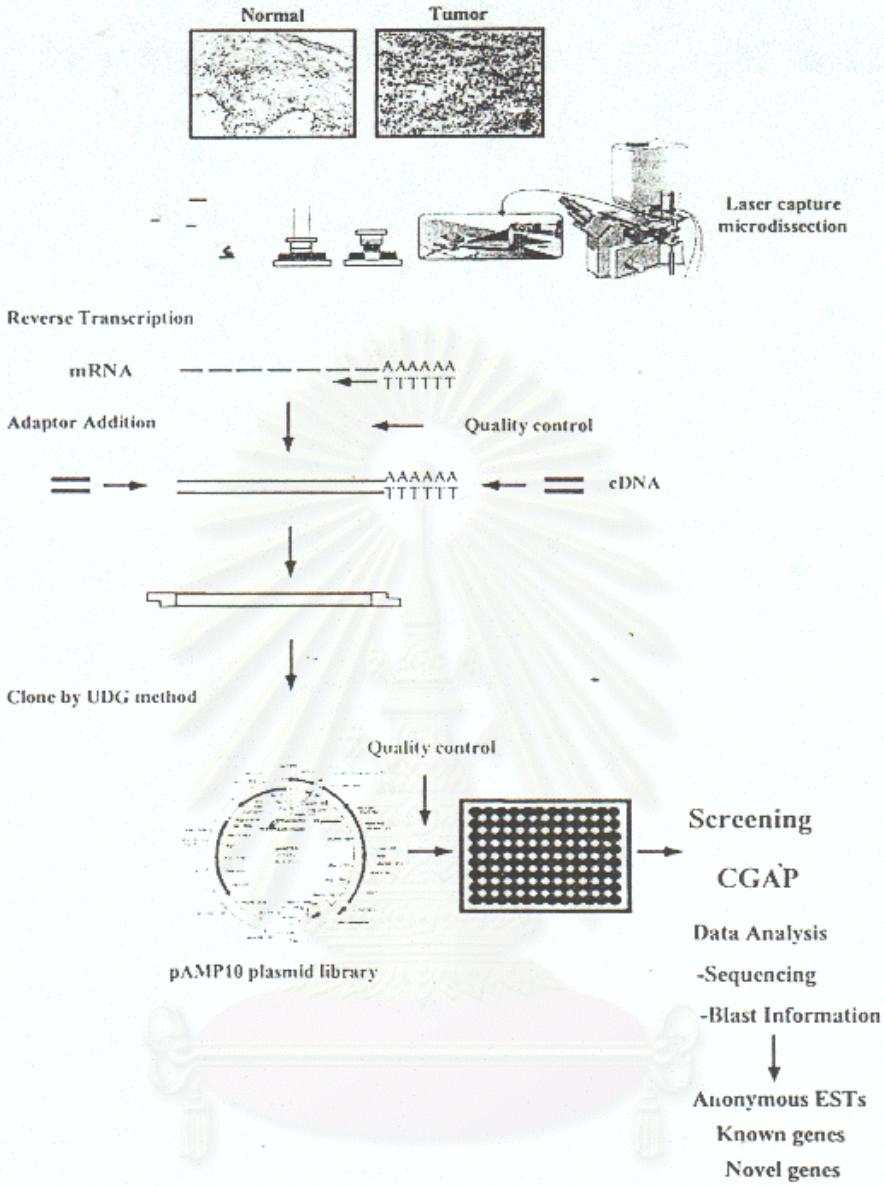


Fig. 1. Scheme illustrating the experimental procedure used in the study. Normal and malignant oral keratinocytes are procured by laser capture microdissection and the RNA extracted. After assessment of the quality, the RNA samples are reverse transcribed and with the resulting cDNAs, adaptors are ligated to enable cloning. Clones from each cDNA library are sequence analyzed and prepared for Cancer Genome Anatomy Project (CGAP). The available data are further analyzed using the Basic Local Alignment Search Tool (BLAST) program to determine whether any of the information matches to anonymous expressed sequence tags (ESTs), known genes or represent novel genes.

each cDNA library were sequence analyzed and prepared for CGAP submission. The sequencing data were analyzed using the BLAST software to assess whether individual clones matched previously identified genes or anonymous ESTs, or represented novel sequences.

3.2. Clinical and histological feature of HNSCC tissues samples

Clinical samples were obtained from HNSCC patients that had undergone surgery. Samples used for this study were from male patients and the age ranged from 46 to

71 years upon presentation of the lesions. The main clinical features of these lesions as diagnosed by a board-certified pathologist are summarized in Table 1. However, before proceeding, we confirmed the pathophysiology of these tissue samples. Their anatomical sites are representative of that of the most frequently detected neoplastic lesions, and include the retromolar trigone region (S1), floor of the mouth (S2) and tongue (WSU 62). The corresponding lesions were carcinoma in situ (CIS), well-differentiated invasive carcinoma and moderate to poorly differentiated carcinoma, respectively. For each tumor sample, adjacent normal tissue comprising squamous epithelium was also provided and in all cases the histopathology was confirmed. Frozen tissue sections (8 μm) stained with H&E are illustrated in Fig. 2. Normal squamous epithelium S1 (A), S2 (B) and WSU 62 (C) that match to the tumor samples, showed an orderly architecture that ranges from immature small round cells in the basal layer to those of mature flattened cells with abundant cytoplasm and small nuclei at the superficial layer (×500). In all cases, the basement membrane was observed to be intact. In contrast, tumor sample S1 (D), a carcinoma in situ, showed a variation in size and shape of cells, nuclear pleomorphism, hyperchromasia, and loss of normal cellular maturation (×500). Tumor sample S2 (E), a well-differentiated squamous cell carcinoma, showed infiltration of the underlying stroma with sheets and islands of cancer cells displaying nuclear pleomorphism and hyperchromasia. Furthermore, occasional keratin pearls were observed (×500). Tumor sample WSU 62 (E), a moderate to poorly differentiated invasive carcinoma, showed infiltration of underlying stroma as sheets and islands of cancer cells showing marked nuclear pleomorphism and hyperchromasia. In this sample, there is distinct absence of keratin pearls (×500). The data illustrates that anatomical and histological differences in tumor samples make them suitable for their use in identifying those genes that may be critical in tumor development.

3.3. Total RNA extractable from microdissected tissues

Before proceeding with the construction of representational cDNAs from the tissue sets, we sought to determine the amount and integrity of total RNA that was extractable from approximately 5000 microdissected cells. As shown in Table 2, the average amount of total RNA that was quantifiable using the fluorometric system, was 16.7 ng and this was demonstrated to be similar (14.7–18.6 ng) among samples. In addition, when 2000 and 10,000 cells were microdissected from each tissue sample, the average amount of total RNA extracted was 7.9 and 21.4 ng, respectively (data not shown).

The integrity of the RNA in all cases was assessed by RT-PCR of GAPDH. As demonstrated in Fig. 3, RNA extracted from all tissue sets (S1, S2, WSU 62) was considered to be of sufficient quality to be reverse transcribed and amplified using specific primers, which yielded a 600 bp PCR product of GAPDH. The data demonstrate that small amounts (ng) of total RNA extracted from microdissected cells are of sufficient integrity to synthesize and construct representational cDNA libraries from oral carcinomas and matching normal epithelium.

3.4. Quality assessment of amplified double-strand cDNAs

As the integrity of total RNA was maintained, we next determined whether this would be reflected in the synthesis of high-quality cDNA libraries. Double-strand cDNA product of each tissue sample was amplified by PCR and the quality determined on an agarose gel. A homogenous smear (200–1500 bp) was observed for all samples, indicating sufficient complexity of transcripts for library construction (data not shown). Amplified cDNAs were purified, subcloned, and randomly picked clones (12) from each library were analyzed to determine the average insert size. In all cases, the average size was observed to be around 600 bp, with a range of 300–1500 bp (data not shown). The data indicate a complexity of cDNAs that may be representational of the genes expressed in oral epithelium.

3.5. Sequence analysis of cDNA libraries

After quality control, 96 clones were randomly picked from each of the six cDNA libraries, expanded,

Table 1
Clinical characteristics of lesions from patients with head and neck squamous cell carcinoma (HNSCC)[a]

Sample	Age	Sex	Origin	Pathology
S1	65	M	Retromolar trigone	Carcinoma in situ
S2	71	M	Floor of mouth	Well-differentiated invasive
WSU 62	46	M	Tongue	Moderate to poorly differentiated invasive

[a] Tissues, including adjacent normal mucosa, were surgically removed from patients with oral carcinomas for clinical classification. Anatomical site and severity of the carcinomas, S1, S2 and WSU 62 are indicated. In all cases, corresponding normal tissues were confirmed as squamous epithelium.

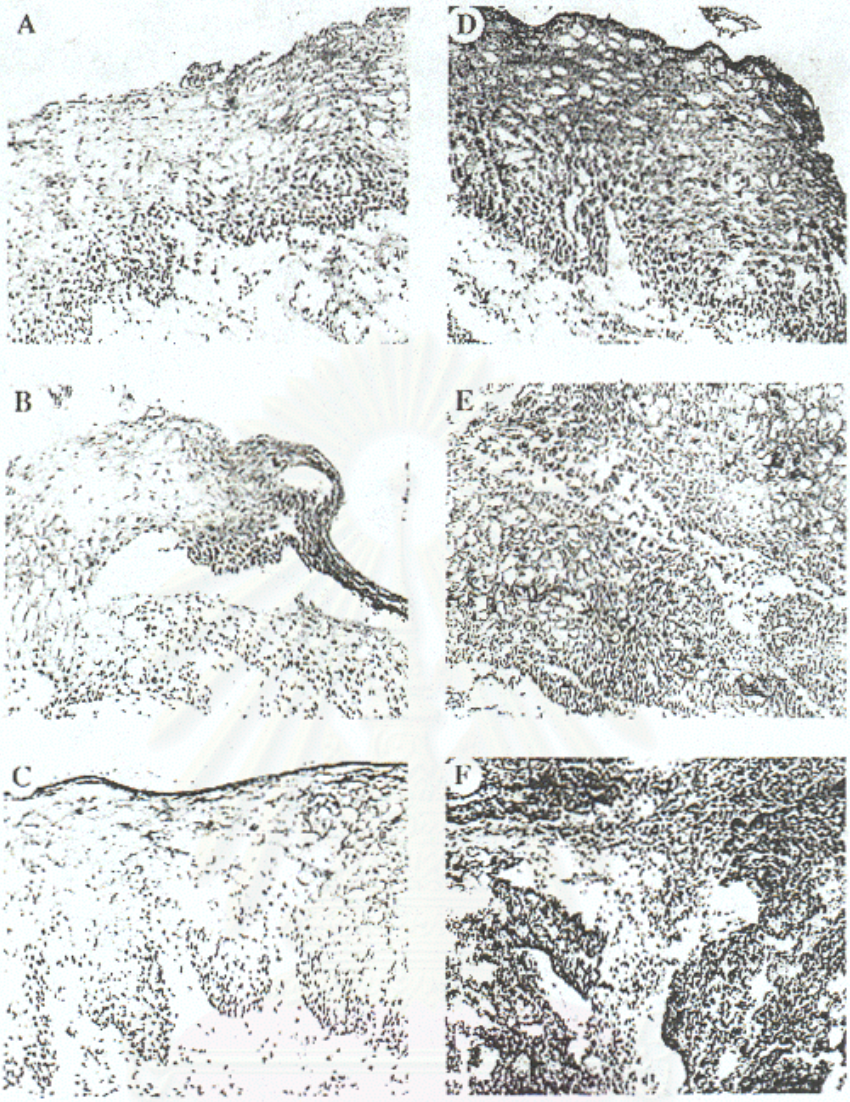


Fig. 2. Histopathological features of head and neck squamous cell carcinoma (HNSCC). Tissue sets, comprising of both normal and pathological lesion from the same HNSCC patient, were snap-frozen and 8μm thick sections were stained with hematoxylin and eosin (H&E). Histopathological features of progression of HNSCC from normal to carcinoma are demonstrated. Representative normal squamous epithelium (A,B,C) from the oral cavity shows an orderly maturation with progressive flattening of the cells going from deep to superficial cell layers, uniform small nuclei, and low nuclear:cytoplasmic ratio (magnification×500). Carcinoma in situ (D) shows a variation in size and shape of cells, nuclear pleomorphism, hyperchromasia and a loss of normal cellular maturation is observed. There is no breach of the basement membrane (magnification×500). The well-differentiated squamous cell carcinoma (E) shows infiltration of the underlying stroma with sheets and islands of cancer cells showing nuclear pleomorphism and hyperchromasia. Occasional keratin pearls are also observed (magnification×500). Moderate to poorly differentiated invasive carcinoma (F) shows infiltration of the underlying stroma as sheets of cancer cells showing marked nuclear pleomorphism and hyperchromasia. However, a notable absence of keratin pearls is noted (magnification×500).

sequenced and their nucleotide sequences analyzed using the BLAST program. The cumulative information is summarized in Table 3. Analysis of the nucleotide information obtained from the successful sequencing of 76–96% of the inserts indicated that these DNAs represent either anonymous ESTs (25–48%), known genes (9–29%) or novel transcripts (27–51%), respectively. Both *Alu* repeats and ribosomal RNA were detected at low frequencies in all six libraries (2–9%) and no sequences of mitochondrial or bacterial origin were recorded. Approximately 7% of the total number of clones analyzed were either without insert or contained short sequences that were uninformative. The data indicate the successful cloning of a large number of

Table 2
Total RNA extracted from microdissected tissues*

Tissue	S1		S2		WSU 62		Average
	N	T	N	T	N	T	
RNA (ng)	18.6	15.6	15.9	17.3	17.8	14.7	16.7

* RNA extracted from 5000 cells microdissected from the three tissue sets (S1, S2, WSU 62) was labeled with a fluorescent dye and quantified using appropriate standards and using a VersaFluor™ Fluorometer. The amount of total RNA extractable for normal (N) and tumor (T) epithelium, using the experimental conditions described in Section 2 and the average for all six samples is indicated. The data are representative of three separate experiments.

novel transcripts that are likely to be representative of those expressed in normal and malignant oral epithelium.

3.6. Identification of known genes in cDNA libraries from oral epithelium

We next determined, by using the BLAST program, whether any of these sequences matched known genes. Those that were identified as being expressed in any of the six libraries are listed in Table 4. Of interest, known genes identified more than once in a library include calgranulin A, kappa casein precursor, elafin precursor, monocyte chemotactic protein 3 precursor, cyclin I, keratin 13, keratin 56 kD, cornifin B and the interleukin-1 receptor antagonist protein precursor. Collectively, of the known genes identified, six were represented in two or more of the libraries and are listed in Table 5. Kappa casein precursor was common in all the libraries and monocyte chemotactic protein 3 precursor was represented in libraries made from all tumor tissues and one normal (62N). The keratin 4 was detected in two normal libraries (S1, S2) whereas the keratin 13, calgranulin A (calcium binding protein A8) and cornifin B (small proline rich protein, SPRR 1B) were identified in two of the libraries.

4. Discussion

Approximately 10% of the total number of genes are suspected to be expressed in a given cell type. Determining their identity is an important first step towards understanding the patterns of gene expression that mediate normal cellular physiology and disease process. In this study, we report the construction of six high-quality cDNA libraries from tissues of oral origin, including normal and malignant epithelium. Previous studies have reported on genes that are expressed in tissue specimens containing squamous epithelium [16]. However, these studies utilized heterogenous tissues as starting material, thus the expression data may not reflect the genes that are specifically active in the epithelium. In contrast, for the present investigation, we procured pure populations of cells (approximately 95% purity) from normal and malignant epithelium using LCM.

A molecular fingerprint of every expressed gene in each cell type is its mRNA, but this usually constitutes approximately 5% of the total RNA [17]. Therefore, the quantity and integrity of RNA extracted from the starting material are important considerations to ensure that the gene expression analysis is representational. In addition, procedures involved in tissue isolation using LCM may themselves limit this type of analysis by facilitating RNA degradation. However, from approximately 5000 cells procured from each tissue sample by LCM, we were able to demonstrate that the RNA extracted was quantifiable and of good integrity as assessed by the ability to amplify GAPDH.

cDNA libraries were prepared from three patient tissue sets, comprised of normal and tumor oral epithelium. Our primary goal from this study was to identify those genes whose expression may be solely confined to the malignant oral epithelium and these may be causal in the development of HNSCC. In addition, we wanted to use these high-quality cDNA libraries to contribute to the HNCGAP, in their gene discovery efforts and those of systematically cataloging cancer genes from different neoplasias. We and others (http://www.ncbi.nlm.nih.

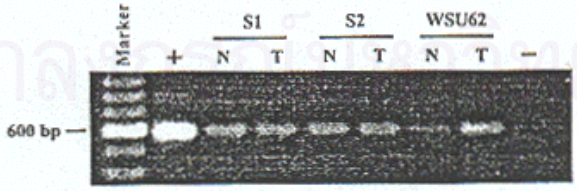


Fig. 3. Assessment of RNA integrity by reverse transcriptase-polymerase chain reaction (RT-PCR). Total RNA was extracted from microdissected tissues (approximately 5000 cells) as described in Section 2. RNA integrity was assessed by RT-PCR of the *GAPDH* gene. Using specific primers for *GAPDH*, a 600 bp fragment is amplified, as observed for all the tissue samples. Appropriate positive and negative controls are indicated. Data shown is representative of four independent experiments.

Table 3
Cumulative sequence analysis of 96 head and neck squamous cell carcinoma (HNSCC)-specific cDNA clones[a]

CDNA library	S1		S2		WSU 62	
	N	T	N	T	N	T
Clones analyzed	92	92	73	84	82	90
Known genes	8	27	22	11	9	14
Anonymous ESTs	44	35	18	25	42	41
Novel sequences	37	24	30	43	22	30
Expressed *Alu* repeats	3	4	3	2	7	4
Ribosomal RNA	0	2	0	3	2	1

[a] Random clones (96) from libraries constructed from normal (N) and tumor (T) epithelium were sequenced and this information was used to search available databases, using the BLAST program. Clones from the six libraries were subsequently categorized as having sequences similar to known genes, those that matched to expressed sequence tags (ESTs) and those that showed no significant homology. Frequency of *Alu* repeats and ribosomal RNA sequences are indicated.

gov) have previously contributed to the HNCGAP initiative. However, cDNA libraries for these efforts were either constructed from bulk tissues or from representative HNSCC cell lines and from normal and immortalized gingival keratinocytes [18]. To date, this report is the first describing the construction of cDNA libraries from microdissected oral epithelium which may represent more accurately those genes expressed in vivo.

From the first round of analysis, sequence information was obtained from approximately 76–96% of clones analyzed, which included anonymous ESTs and known genes. Furthermore, these six libraries collectively demonstrated a remarkable sequence novelty, with most transcripts likely representing unknown genes. However, it is still possible that these novel sequences may be found in one or more of the libraries. An important parameter indicating the quality of cDNA libraries is the presence of contaminating sequences. In this regard, no bacterial or vector sequences were detected, while a very low frequency (0–7.3%) of *Alu* repeats and rRNA sequences were present. Thus, the high percentage of quality sequencing suggests these libraries are representative with a concomitant maintenance of complexity.

The high incidence of detection of novel sequences in only the first round of analysis is, in itself, remarkable and surpasses the figure (4% per library) set out by the gene discovery efforts of CGAP. Additional sequence analysis of these libraries is likely to identify many new transcripts. These yet-to-be-identified transcripts may indeed be uniquely expressed in oral epithelium. It therefore follows that the expression pattern and the function of these genes may help to identify gene products involved in the transformed and/or the metastatic phenotype, as well as additional molecules that, without playing an obvious role in the neoplastic process, can nevertheless be used as clinically useful markers of tumor development.

In this regard, the 55 known genes identified from this first round of sequence analysis is in itself interesting, but whether they may play a role in HNSCC tumor development is not documented. Of particular interest was the sequence match to the monocyte chemotactic protein 3 precursor gene (MCP3), which was readily detected in all three cDNA libraries from tumor tissues (ST1, ST2, 62T). MCP3 is a chemokine known to induce the production of gelatinase B and chemotaxis of monocytes. In addition, MCP3 is also known to be produced by tumor cells and its expression in HNSCC may play a role in tumor progression [19–22]. Additional known genes, while not causal to tumor development, may be useful markers. These include cornifin B, also known as small proline rich protein 1 (SPRR1). Its gene product has been reported to be expressed in the sublingual and tongue epithelium and in malignant oral epithelium. While its precise function is unclear it may be involved in the terminal differentiation status of keratinocytes [23–25]. Its use as a marker of tumor progression has been highlighted in a recent report [26]. Also included is calgranulin A, which has a functional role in epithelium differentiation and whose expression has been reported in a restricted subset of normal stratified squamous epithelium of the tongue and buccal mucosa [27]. In this regard, the expression of cytokeratins (4, 5 and 13) can be used as makers of epithelial differentiation and altered expression of these proteins have been reported in oral malignancies [28]. Sequence match to kappa casein precursor and many ribosomal proteins (S11, S20, S24, L3, L8, L38, L37) were readily detected but whether they have any role in tumor development is currently unclear.

Although data presented in this report involves a limited set of tissue samples, a high frequency of novel transcripts have been already identified from the first round of sequence analysis, thus suggesting the presence of many unknown genes that may play a critical role in the biology of oral epithelium. Indeed, by subsequent sequencing of these cDNA libraries HNCGAP has already identified an additional 57 novel transcripts (http://www.ncbi.nlm.nih.gov/ncicgap). A detailed analysis of the genes identified by the HNCGAP effort in these and additional cDNA libraries constructed from tissues of oral origin, including different sites and stages of malignancies, is already in progress and will be reported shortly. Thus, we can conclude that the construction of representational cDNA libraries form normal and neoplastic oral tissues has resulted in the availability of cDNAs for many novel genes which are expressed in normal oral epithelium and in HNSCCs. These efforts, together with the use of gene array technologies and LCM will soon make it possible to define a pattern of gene expression in a tumor pro-

Table 4
Known genes identified from head and neck squamous cell carcinoma (HNSCC)-specific cDNA libraries

14-3-3 protein σ
Acyl-CoA-binding protein
Annexin I
Antigen peptide transporter2 (transporter, ATP-binding cassette)
Antileukoproteinase 1
ATP synthase A chain
Breast basic conserved protein 1 (breast basic conserved gene 1)
Calgranulin A (calcium binding protein A8)
Calgranulin B (calcium binding protein A9)
Calpain 2 large
CD9 antigen
Cds 1 human SPR2-1 gene for small proline rich protein
Cornifin B (small proline rich protein, SPRR 1B)
Cyclin I
Cystatin A
Cytochrome C oxidase polypeptide via liver, 2, 3
Dolichyl-phosphate β-glucosyltransferase
E25B protein mRNA
Elafin precursor
Erythrocyte adducin α subunit
Erythrocyte adducin β subunit
Ferritin heavy chain
GAP junction β-2 protein (Connexin 26)
Growth relating protein BB1
Histocompatibility antigen, M alpha chain
Histone H3.3
Interferon γ
Interleukin-1 receptor antagonist protein precursor
Kappa casein precursor
Keratin, type I cytoskeletal 13
Keratin, type II cytoskeletal 4, 5, 7, 56kD
L-3 phosphoserine-phosphatase homologue
Leukocyte elastase inhibitor
Mitotic kinesin-like protein-1
Monocyte chemotactic protein 3 precursor
Myosin like chain Alkali smooth muscle Isoform
NADH-Ubiquinone oxidoreductase chain 1 and 4
Protein phosphatase PP1-γ
Ribosomal protein S11, S20, S18, S24, L3, L8, L37, L38
Signal recognition particle receptor βsubunit
Transcriptional coactivator

[a] Sequence analysis of random clones from cDNA libraries constructed from normal (N) and tumor (T) epithelium identified many as having homology to known genes. Some of these known genes were identified more than once in one or more of the six libraries. The 55 known genes that were identified are listed by name alphabetically.

Table 5
Known genes identified in one or more head and neck squamous cell carcinoma (HNSCC)-specific cDNA libraries[a]

Kappa casein precursor	S1 (N, T), S2 (N, T), WSU 62 (N, T)
Monocyte chemotactic protein 3 precursor	ST1, ST2, WSU 62 (N, T)
Keratin, type II cytoskeletal 4	SN1, SN2
Keratin, type I cytoskeletal 13	SN2, ST1
Calgranulin A	S2 (N, T)
Cornifin B	SN2, ST1

[a] Six known genes were identified in two or more HNSCC-specific cDNA libraries (N, normal, T, tumor) and listed. Only one known gene was present in all six libraries analyzed.

gression model of HNSCC. This experimental approach is also expected to facilitate the identification of earlier markers heralding malignancy, thus providing valuable tools of diagnostic and prognostic value to study pre-malignant lesions.

Acknowledgment

We thank Dr. Vladimir Knezevic for helpful advice on quantification of total RNA using Fluorometer system.

References

[1] Trent JM, Bittner M, Zhang J, Wiltshire R, Ray M, Su Y, Gracia E, Meltzer P, De Risi J, Penland L, Brown P. Use of microgenomic technology for analysis of alterations in DNA copy number and gene expression in malignant melanoma. Clin Exp Immunol 1997;107(Suppl 1):33–40.

[2] Urbain JL. Oncogenes, cancer and imaging. J Nucl Med 1999;40:498–504.

[3] Strausberg RL, Dahl CA, Klausner RD. New opportunities for uncovering the molecular basis of cancer (news; comment) (see comments). Nat Genet 15 (Spec No) 1997;415–6.

[4] Strausberg RL, Buetow KH, Emmert-Buck MR, Klausner RD. The Cancer Genome Anatomy Project: building an annotated gene index. Trends Genet 2000;16:103–6.

[5] Emmert-Buck MR, Strausberg RL, Krizman DB. Molecular profiling in humans: feasibility and applications. Am J Pathol 2000 (in press).

[6] Boring CC, Squires TS, Tong T. Cancer statistics. CA Cancer J Clin 1993;43:7–26.

[7] Parkin DM, Pisani P, Ferlay J. Global cancer statistics. CA Cancer J Clin 1999;49:33–64, 2.

[8] Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, (see comments). CA Cancer J Clin 1999;49:8–31, 1.

[9] Califano J, van der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, Corio R, Lee D, Greenberg B, Koch W, Sidransky D. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res 1996;56:2488–92.

[10] Bockmuhl U, Wolf G, Schmidt S, Schwendel A, Jahnke V, Dietel M, Petersen I. Genomic alterations associated with malignancy in head and neck cancer. Head Neck 1998;20:145–51.

[11] Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA. Laser capture microdissection (see comments). Science 1996;274:998–1001.

[12] Bohm M, Wieland I, Schutze K, Rubben H. Microbeam MOMeNT: non-contact laser microdissection of membrane-mounted native tissue. Am J Pathol 1997;151:63–7.

[13] Simone NL, Bonner RF, Gillespie JW, Emmert-Buck M, Liotta LA. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. Trends Genet 1998;14:272–6.

[14] Krizman DB, Chuaqui RF, Meltzer PS, Trent JM, Duray PH, Linehan WM, Liotta LA, Emmert-Buck MR. Construction of a representative cDNA library from prostatic intraepithelial neoplasia. Cancer Res 1996;56:5380–3.

[15] Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. Gene expression profiles in normal and cancer cells. Science 1997;276:1268–72.

[16] van der Velden LA, Manni JJ, Ramaekers FC, Kuijpers W. Expression of intermediate filament proteins in benign lesions of the oral mucosa. Eur Arch Otorhinolaryngol 1999;256:514–9.

[17] Eberwine J. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. Biotechniques 1996;20:584–91.

[18] Shillitoe EJ, May M, Patel V, Leethanakul C, Ensley JF, Strausberg RL, Gutkind JS. Genome-wide analysis of oral cancer — early results from the Cancer Genome Anatomy Project. Oral Oncology 2000;36:8–16.

[19] Mantovani A, Bottazzi B, Colotta F, Sozzani S, Ruco L. The origin and function of tumor-associated macrophages (see comments). Immunol Today 1992;13:265–70.

[20] Opdenakker G, Van Damme J. Chemotactic factors, passive invasion and metastasis of cancer cells (letter; comment). Immunol Today 1992;13:463–4.

[21] Opdenakker G, Froyen G, Fiten P, Proost P, Van Damme J. Human monocyte chemotactic protein-3 (MCP-3): molecular cloning of the cDNA and comparison with other chemokines. Biochem Biophys Res Commun 1993;191:535–42.

[22] Opdenakker G, Fiten P, Nys G, Froyen G, Van Roy N, Speleman F, Laureys G, Van Damme J. The human MCP-3 gene (SCYA7): cloning, sequence analysis, and assignment to the C-C chemokine gene cluster on chromosome 17q11.2-q12. Genomics 1994;21:403–8.

[23] Hohl D, de Viragh PA, Amiguet-Barras F, Gibbs S, Backendorf C, Huber M. The small proline-rich proteins constitute a multigene family of differentially regulated cornified cell envelope precursor proteins. J Invest Dermatol 1995;104:902–9.

[24] Fujimoto W, Nakanishi G, Arata J, Jetten AM. Differential expression of human cornifin alpha and beta in squamous differentiating epithelial tissues and several skin lesions. J Invest Dermatol 1997;108:200–4.

[25] Fischer DF, van Drunen CM, Winkler GS, van de Putte P, Backendorf C. Involvement of a nuclear matrix association region in the regulation of the SPRR2A keratinocyte terminal differentiation marker. Nucleic Acids Res 1998;26:5288–94.

[26] Nacht M, Ferguson AT, Zhang W, Petroziello JM, Cook BP, Gao YH, Maguire S, Riley D, Coppola G, Landes GM, Madden SL, Sukumar S. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer (In Process Citation). Cancer Res 1999;59:5464–70.

[27] Wilkinson MM, Busuttil A, Hayward C, Brock DJ, Dorin JR, Van Heyningen V. Expression pattern of two related cystic fibrosis-associated calcium- binding proteins in normal and abnormal tissues. J Cell Sci 1988;91:221–30.

[28] Vaidya MM, Borges AM, Pradhan SA, Bhisey AN. Cytokeratin expression in squamous cell carcinomas of the tongue and alveolar mucosa. Eur J Cancer B Oral Oncol 1996;32B:333–6.

VITA

Mrs Chidchanok Leethanakul, was born in October 21st, 1966 at Supprasitthiprasong Hospital, Ubonratchathanee, Thailand. Graduated in Dentistry from Faculty of Dentistry, Prince of Songkla University, Hat Yai, Songkhla in 1990 and Master of Sciences (M.Sc.) in Orthodontics from Faculty of Dentistry, Chulalongkorn University, Bangkok in 1996. Three year experience in teaching as the instructor of Preventive Department, Faculty of Dentistry, Prince of Songkla University in 1990-1993. Enrolled in a Ph. D. program in Faculty of Dentistry, Chulalongkorn University during 1996-2001. Fellow scholarship, the Ministry of the University Affairs of Thailand in 1997. Postdoctoral Fellowship 1998-2001, Postdoctoral Travel Award 1999 and Royalty Awards 2000, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, United States. Present position is the instructor of Preventive Department, Faculty of Dentistry, Prince of Songkla University.