View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Helsingin yliopiston digitaalinen arkisto

Helsinki University Biomedical Dissertations No. 109

MICROARRAYS IN MOLECULAR PROFILING OF CANCER

FOCUS ON HEAD AND NECK SQUAMOUS CELL CARCINOMA

Anna-Kaarina Järvinen

Institute of Biomedicine and Biomedicum Biochip Center
University of Helsinki

Department of Otorhinolaryngology - Head and Neck Surgery
Helsinki University Central Hospital

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in Lecture Hall 2, Biomedicum Helsinki, on May 22nd, 2008, at 12 noon.

Helsinki 2008

Supervised by

Docent Outi Monni
Institute of Biomedicine and
Biomedicum Biochip Center
University of Helsinki
Helsinki, Finland

Professor Antti Mäkitie
Department of Otorhinolaryngology -
Head and Neck Surgery
Helsinki University Central Hospital
Helsinki, Finland

Reviewed by

Docent Merja Perälä
Medical Biotechnology
VTT Technical Research
Centre of Finland
Turku, Finland

Docent Ritva Karhu
Laboratory of Cancer Genetics
Institute of Medical Technology
University of Tampere
Tampere, Finland

Official opponent

Professor Tapio Visakorpi
Institute of Medical Technology
University of Tampere
Tampere, Finland

ISSN 1457-8433
ISBN 978-952-10-4654-4 (paperback)
ISBN 978-952-10-4655-1 (PDF)
http://ethesis.helsinki.fi
Yliopistopaino
Helsinki 2008

Äidille

Contents

List of original publications

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I-III).

I Järvinen A-K, Hautaniemi S, Edgren H, Auvinen P, Saarela J, Kallioniemi O-P, Monni O. Are data from different gene expression microarray platforms comparable? Genomics 83: 1164-1168, 2004.

II Järvinen A-K, Autio R, Haapa-Paananen S, Wolf M, Saarela M, Grénman R, Leivo I, Kallioniemi O, Mäkitie AA, Monni O. Identification of target genes in laryngeal squamous cell carcinoma by high-resolution copy number and gene expression microarray analyses. Oncogene 25: 6997-7008, 2006.

III Järvinen A-K, Autio R, Kilpinen S, Saarela M, Leivo I, Grénman R, Mäkitie AA, Monni O. High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx. Genes, Chromosomes and Cancer 47: 500-509, 2008.
Reprinted with permission of John Wiley & Sons, Inc.

Abbreviations

aCGH	array comparative genomic hybridization
ATCC	American Type Culture Collection
BAC	bacterial artificial chromosome
cDNA	complementary deoxyribonucleic acid
CGH	comparative genomic hybridization
CNV	copy number variant
DNA	deoxyribonucleic acid
FDA	U.S. Food and Drug Administration
GO	gene ontology
HNSCC	head and neck squamous cell carcinoma
HPV	human papilloma virus
kb	kilobase pair
IVT	*in vitro* transcription
LOH	loss of heterozygosity
LSCC	laryngeal squamous cell carcinoma
MAS5	Affymetrix Microarray Suite 5
Mb	megabase pair
MGED	Microarray Gene Expression Data group
MIAME	minimum information about a microarray experiment
miRNA	microRNA
MM	mismatch in Affymetrix probe pair
mRNA	messenger RNA
OTSCC	oral tongue squamous cell carcinoma
PCR	polymerase chain reaction
PM	perfect match in Affymetrix probe pair
RMA	robust multi-array average method
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
RT-PCR	reverse transcription - polymerase chain reaction
siRNA	short interfering RNA
SNP	single-nucleotide polymorphism
SOM	self-organizing map
tRNA	transfer RNA
TSG	tumor suppressor gene

All gene symbols can be found at www.ncbi.nlm.nih.gov/sites/entrez?db=gene.

Abstract

Microarrays have a wide range of applications in the biomedical field. From the beginning, arrays have mostly been utilized in cancer research, including classification of tumors into different subgroups and identification of clinical associations. In the microarray format, a collection of small features, such as different oligonucleotides, is attached to a solid support. The advantage of microarray technology is the ability to simultaneously measure changes in the levels of multiple biomolecules. Because many diseases, including cancer, are complex, involving an interplay between various genes and environmental factors, the detection of only a single marker molecule is usually insufficient for determining disease status. Thus, a technique that simultaneously collects information on multiple molecules allows better insights into a complex disease. Since microarrays can be custom-manufactured or obtained from a number of commercial providers, understanding data quality and comparability between different platforms is important to enable the use of the technology to areas beyond basic research. When standardized, integrated array data could ultimately help to offer a complete profile of the disease, illuminating mechanisms and genes behind disorders as well as facilitating disease diagnostics.

In the first part of this work, we aimed to elucidate the comparability of gene expression measurements from different oligonucleotide and cDNA microarray platforms. We compared three different gene expression microarrays; one was a commercial oligonucleotide microarray and the others commercial and custom-made cDNA microarrays. The filtered gene expression data from the commercial platforms correlated better across experiments (r=0.78-0.86) than the expression data between the custom-made and either of the two commercial platforms (r=0.62-0.76). Although the results from different platforms correlated reasonably well, combining and comparing the measurements were not straightforward. The clone errors on the custom-made array and annotation and technical differences between the platforms introduced variability in the data. In conclusion, the different gene expression microarray platforms provided results sufficiently concordant for the research setting, but the variability represents a challenge for developing diagnostic applications for the microarrays.

In the second part of the work, we performed an integrated high-resolution microarray analysis of gene copy number and expression in 38 laryngeal and oral tongue squamous

cell carcinoma cell lines and primary tumors. Our aim was to pinpoint genes for which expression was impacted by changes in copy number. We detected multiple frequent genomic alterations in head and neck squamous cell carcinoma (HNSCC) material. The data revealed that especially amplifications had a clear impact on gene expression. Across the genome, 14-32% of genes in the highly amplified regions (copy number ratio >2.5) had associated overexpression. The impact of decreased copy number on gene underexpression was less clear. This might be partly due to technical as well as biological reasons since the loss of one copy of the gene does not always lead to a detectable change in expression. Using statistical analysis across the samples, we systematically identified hundreds of genes for which an increased copy number was associated with increased expression. For example, our data implied that *FADD* and *PPFIA1* were frequently overexpressed at the 11q13 amplicon in HNSCC. The 11q13 amplicon, including known oncogenes such as *CCND1* and *CTTN*, is well-characterized in different type of cancers, but the roles of *FADD* and *PPFIA1* remain obscure. Taken together, the integrated microarray analysis revealed a number of known as well as novel target genes in altered regions in HNSCC. The identified genes provide a basis for functional validation and may eventually lead to the identification of novel candidates for targeted therapy in HNSCC.

Introduction

The first version of the human genome sequence was published at the beginning of this decade (Lander *et al*., 2001; Venter *et al*., 2001). After the initial draft sequence, the information has been updated (International Human Genome Sequencing Consortium, 2004). The availability of the sequence information has promoted development of a number of high-throughput technologies, including microarrays. The microarrays have played an important role in changing the concept in biological research from investigation of single genes to an omics approach (reviewed by Weinstein, 2002; Ge *et al*., 2003; Liu *et al*., 2006). Omics studies are characterized by the use of high-throughput methods that produce large quantities of data.

DNA microarray technology, which allows the investigation of multiple genes in a single experiment, was developed over 10 years ago (Fodor *et al*., 1991; Schena *et al*., 1995). DNA microarrays can comprise thousands of DNA fragments, such as oligonucleotides, or cDNA clones, robotically arrayed or *in situ* synthesized on a solid support. Initially, custom-made arrays were frequently applied in academic laboratories, but their use has diminished radically due to restrictions in the amounts of time and money allocated to manufacturing and quality control (reviewed by Holloway *et al*., 2002; Gershon, 2004). Currently available commercial whole genome microarrays are composed of over million features representing various transcripts or exons, allowing genome-wide identification of differentially expressed genes and alternatively spliced variants. In addition to well-established gene expression microarrays, the technology can be applied to measure other biological variables, such as copy number and single-nucleotide polymorphisms (SNPs) (Solinas-Toldo *et al*., 1997; Pinkel *et al*., 1998; reviewed by Pinkel and Albertson, 2005; Syvänen, 2005). The development of microarray technology in gene expression profiling as well as its use in various other applications are described in Nature Genetics microarray theme numbers "Chipping Forecasts" published in 1999, 2002, and 2005.

Microarrays can measure RNA, DNA, or protein levels from cells or tissues on a genome-wide scale. These molecular profiles are invaluable in pinpointing genes critical in tumorigenesis (reviewed by Weinstein, 2006). For example, DNA and RNA level alterations measured from the same sample provide information about genes in which expression is altered due to increased or decreased copy number. Copy number alterations represent an important mechanism for cancer cells to promote or suppress the expression

of genes involved in cancer progression. Furthermore, genes deregulated in association with high-level amplifications have been linked to poor outcome of cancer, representing potential drug targets (Chin *et al*., 2006). Thus, the integrated array data can identify therapeutic targets, which might then provide alternative options to surgery and radiation therapy in cancer. Here, we integrated data from gene expression and copy number microarrays and identified target genes for genomic alterations of potential importance in HNSCC pathogenesis.

Large quantities of gene expression microarray data are now deposited in public databases and are thus available to the whole research community. Using the data can be challenging since numerous different microarray platforms exist with different array design, labeling and hybridization protocols, equipment, and analysis software. To enable available data to be utilized effectively, the data should be well-annotated and of high quality irrespective of the platform and experimental methods. The use of published data could potentially replace some experiments in the laboratory. In the present work, our aim was to correlate results from gene expression microarrays to evaluate whether the data from different platforms are comparable and reliable. *In silico* resources, providing data in an electronic format, are not limited to gene expression microarray databases; other data types, such as the antibody-based protein atlas for protein expression and localization patterns (Uhlén *et al*., 2005), are also available. Future drug development could be dependent on the effective integration of *in silico* data from different sources, the application of which requires an in-depth understanding of measurement techniques and analysis methods (reviewed by Searls, 2005; Loging *et al*., 2007).

Review of the literature

1 Genome and gene expression

DNA represents genetic material of the cell. Traditionally, genes are defined as DNA fragments that code proteins. All genetic information of an organism is known as a genome. Of the human genome, only 3.5% is estimated to represent protein-coding DNA, with the majority being nonprotein-coding DNA. The estimated number of human protein-coding genes is currently around 23 000 (www.ensembl.org/Homo_sapiens/index.html, accessed 14.1.2008). However, the exact number remains unclear and estimates vary depending on the method used in predictions. The number of proteins could be more than 50 times higher than the number of genes due to alternative splicing events of mRNA and additional variability created through posttranslational modifications (reviewed by Jensen, 2004). Human DNA, and thus, genes are distributed across 23 chromosome pairs. In cytogenetic nomenclature, chromosomes are divided into arms, the short arm p and the long arm q, arms into regions, and regions into bands and sub-bands. Each cell of an organism has the same DNA. However, different genes are active in different cells. This is dependent on a number of factors, such as developmental stage or environmental factors. Gene expression can be regulated at the level of transcription, posttranscriptionally, or epigenetically.

Information flow from DNA to mRNA and finally to functional proteins has been a central theorem of biology. Even though the majority of genes encode for proteins, some RNAs, such as rRNAs and tRNAs, are not translated into proteins. The emergence of other nonprotein-coding sequences, microRNAs (miRNAs), with a regulatory function (Lee *et al*., 1993; Lagos-Quintana *et al*., 2001; Lee and Ambros, 2001) has broadened our view of genes as well as the functions of RNA. miRNAs negatively regulate genes in two ways. miRNAs that are perfectly or nearly perfectly complementary with their target mRNA direct the cleavage of the target, resulting in degradation of the transcript. This phenomenon works in the same manner as short interfering RNAs (siRNAs) in RNA interference (RNAi) (Fire *et al*., 1998; reviewed by Hannon, 2002; Hannon and Rossi, 2004). miRNAs can also bind imperfectly to the target mRNA's 3′ untranslated regions, regulating target expression at the translational level. The distinction between miRNA and endogenous siRNA molecules is sometimes unclear, but differences exist in their origin,

processing, evolutionary conservation, and the genes that they silence (reviewed by Bartel, 2004). Identification of miRNA's target genes is challenging due to their ability to regulate the target by imperfect binding. It has been estimated that a single miRNA could bind hundreds of target genes (reviewed by Esquela-Kerscher and Slack, 2006). At the moment, 541 human miRNA sequences are reported (microrna.sanger.ac.uk/sequences/, accessed 4.2.2008) (Griffiths-Jones *et al*., 2006). Interestingly, miRNA molecules are implicated as having a role in cancer either as oncogenes or tumor suppressors (reviewed by Calin and Croce, 2006; Esquela-Kerscher and Slack, 2006).

2 Molecular biology of cancer

Cancer is regarded as a genetic disease that occurs due to sequential accumulation of genetic alterations in oncogenes, tumor suppressor genes (TSGs), and stability genes (reviewed in Section 2.1). These alterations cause abnormal activation or inactivation of a number of critical pathways and signaling cascades, resulting in uncontrolled cellular growth (reviewed by Vogelstein and Kinzler, 2004). Environmental, viral, and chemical agents as well as physical substances can promote carcinogenesis (reviewed by Peto, 2001; Wogan *et al*., 2004). After the exposure to the carcinogen, 20-40 years can pass until the clinical detection of a solid tumor (reviewed by Wogan *et al*., 2004). The risk of cancer can therefore be associated with lifestyle and environmental factors, even though hereditary factors also play a role (reviewed by Peto, 2001; Ponder, 2001; Balmain *et al*., 2003).

The majority of tumors are monoclonal since they derive from a single progenitor cell. In a multistep tumorigenesis process, clonal expansions involving genetic and epigenetic alterations follow each other (reviewed by Ponder, 2001; Balmain *et al*., 2003). Within a tumor, different subclones can have distinct alterations caused by simultaneous clonal expansion of different clones as a result of instability in a tumor genome (reviewed by Weinberg, 2006). If genome integrity was not compromised in cancer, the mutation rate would probably be too low to allow cancer progression (reviewed by Loeb, 2001). Instability can be acquired during tumor development or by inherited mutations occurring, for example, in genes that are responsible for genome integrity. Therefore, a person with inherited mutations in critical genes becomes predisposed to cancer (reviewed by Fearon, 1997). Moreover, the accelerated cell proliferation in cancer allows mutations to occur at

an increased rate. A number of factors, such as inflammation, drugs, hormones, chemical and infectious agents, and physical trauma, increase the rate of cell proliferation (reviewed by Weinberg, 2006). Furthermore, communication of different cell types in a tumor microenvironment is important in cancer development and progression. Tumor-surrounding stromal cells, for instance, can contribute to angiogenesis and invasion. Communication between cancerous epithelial cells and stromal cells can also cause changes in stromal cells, differentiating them from the normal state (reviewed by Tlsty and Hein, 2001).

Cancer cells are characterized by acquired functional capabilities: self-sufficiency in exogenous growth signals, insensitivity to antigrowth signals, limitless replicative potential, evasion of apoptosis, sustained angiogenesis, and acquisition of invasiveness and metastatic ability (Hanahan and Weinberg, 2000). The order and mechanistic means to achieve these properties can vary between different tumors. Therefore, an understanding of defective signaling pathways instead of single genes could be vital (reviewed by Vogelstein and Kinzler, 2004). Although recent studies have illuminated genetic changes needed to transform human cells (Sjöblom *et al.*, 2006), the exact number of changes required is still under debate. Certain pathways seem, however, to be often involved when cells are transformed *in vitro*. These include the mitogenic signaling pathway controlled by *RAS*, the cell cycle checkpoint controlled by *RB1*, the apoptosis pathway controlled by *TP53*, telomerase maintenance controlled by *hTERT*, and the signaling pathway controlled by *PP2A*. Whether all cancers have the same type of alterations remains unclear (reviewed by Hahn and Weinberg, 2002). To date, 367 human genes have been causally implicated in cancer development either through mutation, copy number alteration, or rearrangement (www.sanger.ac.uk/genetics/CGP/Census/, accessed 19.1.2008) (Futreal *et al.*, 2004). Recently, cancer genes were mapped by a large-scale sequencing effort aimed at identifying somatic driver mutations in kinases in the cancer genome (Greenman *et al.*, 2007). The list of cancer genes is thus by no means complete.

2.1 Oncogenes and tumor suppressors

Exogenous and endogenous mutagenic molecules as well as chemical and physical factors can modify DNA. Thus, these agents can promote carcinogenesis by affecting critical genes, including proto-oncogenes and TSGs. Proto-oncogenes are genes controlling such

normal cellular functions as proliferation, differentiation, and growth signaling. Initially, they were recognized through viruses (reviewed by Ponder, 2001). A gain-of-function mutation in a proto-oncogene creates an active form, an oncogene, by changing expression (regulatory effect) or protein structure (structural effect). This can happen through viral involvement, point mutation, gene amplification, chromosomal translocation, or some other structural alteration (reviewed by Albertson *et al*., 2003; Weinberg, 2006). Oncogenes can then activate mitogenic signaling pathways and allow cells to become independent of external signals. Thus, oncogenes promote tumorigenesis by giving normal cells properties that allow them to escape from cellular growth control. In different cell types, different pathways involving different genes can regulate cell growth and division (reviewed by Vogelstein and Kinzler, 2004; Weinberg, 2006).

In contrast to oncogenes, TSGs restrain the growth of the cell. Proteins encoded by TSGs can have various functions in the cell, but they all reduce the possibility of cancer development. TSGs can be divided into three groups: gatekeepers, caretakers, and landscapers (Kinzler and Vogelstein, 1997; Kinzler and Vogelstein, 1998). Gatekeepers, such as *RB1* and *TP53*, control cell growth by inhibiting growth or promoting cell death. For example, the loss of *RB1* allows cells to proceed through the cell cycle, leading to deregulated growth and the loss of *TP53* to escape apoptosis. Caretakers, such as *BRCA1* and *BRCA2*, are DNA maintenance genes that take care of genome integrity, thus affecting the rate at which cells accumulate mutations. Landscapers work through less direct mechanisms, affecting the tumor microenvironment.

TSG can be inactivated in cancer cells via genetic or epigenetic mechanisms. Epigenetic changes, such as methylation, do not affect the DNA sequence. According to Knudson's two-hit hypothesis (Knudson, 1971; reviewed by Knudson, 2001), both alleles of a TSG need to be inactivated to have an effect on cell phenotype. Thus, complete loss of function is required. If one allele is inactivated through mutation or promoter methylation, then another allele can be inactivated via loss of heterozygosity (LOH). LOH can be achieved, for instance, by inappropriate chromosomal segregation, mitotic recombination, or loss of a chromosomal segment (reviewed by Balmain *et al*., 2003). In addition to the two-hit hypothesis, a concept of haploinsufficiency has been proposed. Haploinsufficiency is related to a gene dosage effect in which a one-copy loss of a potential TSG could have an impact on cell phenotype (reviewed by Fodde and Smits, 2002).

2.2 Genomic alterations in cancer

Genomic alterations including changes in gene copy number, such as gains and losses, point mutations, and translocations are common in cancer (reviewed by Albertson *et al*., 2003). Together with epigenetic changes (reviewed by Baylin and Ohm, 2006; Jones and Baylin, 2007), they affect gene regulation at the expression level. Many cellular mechanisms, including defects in chromosomal segregation, centrosome dynamics, cell cycle regulation, cellular checkpoints, telomere stability, and DNA damage response, are involved in numerical and structural chromosomal instability. At the gene and protein levels, genetic instability can involve the enzymes that replicate or repair DNA, the proteins that influence chromosomal stability, and the proteins that control apoptosis and cell cycle regulation in response to DNA damage. Mutations in these pathways have been connected to the pathogenesis of cancer in humans and animals (reviewed by Beckman and Loeb, 2005; Gollin, 2005; Bayani *et al*., 2007).

Cancer cells present often aneuploidy with losses or extra copies of whole chromosomes. Changes in chromosome number can be caused by chromosomal instability (CIN). CIN can be strictly defined as the gain or loss of whole chromosomes or chromosomal segments at a higher rate in cancer cells than in normal cells. CIN can be a consequence of mis-segregation of chromosomes during mitosis (reviewed by Lengauer *et al*., 1998; Michor *et al*., 2005). Another type of genomic instability is microsatellite instability (MIN), occurring at the nucleotide level due to defects in mismatch repair. Tumor cells typically present either CIN or MIN (reviewed by Lengauer *et al*., 1998).

Segmental chromosomal gains and losses arise from structural alterations, including translocations, amplifications, and deletions. Chromosome breakage and rearrangement due to defective cell cycle checkpoints, the DNA damage response or loss of telomere integrity can cause structural instability (reviewed by Gollin, 2005). Gene amplification can be initiated by a DNA double-strand break (reviewed by Pierce *et al*., 2001) in cells that progress through the cell cycle with the damaged DNA (reviewed by Albertson, 2006). Double-strand breaks and telomere dysfunction have been suggested to play a role in creating breakage-fusion-bridge cycles, which can lead to amplification (Toledo *et al*., 1992; Hellman *et al*., 2002; reviewed by Albertson, 2006). Amplifications are usually restricted to narrower chromosomal areas than low-level gains. Amplification can be manifested as homogeneously staining regions, double minutes, or distributed at various

locations in the genome (reviewed by Albertson *et al*., 2003). Clinically, copy number changes can have diagnostic or prognostic value. Amplification is one of the basic mechanisms that leads to overexpression of oncogenes in solid tumors. Identification of copy number changes in the cancer genome can therefore help in target identification for therapeutic interventions. Amplification has also been suggested as a mechanism for acquired drug resistance (reviewed by Schwab, 1999; Albertson *et al*., 2003; Albertson, 2006; Myllykangas and Knuutila, 2006). Until recently, recurrent specific translocations were reported mostly in hematological malignancies, such as formation of the *BCR-ABL* fusion gene in leukemia (de Klein *et al*., 1982). In prostate cancer, a recurrent fusion of *TMPRSS2* to *ERG* or *ETV1* was identified in 2005, suggesting that causal gene rearrangements frequently also occur in epithelial cancers (Tomlins *et al*., 2005).

As described above, alterations of different types and sizes occur in the human genome. These can be investigated using a number of methods such as chromosomal banding, fluorescent *in situ* hybridization (FISH), spectral karyotyping (SKY), and comparative genomic hybridization (CGH) (reviewed by Speicher and Carter, 2005). Chromosome-banding techniques, which first allowed the investigation of microscopically observed alterations, are based on a specific banding pattern of each chromosome. In the 1990's, such techniques as SKY (Schröck *et al*., 1996), which yields information about chromosome numbers and structural changes, and chromosomal CGH (Kallioniemi *et al*., 1992) were introduced. Chromosomal CGH is based on detecting relative copy number changes between samples using metaphase chromosomes as hybridization targets. The resolution of chromosomal CGH and conventional cytogenetic methods is limited (~2-10 Mb) (reviewed by Speicher and Carter, 2005). Introduction of array-based CGH, reviewed in Section 5, allowed copy number detection at a gene level, facilitating high-resolution studies of the genome.

3 Head and neck squamous cell carcinoma (HNSCC)

Head and neck cancers represent a heterogeneous group of tumors in the upper aerodigestive tract. The majority of head and neck cancers are squamous cell carcinomas (SCCs), which arise from epithelial cells forming protective layers for cell populations underneath. HNSCC generally includes cancers of the oral cavity, nasal cavity and paranasal sinuses, pharynx, and larynx. Other malignancies in the head and neck area

include salivary gland cancer, thyroid cancer, soft tissue or bone sarcomas, and lymphomas. Ideally, HNSCC tumors should be categorized into specific subgroups to allow an appropriate treatment to be selected and to facilitate prognostication. The prognosis of HNSCC is affected by various factors, including tumor stage (reviewed by Diaz *et al*., 2003). Staging is based on tumor, node, and metastasis (TNM) classification (Wittekind *et al*., 2005). Current methods fail to classify and prognosticate HNSCC in an adequate fashion. Thus, a large research effort is focused on identification of biomarkers to improve these aspects (reviewed by Rodrigo *et al*., 2005).

In 2002, altogether 274 000 and 159 000 new cases of oral cavity and laryngeal cancers were diagnosed worldwide (Parkin *et al*., 2005). The incidence of HNSCC is higher in men than in women. According to the Finnish Cancer Registry, approximately 600 new cases are diagnosed each year in Finland (Finnish Cancer Registry, Cancer Statistics at www.cancerregistry.fi, accessed 14.1.2008). The incidences of the most common cancer in women, breast cancer, and the most common cancer in men, prostate cancer, are over six and eight times higher. In HNSCC, several etiological factors have been identified, the most important of which are tobacco and alcohol (reviewed by Forastiere *et al*., 2001). Viral contribution by the Epstein-Barr virus (EBV) and human papillomavirus (HPV) (reviewed by Syrjänen, 2007) has been established, as has a role for occupational exposure. In recent years, the incidence of oral tongue squamous cell carcinoma (OTSCC) has increased among young patients, raising questions about the potential risk factors (Annertz *et al*., 2002).

The current management for HNSCC includes radiation therapy and surgery, either alone or in combination with chemotherapy (reviewed by Vokes *et al*., 1993; Forastiere *et al*., 2001; Brockstein and Vokes, 2004). Modern surgical and chemoradiation techniques seem effective in improving local control and providing reduction in patient morbidity, thus offering better quality of life. Radical management of HNSCC, especially laryngeal squamous cell carcinoma (LSCC), can still dramatically affect the quality of life of those patients who survive. The overall five-year survival rates for HNSCC have remained low over the past few decades. Every year, 127 000 people worldwide die due to oral cavity cancer, and 90 000 due to laryngeal cancer (Parkin *et al*., 2005). In the US, the five-year relative survival rate for laryngeal cancer is 65%, placing it among the cancers that have shown no significant improvement in survival over the past 25 years (Jemal *et al*., 2007).

3.1 Copy number alterations in HNSCC

Like many other solid tumors, HNSCC is characterized by recurrent patterns of both structural and numerical aberrations (reviewed by Mao *et al*., 2004; Hunter *et al*., 2005; Perez-Ordonez *et al*., 2006). Table 1 summarizes the results of chromosomal CGH review studies (Gollin, 2001; Patmore *et al*., 2005) reporting frequent genomic aberrations and examples of amplifications (Singh *et al*., 2001; Wreesmann *et al*., 2004). CGH data are also available in online databases (www.progenetix.com; cgap.nci.nih.gov/Chromosomes/Mitelman), which are frequently updated as new data are published. Based on chromosomal CGH data collected in the database (Progenetix database; Baudis and Cleary, 2001), in LSCC frequent overrepresentations of 3q, 5p, 7q21-q31, 8q, 11q13, and 18p and underrepresentations of 3p, 5q, 9p, 11q22-q25, 13, and 18q were reported. Similarly, for OTSCC, overrepresentations of 3q, 5p, 7q21, 8q, 9p21-p23, and 11q13 and underrepresentations of 3p, 8p, 18q, and 21q occurred.

Table 1 Common aberrations detected by chromosomal comparative genomic hybridization in head and neck squamous cell carcinoma.

Study	Gains	Losses/Deletions
Gollin, 2001	3q, 5p, 7p, 8q, 9q, 11q13, 20q	3p, 5q, 8p, 9p, 13q, 18q, 21q
Patmore *et al*., 2005	1q, 3q, 5p, 7, 8q, 9q, 10q, 11q13, 14q, 15q, 16, 19, 20, 22q	1p, 3p, 4q, 5q, 8p, 9p, 11q, 18q
	Amplifications	
Singh *et al*., 2001	3q13, 3q25-q26, 5q22-q23, 7q21, 8q24, 11q13-q14, 12p13, 14q24, 20q13.1	
Wreesmann *et al*., 2004	2q32, 3q26, 4p15.3-p16, 5p15, 7q11.2-p12, 7q21, 8p11, 8q24, 9p22-p24, 11q13, 12p13, 18p, 19p	

Some of the copy number alterations also have prognostic significance in HNSCC (reviewed by Wreesmann and Singh, 2005; Akervall, 2006). Gains of 3q21-q29 and 11q13, and loss of 8p21-p22 (Bockmuhl *et al*., 2000) as well as amplification at 11q13, gain of 12q24, and losses at 5q11, 6q14, and 21q11 have been consistently associated with poor prognosis (Wreesmann *et al*., 2004). Overrepresentations of 2q12, 3q21-q29, 6p21.1, 11q13, 14q23, 14q24, 14q31, 14q32, 15q24, and 16q22, and deletions of 8p21-p22 and 18q11.2 have been significantly associated with both shorter disease-free interval and

shorter disease-specific survival (Bockmuhl *et al*., 2000). Here, the nomenclature for alterations (gain, amplification, loss, deletion) is used as presented in the studies.

Wreesmann and Singh (2005) proposed 53 target genes for chromosomal aberrations in SCCs based on multiple chromosomal CGH studies. The genes included *FHIT* (3p14), *PIK3CA* (3q26.3), *APC* (5q21), *EGFR* (7p11), *CDKN2A* (9p21), *PTEN* (10q23), *CCND1*, *TAOS1, EMS1* (11q13), and *DPC4* (18q21.1) for HNSCC (Wreesmann *et al*., 2005). From these genes, *FHIT*, *CDKN2A*, *CCND1*, and *EGFR* as well as *TP53* (17p13) have been linked to genetic progression of HNSCC (Section 3.2).

3.2 Genetic progression model for HNSCC

Slaughter and coworkers (1953) proposed the concept of field cancerization in which changes are induced by carcinogens, such as tobacco and alcohol, throughout the mucosal surfaces of the upper aerodigestive tract. This increases the possibility of future or concurrent disease. In 1996, Califano and coworkers introduced their preliminary genetic progression model for HNSCC (Figure 1). The authors suggested that areas of histopathological abnormality surrounding malignant and premalignant lesions are all generally derived from a single common progenitor clone. Subsequent genetic events in various subclones produce different phenotypic alterations, resulting in histopathologically different regions in a local anatomical area. A subclone that has acquired a particular selective growth advantage through clonal expansions may obtain a dominant position (Califano *et al*., 1996).

The exact role of multiple foci of independent alterations versus clonal expansions in the phenomenon of field cancerization is being debated (reviewed by Ha and Califano, 2003). A classification system where second or subsequent HNSCCs comprise three types has been proposed. Tumors might be derived from the primary SCC itself (recurrence); they might have a different but overlapping spectrum of genetic changes, having developed from an intervening field of abnormal oral mucosa (second field tumors); or they might be true second primary tumors with an independent origin (Tabor *et al*., 2001; reviewed by Braakhuis *et al*., 2003; Hunter *et al*., 2005).

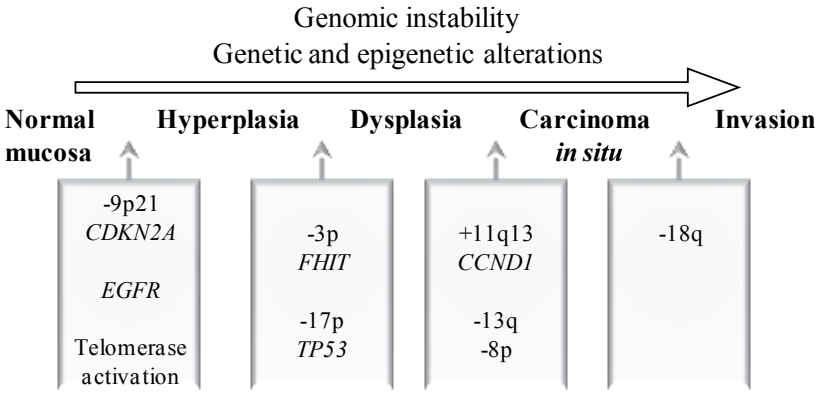

Figure 1 A genetic progression model for head and neck squamous cell carcinoma (modified from Califano *et al*., 1996; Mao *et al*., 2004; Perez-Ordonez *et al*., 2006). Genetic changes can also be present in a normal or benign appearing tissue.

4 Gene expression profiling

Multiple techniques, such as *in situ* hybridization, Northern blot, and reverse transcription-polymerase chain reaction (RT-PCR), allow measurement of gene expression levels. Genome-wide gene expression measurement techniques include microarrays, differential display, and serial analysis of gene expression (SAGE). Differential display allows monitoring of previously unknown genes using PCR primers arbitrary in sequence (Liang and Pardee, 1992). The method is applied to compare gene expression levels between samples. SAGE is a sequencing-based method for identifying expressed genes in a cell and revealing their number (Velculescu *et al*., 1995). DNA microarrays provide a tool for measuring relative differences in RNA levels between samples. Development of microarray technology was preceded by lower-resolution dot blots and nylon filter arrays with radioactive labeling. DNA microarray technology is based on the complementary base pairing property of nucleic acids. In the hybridization reaction, two complementary nucleic acid strands form molecules according to base pairing rules. In the traditional hybridization methods, a specific labeled probe is applied to detect complementary target sequence in the mixture (reviewed by Southern, 2001). In the array technology, a target is a labeled sample hybridized onto the surface. A probe is a known nucleic acid bound to that solid surface. Gene expression microarrays can be categorized according to probe type as cDNA and oligonucleotide arrays (Fodor *et al*., 1991; Schena *et al*., 1995).

The microarray experimental process can be divided into array acquisition, sample preparation, hybridization, image analysis, and data interpretation (reviewed by Holloway *et al*., 2002; Hariharan, 2003). Depending on the microarray platform, one can either

compare expression from two samples, test and reference, on the same array, or use only one array per sample. Accordingly, the result is presented either as a ratio between the expression levels from test and reference samples or as an estimate of transcript levels in one sample (reviewed by Hardiman, 2004). The resulting ratio or intensity data can be analyzed by different softwares.

4.1 cDNA microarrays

cDNA microarray manufacturing requires many steps. To obtain sufficient material for array construction, a collection of well-annotated and characterized cDNA clones is amplified by PCR after culturing. Usually, a glass slide, which can be coated (*e.g.*, poly-L-lysine) to increase the binding efficiency of the probe cDNA, is applied as an array surface. The amplified and purified cDNA sequences are spotted on the slides by a robotic arrayer. The pins of the arrayer collect the probe and deposit small aliquots each time they touch the surface in a contact printing process. A single loading of a pen can provide around two hundred spots (reviewed by Holloway *et al*., 2002). cDNA microarrays can also be manufactured by a noncontact printing method, such as inkjet technology (Blanchard *et al*., 1996), in which electrical pulse is applied to expel a drop of liquid onto the surface.

4.2 *In situ* synthezised oligonucleotide microarrays

Oligonucleotide arrays can be manufactured in a similar fashion as described above by synthesizing individual oligonucleotides prior to spotting. In addition to these delivery-based methods, probes can be *in situ* synthesized nucleotide by nucleotide on a solid support. *In situ* synthesis provides a number of advantages over delivery such as consistent and high yields over the surface of the support (Southern *et al*., 1999). *In situ* synthesis is used by most microarray companies, including Nimblegen (Nuwaysir *et al*., 2002) as well as Affymetrix (Fodor *et al*., 1991) and Agilent Technologies (Blanchard *et al*., 1996).

One of the biggest and oldest manufacturers of *in situ* microarrays is Affymetrix. In array manufacturing, Affymetrix uses light-directed chemical synthesis relying on photolithography adapted from the semiconductor industry. The surface of a solid support, containing photolabile-protecting groups attached to the linker, is illuminated through a

photolithographic mask, creating reactive hydroxyl groups. 3´-O-phosphoramidite-activated deoxynucleosides are then added and coupled to reactive hydroxyl groups. The coupled deoxynucleosides are also 5´-protected. Then, a new mask is applied over the surface and a second round of deoxynucleosides is coupled in regions exposed to light. The cycles are repeated until the desired probes are obtained (Pease *et al*., 1994). The physical size of the array and the lithographic resolution set limits on the technique. The synthesis of probes is simultaneous, and the synthesis time is therefore dependent on the length of the oligonucleotides (Lipshutz *et al*., 1999). In the manufacturing process, multiple arrays are synthesized on a quartz glass wafer, which is then diced. The individual arrays are packaged in cartridges, which protect the arrays and serve as hybridization chambers.

In the early days of Affymetrix array development, Lockhart and coworkers (1996) investigated whether short *in situ* synthesized oligonucleotides quantitatively detected RNA in a cellular population. Array layout was based on probe pair strategy. Each probe pair was composed of a 20-mer that was perfectly complementary to the transcript (perfect match, PM) and a 20-mer that had a single base pair difference in a central position (mismatch, MM). The obtained hybridization signals were specific and quantitatively related to target concentration, and PM hybridizations were distinguished from MM hybridizations. At the higher RNA target concentrations, the hybridization intensity was nonlinearly related to concentration because of probe site saturation (Lockhart *et al*., 1996). The use of MM oligonucleotides allows direct subtraction of background and cross-hybridization signals. Affymetrix's standard GeneChip® gene expression arrays currently contain around 11 different 25-mer oligonucleotide probe pairs, representing one probe set specific for a gene or transcript. Thus, different probe pairs hybridize at different sites of the same RNA transcript. The array itself therefore provides a kind of replication, although it simultaneously presents a challenge for the probe design. The *in silico* design is based on the probe's ability to hybridize with the transcript of interest, its uniqueness, and its lack of similarity to any other highly abundant RNAs in the sample. When probes on the array change, a new set of photolithographic masks is designed (Lipshutz *et al*., 1999). The Affymetrix Human Genome U133 Plus 2.0 array contains over 54 000 probe sets and over 1.3 million features. The more recently released Human Exon 1.0 ST array comprises 1.4 million probe sets and over 5.5 million features (www.affymetrix.com, accessed 14.1.2008).

Longer 60-mer oligonucleotide microarrays can be *in situ* synthesized by, for instance, inkjet technology (Blanchard *et al*., 1996), which is currently used by Agilent Technologies. The principle of the technology was described by Blanchard and colleagues (1996) when they constructed a high-density oligonucleotide array utilizing surface tension effects and inkjet pumps. In this method, an array containing thousands of wells on the surface of oxidized silicon wafer is produced. A hydrophobic coating is applied to the area surrounding the wells. Small amounts of nucleotides are delivered to the hydrophilic wells by inkjet pumps, which utilize capillary action and electrical pulse through a piezoelectric element to transfer the liquid. The machine resembles a four-color inkjet printer. The excess of monomer is rinsed away, followed by acid treatment to the entire array surface to deprotect the new end of the oligonucleotide for the next synthesis step. Depending on how many inkjets are utilized in the process, the manufacturing can be very quick and flexible (Blanchard *et al*., 1996). Inkjet-manufactured oligonucleotide microarrays were examined for sensitivity and specificity (Hughes *et al*., 2001), which are dependent on, for instance, oligonucleotide length. Long oligonucleotides were optimal due to both steric and nonsteric factors fulfilling specificity and sensitivity requirements. In complex cellular populations, transcript ratios at one copy per cell were reliably detected by a 60-mer probe. Thus, a single, carefully chosen oligonucleotide can be more specific than several oligonucleotides, among which cross-hybridization can be challenging to control. The resolution achievable by inkjet technology is high, as demonstrated by the Agilent Technologies human genome CGH oligonucleotide microarray containing 244 000 probes on a single microscope slide. The resolution also enables the profiling of multiple samples on a single array (4x44 000; www.agilent.com, accessed 14.1.2008).

4.3 Experimental aspects of microarrays

To prepare a target for hybridization on microarray, high-quality mRNA or total RNA is extracted from a sample such as a tissue or cell line. Amplification methods can be utilized for small RNA amounts (*e.g.*, Van Gelder *et al*., 1990; Eberwine *et al*., 1992). In the direct two-color labeling method, extracted RNA is reverse-transcribed into cDNA and labeled with fluorochromes such as Cy3 and Cy5. Alternatively, in the indirect labeling method, amino-allyl conjugated nucleotides are incorporated into the first-strand cDNA, followed by chemical coupling of fluorochromes. Indirect labeling does not suffer from a dye bias

effect caused by the test and reference sample being labeled with two fluorochromes of different properties. To compensate the possible dye bias effect of the direct labelling, dye-swap replicates with reversed labeling between replicate arrays can be performed (reviewed by Hardiman, 2004). Affymetrix oligonucleotide arrays utilize a one-color labeling system (Affymetrix, 2004). RNA is reverse-transcribed to cDNA, followed by *in vitro* transcription (IVT) -based amplification of cDNA to cRNA. Biotinylated nucleotides are incorporated to cRNA during IVT, and streptavidin-phycoerythrin conjugates are used in the detection and visualization. Compared with arrays where two samples labeled with different fluorochromes are hybridized simultaneously, Affymetrix oligonucleotide arrays provide an estimate of transcript levels in one sample. Thus, they allow flexibility in sample comparison, which can be performed afterwards by computer. Nowadays, the Agilent two-color labeling system applies the same procedure, but instead of biotin, cRNA from test and reference samples is labeled with Cy5 and Cy3 (Agilent Technologies, 2007a). Recently, Agilent also introduced a one-color Cy3-labeling protocol (Agilent Technologies, 2007b). High-quality commercial manufacturing processes have decreased variability due to microarray production, improving the consistency of microarray results at both the signal and ratio levels. Thus, the choice between one- and two-color arrays is considered more a personal one (Patterson *et al*., 2006).

Labeled targets are hybridized on a microarray, usually overnight, and either manual or automatic washing procedures are used to remove the unbound target. Microarray is then scanned with device that excites fluorescently labeled targets by laser and produces an image of the array. The acquired images are analyzed by software, which places a grid, usually automatically, on the array. In the segmentation and quantification process, software provides measures of the signal for each feature and the background. Signals intensities from the features are background-corrected if deemed necessary. Most image analysis programs also flag low-quality measurements for filtering (reviewed by Hariharan, 2003; Smyth *et al*., 2003a). A number of different algorithms to preprocess and analyze microarray data (reviewed in Section 4.4) have been developed. However, no consensus exists as to which method is the gold standard (Allison *et al*., 2006). When performing image and data analysis, it is important for the experimenter to understand the nature of the data and thereby choose appropriate analysis methods to achieve meaningful results (reviewed by Tilstone, 2003).

4.4 Microarray data analysis

Since experimental design has a clear impact on data analysis, it should be carefully considered already when starting array experiments. The design of the microarray experiment is dependent on the hypothesis, array platform, number of samples, number of biological and technical replicates, amount of RNA, and cost (reviewed by Churchill, 2002; Dobbin *et al*., 2003). The experimental design and other key steps involved in microarray analysis are reviewed, for example, in Imbeaud and Auffray (2005) and an overview is presented in the next paragraphs.

To remove sources of systematic nonbiological variation in the microarray data, a mathematical adjustment, normalization, is performed. Variation can be caused by many factors such as dye bias, experimental conditions, or unequal starting amounts of RNA. For spotted arrays, print-tip group information can be utilized in normalization because a systematic difference may exist between subarrays due to variation in pins of the arrayer. Normalization can be performed within an array, between a pair of arrays, or between multiple arrays (reviewed by Quackenbush, 2002; Yang *et al*., 2002; Hariharan, 2003; Smyth *et al*., 2003a). If multiple arrays are compared, adjustment of scale differences between arrays may be necessary (Yang *et al*., 2002; Smyth and Speed, 2003b). Many of the normalization methods assume that most of the genes on the array, some subset of genes such as housekeeping genes, or a set of exogenous controls have constant expression values. Based on this assumption, the normalization factor is calculated to adjust the data to compensate for systematic variability (reviewed by Quackenbush, 2002; Yang *et al*., 2002; Hariharan, 2003). Multiple different advanced normalization methods are available, one of the most popular being lowess (locally weighted scatter plot smoothing) (Cleveland, 1979), which is suitable for nonlinearly dependent data. Normalization methods can be adapted on different array platforms (reviewed by Quackenbush, 2002).

Since the Affymetrix array design is unique, arrays are usually preprocessed with the company's own Microarray Suite 5 (MAS5) software or current Gene Chip Operating software (GCOS). Software performs the background subtraction, calculates PM and MM probe values, calculates of the probe set value, and scales the data (Affymetrix, 2005). Alternative approaches performing data preprocessing differently at probe level, such as robust multi-array average (RMA) (Irizarry *et al*., 2003) or dChip (Li and Wong, 2001),

are also available. Affymetrix preprocessing algorithms have been reviewed in detail in Hariharan (2003) and are compared, for example, in Irizarry *et al*. (2006).

Preprocessed microarray data are analyzed to identify differentially expressed genes between samples. The analysis can be simply performed using fold change with an arbitrarily determined fixed cut-off. Standard and advanced statistical tests, such as t-test and analysis of variance, can also be applied to the gene expression data, but they often require normal data distribution. Gene expression data are often assumed to be normally distributed after logarithmic transformation (reviewed by Hariharan, 2003; Imbeaud and Auffray, 2005; Allison *et al*., 2006).

Unsupervised classification, clustering, can be used for class discovery in microarray experiments. Clustering analysis can reveal genes that show similar expression patterns as well as illustrate multidimensional gene expression data effectively. Most clustering techniques applied in gene expression analysis are hierarchical resulting a tree structure (Eisen *et al*., 1998), but nonhierarchical ones, such as k-means (Tavazoie *et al*., 1999) or self-organizing map (SOM) (Kohonen, 2001), have also been utilized. When applying clustering methods, the user must decide several parameters, which have an effect on the results. In addition to the unsupervised approaches, supervised algorithms can be used if some information on how to group data is available. These class prediction methods preferably include independent training and test data sets to build and evaluate the classifier (reviewed by Ringner *et al*., 2002; Allison *et al*., 2006; Dupuy and Simon, 2007).

To reveal which biological processes are active among differentially expressed genes, gene ontology (GO) (www.geneontology.org) (The Gene Ontology Consortium, 2008) or pathway analysis can be performed. GO analysis groups genes into functionally meaningful classes using different GO terms. The terms are divided into three main branches: biological process, cellular component, and molecular function. Gene set enrichment analysis (GSEA) focuses on differentially expressed gene sets instead of single genes (Mootha *et al*., 2003; Subramanian *et al*., 2005). In GSEA, gene sets can be created based on GO or pathway information, highlighting the altered biological processes.

4.5 Validation of microarray data

Developments of microarray technology have reduced the need for additional validation at gene expression level to rule out false-positive results. Typically, gene expression validation has been carried out using methods such as Northern blot or real-time RT-PCR (*e.g.*, Yuen *et al*., 2002; Ginos *et al*., 2004; Larkin *et al*., 2005; Canales *et al*., 2006). Canales and coworkers (2006) demonstrated a good correlation between quantitative gene expression platforms and microarrays. Discrepancies in expression measurements were mostly dependent on the detection limit of the array platform and differences in the probe sequence. Both methods gave accurate results for the specific sequence they were measuring (Canales *et al*., 2006). Inherent limitations of microarray technology due to sensitivity, especially at lower expression levels, and accuracy as compared with RT-PCR methods exist (reviewed by Draghici *et al*., 2006).

To validate functional relevance of the gene expression microarray results, protein level studies using immunohistochemistry, Western blotting, 2-D gel electrophoresis, and mass spectrometry or RNAi experiments can also performed (*e.g.*, Chung *et al*., 2004; Choi *et al*., 2005; Tonon *et al*., 2005). As evolving standards and comparability studies allow effective use of publicly available data, meta-analysis provides an attractive option for further studies (Rhodes *et al*., 2004; Rhodes *et al*., 2007).

5 Copy number profiling

Copy number profiling by CGH can be performed on microarrays containing oligos, cDNA, or bacterial artificial chromosome (BAC) clones instead of chromosomal targets described in Kallioniemi *et al*. (1992). The resolution was significantly improved by the substitution of the metaphase chromosomes by an array with large inserts from genomic sequences (Solinas-Toldo *et al*., 1997). Pinkel and coworkers (1998) described the implementation of array CGH (aCGH) using genomic clones (BAC, P1) as probes on an array. In 2004, a tiling resolution array consisting of 32 433 overlapping BAC clones covering the entire human genome was constructed (Ishkanian *et al*., 2004).

cDNA arrays in CGH were introduced almost at the same time as BAC arrays (Pollack *et al*., 1999). An oligonucleotide-based aCGH method, representational oligonucleotide microarray analysis (ROMA), was described a few years later (Lucito *et al*., 2003). The

principle of ROMA is similar to that of cDNA- and BAC-based methods, but it includes complexity reduction of DNA samples by representations to increase the signal to noise. This is accomplished by PCR performed in a specific size range after digestion by restriction enzymes. Complexity reduction is applied also on Affymetrix SNP oligonucleotide arrays (Bignell *et al*., 2004; Zhao *et al*., 2004), which allow simultaneous detection of LOH and copy number. Oligonucleotide-based aCGH without complexity reduction was introduced only four years ago (Barrett *et al*., 2004; Brennan *et al*., 2004; Carvalho *et al*., 2004).

In aCGH, different labeling and hybridization protocols can be used (Pollack *et al*., 1999; Snijders *et al*., 2001; Barrett *et al*., 2004). The sample material is total genomic DNA, which is digested by restriction enzymes. Test and reference samples can be then differentially labeled using random priming and cohybridized on array with blocking DNA, such as Cot-1 DNA, which suppresses signals from repetitive sequences. Depending on the array platform, labeling protocols vary, but typically test and reference samples are labeled with Cy3 and Cy5. The data can be analyzed with commercial (*e.g.*, Agilent Technologies) or custom-developed programs (*e.g.*, Autio *et al*., 2003).

DNA of phenotypically normal individuals used as a reference sample in aCGH can possess marked genomic variability, affecting data interpretation. One of the most common forms of variability in the human genome is in SNPs (*e.g.*, HapMap project, www.hapmap.org), but a few years ago, two studies reported wide-spread copy number variants (CNVs) in a normal population (Iafrate *et al*., 2004; Sebat *et al*., 2004). CNV can be defined as a segment of DNA that is one kb or larger and is present in variable copy number in comparison with a reference genome. CNVs include insertions, deletions, and duplications as well as large-scale CNVs, which are variants involving DNA segments over 50 kb that are detectable by CGH (reviewed by Feuk *et al*., 2006). Efforts to collect information about CNVs are on-going, *e.g.* the Database of Genomic Variants (projects.tcag.ca/variation/) (Iafrate *et al*., 2004). Interpretation of copy number changes between samples from different projects and standardization of databases of structural variants can be complicated due to various reference samples used in the projects (reviewed by Feuk *et al*., 2006).

Compared with aCGH, chromosomal CGH fails to produce high-resolution data because metaphase chromosomes are utilized as hybridization targets. Therefore, the boundaries

and sizes of the genomic alterations are impossible to define accurately. aCGH with cDNAs, BACs, or oligos as targets provides high-resolution data with a good dynamic range (reviewed by Pinkel and Albertson, 2005; Ylstra *et al*., 2006). Compared with BAC arrays, oligonucleotide arrays allow better resolution because of the size of the BAC clones. However, in many oligonucleotide platforms, 3-5 adjacent oligonucleotides are used to define gain or loss (reviewed by Ylstra *et al*., 2006). Interestingly, allele-specific copy number profiling on Affymetrix SNP arrays has been described (LaFramboise *et al*., 2005). As a method, CGH can detect only unbalanced rearrangements and polyploidy or balanced rearrangements remain undetected.

6 Molecular profiling of cancer by array technologies

Gene expression microarrays have a wide variety of applications in biomedical research, but 80% of the publications have been reported to rise from the field of cancer research (reviewed by Ewis *et al*., 2005). These applications include classification of tumors into biological and diagnostic subgroups, identification of molecular mechanisms behind cancer, identification of clinical associations such as prognosis, response to treatment, and disease progression, and identification of drug targets (reviewed by Gerhold *et al*., 2002; Ewis *et al*., 2005). As an example, in a study of 17 HNSCC patients, 375 genes discriminating between two genotypic subtypes of HNSCC were identified with different clinical outcome (Belbin *et al*., 2002). Another study of 41 HNSCCs reported a gene expression signature associated with recurrent disease (Ginos *et al*., 2004). A study of 60 HNSCCs revealed four distinct subtypes of HNSCC with differences in recurrence-free survival as well as overall survival based on gene expression. Furthermore, a set of genes was proposed to be predictive of lymph node metastases (Chung *et al*., 2004). Roepman and coworkers (2005) also identified a predictor of lymph node metastases based on the differential expression of 102 genes using 82 tumors. These results and over 20 other HNSCC gene expression microarray studies reviewed in Choi and Chen (2005) revealed considerable heterogeneity in experimental design, number of samples used, site and stage of disease, ratio of tumor-to-stromal cells analyzed, microarray platform used, and validation of results by other methodologies. The heterogeneity makes it challenging to compare and integrate the results from different studies. Microarrays have also been used in copy number profiling of cancer (*e.g.*, Pollack *et al*., 1999; Snijders *et al*., 2001; Zhao *et*

al., 2004; LaFramboise *et al*., 2005; reviewed by Pinkel and Albertson, 2005). aCGH studies of HNSCC are described in detail in Section 8.1.1.

In addition to expression and copy number analysis, microarrays have been applied in genotyping and sequencing applications in cancer (reviewed by Fan *et al*., 2006a). Since the concentration of mRNA can be poorly correlated with the protein levels in the cell, high-throughput array formats to measure the actual protein levels have also been actively developed (Zhu *et al*., 2001; reviewed by MacBeath, 2002; Mitchell, 2002; LaBaer and Ramachandran, 2005). To verify molecular markers, a tissue microarray (TMA) can be applied (Kononen *et al*., 1998). In TMA, tissue from paraffin-embedded tumor blocks is arrayed into a new paraffin block until the desired set of tumors is printed on the new array block. Hundreds of sections could be cut from each tumor array block. TMA enables the investigation of changes in DNA, RNA, or protein levels from a large set of tumors on one slide. Further microarray applications include chromatin immunoprecipitation followed by array detection (ChIP-on-chip) to examine the interactions between DNA and proteins, epigenetic studies of methylation, and cell array to perform functional screenings (reviewed by Hoheisel, 2006). Thus, array-based methods can provide molecular profiles which can be integrated for more comprehensive understanding of complex diseases such as cancer (Figure 2).

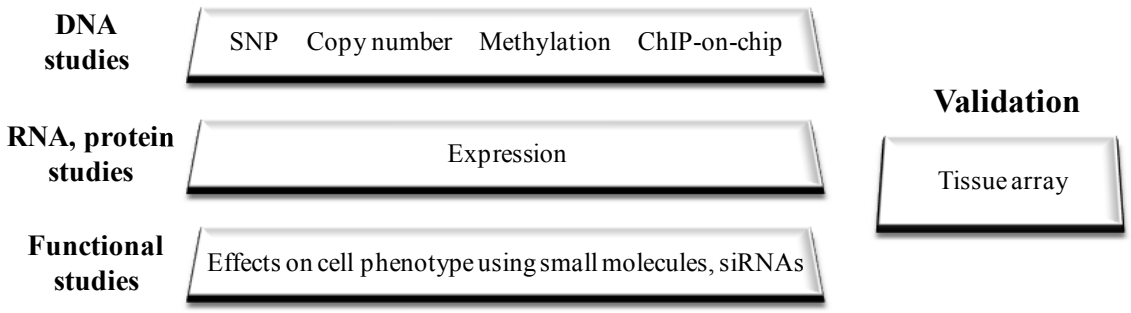

Figure 2 Array-based approaches for molecular profiling (modified from Weinstein, 2006).

Aims of the study

The aims of the study were the following:

1. To compare data from different gene expression microarray platforms.

2. To integrate copy number and gene expression microarray data to identify genes in which expression is changed in association with underlying genetic alteration.

    This genome-wide approach was applied to discover genes potentially important in HNSCC pathogenesis.

Materials and methods

The materials and methods used are listed in Table 2 and described in detail in the original publications. In addition to methods presented in Table 2, standard molecular biology methods, such as gel electrophoresis and spectrophotometry, were used. The UT-SCC cell lines were kindly provided by the Department of Otorhinolaryngology-Head and Neck Surgery at Turku University Central Hospital (TUCH), University of Turku, Finland. Other cell lines were from the American Type Culture Collection (ATCC, VA). The use of LSCC clinical sample material was approved by the Research Ethics Board at the Department of Otorhinolaryngology, Helsinki University Central Hospital, and the Joint Ethics Committee of TUCH and the University of Turku.

Table 2 Materials and methods used in Studies I-III.

Cell lines and tumor samples	**Study**
Breast cancer cell lines: MDA-MB-361, MDA-MB-436, BT-474, MCF-7, HBL-100	I
LSCC cell lines: UT-SCC-8, -11, -19A, -19B, -29, -34, -38, -42A, -49, -75	II, III
OTSCC cell lines: UT-SCC-16A, -16B, -21, -24A, -24B, -30, -40, -67, -73, -76A, -76B, -81, -87, -95	III
SCC-4, SCC-9, SCC-15, SCC-25	
10 LSCC tumor samples	II
Methods	
Cell culturing	I, II, III
DNA extraction	II, III
Total RNA extraction	I, II, III
Lab-on-a-Chip/Agilent Bioanalyzer	I, II, III
Expression profiling on Affymetrix microarray	I
Expression profiling on custom-made microarray	I
Expression profiling on Agilent microarrays	I, II, III
Copy number profiling on Agilent microarrays	II, III
Microarray data analysis by custom-developed and commercial programs	I, II, III
Gene ontology analysis	II, III
Pathway analysis	II, III
Real-time quantitative RT-PCR	III
Immunohistochemistry	II
Sequencing	I

Results and Discussion

7 Comparison of the data from different gene expression microarray platforms (I)

7.1 Gene expression profiling on microarrays

We compared data from three different microarray platforms, Affymetrix 13k HG-U95v2 *in situ* oligonucleotide array, Agilent Technologies 13k Human 1 cDNA array, and custom-made 13k cDNA array. Four ATCC breast cancer cell lines were used as test samples and the HBL-100 cell line as a reference sample. The laboratory protocols and data analysis methods provided by commercial manufacturers were applied without modifications to reflect the situation of the average biologically oriented user when starting array experiments. For custom-made arrays, two different filtering-normalization options were applied. On Affymetrix arrays, we performed data analysis using both MAS5 and RMA because of the wide acceptance of RMA by the research community.

To determine the concordance between the three different microarray platforms, correlation coefficients were calculated. Using common probes between platforms, Pearson and Spearman correlation coefficients gave similar results, Pearson correlations being slightly better. The unfiltered gene expression data (n=2340) from the commercial platforms correlated well across experiments (r=0.70-0.83), whereas correlations between the custom-made and either of the two commercial platforms were consistently lower (r=0.53-0.63). SOM analysis for filtered data confirmed the results, illustrating better concordance on commercial platforms than on the custom-made platform. Our results demonstrated that filtering improved the correlations. Filtered data (n=1093-1206) between commercial platforms showed correlation coefficients from 0.78 to 0.86, whereas between custom-made and commercial platforms the correlations ranged from 0.62 to 0.76 (Table 3). We observed better correlations in unfiltered data between Affymetrix and other platforms when using RMA as compared with MAS5. The two data analysis options for custom-made arrays produced comparable outcomes when used in unfiltered and filtered data.

Table 3 Range of Pearson and Spearman correlation coefficients between different microarray platforms in filtered data.

Comparison	Correlation
Affymetrix - Agilent	0.78-0.86
Affymetrix - custom-made	0.66-0.76
Agilent - custom-made	0.62-0.73

Differences between microarray platforms can arise from a variety of factors, including the wrong probe or incomplete annotation on the array, splice variants, different kinds of hybridization properties between target and oligonucleotide compared with target and full-length cDNA clone, low-quality spots on the custom-made array, homologies between genes, or a suboptimal image or data analysis algorithm. We were able to verify 16 incorrect sequences on the custom-made array by sequencing 28 cDNA clones giving discrepant results between arrays. Sequencing of an additional unselected set of 90 clones from one random plate in the cDNA library revealed a 16% error rate. Despite these results, hierarchical clustering (n=877) grouped the same cell lines, independent on the platform, together, implying that biological differences still dominanted over technical ones.

In general, cDNA arrays are likely to be prone to errors because of the handling of thousands of cDNA clones through multiple processing steps. Up to a 30% error rate has been reported in cDNA libraries (Halgren *et al*., 2001; Knight, 2001). Thus, cDNA clones should be resequenced as a quality procedure even those from sequence-verified libraries, and the local error rate should always be reported. Unfortunately, the sequencing of thousands of clones is often an impossible task due to time constraints and cost. Potential error rate should be also kept in mind if old cDNA microarray data are reanalyzed, for example, in meta-analysis studies. When manufacturing oligonucleotide arrays, much of the tedious preparation can be avoided. On the other hand, manufacturing of oligonucleotide arrays requires knowledge of the gene sequence to perform the synthesis. The synthesized oligonucleotides are as reliable as their sequence data. In addition to manufacturing differences, the hybridization properties as well as the targets, such as different splice variants of the same gene, of longer cDNA clones and shorter oligonucleotides can be different.

Our results revealed challenges in manufacturing and quality control of custom-made microarrays. Such printing conditions as temperature and humidity can affect spot quality. Furthermore, pins can be blocked during the printing process in the case of dust on the slides or impurities in a probe sample. This can lead to spot bleeding, nonuniform pixel distribution, and missing spots. The variable quality of spots on the custom-made array presents a challenge for image analysis software and data filtering. In commercial arrays, the quality was more uniform. Therefore, commercial arrays could provide a platform where technical replication may not be as crucial as on custom-made arrays. On the custom-made microarrays, however, the within-platform correlation of arrays belonging to the same print batch was better (average 0.9; n=7412-9982) than on Agilent cDNA arrays (dye-swap average 0.77; n=10 080-10 596). Unfortunately, the concordance between custom-made and commercial platforms was poorer than that between commercial platforms.

Annotating of the probes on the array is a demanding and dynamic process that may cause variability in the results obtained from different platforms. Many manufacturers provide detailed information about the probes on the array. These annotations might not, however, be up-to-date, and thus, they should be updated before analysis. For example, the latest information of the Human Genome Project is deposited in public databases (genome.ucsc.edu; www.ensembl.org; www.ncbi.nlm.nih.gov) and is readily available for annotation purposes, even at sequence level. Correct annotations are crucial for identifying valid target genes and drawing correct conclusions from an array study. Here, we annotated probes using UniGene cluster ID as the common identifier between different platforms. Annotation at the transcript level would be more optimal, especially in the case of the oligonucleotide arrays. Furthermore, as manufacturers nowadays usually provide the probe sequence information, probes should be mapped to the most up-to-date genome sequence for annotation. A recent study reported problems in probe annotations on Affymetrix arrays caused by probe design being older than the current knowledge of genome sequence (Dai *et al*., 2005). This had a profound effect on results, underlining the importance of keeping probe annotations updated. Moreover, multiple available analysis options for each platform make comparisons challenging. In addition to manufacturer's recommended algorithms and programs, data analysis methods are numerous, which may influence the correlation between platforms. For example, here, we noticed a difference in

the results when using MAS5 and RMA for Affymetrix data. Overall, the original raw data should be available for other researchers for reanalysis.

In our study, the discrepant findings between platforms were due to clone errors on the custom-made microarrays, old annotations, or unknown causes. The results illustrate points that should be taken into account when comparable data from microarrays are desired in research settings as well as in clinical settings. In concordance with our results, when currently profiling common organisms, such as human, custom-made arrays are less popular than commercial ones, and cDNA arrays are less popular than oligonucleotide arrays (reviewed by Gershon, 2004). However, cDNA arrays might still be useful in cases in which sequence information is not available, and custom-made arrays when no commercial option exists. Diagnostic and clinical applications of the DNA microarray technology will be dependent on generally acceptable and comparable platforms, placing far more stringent demands on quality control than in research settings (reviewed by Petricoin *et al*., 2002).

Despite rapid development of microarray technology and availability of a vast amount of gene expression data, we still face challenges in the clinically orientated applications of microarrays (reviewed by Abdullah-Sayani *et al*., 2006; Tinker *et al*., 2006). Encouragingly, the results from breast cancer research suggest that gene expression profiles can predict the outcome of disease as accurately as the currently used clinical parameters (van't Veer *et al*., 2002). Importantly, the results have been translated to a diagnostic test approved by the U.S. Food and Drug Administration (FDA) (Glas *et al*., 2006). Agendia's MammaPrint® breast cancer prognosis test (Agilent oligonucleotide array) is the second microarray-based test after Roche's pharmacogenetic AmpliChip® CYP450 test (Affymetrix oligonucleotide array) approved for clinical use. Despite these few successful examples, criticism about the usefulness and added value of gene expression profiles in the clinics compared with conventional prognostic factors has been voiced (reviewed by Michiels *et al*., 2007).

7.2 Comparability of gene expression microarray data

When starting this project in 2002, scarce publications comparing results from different gene expression microarray platforms existed, and only a few more were published during the project. The results were inconsistent, with some studies reporting agreement (Kane *et*

al., 2000; Hughes *et al*., 2001; Yuen *et al*., 2002; Barczak *et al*., 2003; Wang *et al*., 2003) and others not (Kothapalli *et al*., 2002; Kuo *et al*., 2002; Li *et al*., 2002; Tan *et al*., 2003). Nowadays, microarray technology is more mature and multiple systematic comparisons between different array platforms as well as laboratories have been published. Still comparability and reproducibility issues arise regularly, as evidenced by two recent large-scale efforts presented in the next paragraphs (Bammler *et al*., 2005; Irizarry *et al*., 2005; Larkin *et al*., 2005; MAQC Consortium, 2006).

In the Nature Methods Journal three articles published in 2005 (Bammler *et al*., 2005; Irizarry *et al*., 2005; Larkin *et al*., 2005) brought together work by researchers from 17 laboratories, using more than 15 microarray platforms. Larkin and coworkers (2005) used both Affymetrix and spotted cDNA arrays to examine gene expression changes in a mouse heart. The obtained expression data were consistent and independent of platform for most of the common genes, as biological effect dominated over platform effect. Only a small subset of genes, 9% (n=504), from the two platforms gave discordant results. Based on quantitative RT-PCR results, the discordance could be due to different splice variants of the same gene measured by two platforms. In the second study by Irizarry and coworkers (2005), Affymetrix, spotted cDNA, and spotted long oligonucleotide arrays using identical RNA samples were applied in different laboratories. Researchers observed that laboratory had a larger effect on many parameters than platform. The results from the best laboratories, which presented repeatable data within the laboratory, agreed well with the highest cross-platform correlation between Affymetrix and cDNA array, 0.48, and between Affymetrix and spotted long oligonucleotide array, 0.57. In the third study by Bammler and coworkers (2005), expression data were generated in eight laboratories that compared two standard RNA samples using different cDNA and oligonucleotide microarray platforms. Although intra-platform reproducibility within a single laboratory was good, reproducibility across laboratories was generally poor for the same platform. The implementation of standardized protocols for all aspects of the study including RNA labeling, hybridization, microarray processing, data acquisition, and normalization, however, increased inter-laboratory reproducibility. After standardization, the correlation for gene expression ratios on the commercial oligonucleotide array varied from 0.79 to 0.93 between laboratories. To compare correlations between seven different array platforms, 502 common genes were identified. The obtained cross-platform correlations varied both within and between laboratories (r=0.11-0.76). However, the identified

biological themes, as defined by GO analysis, were quite consistent. Based on the result, the authors concluded that microarray platform can be a source of gene expression variability and standardization is essential. In these three studies, the importance of annotation and preprocessing were acknowledged. Here, the relative expression measures seemed to give more reliable results than intensities. Overall, the results imply that microarray data can be reproducible and comparable between different platforms as well as between laboratories, but standardization and user experience are important issues affecting data quality, especially when data from different laboratories are compared.

Another recent effort is the MicroArray Quality Control project (MAQC) (www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/index.htm) led by FDA scientists and involving 137 participants from 51 organizations (MAQC Consortium, 2006). In this project, gene expression levels of two RNA samples were measured on the following seven microarray platforms: Applied Biosystems, Affymetrix, Agilent Technologies (both one- and two-color), GE Healthcare, Illumina, spotted microarrays using Operon (NCI) oligonucleotides, and Eppendorf. Each microarray platform was used at three test sites, and five replicates were performed at each site. The study indicated that microarray results were generally repeatable within a test site, reproducible between test sites, and comparable across platforms, despite probe sequence differences and unique protocols for labeling and detection of expression. Between platforms, the median rank correlation of log ratios was 0.87, and the smallest rank correlation was 0.69. Five accompanying articles concerning issues of evaluation of microarray results compared with quantitative gene expression platforms (Canales *et al*., 2006), use of titration pools to assess microarray performance and normalization issues (Shippy *et al*., 2006), evaluation of external RNA controls to assess microarray performance (Tong *et al*., 2006), comparison of one- and two-color platforms (Patterson *et al*., 2006), and a rat toxicogenomic study that validated the findings of the MAQC main data set (Guo *et al*., 2006) were published in the same issue. The complete data set created in the MAQC study is the most extensive published to date and is freely available to the research community.

Taken together, recent studies have demonstrated concordant results between different gene expression microarray platforms. We and others have also reported that biological differences exceed technical ones. The terms “concordant” and “discordant” can also be quite subjective and dependent on perspective. For example, in our study, we found

reasonably good correlations after filtering between all platforms (r=0.62-0.86), especially between commercial ones (r=0.78-0.86). These results are in line with those of large-scale studies (Bammler *et al*., 2005; Irizarry *et al*., 2005; Larkin *et al*., 2005; MAQC Consortium, 2006). In a clinical diagnostic setting, however, these correlations are inadequate. One measure of comparability of microarray results has been suggested to be an accurate and consistent prediction of individual patient outcome. In a study of breast cancer, the authors demonstrated that even though different gene sets from different microarray studies were used for prognostication of patients, four of the five tests showed significant agreement in outcome predictions for individual patients. Thus, they probably track a common set of biological characteristics (Fan *et al*., 2006b).

7.3 Reliability and reproducibility – towards standardization

Standardized laboratory practices allow better comparability of microarray data from different platforms and laboratories (Bammler *et al*., 2005). Reliable results are a prerequisite for clinical applications (reviewed by Petricoin *et al*., 2002). The Microarray Gene Expression Data (MGED) group was established in 1999 to develop standards for describing microarray experiments. Gene expression data are meaningful only when sufficient background information of data is provided in a standardized fashion. In 2001, MGED published a recommendation, the Minimum Information About a Microarray Experiment (MIAME), to define the information needed in association with a microarray experiment (Brazma *et al*., 2001). The purpose of MIAME is not to promote any particular format, but to outline certain common principles. MIAME has not been adopted by all scientific journals.

Many steps occur between a raw image and final gene expression values, but most of these steps are ill-defined in publications. According to MIAME guideline, the following six sections should be described in any microarray experiment: the raw data for each hybridization, the final processed data for the set of hybridizations in the experiment, the essential sample annotation, the experiment design, sufficient annotation of the array design, and essential experimental and data processing protocols (www.mged.org/Workgroups/MIAME/miame.html, accessed 14.1.2008). The massive amounts of original data created in microarray experiments can be stored in a MIAME-compliant fashion in publicly available databases such as Gene Expression Omnibus

(www.ncbi.nlm.nih.gov/geo/) (Barrett *et al*., 2007) and ArrayExpress (www.ebi.ac.uk/microarray/ArrayExpress/arrayexpress.html) (Parkinson *et al*., 2007). In addition to MIAME, other ongoing standardization efforts, such as the External RNA Controls Consortium (ERCC) (Baker *et al*., 2005), best practices for Affymetrix arrays (Tumor Analysis Best Practices Working Group, 2004), and MAQC (MAQC Consortium, 2006), exist.

8 Copy number and gene expression profiling on microarrays (II, III)

In Studies II and III, we investigated genes targeted by genomic alterations in HNSCC by integrating copy number and gene expression microarray data. We concentrated on two HNSCC sites: OTSCC, which is anatomically part of the oral cavity, and LSCC. Based on results from the first study, we applied commercial cDNA and oligonucleotide microarrays from Agilent Technologies in copy number and expression profiling, respectively.

Head and neck SCCs present a number of recurrent gains and losses as well as well-defined amplifications and homozygous deletions. Previous HNSCC microarray studies have focused on either copy number or gene expression profiling on a genome-wide scale, performing further analysis for a few selected genes by such methods as quantitative PCR or RT-PCR (Cromer *et al*., 2004; Baldwin *et al*., 2005; Snijders *et al*., 2005). None of the studies have integrated genome-wide data in a systematic manner to achieve more accurate information about genes that are activated or inactivated by copy number alteration with focus on specific HNSCC site. Such studies have been published in various other cancer types (Guo *et al*., 2002; Hyman *et al*., 2002; Pollack *et al*., 2002; Wolf *et al*., 2004; Heidenblad *et al*., 2005; Tonon *et al*., 2005). In bladder cancer, a converse integrative microarray approach to examine genome-wide copy number independent transcription, possibly regulated by epigenetic mechanisms, has also been taken (Stransky *et al*., 2006).

8.1 Copy number profiling of HNSCC

We copy number profiled 10 LSCC cell lines and 10 LSCC tumor samples on cDNA microarrays providing a theoretical resolution of ~400 kb across the genome (II). We

identified copy number alterations in several regions, including gains at 3q, 5p, 7p, 8q, 9q, 11q, 14q, 15q, 16q, and 20, as well as losses at 3p, 4q, 5, 6q, 8p, 9p, 10p, 11q, 13q, 18q, and 21. The analysis was followed by similar profiling of 18 OTSCC cell lines (III). Frequent copy number gains in OTSCC cell lines were identified at 3q, 5p, 7p, 8q, 9, 14, and 20 and losses at 3p, 4q, 8p, 10p, 13q, and 18q. Thus, these two HNSCC sites seemed to have similar alterations, which were also in concordance with earlier chromosomal CGH studies (Section 3.1). Figure 3 summarizes the copy number alterations across all LSCC and OTSCC samples.

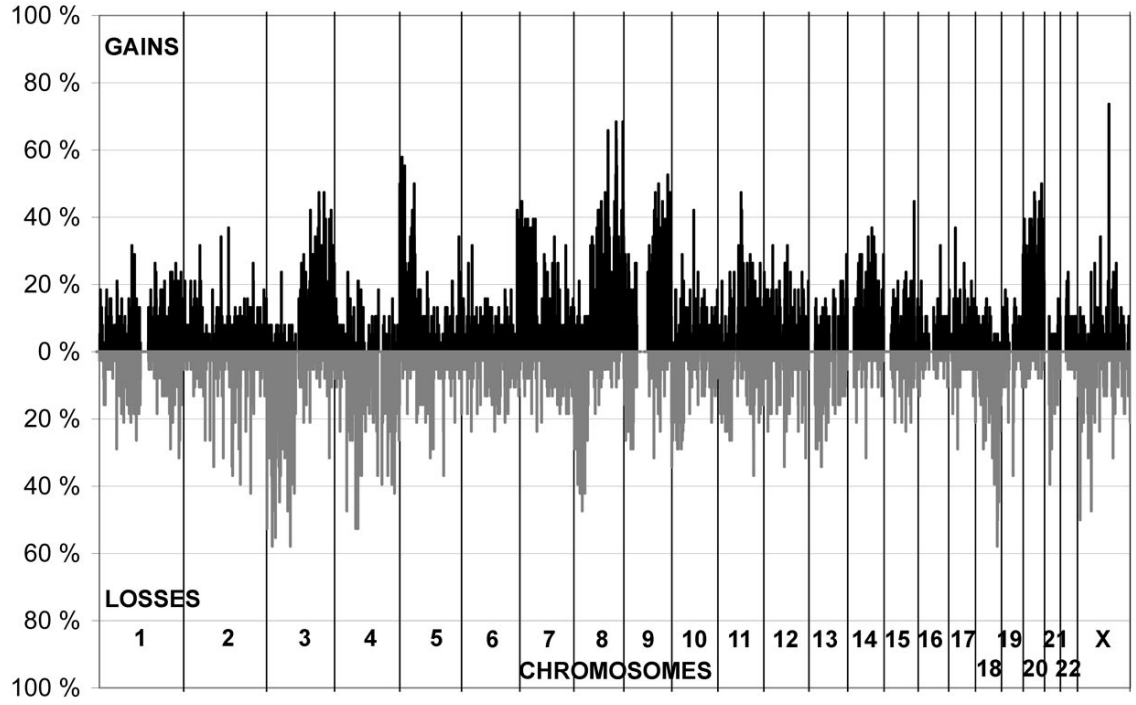

Figure 3 Frequency of copy number alterations (gain>1.3, loss<0.7) in 38 HNSCC samples. All copy number data were processed similarly and were plotted using 750 kb smoothing with CGH-Plotter (Autio *et al*., 2003). The clone coverage in the chromosome Y was poor and thus, it was omitted from the analysis.

We and others have reported that homozygous deletions are challenging to distinguish from one copy losses on a cDNA platform. Therefore, the LSCC cell line UT-SCC-75 and OTSCC cell lines UT-SCC-24B, UT-SCC-30, UT-SCC-67, and UT-SCC-76A, all of which presented multiple genomic alterations on cDNA arrays, were evaluated on 185k oligonucleotide CGH arrays. The array provided a resolution of dozens of kbs and was capable of detecting intragenic deletions. We identified multiple homozygously deleted regions (both copies of the gene lost) containing TSG previously indicated in HNSCC such as *LRP1B* at 2q21.2 (Nakagawa *et al*., 2006), *FHIT* at 3p14.2 (Mao *et al*., 1996;

Virgilio *et al*., 1996), *CSMD1* at 8p23.2 (Sun *et al*., 2001), *MTAP/CDKN2A/CDKN2B* at 9p21 (Reed *et al*., 1996; Worsham *et al*., 2006), and *MRO/ME2/ELAC1/SMAD4* at 18q21 (Kim *et al*., 1996). In addition, previously unrecognized homozygously deleted targets in HNSCC, such as *ITGAV* at 2q31-q32, *PDE4D* at 5q12, and *IL1RAPL2* at Xq22.2-q22.3, were located. From these genes, we observed simultaneous underexpression of *ITGAV*, which codes integrin alpha chain V. Although no decreased expression of *PDE4D* in association with homozygous deletion was detected, the gene has been reported by others homozygously deleted in lung adenocarcinoma (Weir *et al*., 2007). The authors located no somatic mutations in *PDE4D*. However, when combining the results of these individual studies, some role may exist for *PDE4D*, which participates in degradation of cAMP, in different type of cancers. These CGH results from oligonucleotide array illustrate the potential of microarrays in discovering new TSGs. Overall, our data showed that genomic alterations in an individual HNSCC sample can be composed of low-level gains and losses as well as narrower amplifications and homozygously deleted regions, all of which can be identified by array-based methods.

8.1.1 Genome-wide copy number studies on microarrays

Our results as well as previous data show that genomic alterations are recurrent in HNSCC and are likely to contain genes with pathogenetic relevance in the disease. Thus far, only a few comprehensive high-resolution genome-wide aCGH studies of HNSCC have been published (Table 4).

Table 4 Genome-wide aCGH studies in HNSCC. Smaller specific altered areas distinguished by the studies are presented in italics. Definitions of gain, loss, amplification, and deletion vary between Studies.

Study	Gains/Amplifications	Losses/Deletions
Baldwin *et al*., 2005	3q, 8q, 9q, 11q, 14q, 20q	3p, 4, 5q, 8p, 9p, 10q, 11, 18q, 21q
	3q23, 5p15.2, 7p11.2, 7p12.3-p13, 7q21.2, 7q35, 11q13.3, 11q22.2-q22.3	*2p15, 4q34.3, 8p23.2, 16q23.2*
Järvinen *et al*., 2006 (II)	3q, 5p, 7p, 8q, 9q, 11q, 14q, 15q, 16q, 20	3p, 4q, 5, 6q, 8p, 9p, 10p, 11q, 13q, 18q, 21
	5p12, 5p15.33, 8q24.12-q24.13, 11q13.2-q13.4, 14q24.3, 20p13, 20q13.13-q13.31, 20q13.33	*4q13.3, 4q31.3-q32.1, 9p21.3 10p12.31-p13, 11q22.3, 18q21.33-q22.3*
Järvinen *et al*., 2008 (III)	3q, 5p, 7p, 8q, 9, 14, 20	3p, 4q, 8p, 10p, 13q, 18q
	6q12-q14, 9p13.3, 9p22.3-p24.3, 10q11.21, 11p11.2-p13, 11q12.1-q12.2, 11q13.2-q13.4, 14q24.3, 22q11.21	*2q21.2, 2q31-q32, 3p14.2, 5q12, 8p11-p12, 8p23.2, 9p21, 18q21, Xq22.2-q22.3*
Smeets *et al*., 2005	3q22.2-qter, 5p15.2-pter, 8p11.2-qter, 9q22-q34.1, 20	11q14.1-qter, 13q11-q33
Snijders *et al*., 2005	3q, 8q, 11q, 20	3p, 4, 5q, 8p, 9p, 18, 21
	2q14.2, 3q24-q25, 4p15.2, 5p13.2, 6q12, 7p11.2, 7q21.2, 8p12, 9p13.3, 9p24.1, 11p11.2, 11p13, 11q13.3, 11q13.5, 11q22, 12q15, 18q11.2, 20p12.2	
Sparano *et al*., 2006	3q, 5p, 8q, 9q, 20q	3p, 8p, 13q, 18q
	3q23, 3q25.2, 3q26.31, 3q26.33, 3q27.1-q27.3, 5p13.1-p13.2, 5p15.33, 8q11.21, 8q13.3-q21.11, 8q21.3-q22.1, 8q24.11-q24.12, 8q24.21, 8q24.23-q24.3, 9q34.3-qter, 20q13.33	*3p12.3-p13, 3p21.32, 3p22.1-p22.3, 3p23-p24.1, 3p24.2-p24.3, 3p25.3, 3p26.3, 3q22.1-q22.2, 8p12, 8p23.2, 13q12.3, 13q13.3, 13q13.3-q14.11, 13q21.1, 18q12.1*

Common alterations, such as copy number increase of 3q, 8q, 9q, 11q, 14q, and 20q as well as copy number decrease of 3p, 4q, 8p, 9p, 13q, and 18q, have been reported. The smaller, specific regions show more variability, even though alterations such as 5p15, 8p23.2, 11q13, and 11q22 are reported by multiple studies. Additionally, Smeets and coworkers (2005) performed analyses of HNSCC with and without HPV16 involvement.

HPV-negative tumors presented losses at 3p11.2-p26.3, 5q11.2-q35.2, and 9p21.1-p24 and gains/amplifications at 11q12.1-q13.4 which were missing in HPV-positive tumors. Furthermore, HPV-negative tumors presented loss at 18q12.1-q23 and HPV-positive tumors presented gain at the same region. Regions altered at high frequency (>33%) in both groups are reported in Table 4. Although, we detected these HPV-associated alterations in our material, HPV status was unavailable, and thus, the relationship between alterations and HPV could not be assessed.

An additional layer of complexity of interpretation of CGH data is added by CNVs, which could account for a large amount of normal phenotypic variation. CNVs can directly cause or predispose to disease, or they might also function as susceptibility alleles. Some large variants might seem benign, but in combination with other genetic and environmental factors, they might influence on a disease phenotype (reviewed by Feuk *et al*., 2006; Freeman *et al*., 2006). Here, we have not systematically taken into account CNV regions when analyzing CGH data. CNVs can affect gene dosage at the transcriptional level, either directly or indirectly through position effects (reviewed by Feuk *et al*., 2006; Freeman *et al*., 2006).

8.2 Integration of copy number and gene expression data in HNSCC

Copy number alterations represent an important mechanism for cancer cells to promote or suppress the expression of target genes. Since HNSCC contains genome-wide expression (Section 6) as well genomic alterations with a poorly known connection, it provides a good model to study the impact of copy number on gene expression. To illustrate the impact, we performed gene expression profiling on oligonucleotide microarrays (~200 kb resolution) of all 38 copy number profiled LSCC and OTSCC samples (II, III). Because we had different array platforms for genome-wide copy number (cDNA) and gene expression (oligonucleotide) analysis, we applied and developed bioinformatic procedures to analyze and integrate the data (Hyman *et al*., 2002; Autio *et al*., 2003; Hautaniemi *et al*., 2004). We produced a novel interpolation option in the CGH-Plotter, which is a MATLAB toolbox for identifying and visualizing altered regions in aCGH data (Autio *et al*., 2003). In the interpolation process, if no matching cDNA clone for a gene or transcript present on the oligonucleotide microarray was located, we used location-based linear interpolation algorithm to interpolate CGH values from the cDNA data with a 750 kb

window. Here, we assumed that aCGH values from a certain genomic region are linearly dependent on the adjacent values. The interpolation option enables integration of data from different microarray platforms lacking an extensive probe sequence overlap. We were able to map over 11 000 genes between copy number and gene expression microarrays. We also improved the visualization and data smoothing options of the CGH-Plotter. Furthermore, the direct integration of processed aCGH data to the Expression annotation of Copy Number (ECN) tool was made possible. The MATLAB-based ECN tool allows visualization of expression data with processed aCGH data. The developed tools are freely available at www.cs.tut.fi/~bsmg/download.html.

8.2.1 Impact of copy number on gene expression

Using integrated microarray data, we first investigated the global impact of copy number on gene expression in HNSCC. Especially amplifications affected gene expression. Highly amplified regions (>2.5) had on average 5.2-fold more overexpressed genes in the cell lines and 2.5-fold in the primary tumors as compared with normal copy number regions (0.7-1.3). This is in line with studies of different cancers reporting that amplified regions can have many overexpressed genes (*e.g.*, Hyman *et al*., 2002; Wolf *et al*., 2004; Heidenblad *et al*., 2005). The association between decreased copy number and underexpression was less clear. It was, however, detectable in the cell lines with, on average, 3.0-fold more genes being underexpressed in the regions with decreased copy number (<0.4) (Figure 4). The actual percentages varied between samples. The weak association in primary tumors compared with cell lines could be due to other cell populations, such as normal cells, present in tissue samples. To study a homogenous cell population, one option is to apply laser capture microdissection, which allows the selection of a certain cell population from the primary tumor (Luo *et al*., 1999; reviewed by Espina *et al*., 2006). Overall, our results illustrated that in more than half of the cases copy number change was not associated with the respective change in gene expression, emphasizing the presence of multiple important gene expression regulation mechanisms in the cell.

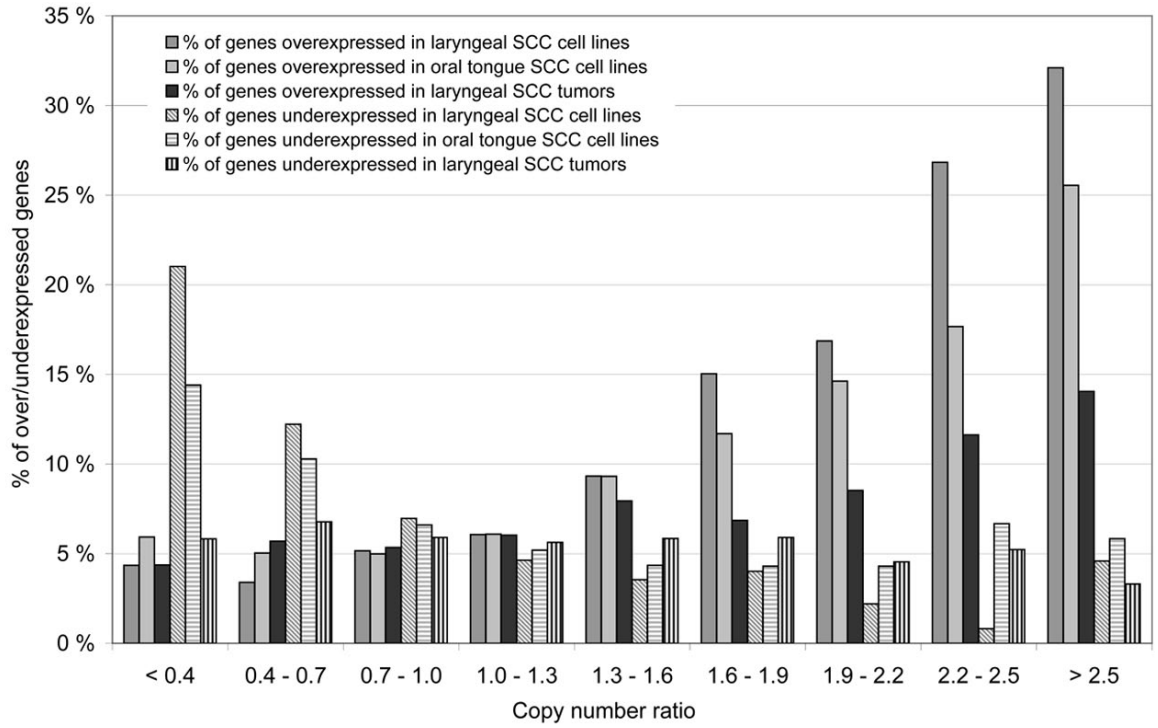

Figure 4 Global influence of copy number on gene expression in HNSCC samples. Copy number class from deletion to amplification is plotted from left to right on the X-axis. Percentage of over- and underexpressed genes in each copy number class (Y-axis) is illustrated separately for the sample groups. Threshold values for over- and underexpression for OTSCC cell lines were 1.51 and 0.69 (global upper and lower 7th percentile of median centered expression ratios including interpolated values), for LSCC cell lines 1.50 and 0.69, and for LSCC tumors 1.95 and 0.56.

In our study, the size and structure of the amplicons varied; some included multiple overexpressed genes, whereas others had only a few candidates. In LSCC, altered regions with overexpressed genes included 7p11-p15 (*EGFR*), 8p11-p12 (*BRF2*, *ASH2L*, *WHSC1L1*, *sFRP1*, *GCP16*, *FNTA*), 9p24 (*CDC37L1*, *RCL1*, *JAK2*, *RLN1*, *AK3*, *SLC1A1*), 11q13 (*FADD*, *PPFIA1*), 11q22 (*BIRC2*, *BIRC3*), 12q15 (*HMGA2*, *MDM1*, *MDM2*, *DYRK2*), and 17q23 (*PPM1D*, *APPBP2*) (examples of target genes shown in parentheses). Alterations at 11q13 and 9p24 were present also in the OTSCC cell lines, implying importance of these regions in HNSCC pathogenesis. In OTSCC, the identified alterations and candidate target genes included 6q12-q14 (*CD109*, *MTO1*), 11p11.2-p13 (*TRAF6*, *COMMD9*, *TRIM44*, *FJX1*, *PDHX*, *APIP*), 14q24.3 (*ABCD4*, *HBLD1*, *LTBP2*, *ZNF410*, *COQ6*, *JDP2*, *EIF2B2*, *ACYP*), and 22q11.21 (*SNAP29*). An association between chromosomal alterations at 11q13 (Bockmuhl *et al*., 2000; Wreesmann *et al*., 2004) or 14q24 (Bockmuhl *et al*., 2000) and poor prognosis of HNSCC patients has been reported, but specific target genes have remained obscure.

In OTSCC, we observed that 9-64% of genes located at the highly amplified regions (ratio >2.5) containing overexpressed genes (6q12-q14, 9p13.3, 9p22.3-p24.3, 10q11.21, 11p11.2-p13, 11q12.1-q12.2, 11q13.2-q13.4, 14q24.3, 22q11.21) had an associated increase in gene expression. Therefore, identification of the driver gene, even in highly amplified region, can be challenging. Instead of a single driver gene, a number of genes could be providing a growth advantage for cancer cells, as speculated in connection with the 11q13 amplicon in oral SCC (Huang *et al*., 2006). In HNSCC aCGH studies (Table 4), the expression of only a few genes in the altered region was determined by RT-PCR in selected samples (Baldwin *et al*., 2005; Snijders *et al*., 2005). RT-PCR analysis for selected genes can be an effective approach if a large and well-defined sample material in copy number analysis narrows the altered region clearly, and thus allowing identification of a manageable number of candidate genes. Integration of high-resolution genome-wide DNA- and RNA-level data can help to prioritize candidate genes also in a smaller material. However, especially in very complex and large gains, many bystander genes could have an altered expression, and distinguishing relevant genes can be difficult despite the methods used. For example, when previously reported chromosomal CGH data were correlated with Affymetrix data in HNSCC, Masayesva and coworkers (2004) described concordance between direction of copy number alteration and average expression values on chromosomal arms 1q, 2p, 3q, 5p, 7p, 8q, 9q, 20p, and 20q. Chromosomal aberrations resulted in alterations of both significantly and nonsignificantly expressed genes over a large chromosomal region. Therefore, alterations in expression might occur in bystander genes that are located close to target genetic or chromosomal alteration. Alternatively, in a larger region of chromosomal loss, a selective growth advantage could be achieved by summing expression alterations in multiple genes that individually have little or no effect (Masayesva *et al*., 2004).

8.2.2 Statistical analysis of target genes in altered regions

We further performed a statistical analysis (Hyman *et al*., 2002; Hautaniemi *et al*., 2004) to systematically identify genes with an association between copy number and gene expression. Across the genome, the overexpression of 739 genes could be attributed to copy number increase in 10 LSCC cell lines, with 325 genes showing the same association in 10 LSCC primary tumors (II). The majority of the genes, 62% in cell lines and 77% in tumors, were identified only in one sample. The cross-section of identified target gene

groups consisted of 40 genes including *FADD* and *PPFIA1* at the 11q13 amplicon. 11q13 amplicon has been studied in HNSCC extensively with *CCND1* and *CTTN* as potential target genes (reviewed by Schuuring, 1995). Using paraffin-embedded tissue sections from eight samples in immunohistochemistry, we further detected protein-level changes of FADD and CCND1 in the same primary tumor sample. Thus, we were able to demonstrate not only the association between DNA and RNA levels but also with protein level. The fact that the 11q13 amplicon spanned over several Mbs with overexpressed genes suggests that multiple target genes in the region could have a role in cancer pathogenesis. To clarify the function of each gene, they need to be tested by protein and functional level analyses. After the publication of our results, a study verified *FADD* as one of the potential driver genes of the 11q13 amplicon in laryngeal and pharyngeal SCCs (Gibcus *et al*., 2007). The authors also hypothesized that cells expressing high levels of the phosphorylated isoform of FADD (Ser194) could be more sensitive to Taxol-induced cell cycle arrest. This suggests that patients with FADD overexpression might benefit from therapy.

The number of common genes with a statistically significant association between copy number and expression in LSCC tumors and cell lines was small (40). However, GO analysis of all identified genes revealed that many top GO terms were shared by tumors and cell lines, such as ones involved in ion binding and transcription factor activities as well as signal transduction and proteolysis and peptidolysis. Therefore, the important processes involved in LSCC might be identified, although different pathways and genes could be activated. We also applied statistical analysis to identify genes with an association between copy number decrease and underexpression in LSCC. We identified 502 genes in the cell lines, including also a known TSG *CDKN2A*, and 223 genes in the primary tumors, with a cross-section of groups containing 22 genes.

Since LSCC and OTSCC cell lines presented for the most part a common spectrum of genomic alterations with similar effects on gene expression, we also performed statistical analysis (Hyman *et al*., 2002; Hautaniemi *et al*., 2004) for the LSCC and OTSCC cell lines together (III). This analysis yielded 1192 genes, 32 of which had a ratio higher than two between expression levels in groups with and without copy number increase in at least 15% of the cell lines (Table 5). These 32 genes included such candidates as *TRIO* (5p15), *FANCG* and *STOML2* (9p13), *CD44* (11p13), *FADD* (11q13), and *BIRC3* (11q22) (Table 5, Figure 5), which have earlier been implicated in HNSCC biology.

Table 5 Statistical analysis of HNSCC cell lines yielded 1192 genes with an association between copy number increase and gene expression. Here, 32 genes were selected based on the expression levels and detection frequency. Genes presented in bold are illustrated in Figure 5.

Ensembl ID (ENSG)	Gene symbol	Gene name	Chr
00000135919	*SERPINE2*	serpin peptidase inhibitor, clade E, member 2	2q36.1
00000163710	*PCOLCE2*	procollagen C-endopeptidase enhancer 2	3q23
00000169908	*TM4SF1*	transmembrane 4 L six family member 1	3q25.1
00000114248	*LRRC31*	leucine rich repeat containing 31	3q26.2
00000175166	*PSMD2*	proteasome 26S subunit, non-ATPase, 2	3q27.1
00000038382	*TRIO*	triple functional domain (PTPRF interacting)	5p15.2
00000166508	*MCM7*	minichromosome maintenance complex component 7	7q22.1
00000104368	*PLAT*	plasminogen activator, tissue	8p11.21
00000067167	*TRAM1*	translocation associated membrane protein 1	8q13.3
00000164695	*CHMP4C*	chromatin modifying protein 4C	8q21.13
00000086065	*CHMP5*	chromatin modifying protein 5	9p21.1
00000165281	***FANCG***	**Fanconi anemia, complementation group G**	**9p13.3**
00000165283	***STOML2***	**stomatin (EPB72)-like 2**	**9p13.3**
00000099139	*PCSK5*	proprotein convertase subtilisin/kexin type 5	9q21.2
00000135069	*PSAT1*	phosphoserine aminotransferase 1	9q21.31
00000169583	*CLIC3*	chloride intracellular channel 3	9q34.3
00000026508	***CD44***	**CD44 molecule**	**11p13**
00000168040	***FADD***	**Fas (TNFRSF6)-associated via death domain**	**11q13.3**
00000172893	*DHCR7*	7-dehydrocholesterol reductase	11q13.4
00000172890	*NADSYN1*	NAD synthetase 1	11q13.4
00000175581	*MRPL48*	mitochondrial ribosomal protein L48	11q13.4
00000175575	*PAAF1*	proteasomal ATPase-associated factor 1	11q13.4
00000023445	*BIRC3*	baculoviral IAP repeat-containing 3	11q22.2
00000111700	*SLCO1B3*	solute carrier organic anion transporter family, member 1B3	12p12.2
00000178695	*KCTD12*	potassium channel tetramerisation domain containing 12	13q22.3
00000072110	*ACTN1*	actinin, alpha 1	14q24.1
00000182351	*CRIP1*	cysteine-rich protein 1 (intestinal)	14q32.33
00000184254	*ALDH1A3*	aldehyde dehydrogenase 1 family, member A3	15q26.3
00000128482	*ZNF179*	zinc finger protein 179	17p11.2
00000125846	*ZNF133*	zinc finger protein 133	20p11.23
00000178726	*THBD*	thrombomodulin	20p11.21
00000101443	*WFDC2*	WAP four-disulfide core domain 2	20q13.12

At 9p13, DNA repair gene *FANCG* has been described as amplified in HNSCC (Sparano *et al*., 2006), and the overexpression of *STOML2* has been identified in LSCC with a correlation with clinical stage (Cao *et al*., 2007). 11p13 amplification, including *CD44* and overexpressed *FJX1*, was detected in oral SCC by genome-wide CGH analysis on BAC arrays (Snijders *et al*., 2005). In aCGH study by Baldwin and coworkers (2005), RT-PCR using a subset of samples was performed for *TRIO* (5p15.2) and *CDK6* (7q21.2), confirming the overexpression of these genes in association with segmental amplification in oral cancer. In concordance, we detected a frequent increase in copy number at 5p15, including overexpressed *TRIO*. Baldwin and coworkers (2005) also reported amplification of 11q22 region (*MMP1*, *3*, *7*, *8*, *10*, *12*, *13*, *20*, and *27*; *BIRC2* and *3*) with no integration with expression data. Thus, no target genes could be pinpointed. Since our data allowed a direct integration of genome-wide aCGH and expression data, we could distinguish the potential target genes, including *BIRC3* at the 11q22 amplicon, which are deregulated through copy number alteration and are not just bystanders in the region.

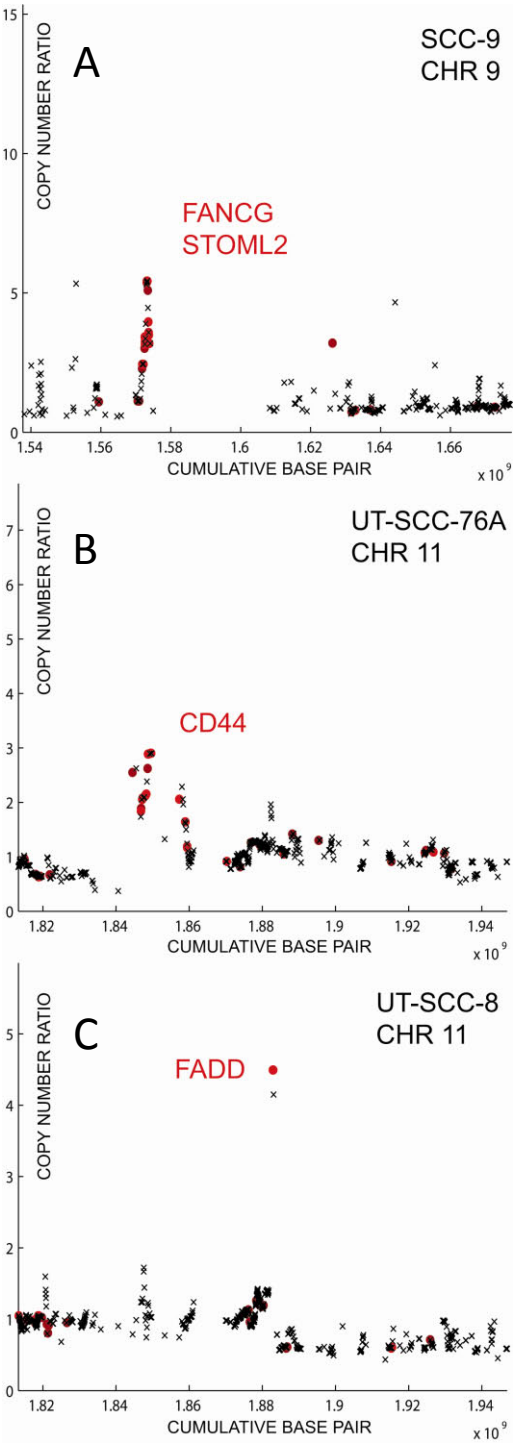

Figure 5 Altered regions with overexpressed genes (presented in Table 5) visualized by the ECN-tool. A) 9p13 region in OTSCC cell line SCC-9, B) 11p13 region in OTSCC cell line UT-SCC-76A, and C) 11q13 region in LSCC cell line UT-SCC-8.

Our results also demonstrated frequent, statistically significant overexpression of *FADD* in association with 11q13 amplification in OTSCC cell lines. This indicates that OTSCC and LSCC patients possessing the 11q13 amplicon with overexpressed *FADD* might benefit from similar treatment options. Of the identified 1192 genes, 316 mapped to biological pathways. Most genes did not map to known pathways, implicating that their exact function is still unclear.

Overall, the results showed that we were able to detect genes with previously reported relevance in HNSCC. This implies that also the genes identified in this study with no previous association with HNSCC could be important in disease pathogenesis. Taken together, the integrated microarray data revealed known as well as novel target genes in altered regions. The data set created in this work will provide a basis for protein and functional studies that could eventually lead to clinically relevant findings.

8.3 Targeted therapy in cancer

During the last few years, targeted therapy options, such as monoclonal antibodies and small molecule drugs, have been introduced for cancer. Examples of these therapies include Gleevec (BCR-ABL), Tarceva (EGFR), Erbitux (EGFR), Herceptin (ERBB2), Rituxan (CD20), and Avastin (VEFG-A). Furthermore, the drug Tykerb (lapatinib), which is a dual tyrosine kinase inhibitor (EGFR, ERBB2), has recently been approved as a breast cancer treatment. As a gene dosage effect is a common mechanism for gene regulation, combined copy number and gene expression data can give clues about potential drug targets. Currently, a number of drugs on the market or being developed target proteins encoded by genes that are often amplified in solid tumors, such as ERBB2. In HNSCC, EGFR small molecule tyrosine kinase inhibitors as well as monoclonal antibodies, either alone or in combination with other treatments, have been applied as targeted therapy option (reviewed by Forastiere and Burtness, 2007). Thus, in addition to targeted therapy alone, the synergistic combination of targeted therapy and conventional cancer treatments could be applied.

Unfortunately, in many cases, patient can develop resistance to the treatment, necessitating a second-line therapy. Thus, therapy affecting cells that provide support to the cancer cells might be beneficial since these genetically normal cells are not as prone to developing drug resistance as cancer cells (reviewed by Sawyers, 2004). Recently, a specific cell

population, cancer stem cells, has been proposed as a true target for genetic alterations and clonal selection, thus presenting also an attractive option for targeted therapy (reviewed by Reya *et al*., 2001; Tan *et al*., 2006).

Completion of genome sequencing project (International Human Genome Sequencing Consortium, 2004) has already led to the development of novel genome-wide technologies. Therefore, in the future, new examples of targeted therapies for specific genetic alterations are likely to appear for HNSCC as well as for other cancer types. Patient stratification through accurate molecular diagnostics will be crucial in the development and success of new targeted drugs (reviewed by Sawyers, 2004).

Conclusions and Future prospects

A new generation of microarrays providing a more detailed, higher-resolution view of the genome is launched almost every year. The constant developments have improved microarray technology. The reasons for the initial problems with reproducibility and comparability (reviewed by Marshall, 2004) within and between microarray platforms are understood better today than at the beginning of the decade, as demonstrated by the emerging clinical applications. In Study I, we showed a reasonably good concordance between the results from different gene expression microarray platforms despite challenges in comparing the data. Our results illustrated points that can create variability between platforms, such as annotation differences, clone errors, and differences in data preprocessing and analysis methods. Comparability of different gene expression data sets has become increasingly important with the growing amount of data available in public databases. Tools to carry out meta-analysis with gene expression microarray data are now available (www.genelogic.com; www.oncomine.org, Rhodes *et al*., 2007). As standards to describe other array-based data develop (*e.g.*, Scheinin *et al*., 2008), the collected high-throughput data will have a more central role in biological discoveries.

The research community has invested heavily in molecular profiling of the cancer genome, *e.g.* The Cancer Genome Atlas -project (TCGA; cancergenome.nih.gov). The TCGA project aims to improve our understanding of the molecular basis of cancer by genome analysis technologies, such as array-based gene expression, copy number, SNP, and miRNA profiling. Additionally, the project is focused on application of large-scale genome sequencing technologies, which have rapidly developed during the last few years (Wold and Myers, 2008). In the future, high-throughput sequencing may partly replace some applications of microarrays.

In Studies II and III, we integrated microarray-based copy number and gene expression data from 38 HNSCC cell lines and tumor samples. The integrated microarray data revealed the impact of copy number changes on gene expression in HNSCC, especially with regard to high-level amplifications. We identified a number of amplified regions, including 11q13 which contains overexpressed *FADD*. Since cancer cell lines have been reported to reflect the genomic and resulting transcriptional abnormalities of primary tumors (Neve *et al*., 2006), cell lines are an important model system to identify genes of pathogenetic relevance. Functional experiments with identified genes by RNAi will

provide further information of the effect of these genes on cancer phenotype. In addition, integration of data from other genome-wide methods, such as protein or lysate arrays, could reveal a more comprehensive molecular profile of HNSCC.

Acknowledgments

This work was carried out at the Institute of Biomedicine and Biomedicum Biochip Center, University of Helsinki, and the Department of Otorhinolaryngology - Head and Neck Surgery, Helsinki University Central Hospital, during 2002-2007. I thank the former and present heads of all institutes for providing excellent research facilities. I gratefully acknowledge the financial support received from Helsinki Biomedical Graduate School, University of Helsinki, Korvatautien tutkimussäätiö, the Ida Montin Foundation, the Helsinki University Central Hospital Research Fund, Finnish Concordia Fund, and the Finnish Cancer Organisations.

I have been privileged to work and meet various people from Helsinki, Tampere, and Turku during the years. Especially, I express my gratitude to:

My supervisors Outi Monni and Antti Mäkitie for providing opportunity to work with interesting projects and to present the results in various conferences.

Members of my thesis committee, Jaakko Hollmén and Merja Perälä, for their guidance and time. The yearly committee meeting was always a pleasant and encouraging event. I also thank Merja and Ritva Karhu for reviewing my thesis efficiently and for providing the relevant comments and criticism.

All collaborators and coauthors, whose contributions have been crucial for the publications. Especially, Reija Autio who made the last projects much smoother. You are a great colleague as well as a friend. Matti Saarela for filling Reija's place very efficiently. Sampsa Hautaniemi for providing an excellent data analysis as well as critical down-to-earth attitude during the years. Thank you, Sampsa and Lumi, for keeping me company on trips to Boston.

Reidar Grénman and Marita Potila for providing the essential UT-SCC cell lines along with valuable advice. Ilmo Leivo for helping with HNSCC tumor material and immunohistochemistry.

Tuula Airaksinen for guiding me in the laboratory almost from beginning to end. Päivi Tainola for particularly appreciated help first with Affymetrix and then with real-time RT-PCR. Janna Saarela and Petri Auvinen for their valuable contribution to the first publication.

Olli Kallioniemi for his advice and expertise, especially at the beginning of the project. Maija Wolf for her ever patient guidance and support. Saija Haapa-Paananen for sharing the enthusiasm of hundreds of interesting hits yielded by different alpha list analyses. Sami Kilpinen for helping with various data analysis issues. Henrik Edgren for being a realiable coauthor and coordinator along with being a true friend.

Siina Junnila and Henna Nieminen for bringing more life and action to our lab. All the people who visited and worked in our group during these years. Special thanks go to all the members of the Department of Obstetrics and Gynecology and the Clinical Chemistry for being our friendly neighbourghs for all these years. Especially, my very first roommates from A417a for creating an enjoyable work atmosphere and Outi K for facing many things together along the PhD way.

Kiitän kaikkia ystäviäni vuosien varrelta, joiden seura on tuonut kaivattua vaihtelua ja vienyt ajatukset pois työasioista. Yhteiset illanistujaiset, lomat, juhlat ja harrastukset ovat tehneet rentouttavan tehtävänsä. Olen kiitollinen perheelleni, jolla on aina riittänyt ymmärrystä, uskoa ja hyviä neuvoja. Toiset kannustavat, toiset näyttävät omalla esimerkillään, että asiat voi tehdä toisinkin ja että aina ei tarvitse olla niin kuin muut ollakseen onnellinen.

Espoo, April 2008

Anna-Kaarina Järvinen

References

Abdullah-Sayani A, Bueno-de-Mesquita JM, van de Vijver MJ. Technology Insight: tuning into the genetic orchestra using microarrays-limitations of DNA microarrays in clinical practice. Nat Clin Pract Oncol 3:501-516, 2006.

Affymetrix. GeneChip expression analysis technical manual. Technical Report 701021 Rev. 5:2004.

Affymetrix. Affymetrix GeneChip operating software. User's Guide Version 1.3. Technical Report 701439 Rev. 4:2005.

Agilent Technologies. Two-color microarray-based gene expression analysis v.5.5. Technical Report G4140-90050:2007a.

Agilent Technologies. One-color microarray-based gene expression analysis v.5.5. Technical Report G4140-90040:2007b.

Akervall J. Genomic screening of head and neck cancer and its implications for therapy planning. Eur Arch Otorhinolaryngol 263:297-304, 2006.

Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nat Genet 34:369-376, 2003.

Albertson DG. Gene amplification in cancer. Trends Genet 22:447-455, 2006.

Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet 7:55-65, 2006.

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

Autio R, Hautaniemi S, Kauraniemi P, Yli-Harja O, Astola J, Wolf M, Kallioniemi A. CGH-Plotter: MATLAB toolbox for CGH-data analysis. Bioinformatics 19:1714-1715, 2003.

Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, Foy C, Fuscoe J, Gao X, Gerhold DL, Gilles P, Goodsaid F, Guo X, Hackett J, Hockett RD, Ikonomi P *et al*. The External RNA Controls Consortium: a progress report. Nat Methods 2:731-734, 2005.

Baldwin C, Garnis C, Zhang L, Rosin MP, Lam WL. Multiple microalterations detected at high frequency in oral cancer. Cancer Res 65:7561-7567, 2005.

Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet 33 Suppl:238-244, 2003.

Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P *et al.* Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods 2:351-356, 2005.

Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP, Erle DJ. Spotted long oligonucleotide arrays for human gene expression analysis. Genome Res 13:1775-1785, 2003.

Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS, Yakhini Z, Bruhn L, Laderman S. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. Proc Natl Acad Sci U S A 101:17765-17770, 2004.

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R. NCBI GEO: mining tens of millions of expression profiles-database and tools update. Nucl Acids Res 35:D760-765, 2007.

Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell 116:281-297, 2004.

Baudis M, Cleary ML. Progenetix.net: an online repository for molecular cytogenetic aberration data. Bioinformatics 17:1228-1229, 2001.

Bayani J, Selvarajah S, Maire G, Vukovic B, Al-Romaih K, Zielenska M, Squire JA. Genomic mechanisms and measurement of structural and numerical instability in cancer cells. Semin Cancer Biol 17:5-18, 2007.

Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 6:107-116, 2006.

Beckman RA, Loeb LA. Genetic instability in cancer: Theory and experiment. Semin Cancer Biol 15:423-435, 2005.

Belbin TJ, Singh B, Barber I, Socci N, Wenig B, Smith R, Prystowsky MB, Childs G. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. Cancer Res 62:1184-1190, 2002.

Bignell GR, Huang J, Greshock J, Watt S, Butler A, West S, Grigorova M, Jones KW, Wei W, Stratton MR, Futreal PA, Weber B, Shapero MH, Wooster R. High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Res 14:287-295, 2004.

Blanchard AP, Kaiser RJ, Hood LE. High-density oligonucleotide arrays. Biosens Bioelectron 11:687-690, 1996.

Bockmuhl U, Schluns K, Kuchler I, Petersen S, Petersen I. Genetic imbalances with impact on survival in head and neck cancer patients. Am J Pathol 157:369-375, 2000.

Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. Cancer Res 63:1727-1730, 2003.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365-371, 2001.

Brennan C, Zhang Y, Leo C, Feng B, Cauwels C, Aguirre AJ, Kim M, Protopopov A, Chin L. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. Cancer Res 64:4744-4748, 2004.

Brockstein B, Vokes EE. Concurrent chemoradiotherapy for head and neck cancer. Semin Oncol 31:786-793, 2004.

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 56:2488-2492, 1996.

Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 6:857-866, 2006.

Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, Ma Y, Maqsodi B, Papallo A, Peters EH, Poulter K, Ruppel PL, Samaha RR, Shi L, Yang W, Zhang L *et al.* Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotech 24:1115-1122, 2006.

Cao WF, Zhang LY, Liu MB, Tang PZ, Liu ZH, Sun BC. Prognostic significance of stomatin-like protein 2 overexpression in laryngeal squamous cell carcinoma: clinical, histologic, and immunohistochemistry analyses with tissue microarray. Hum Pathol 38:747-752, 2007.

Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B. High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. J Clin Pathol 57:644-646, 2004.

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10:529-541, 2006.

Choi P, Chen C. Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma. Cancer 104:1113-1128, 2005.

Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, Shockley WW, Weissler MC, Dressler LG, Shores CG, Yarbrough WG, Perou CM. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 5:489-500, 2004.

Churchill GA. Fundamentals of experimental design for cDNA microarrays. Nat Genet 32:490-495, 2002.

Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829-836, 1979.

Cromer A, Carles A, Millon R, Ganguli G, Chalmel F, Lemaire F, Young J, Dembele D, Thibault C, Muller D, Poch O, Abecassis J, Wasylyk B. Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. Oncogene 23:2484-2498, 2004.

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33:e175, 2005.

de Klein A, van Kessel AG, Grosveld G, Bartram CR, Hagemeijer A, Bootsma D, Spurr NK, Heisterkamp N, Groffen J, Stephenson JR. A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. Nature 300:765-767, 1982.

Diaz E, Sturgis E, Laramore G, Sabichi A, Lippman S, Clayman G. Neoplasms of the head and neck. Kufe D, Pollock R, Weichselbaum R, Bast Jr R, Gansler T, Holland J, Frei III E. Cancer Medicine 6. BC Decker Inc, Hamilton, Ontario, 2003, 1325-1371.

Dobbin K, Shih JH, Simon R. Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. J Natl Cancer Inst 95:1362-1369, 2003.

Draghici S, Khatri P, Eklund AC, Szallasi Z. Reliability and reproducibility issues in DNA microarray measurements. Trends Genet 22:101-109, 2006.

Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 99:147-157, 2007.

Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P. Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A 89:3010-3014, 1992.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-14868, 1998.

Espina V, Wulfkuhle JD, Calvert VS, VanMeter A, Zhou W, Coukos G, Geho DH, Petricoin EF, Liotta LA. Laser-capture microdissection. Nat Protocols 1:586-603, 2006.

Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 6:259-269, 2006.

Ewis AA, Zhelev Z, Bakalova R, Fukuoka S, Shinohara Y, Ishikawa M, Baba Y. A history of microarrays in biomedicine. Expert Rev Mol Diagn 5:315-328, 2005.

Fan JB, Chee MS, Gunderson KL. Highly parallel genomic assays. Nat Rev Genet 7:632-644, 2006a.

Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DSA, Nobel AB, van't Veer LJ, Perou CM. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355:560-569, 2006b.

Fearon ER. Human cancer syndromes: clues to the origin and nature of cancer. Science 278:1043-1050, 1997.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 7:85-97, 2006.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391:806-811, 1998.

Fodde R, Smits R. Cancer biology. A matter of dosage. Science 298:761-763, 2002.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 251:767-773, 1991.

Forastiere A, Koch W, Trotti A, Sidransky D. Head and neck cancer. N Engl J Med 345:1890-1900, 2001.

Forastiere AA, Burtness BA. Epidermal growth factor receptor inhibition in head and neck cancer--more insights, but more questions. J Clin Oncol 25:2152-2155, 2007.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. Copy number variation: new insights in genome diversity. Genome Res 16:949-961, 2006.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nat Rev Cancer 4:177-183, 2004.
Ge H, Walhout AJ, Vidal M. Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet 19:551-560, 2003.
Gerhold DL, Jensen RV, Gullans SR. Better therapeutics through microarrays. Nat Genet 32 Suppl:547-551, 2002.
Gershon D. Microarrays go mainstream. Nat Meth 1:263-270, 2004.
Gibcus JH, Menkema L, Mastik MF, Hermsen MA, de Bock GH, van Velthuysen ML, Takes RP, Kok K, Alvarez Marcos CA, van der Laan BF, van den Brekel MW, Langendijk JA, Kluin PM, van der Wal JE, Schuuring E. Amplicon mapping and expression profiling identify the fas-associated death domain gene as a new driver in the 11q13.3 amplicon in laryngeal/pharyngeal cancer. Clin Cancer Res 13:6257-6266, 2007.
Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, Pambuccian SE, Ondrey FG, Adams GL, Gaffney PM. Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. Cancer Res 64:55-63, 2004.
Glas A, Floore A, Delahaye L, Witteveen A, Pover R, Bakx N, Lahti-Domenici J, Bruinsma T, Warmoes M, Bernards R, Wessels L, Van 't Veer, L. Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 7:278, 2006.
Gollin SM. Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease. Head Neck 23:238-253, 2001.
Gollin SM. Mechanisms leading to chromosomal instability. Semin Cancer Biol 15:33-42, 2005.
Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J *et al.* Patterns of somatic mutation in human cancer genomes. Nature 446:153-158, 2007.
Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucl Acids Res 34:D140-144, 2006.
Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, Hurban P, Phillips KL, Xu J, Deng X, Sun YA, Tong W, Dragan YP, Shi L. Rat toxicogenomic study reveals analytical consistency across microarray platforms. Nat Biotech 24:1162-1169, 2006.
Guo X, Lui WO, Qian CN, Chen JD, Gray SG, Rhodes D, Haab B, Stanbridge E, Wang H, Hong MH, Min HQ, Larsson C, Teh BT. Identifying cancer-related genes in nasopharyngeal carcinoma cell lines using DNA and mRNA expression profiling analyses. Int J Oncol 21:1197-1204, 2002.
Ha PK, Califano JA. The molecular biology of mucosal field cancerization of the head and neck. Crit Rev Oral Biol Med 14:363-369, 2003.
Hahn WC, Weinberg RA. Rules for making human tumor cells. N Engl J Med 347:1593-1603, 2002.
Halgren RG, Fielden MR, Fong CJ, Zacharewski TR. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. Nucl Acids Res 29:582-588, 2001.
Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 100:57-70, 2000.
Hannon GJ. RNA interference. Nature 418:244-251, 2002.
Hannon GJ, Rossi JJ. Unlocking the potential of the human genome with RNA interference. Nature 431:371-378, 2004.
Hardiman G. Microarray platforms-comparisons and contrasts. Pharmacogenomics 5:487-502, 2004.
Hariharan R. The analysis of microarray data. Pharmacogenomics 4:477-497, 2003.
Hautaniemi S, Ringnér M, Kauraniemi P, Autio R, Edgren H, Yli-Harja O, Astola J, Kallioniemi A, Kallioniemi O. A strategy for identifying putative causes of gene expression variation in human cancers. J Franklin Inst 341:77-88, 2004.
Heidenblad M, Lindgren D, Veltman JA, Jonson T, Mahlamäki EH, Gorunova L, van Kessel AG, Schoenmakers EF, Höglund M. Microarray analyses reveal strong influence of DNA copy

number alterations on the transcriptional patterns in pancreatic cancer: implications for the interpretation of genomic amplifications. Oncogene 24:1794-1801, 2005.
Hellman A, Zlotorynski E, Scherer SW, Cheung J, Vincent JB, Smith DI, Trakhtenbrot L, Kerem B. A role for common fragile site induction in amplification of human oncogenes. Cancer Cell 1:89-97, 2002.
Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 7:200-210, 2006.
Holloway AJ, van Laar, Ryan K, Tothill RW, Bowtell DDL. Options available - from start to finish - for obtaining data from DNA microarrays II. Nat Genet 37:481-489, 2002.
Huang X, Godfrey TE, Gooding WE, McCarty KS,Jr, Gollin SM. Comprehensive genome and transcriptome analysis of the 11q13 amplicon in human oral cancer and synteny to the 7F5 amplicon in murine oral carcinoma. Genes Chromosomes Cancer 45:1058-1069, 2006.
Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19:342-347, 2001.
Hunter KD, Parkinson EK, Harrison PR. Profiling early head and neck cancer. Nat Rev Cancer 5:127-135, 2005.
Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi O, Kallioniemi A. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 62:6240-6245, 2002.
Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet 36:949-951, 2004.
Imbeaud S, Auffray C. 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments. Drug Discov Today 10:1175-1182, 2005.
International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 431:931-945, 2004.
Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249-264, 2003.
Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J *et al.* Multiple-laboratory comparison of microarray platforms. Nat Methods 2:345-350, 2005.
Irizarry RA, Wu Z, Jaffee HA. Comparison of Affymetrix GeneChip expression measures. Bioinformatics 22:789-794, 2006.
Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 36:299-303, 2004.
Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. CA Cancer J Clin 57:43-66, 2007.
Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol 8:33-41, 2004.
Jones PA, Baylin SB. The epigenomics of cancer. Cell 128:683-692, 2007.
Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818-821, 1992.
Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucl Acids Res 28:4552-4557, 2000.
Kim SK, Fan Y, Papadimitrakopoulou V, Clayman G, Hittelman WN, Hong WK, Lotan R, Mao L. DPC4, a candidate tumor suppressor gene, is altered infrequently in head and neck squamous cell carcinoma. Cancer Res 56:2519-2521, 1996.
Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature 386:761-763, 1997.

Kinzler KW, Vogelstein B. Oncogenesis: Landscaping the cancer terrain. Science 280:1036-1037, 1998.
Knight J. When the chips are down. Nature 410:860-861, 2001.
Knudson AG,Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 68:820-823, 1971.
Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer 1:157-162, 2001.
Kohonen T. Self-organizing maps. Springer-Verlag, New York, NJ, 2001, 501.
Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 4:844-847, 1998.
Kothapalli R, Yoder S, Mane S, Loughran T. Microarray results: how accurate are they? BMC Bioinformatics 3:22, 2002.
Kuo WP, Jenssen T, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 18:405-412, 2002.
LaBaer J, Ramachandran N. Protein microarrays as tools for functional proteomics. Curr Opin Chem Biol 9:14-19, 2005.
LaFramboise T, Weir BA, Zhao X, Beroukhim R, Li C, Harrington D, Sellers WR, Meyerson M. Allele-specific amplification in cancer revealed by SNP array analysis. PLoS Comput Biol 1:e65, 2005.
Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 294:853-858, 2001.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K *et al.* Initial sequencing and analysis of the human genome. Nature 409:860-921, 2001.
Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. Independence and reproducibility across microarray platforms. Nat Methods 2:337-344, 2005.
Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843-854, 1993.
Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294:862-864, 2001.
Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 396:643-649, 1998.
Li C, Wong WH. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. Proc Natl Acad Sci U S A 98:31-36, 2001.
Li J, Pankratz M, Johnson JA. Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. Toxicol Sci 69:383-390, 2002.
Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967-971, 1992.
Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet 21:20-24, 1999.
Liu ET, Kuznetsov VA, Miller LD. In the pursuit of complexity: systems medicine in cancer biology. Cancer Cell 9:245-247, 2006.
Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675-1680, 1996.
Loeb LA. A mutator phenotype in cancer. Cancer Res 61:3230-3239, 2001.
Loging W, Harland L, Williams-Jones B. High-throughput electronic biology: mining information for drug discovery. Nat Rev Drug Discov 6:220-230, 2007.
Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KCQ, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. Genome Res 13:2291-2305, 2003.

Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG. Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat Med 5:117-122, 1999.

MacBeath G. Protein microarrays and proteomics. Nat Genet 32:526-532, 2002.

Mao L, Fan YH, Lotan R, Hong WK. Frequent abnormalities of FHIT, a candidate tumor suppressor gene, in head and neck cancer cell lines. Cancer Res 56:5128-5131, 1996.

Mao L, Hong WK, Papadimitrakopoulou VA. Focus on head and neck cancer. Cancer Cell 5:311-316, 2004.

MAQC Consortium: Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 24:1151-1161, 2006.

Marshall E. Getting the noise out of gene arrays. Science 306:630-631, 2004.

Masayesva BG, Ha P, Garrett-Mayer E, Pilkington T, Mao R, Pevsner J, Speed T, Benoit N, Moon C, Sidransky D, Westra WH, Califano J. Gene expression alterations over large chromosomal regions in cancers include multiple genes unrelated to malignant progression. Proc Natl Acad Sci U S A 101:8715-8720, 2004.

Michiels S, Koscielny S, Hill C. Interpretation of microarray data in cancer. Br J Cancer 96:1155-1158, 2007.

Michor F, Iwasa Y, Vogelstein B, Lengauer C, Nowak MA. Can chromosomal instability initiate tumorigenesis? Semin Cancer Biol 15:43-49, 2005.

Mitchell P. A perspective on protein microarrays. Nat Biotechnol 20:225-229, 2002.

Mootha VK, Lindgren CM, Eriksson K, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D *et al.* PGC-1[alpha]-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34:267-273, 2003.

Myllykangas S and Knuutila S. Manifestation, mechanisms and mysteries of gene amplifications. Cancer Lett 232:79-89, 2006.

Nakagawa T, Pimkhaokham A, Suzuki E, Omura K, Inazawa J, Imoto I. Genetic or epigenetic silencing of low density lipoprotein receptor-related protein 1B expression in oral squamous cell carcinoma. Cancer Sci 97:1070-1074, 2006.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10:515-527, 2006.

Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR *et al.* Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res 12:1749-1755, 2002.

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

Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A. ArrayExpress-a public database of microarray experiments and gene expression profiles. Nucl Acids Res 35:D747-750, 2007.

Patmore HS, Cawkwell L, Stafford ND, Greenman J. Unraveling the chromosomal aberrations of head and neck squamous cell carcinoma: a review. Ann Surg Oncol 12:831-842, 2005.

Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu T, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD. Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC) project. Nat Biotech 24:1140-1150, 2006.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 91:5022-5026, 1994.

Perez-Ordonez B, Beauchemin M, Jordan RC. Molecular biology of squamous cell carcinoma of the head and neck. J Clin Pathol 59:445-453, 2006.

Peto J. Cancer epidemiology in the last century and the next decade. Nature 411:390-395, 2001.

Petricoin EF, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, Woodcock J, Feigal DW, Zoon KC, Sistare FD. Medical applications of microarray technologies: a regulatory science perspective. Nat Genet 32:474-479, 2002.

Pierce AJ, Stark JM, Araujo FD, Moynahan ME, Berwick M, Jasin M. Double-strand breaks and tumorigenesis. Trends Cell Biol 11:S52-59, 2001.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20:207-211, 1998.

Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. Nat Genet 37 Suppl:S11-17, 2005.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23:41-46, 1999.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale A, Brown PO. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 99:12963-12968, 2002.

Ponder BAJ. Cancer genetics. Nature 411:336-341, 2001.

Quackenbush J. Microarray data normalization and transformation. Nat Genet 32:496-501, 2002.

Reed AL, Califano J, Cairns P, Westra WH, Jones RM, Koch W, Ahrendt S, Eby Y, Sewell D, Nawroz H, Bartek J, Sidransky D. High frequency of p16 (CDKN2/MTS-1/INK4A) inactivation in head and neck squamous cell carcinoma. Cancer Res 56:3630-3633, 1996.

Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 414:105-111, 2001.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 6:1-6, 2004.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 9:166-180, 2007.

Ringner M, Peterson C, Khan J. Analyzing array data using supervised methods. Pharmacogenomics 3:403-415, 2002.

Rodrigo JP, Ferlito A, Suarez C, Shaha AR, Silver CE, Devaney KO, Bradley PJ, Bocker JM, McLaren KM, Grenman R, Rinaldo A. New molecular diagnostic methods in head and neck cancer. Head Neck 27:995-1003, 2005.

Roepman P, Wessels LFA, Kettelarij N, Kemmeren P, Miles AJ, Lijnzaad P, Tilanus MGJ, Koole R, Hordijk G, van der Vliet, Peter C, Reinders MJT, Slootweg PJ, Holstege FCP. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. Nat Genet 37:182-186, 2005.

Sawyers C. Targeted cancer therapy. Nature 432:294-297, 2004.

Scheinin I, Myllykangas S, Borze I, Böhling T, Knuutila S, Saharinen J. CanGEM: mining gene copy number changes in cancer. Nucl Acids Res 36:D830-835, 2008.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470, 1995.

Schröck E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, Garini Y, Ried T. Multicolor spectral karyotyping of human chromosomes. Science 273:494-497, 1996.

Schuuring E. The involvement of the chromosome 11q13 region in human malignancies: Cyclin D1 and EMS1 are two new candidate oncogenes-a review. Gene 159:83-96, 1995.

Schwab M. Oncogene amplification in solid tumors. Semin Cancer Biol 9:319-325, 1999.

Searls DB. Data integration: challenges for drug discovery. Nat Rev Drug Discov 4:45-58, 2005.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A *et al.* Large-scale copy number polymorphism in the human genome. Science 305:525-528, 2004.

Shippy R, Fulmer-Smentek S, Jensen RV, Jones WD, Wolber PK, Johnson CD, Pine PS, Boysen C, Guo X, Chudin E, Sun YA, Willey JC, Thierry-Mieg J, Thierry-Mieg D, Setterquist RA, Wilson M, Lucas AB, Novoradovskaya N, Papallo A, Turpaz Y *et al.* Using RNA sample titrations to assess microarray platform performance and normalization techniques. Nat Biotech 24:1123-1131, 2006.

Singh B, Gogineni SK, Sacks PG, Shaha AR, Shah JP, Stoffel A, Rao PH. Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. Cancer Res 61:4506-4513, 2001.

Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JKV, Gazdar AF, Hartigan J *et al.* The consensus coding sequences of human breast and colorectal cancers. Science 314:268-274, 2006.

Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 6:963-968, 1953.

Smeets SJ, Braakhuis BJM, Abbas S, Snijders PJF, Ylstra B, van de Wiel, M A, Meijer GA, Leemans CR, Brakenhoff RH. Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. Oncogene 25:2558-2564, 2005.

Smyth GK, Yang YH, Speed T. Statistical issues in cDNA microarray data analysis. Methods Mol Biol 224:111-136, 2003a.

Smyth GK, Speed T. Normalization of cDNA microarray data. Methods, 31:265-273, 2003b.

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 29:263-264, 2001.

Snijders AM, Schmidt BL, Fridlyand J, Dekker N, Pinkel D, Jordan RC, Albertson DG. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. Oncogene 24:4232-4242, 2005.

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 20:399-407, 1997.

Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. Nat Genet 21:5-9, 1999.

Southern EM. DNA microarrays: History and overview. Rampal JB. Methods in Molecular Biology. DNA arrays: Methods and protocols. Humana Press, Totowa, NJ, 2001, 1-15.

Sparano A, Quesnelle KM, Kumar MS, Wang Y, Sylvester AJ, Feldman M, Sewell DA, Weinstein GS, Brose MS. Genome-wide profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization. Laryngoscope 116:735-741, 2006.

Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet 6:782-792, 2005.

Stransky N, Vallot C, Reyal F, Bernard-Pierrot I, de Medina SG, Segraves R, de Rycke Y, Elvin P, Cassidy A, Spraggon C, Graham A, Southgate J, Asselain B, Allory Y, Abbou CC, Albertson

DG, Thiery JP, Chopin DK, Pinkel D, Radvanyi F. Regional copy number-independent deregulation of transcription in cancer. Nat Genet 38:1386-1396, 2006.
Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545-15550, 2005.
Sun PC, Uppaluri R, Schmidt AP, Pashia ME, Quant EC, Sunwoo JB, Gollin SM, Scholnick SB. Transcript map of the 8p23 putative tumor suppressor region. Genomics 75:17-25, 2001.
Syrjänen S. Human papillomaviruses in head and neck carcinomas. N Engl J Med 356:1993-1995, 2007.
Syvänen AC. Toward genome-wide SNP genotyping. Nat Genet 37 Suppl:S5-10, 2005.
Tabor MP, Brakenhoff RH, van Houten VM, Kummer JA, Snel MH, Snijders PJ, Snow GB, Leemans CR, Braakhuis BJ. Persistence of genetically altered fields in head and neck cancer patients: biological and clinical implications. Clin Cancer Res 7:1523-1532, 2001.
Tan PK, Downey TJ, Spitznagel EL,Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. Evaluation of gene expression measurements from commercial microarray platforms. Nucl Acids Res 31:5676-5684, 2003.
Tan BT, Park CY, Ailles LE, Weissman IL. The cancer stem cell hypothesis: a work in progress. Lab Invest 86:1203-1207, 2006.
Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. Nat Genet 22:281-285, 1999.
The Gene Ontology Consortium. The Gene Ontology project in 2008. Nucl Acids Res 36:D440-444, 2008.
Tilstone C. DNA microarrays: Vital statistics. Nature 424:610-612, 2003.
Tinker AV, Boussioutas A, Bowtell DDL. The challenges of gene expression microarrays for the study of human cancer. Cancer Cell 9:333-339, 2006.
Tlsty TD, Hein PW. Know thy neighbor: stromal cells can contribute oncogenic signals. Curr Opin Genet Dev 11:54-59, 2001.
Toledo F, Le Roscouet D, Buttin G, Debatisse M. Co-amplified markers alternate in megabase long chromosomal inverted repeats and cluster independently in interphase nuclei at early steps of mammalian gene amplification. EMBO J 11:2665-2673, 1992.
Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310:644-648, 2005.
Tong W, Lucas AB, Shippy R, Fan X, Fang H, Hong H, Orr MS, Chu T, Guo X, Collins PJ, Sun YA, Wang S, Bao W, Wolfinger RD, Shchegrova S, Guo L, Warrington JA, Shi L. Evaluation of external RNA controls for the assessment of microarray performance. Nat Biotech 24:1132-1139, 2006.
Tonon G, Wong K, Maulik G, Brennan C, Feng B, Zhang Y, Khatry DB, Protopopov A, You MJ, Aguirre AJ, Martin ES, Yang Z, Ji H, Chin L, DePinho RA. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A 102:9625-9630, 2005.
Tumor Analysis Best Practices Working Group. Expression profiling-best practices for data generation and interpretation in clinical trials. Nat Rev Genet 5:229-237, 2004.
Uhlén M, Björling E, Agaton C, Szigyarto CA, Amini B, Andersen E, Andersson A, Angelidou P, Asplund A, Asplund C, Berglund L, Bergström K, Brumer H, Cerjan D, Ekström M, Elobeid A, Eriksson C, Fagerberg L, Falk R, Fall J *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics 4:1920-1932, 2005.
Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A 87:1663-1667, 1990.
van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R,

Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-536, 2002.
Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 270:484-487, 1995.
Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M *et al.* The sequence of the human genome. Science 291:1304-1351, 2001.
Virgilio L, Shuster M, Gollin SM, Veronese ML, Ohta M, Huebner K, Croce CM. FHIT gene alterations in head and neck squamous cell carcinomas. Proc Natl Acad Sci U S A 93:9770-9775, 1996.
Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 10:789-799, 2004.
Vokes EE, Weichselbaum RR, Lippman SM, Hong WK. Head and neck cancer. N Engl J Med 328:184-194, 1993.
Wang H, Malek R, Kwitek A, Greene A, Luu T, Behbahani B, Frank B, Quackenbush J, Lee N. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. Genome Biol 4:R5, 2003.
Weinberg RA. Biology of cancer. Garland Science, Taylor and Francis Group, New York, 2006, 864.
Weinstein JN. 'Omic' and hypothesis-driven research in the molecular pharmacology of cancer. Curr Opin Pharmacol 2:361-365, 2002.
Weinstein JN. Spotlight on molecular profiling: "Integromic" analysis of the NCI-60 cancer cell lines. Mol Cancer Ther 5:2601-2605, 2006.
Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J *et al.* Characterizing the cancer genome in lung adenocarcinoma. Nature 450:893-898, 2007.
Wittekind CH, Greene FL, Hutter RVP, Klimpfinger M, Sobin LH. TNM Atlas: Illustrated guide to the TNM classification of malignant tumours. Wiley, Hoboken, NJ, 2005, 448.
Wogan GN, Hecht SS, Felton JS, Conney AH, Loeb LA. Environmental and chemical carcinogenesis. Semin Cancer Biol 14:473-486, 2004.
Wold B, Myers RM. Sequence census methods for functional genomics. Nat Meth 5:19-21, 2008.
Wolf M, Mousses S, Hautaniemi S, Karhu R, Huusko P, Allinen M, Elkahloun A, Monni O, Chen Y, Kallioniemi A, Kallioniemi OP. High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression. Neoplasia 6:240-247, 2004.
Worsham MJ, Chen KM, Tiwari N, Pals G, Schouten JP, Sethi S, Benninger MS. Fine-mapping loss of gene architecture at the CDKN2B (p15INK4b), CDKN2A (p14ARF, p16INK4a), and MTAP genes in head and neck squamous cell carcinoma. Arch Otolaryngol Head Neck Surg 132:409-415, 2006.
Wreesmann VB, Shi W, Thaler HT, Poluri A, Kraus DH, Pfister D, Shaha AR, Shah JP, Rao PH, Singh B. Identification of novel prognosticators of outcome in squamous cell carcinoma of the head and neck. J Clin Oncol 22:3965-3972, 2004.
Wreesmann VB, Singh B. Chromosomal aberrations in squamous cell carcinomas of the upper aerodigestive tract: biologic insights and clinical opportunities. J Oral Pathol Med 34:449-459, 2005.
Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucl Acids Res 30:e15, 2002.
Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). Nucleic Acids Res 34:445-450, 2006.

Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. Nucl Acids Res 30:e48, 2002.

Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen T, Girard L, Minna J, Christiani D, Leo C, Gray JW, Sellers WR, Meyerson M. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res 64:3060-3071, 2004.

Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using proteome chips. Science 293:2101-2105, 2001.