

ONCOGENOMICS

Identification of target genes in laryngeal squamous cell carcinoma by high-resolution copy number and gene expression microarray analyses

A-K Järvinen^{1,2}, R Autio³, S Haapa-Paananen⁴, M Wolf⁴, M Saarela³, R Grénman^{5,6}, I Leivo⁷, O Kallioniemi⁴, AA Mäkitie² and O Monni¹

¹Biomedicum Biochip Center and Institute of Biomedicine, University of Helsinki, Helsinki, Finland; ²Department of Otorhinolaryngology – Head and Neck Surgery, Helsinki University Central Hospital, Helsinki, Finland; ³Institute of Signal Processing, Tampere University of Technology, Tampere, Finland; ⁴Medical Biotechnology, VTT Technical Research Centre of Finland and University of Turku, Turku, Finland; ⁵Department of Otorhinolaryngology – Head and Neck Surgery, Turku University Central Hospital, Turku, Finland; ⁶Department of Medical Biochemistry, Turku University, Turku, Finland and ⁷Department of Pathology, Haartman Institute and Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland

Molecular mechanisms contributing to initiation and progression of head and neck squamous cell carcinoma are still poorly known. Numerous genetic alterations have been described, but molecular consequences of such alterations in most cases remain unclear. Here, we performed an integrated high-resolution microarray analysis of gene copy number and expression in 20 laryngeal cancer cell lines and primary tumors. Our aim was to identify genetic alterations that play a key role in disease pathogenesis and pinpoint genes whose expression is directly impacted by these events. Integration of DNA level data from array-based comparative genomic hybridization with RNA level information from oligonucleotide microarrays was achieved with custom-developed bioinformatic methods. High-level amplifications had a clear impact on gene expression. Across the genome, overexpression of 739 genes could be attributed to gene amplification events in cell lines, with 325 genes showing the same phenomenon in primary tumors including *FADD* and *PPF1A1* at 11q13. The analysis of gene ontology and pathway distributions further pinpointed genes that may identify potential targets of therapeutic intervention. Our data highlight genes that may be critically important to laryngeal cancer progression and offer potential therapeutic targets.

Oncogene advance online publication, 22 May 2006; doi:10.1038/sj.onc.1209690

Keywords: head and neck cancer; cDNA microarrays; oligonucleotide microarrays; comparative genomic hybridization

Introduction

Squamous cell carcinoma of the upper aerodigestive tract is the most common type of malignant head and neck tumors. The major etiological factors for head and neck squamous cell carcinoma (HNSCC) are tobacco and alcohol consumption, and its incidence is substantially higher in men than in women. HNSCC can be located in various anatomical sites and forms therefore a heterogeneous group of tumors. Current methods to predict the outcome of HNSCC patients include clinicopathological parameters such as primary tumor, regional node, distant metastasis (TNM) – stage, depth of invasion, and differentiation grade. However, these parameters do not accurately reflect prognosis of the patients and additional predictors and biomarkers would be useful for patient management. An improved understanding of disease pathogenesis could open therapeutic possibilities.

Novel technologies offer a way to identify new markers for HNSCC management. High-resolution genomewide technologies have been crucial in identifying molecular genetic alterations, and already gene expression profiling by microarrays have distinguished novel biological and prognostic subgroups of HNSCC tumors (e.g. Belbin *et al.*, 2002; Chung *et al.*, 2004; Cromer *et al.*, 2004; Ginos *et al.*, 2004; Choi and Chen, 2005). Although these studies have revealed the molecular portraits of various tumor subgroups based on their gene expression signature, they are not optimal for identification of the key molecular drivers for the disease.

The activation of oncogenes, such as *CCND1* at 11q13 and *EGFR* at 7p12 (reviewed by Mao *et al.*, 2004), and loss of tumor-suppressor genes play an important role in the development of HNSCC. Multiple chromosomal regions, identified by chromosomal comparative genomic hybridization (CGH) as well as array-based CGH (aCGH), are reported to be amplified or deleted in HNSCC. Based on the review by Gollin (2001), the most common gains in HNSCC mapped to 3q, 5p, 7p, 8q, 9q, 11q13 and 20q, whereas the most frequent losses were at

Correspondence: Dr O Monni, Biomedicum Biochip Center and Institute of Biomedicine, PO Box 63, Rm A415b, Haartmaninkatu 8, 00014 University of Helsinki, Finland.

E-mail: outi.monni@helsinki.fi

Received 23 November 2005; revised 23 March 2006; accepted 18 April 2006

3p, 5q, 8p, 9p, 13q, 18q and 21q. In laryngeal carcinomas, alterations in multiple regions have been reported, commonly at +2p, +3q, +5p, +8q, +11q13, +18p, -3p, -5q, -9p, -13q and -18q (Progenetix database; Baudis and Cleary, 2001). Improved understanding of the tumorigenesis of HNSCC could be achieved by studying the impact of copy number on gene expression as has been already demonstrated in several other cancers (Hyman *et al.*, 2002; Pollack *et al.*, 2002; Wolf *et al.*, 2004). To date, previous microarray studies in HNSCC have mainly described separately either copy number or expression alterations. None of the studies have integrated genomewide data to achieve more accurate information about the target genes that are activated or inactivated by amplification or deletion with focus only on laryngeal cancer. Genetic alterations have been critically important in identification of therapeutic targets in cancer, such as HER2 (Trastuzumab), BCR-ABL and c-KIT (Gleevec). EGFR tyrosine kinase inhibitors (Cohen *et al.*, 2003; Soulieres *et al.*, 2004) as well as monoclonal antibodies against the extracellular domain of the receptor (reviewed by Baselga and Arteaga, 2005) represent first attempts to use targeted therapy in HNSCC.

In this study, owing to considerable heterogeneity of HNSCC tumors, we focused on a single well-defined anatomical location, larynx. Larynx is among the cancer sites where no significant improvement in survival rates during the past 25 years has been reported (Jemal *et al.*, 2005). Many aspects of the current management such as the benefit and optimal timing of combining chemotherapy to the radiation are still unknown. Here, we performed an integrated high-resolution copy number and gene expression profiling with genomewide cDNA

and oligonucleotide microarrays. We aimed to illustrate mechanisms behind this disease and more importantly, by focusing on those genes whose expression was altered because of copy number changes, to point out genes, which could be potential targets for new treatment modalities. The approach used in this study offers a general strategy to identify and prioritize targets for functional studies.

Results

Mapping of copy number alterations in laryngeal squamous cell carcinoma

A detailed copy number analysis was performed for each individual laryngeal squamous cell carcinoma (LSCC) sample (Table 1) using the CGH-Plotter (Figure 1a) to define the boundaries of copy number alterations as described in Materials and methods section. In Supplementary Table 1, altered regions present in at least 10% of LSCC samples are reported. In the cell lines, we observed multiple common alterations such as gains of 8q24.12–q24.13, 11q13.2–q13.3, 20p13, 20q11.22, 20q13.13 and 20q13.33 and loss of 18q21.33–q22.3, and in the tumors gains of 5p15.33, 5p12, 11q13.2–q13.4 and 14q24.3 as well as loss of 4q13.3. Several frequent alterations occurring both in cell lines and tumors were defined with high-resolution, such as gains of 8q24.12–q24.13, 11q13.2–q13.4 and 14q24.3 and loss of 10p12.31–p13 (Supplementary Table 1). The correlation coefficient for the frequency of alterations between tumors and cell lines was 0.44 for gains and 0.45 for losses. Overall, our LSCC material presented multiple copy number changes such as gains

Table 1 Clinicopathological properties of LSCC samples

Sample	Sex	Age	TNM stage ^a	Grade	Type of lesion	Previous treatment	Tobacco	Alcohol	Passage
<i>Cell lines</i>									
UT-SCC-8	M	42	T2N0M0	G1	Pri	No	Yes	Yes	p13
UT-SCC-11	M	58	T1N0M0	G2	Rec ^b	RT ^c	Yes	Yes	p8
UT-SCC-19A	M	44	T4N0M0	G2	Pri	No	Yes	Yes	p29
UT-SCC-19B	M	44	T4N0M0	G2	Pri(per ^c)	RT ^c	Yes	Yes	p26
UT-SCC-29	M	82	T2N0M0	G1	Pri	No	(Yes)	(Yes)	p16
UT-SCC-34	M	63	T4N0M0	G1	Pri	No	Yes	(Yes)	p6
UT-SCC-38	M	66	T2N0M0	G2	Pri	No	Yes	Yes	p16
UT-SCC-42A	M	43	T4N3M0	G3	Pri	No	Yes	Yes	p18
UT-SCC-49	M	76	T2N0M0	G2	Pri	No	Yes	Yes	p18
UT-SCC-75	M	56	T2N2bM0	G2	Pri	No	Yes	Yes	p23
<i>Primary tumors</i>									
Tumor 18	F	73	T3N1M0	G3	Pri	No	Yes	No	
Tumor 39	M	74	T3N0M0	G2	Pri	No	Yes	Yes	
Tumor 49	M	61	T3N1M0	G2	Pri	No	Yes	(Yes)	
Tumor 104	M	65	T2N0M0	G3	Rec ^d	No	Yes	No	
Tumor 132	M	77	T4N0M0	G2	Pri	No	No	No	
Tumor 202	M	57	T2N2aM0	G4	Pri	CRT ^f	Yes	Yes	
Tumor 224	M	69	T3N2bM0	G1	Pri	No	Yes	No	
Tumor H1	M	68	T3N0M0	G3	Pri	No	Yes	Yes	
Tumor T1	M	52	T1N0M0	G2	Pri	No	Yes	No	
Tumor T2	M	59	T3N2M0	G2	Pri	No	Yes	Yes	

^aPathological. ^bRec=primary tumor recurrence. ^cPer=persistent disease. ^dRec=neck recurrence, primary tumor removed one year earlier. ^eRT=Radiation therapy. ^fCRT=Chemoradiation therapy 2 years earlier. Abbreviations: LSCC, laryngeal squamous cell carcinoma; SCC, squamous cell carcinoma; TNM, tumor, regional node, distant metastasis.

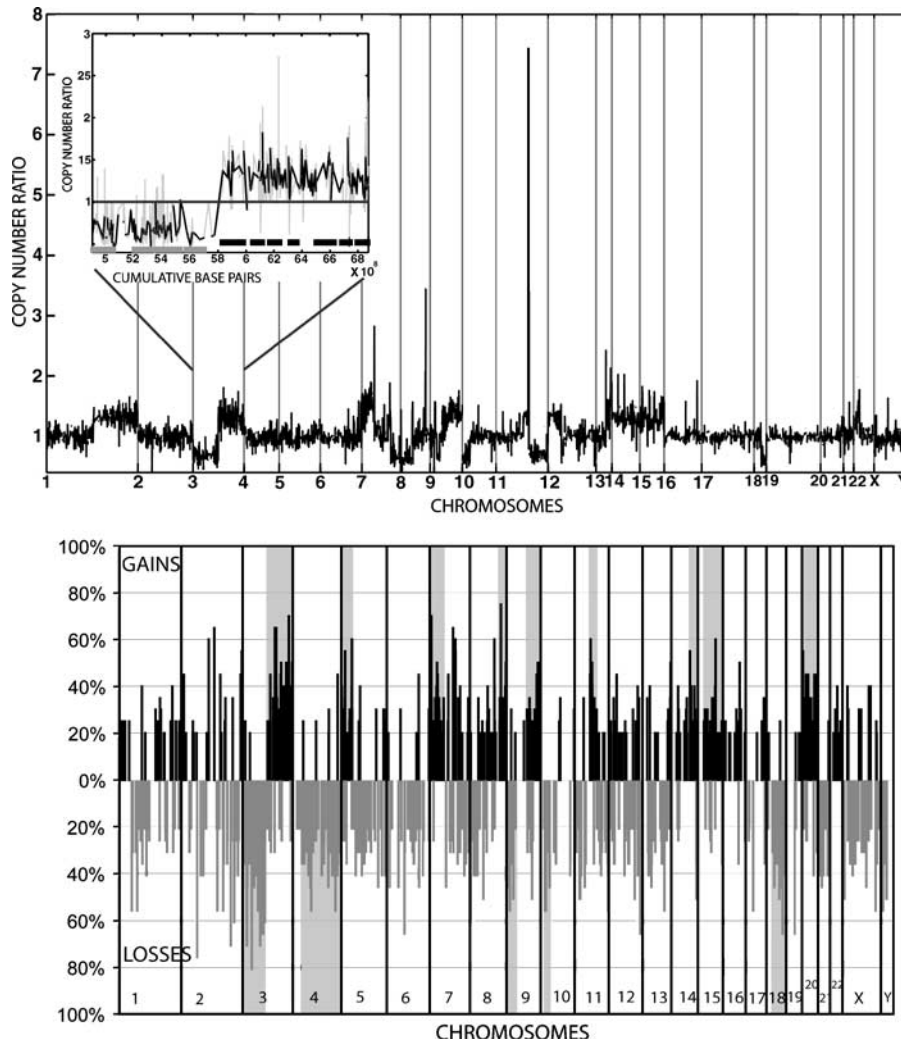


Figure 1 (a) Smoothed copy number profile from 1p to Yq for UT-SCC-8 cell line using a 750 kb window in CGH-Plotter for smoothing. This cell line featured several characteristic copy number alterations for HNSCC, such as gains at 3q, 7p and 11q13 as well as losses at 3p and 18q. Many of the changes featured in the UT-SCC-8 cell line were recurrent in the rest of the material. In a closer view from chromosome 3, the different visualizing options from CGH-Plotter are illustrated: black solid line for smoothed CGH data and gray line for original raw data. Furthermore, amplified and deleted regions in chromosome 3 identified by CGH-Plotter are denoted by bars below the copy number profile (b). Frequency of copy number alterations in 20 LSCC cell lines and primary tumors. Changes present in at least four samples are plotted applying 750 kb smoothed CGH data with same cutoffs for alterations as described earlier. Gray shading indicates areas with frequent alterations.

at 3q, 5p, 7p, 8q, 9q, 11q, 14q, 15q and 20, as well as losses at 3p, 4q, 9p, 10p and 18q, identified in more than 20% of the cases (Figure 1b). We also observed alterations such as gains at 16q and losses at 5, 6q, 8p, 11q, 13q and 21 with lower frequency.

Impact of copy number on gene expression in LSCC

High-level amplifications had a clear impact on the expression of genes in the altered genomic regions (Figure 2). Out of the genes that were in the highest copy number class (>2), 39% showed overexpression in the cell lines and 18% in the primary tumors. The impact of deletions on the reduced expression was less clear, although still detectable in the cell lines, with 14% of genes being underexpressed with copy number ratio <0.7. For comparison, in genomic sites with

no apparent copy number changes (CGH ratios 0.9–1.1), around 6.5% of genes were either over- or underexpressed.

To investigate expression and copy number data in individual samples, we visualized copy number alterations with expression values using custom-developed Expression annotation of Copy Number (ECN)-tool. We detected two different types of changes: (1) amplicons that were complex and included multiple genes (see Figure 3a and b) of which up to 40% were also overexpressed, (2) amplicons that composed of only few candidate target genes (Figure 3c and d). For example, an amplicon at 17q23 clearly led to the upregulation of two genes, *PPM1D* and *APPBP2*, which were reported as amplified and overexpressed in breast cancer (Hyman *et al.*, 2002; Sinclair *et al.*, 2003). Amplification at 11q13, presented in Figure 3a in

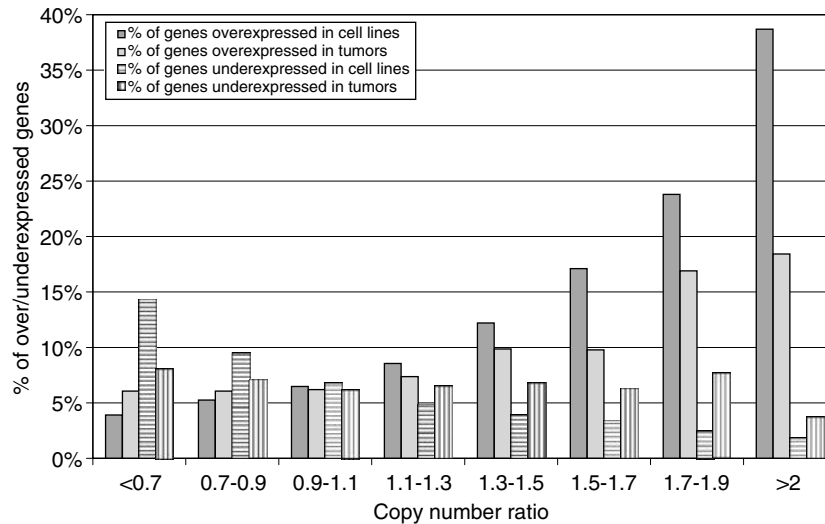


Figure 2 Global influence of copy number on gene expression in LSCC cell lines and primary tumors. Copy number classes from deletions to amplifications are placed from left to right in *X* axis. Percentages of over- and underexpressed genes in each copy number class (*Y* axis) are illustrated separately for both the cell lines and primary tumors. Threshold values for over- and underexpression for cell lines were 1.46 and 0.72 (global upper and lower 7th percentile of median centered expression ratios). The corresponding values for tumors were 1.85 and 0.59. Those gene expression values with no corresponding aCGH ratio were removed from global expression limit calculations.

UT-SCC19B cell line, was one of the most frequent alterations in our material (Supplementary Table 1) in concordance with previous reports in HNSCC. This region presented different sizes of amplicons with number of genes altered in expression. Especially *FADD* and *PPF1A1* demonstrated a good correlation between RNA and DNA values, whereas for example *CCND1* and *EMS1* were highly amplified but not necessarily highly overexpressed in the same sample. The amplification at 12q15 in UT-SCC-75 cell line composed of multiple target genes, such as *HMG2*, *MDM1*, *MDM2* as well as *DYRK2* that had the highest copy number as well as clear overexpression. Additionally, we found multiple other interesting amplified regions, which harbored genes showing simultaneous overexpression (examples in brackets), such as 2q13–q14 (*RALB*), 2q33 (*WDR12*, *NOP5*, *ALS2*, *BMPR2*, *CYP20A1*, *SUMO1*), 3q22 (*RAB6B*), 3q26 (*PDCD10*, *GPCR1*), 3q27 (*MCCC1*, *LAMP3*), 3q29 (*TFRC*), 5p15 (*TRIP13*, *CCT5*, *TRIO*), 7p11–p15 (*EGFR*), 8p11–p12 (*BRF2*, *ASH2L*, *WHSC1L1*, *sFRP1*, *GCP16*, *FNTA*), 8q24 (*SLA*, *WISPI*, *NDRG1*), 9p24 (*HARC*, *RNAC*, *JAK2*, *INSL6*, *RLN1*, *AK3*), 9q22 (*FBP1*), 9q31 (*EDG2*), 10q23.33 (*HHEX*), 11q22 (*BIRC 2* and *3*), 12p12 (*SLC21A8*), 12q13 (*SENPI*, *PFKM*), 13q22, 18q11, 19p13.13–p13.2 (*RAD23A*) and 22q13 (*PLXNB2*) (Supplementary Figures 1–8).

Statistical analysis of the impact of gene copy number on gene expression

We applied statistical analysis (see Materials and methods for details) to systematically identify all genes whose expression was attributable to copy number alteration across the samples. Expression of 739 genes was significantly influenced by copy number increase across all 10 cell lines (Supplementary Table 2).

For these genes, the Pearson correlation between DNA and RNA levels was on average 0.45 (range 0.26–0.68). As LSCC tumors were more heterogeneous than the cell line material, we performed the same analysis separately for the 10 tumor samples and identified 325 genes whose expression was increased due to copy number alteration (Supplementary Table 3). The average DNA-RNA level correlation for these genes was 0.35 (range 0.13–0.61). Using this statistical method, we identified genes that have been previously implicated to have significance in SCC, including the *CCNL1* (Redon *et al.*, 2002) and *CTSL2* genes (Nawata *et al.*, 2003). Of the deregulated genes, 62% in the cell lines and 77% in the tumors, albeit located at clearly changed regions such as 12q14–q15 in UT-SCC-75 cell line, were present only in one sample. Altogether, we identified 40 genes whose expression was systematically influenced by DNA amplifications in both the tumors and the cell lines (Table 2). Some of these common genes were at well-characterized HNSCC amplicons, such as *FADD* and *PPF1A1* at 11q13 and *DVL3* and *MCCC1* at 3q27. One of the amplified and overexpressed genes was *COPS5* (*JAB1*) that maps to 8q13.2, a novel region recently pinpointed to be amplified in HNSCC (Lin *et al.*, 2006). This suggests a strong genetic influence for the activation of *JAB1* in LSCC, where it has previously been reported to be overexpressed at the protein level and also associated with unfavorable clinicopathologic variables (Dong *et al.*, 2005). Our data also implicate, for the first time, *C8orf2*, *BRF2* and *ASH2L* genes at 8p11–p12 in the pathogenesis of LSCC tumors. These genes were recently reported to be targets at the 8p11.21–p12 amplicon in breast cancer (Garcia *et al.*, 2005).

As the deletions seemed to have an association with gene underexpression in LSCC, we applied statistical

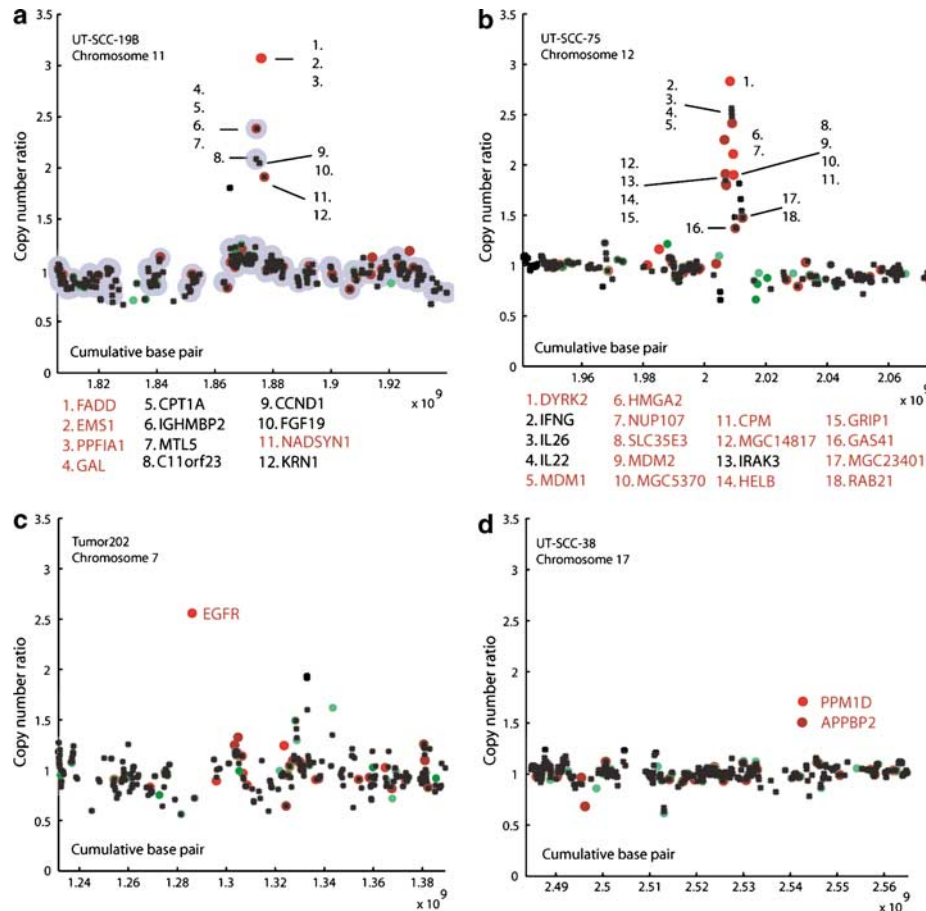


Figure 3 Expression annotated copy number plots for four LSCC amplicons visualized by ECN-tool. Color coding indicates the level of expression as follows: dark red dots indicate the upper 2nd percentiles and bright red dots indicate the next 5th percentiles of the gene expression ratios in the analysed sample (overexpressed genes). Green color denotes for downregulated genes. Gene symbols representing overexpressed genes are color coded in red. (a) 11q13 amplicon in UT-SCC-19B cell line. *FADD*, *PPF1A1* and *GAL* genes mapping to this region were both highly amplified and overexpressed. The interpolated values for genes that were present on an oligo array but no corresponding cDNA clone existed on a CGH array, are illustrated here with blue dots using a built-in option in ECN-tool. (b) 12q14–q15 amplicon in UT-SCC-75 cell line featuring many amplified and overexpressed genes such as *DYRK2*, *HMGA2*, *NUP107*, *SLC35E3*, *MDM2* and *GAS41*. Genes with low level gain and no overexpression are not annotated in the picture. (c) 7p12 amplicon in a primary tumor featuring *EGFR* as the most amplified and overexpressed gene in this region. In this case, *SEC61G* located next to *EGFR* was overexpressed but did not have analysable CGH value. In some cases with 7p amplicons, other genes close to *EGFR*, including *LANCL2*, *HUS1* and *UPP1*, were deregulated. (d) 17q23 amplicon in UT-SCC-38 cell line showing distinct amplification and overexpression of only two target genes, *APPBP2* and *PPM1D*.

approach to find those target genes. In the cell lines, 502 genes had this association including genes that have earlier implicated to have tumor suppressive function in HNSCC, such as *CST3* (20p11; Strojjan *et al.*, 2004), *APC* (5q22; Mao *et al.*, 1998), and *CDKN2A/p16* (9p21; Reed *et al.*, 1996). In the primary tumors, 223 genes were identified including 22 genes that were present also in the cell lines (data in Supplementary Tables 4 and 5). These results suggest that the method is able to detect pathogenetically relevant genes, and other genes pinpointed by our study could similarly turn out to be important in further functional studies.

Immunohistochemical validation of the impact of gene copy number and expression on protein level

Immunohistochemical staining was applied to validate the impact of detected gene deregulation on protein

expression levels. Antibodies for *CCND1* and *FADD* were used as examples to illustrate the concordance between DNA, RNA and protein levels. Both *CCND1* and *FADD* reside in the 11q13 region, commonly amplified in various cancers, including HNSCC. In addition to *CCND1*, our results show that *FADD*, chosen as an example among the set of genes statistically proven to be upregulated by DNA copy number increase in a subset of samples, had also strong protein level expression (Figure 4).

Biological annotation of the genes deregulated by DNA amplifications

We further focused our analysis on overexpressed and amplified genes because amplification had a clear impact on gene deregulation. We applied gene ontology (GO) information (Gene Ontology Consortium, 2000) using

Table 2 Statistically identified genes, whose expression was increased due to copy number gain both in the cell lines and primary tumors

Locus link	Gene symbol	Name
143	<i>ADPRTL1</i>	Poly (ADP-ribose) polymerase family, member 4
954	<i>ENTPD2</i>	Ectonucleoside triphosphate diphosphohydrolase 2
1857	<i>DVL3</i>	Dishevelled, dsh homolog 3 (<i>Drosophila</i>)
2099	<i>ESR1</i>	Estrogen receptor 1
4711	<i>NDUFB5</i>	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16 kDa
5693	<i>PSMB5</i>	Proteasome (prosome, macropain) subunit, beta type, 5 ^a
5927	<i>RBBP2</i>	Jumonji, AT rich interactive domain 1A (RBBP2-like)
6634	<i>SNRPD3</i>	Small nuclear ribonucleoprotein D3 polypeptide 18 kDa
7597	<i>ZNF46</i>	Zinc finger protein 46 (KUP)
8000	<i>PSCA</i>	Prostate stem cell antigen
8443	<i>GNPAT</i>	Glyceronephosphate O-acyltransferase
8500	<i>PPF1A1</i>	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1
8772	<i>FADD</i>	Fas (TNFRSF6)-associated via death domain
8972	<i>MGAM</i>	Maltase-glucoamylase (alpha-glucosidase)
9070	<i>ASH2L</i>	Ash2 (absent, small, or homeotic)-like (<i>Drosophila</i>)
9319	<i>TRIP13</i>	Thyroid hormone receptor interactor 13
9651	<i>KIAA0450</i>	Phospholipase C-like 4
9970	<i>NR1I3</i>	Nuclear receptor subfamily 1, group I, member 3 ^a
10987	<i>COP55</i>	COP9 constitutive photomorphogenic homolog subunit 5 (<i>Arabidopsis</i>)
11073	<i>TOPBP1</i>	Topoisomerase (DNA) II binding protein 1
11160	<i>C8orf2</i>	Chromosome 8 open reading frame 2 ^a
22948	<i>CCT5</i>	Chaperonin containing TCPI, subunit 5 (epsilon)
23077	<i>PAM</i>	MYC binding protein 2
26005	<i>DKFZP586P0123</i>	DKFZP586P0123 protein
27257	<i>LSM1</i>	LSM1 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>)
29083	<i>HSPC135</i>	HSPC135 protein
51117	<i>COQ4</i>	Coenzyme Q4 homolog (yeast)
55290	<i>BRF2</i>	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
55668	<i>C14orf118</i>	Chromosome 14 open reading frame 118
55848	<i>MDS030</i>	Chromosome 9 open reading frame 46
55975	<i>SBB126</i>	Kelch-like 7 (<i>Drosophila</i>)
56922	<i>MCCC1</i>	Methylcrotonoyl-coenzyme A carboxylase 1 (alpha)
64794	<i>DDX31</i>	DEAD (Asp-Glu-Ala-Asp) box polypeptide 31
79906	<i>FLJ13941</i>	Hypothetical protein FLJ13941
80824	<i>DUSP16</i>	Dual specificity phosphatase 16
80896	<i>NPL</i>	<i>N</i> -acetylneuraminidase pyruvate lyase (dihydrodipicolinate synthase)
83638	<i>P5326</i>	Basophilic leukemia expressed protein BLES03
84792	<i>MGC12966</i>	Hypothetical protein MGC12966
90809	<i>C14orf9</i>	Chromosome 14 open reading frame 9
128061	<i>DKFZp547B1713</i>	Hypothetical protein DKFZp547B1713

^aNo annotation in original microarray files.

gene ontology categorizer (GOC) (Joslyn *et al.*, 2004) and focused on molecular function (MF) and biological processes (BP) branches to investigate the biology behind deregulated genes identified by statistical analysis. Table 3 displays the top 15 ontology terms according to the score defined by GOC. Interestingly, many top GO terms were shared between tumors and cell lines, such as ones involved in ion binding and transcription factor activities as well as signal transduction and proteolysis and peptidolysis.

Genes that are overexpressed and amplified could be potential targets for therapeutic intervention. We therefore took a closer look on those ontology terms, which contained genes that might be potential drug targets by gene function. Using MF ontology term, we searched for kinases (GO:0016301) and phosphoprotein phosphatases (GO:0004721), which are critical in many signal transduction pathways; transmembrane receptors (GO:0004888), which are currently targets of many marketed drugs (e.g. G-protein-coupled receptors); peptidases (GO:0008233), or genes with catalytic

activity (GO:0003824, enzymes) (Hopkins and Groom, 2002) (Supplementary Tables 6 and 7). In GOC analysis, none of these terms were among the most enriched ones even though all of them were in upper 22nd percentile of terms, which scored above zero. Ambiguous overrepresentation of druggable terms in the data might be expected since genes that are repressed or activated by copy number alteration do not necessarily belong to a functionally same group. However, for example 34 and 13 kinases (GO:0016301) were detected as amplified and overexpressed in cell lines or tumors including interesting potential target genes such as *DYRK2* (12q15), *TK1* (17q25.3), *TRIO* (5p14–15.1) and *CSNK1E* (22q13.1).

We analysed pathway distribution of deregulated genes using available 717 and 309 LocusLink IDs and MetaCore software that mapped genes in relevant signaling pathways. Deregulated genes mapped at multiple pathways ($n = 161$ in cell lines and $n = 79$ in tumors). In MetaCore, the most enriched pathway classes ($P < 0.05$) in cell lines were involved in amino-acid metabolism and apoptosis, whereas in tumor

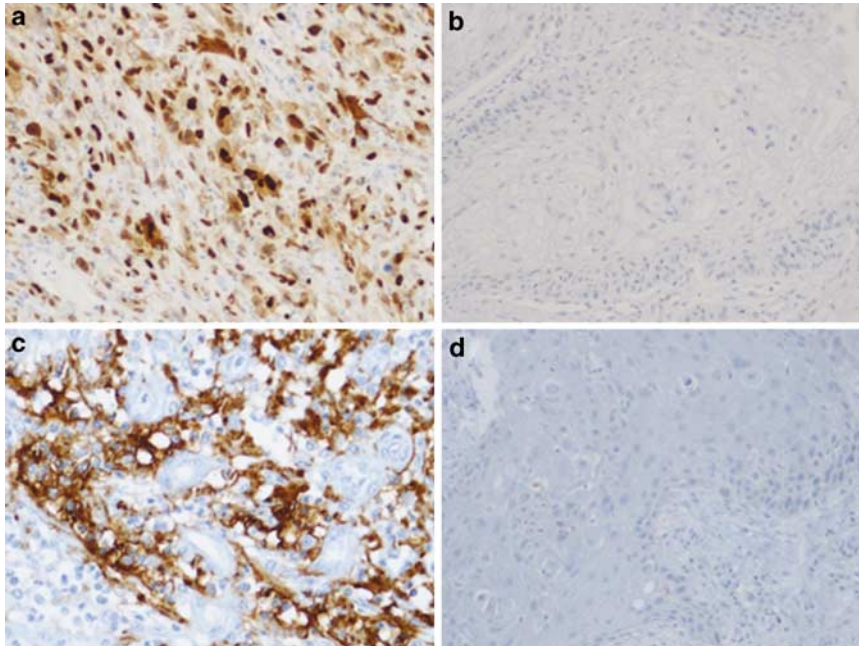


Figure 4 (a) Immunohistochemistry for CCND1 protein in tumor sample T224. Almost all tumor nuclei are strongly stained with some weaker staining also in the cytoplasm. The staining is tumor cell specific as cells of connective tissue stroma, blood vessels and inflammatory cells are negative. In comparison with other tumors of our material, this tumor had the most intense staining that was also most extensively distributed within the tumor. It was the only tumor sample where *CCND1* was identified both overexpressed and amplified in microarray analyses. (b) In CCND1 negative samples, such as T39, no or only negligible staining is detected. This tumor showed no overexpression or amplification of *CCND1* in microarray analyses. (c) Immunohistochemistry for FADD protein in positive tumor sample T224, where tumor tissue is stained strongly and extensively. This sample showed the highest *FADD* expression and copy number values in microarray analyses. (d) T132 as an example of a negative case where no staining of FADD is observed. This tumor showed no overexpression or amplification of *FADD* in microarray analyses.

samples proteolysis and apoptosis were mostly represented (Supplementary Table 8).

Discussion

In the present study, we successfully identified genes that are deregulated by DNA copy number alterations in LSCC. Genomewide CGH and expression microarray analyses revealed known and novel amplicons that showed concomitant increase of copy number and expression of target genes. This set of genes will provide a good basis for functional studies that might potentially lead to identification of novel drug targets in HNSCC. By applying bioinformatics tools, we were able to pinpoint a subset of target genes that could be druggable by function as well as identify processes involving the deregulated genes.

We detected several frequent copy number alterations in LSCC, such as gains at 3q, 5p, 7p, 8q, 9q, 11q, 14q, 15q and 20, as well as losses at 3p, 4q, 9p, 10p and 18q. Copy number increase had a clear impact on gene expression in LSCC, both in the cell lines as well as in the primary tumors. Previous studies by aCGH have reported similar results in breast, prostate and pancreatic cancers (Hyman *et al.*, 2002; Pollack *et al.*, 2002; Wolf *et al.*, 2004; Heidenblad *et al.*, 2005). LSCC featured different types of amplicons, ones that included

multiple target genes, such as the one at the 11q13, but also those that only contained few target genes such as at 17q23 or 7p12.

One of the most common highly amplified regions in our data set, manifesting deregulated genes both in the cell lines and the tumors, was 11q13. This amplicon is well known in HNSCC (Gollin, 2001), and *CCND1* is often considered as one of its target genes. In our integrated data, two genes, *FADD* and *PPF1A1*, featured statistically significant correlation between DNA and RNA levels in four cell lines and three primary tumors at this region. Traditionally, *FADD* has been reported as an adaptor molecule transmitting apoptotic signals of multiple death receptors with tumor-suppressive function (Newton *et al.*, 2000). To our knowledge, the function of *PPF1A1* in cancer is unclear. It is a LAR protein-tyrosine phosphatase interacting protein, and it binds to the intracellular region of the LAR and they both localize to the focal adhesions (Serra-Pagès *et al.*, 1995). Other genes in the region, such as *CCND1*, *EMSI* and *FGF19*, lacked statistically significant correlation between copy number and gene expression, and for *FGF4* and *FLJ10261* (*TMEM16A*) clone was not represented on the cDNA microarray. However, these genes had elevated gene expression levels combined with amplification in some of the samples, but the correlation was not statistically significant when all the samples were analysed. Across the samples, 11q13 region presented amplicons of

Table 3 Fifteen most enriched ontology terms from BP and MF branches in GOC

<i>Biological process</i>					
<i>Cell lines</i>			<i>Tumors</i>		<i>Examples of common genes^a</i>
<i>GO term</i>	<i>GO description</i>	<i>GO term</i>	<i>GO description</i>		
1	GO:0006355	<i>Regulation of transcription, DNA-dependent</i>	GO:0006355	<i>Regulation of transcription, DNA-dependent</i>	<i>ASH2L, RBBP2</i>
2	GO:0006508	<i>Proteolysis and peptidolysis</i>	GO:0000004	<i>Biological process unknown</i>	
3	GO:0000004	<i>Biological process unknown</i>	GO:0007275	<i>Development</i>	<i>ASH2L, DVL3</i>
4	GO:0006468	<i>Protein amino acid phosphorylation</i>	GO:0006508	<i>Proteolysis and peptidolysis</i>	
5	GO:0007165	<i>Signal transduction</i>	GO:0006118	<i>Electron transport</i>	
6	GO:0007275	<i>Development</i>	GO:0007165	<i>Signal transduction</i>	<i>DVL3, DUSP16, FADD</i>
7	GO:0015031	<i>Protein transport</i>	GO:0007166	<i>Cell surface receptor linked signal transduction</i>	
8	GO:0006118	<i>Electron transport</i>	GO:0006281	<i>DNA repair</i>	
9	GO:0006886	<i>Intracellular protein transport</i>	GO:0007267	<i>Cell-cell signaling</i>	
10	GO:0000398	<i>Nuclear mRNA splicing, via spliceosome</i>	GO:0043123	<i>Positive regulation of I-kappaB kinase/NF-kappaB cascade</i>	
11	GO:0016567	<i>Protein ubiquitination</i>	GO:0006810	<i>Transport</i>	
12	GO:0006512	<i>Ubiquitin cycle</i>	GO:0007242	<i>Intracellular signaling cascade</i>	
13	GO:0007155	<i>Cell adhesion</i>	GO:0006366	<i>Transcription from RNA polymerase II promoter</i>	
14	GO:0008283	<i>Cell proliferation</i>	GO:0007399	<i>Neurogenesis</i>	
15	GO:0006810	<i>Transport</i>	GO:0007264	<i>Small GTPase mediated signal transduction</i>	
<i>Molecular function</i>					
1	GO:0008270	<i>Zinc ion binding</i>	GO:0005524	<i>ATP binding</i>	<i>TRIP13, MCCC1</i>
2	GO:0005524	<i>ATP binding</i>	GO:0008270	<i>Zinc ion binding</i>	
3	GO:0005515	<i>Protein binding</i>	GO:0003700	<i>Transcription factor activity</i>	<i>RBBP2</i>
4	GO:0003677	<i>DNA binding</i>	GO:0003677	<i>DNA binding</i>	<i>ASH2L, RBBP2</i>
5	GO:0003700	<i>Transcription factor activity</i>	GO:0005515	<i>Protein binding</i>	<i>COP9, DVL3, FADD</i>
6	GO:0005554	<i>Molecular function unknown</i>	GO:0005554	<i>Molecular function unknown</i>	
7	GO:0003723	<i>RNA binding</i>	GO:0005525	<i>GTP binding</i>	
8	GO:0003676	<i>Nucleic acid binding</i>	GO:0004871	<i>Signal transducer activity</i>	
9	GO:0004872	<i>Receptor activity</i>	GO:0004872	<i>Receptor activity</i>	
10	GO:0004842	<i>Ubiquitin-protein ligase activity</i>	GO:0000287	<i>Magnesium ion binding</i>	
11	GO:0005509	<i>Calcium ion binding</i>	GO:0005509	<i>Calcium ion binding</i>	
12	GO:0005488	<i>Binding</i>	GO:0003924	<i>GTPase activity</i>	
13	GO:0004713	<i>Protein-tyrosine kinase activity</i>	GO:0003676	<i>Nucleic acid binding</i>	
14	GO:0005525	<i>GTP binding</i>	GO:0004842	<i>Ubiquitin-protein ligase activity</i>	
15	GO:0000287	<i>Magnesium ion binding</i>	GO:0005488	<i>Binding</i>	

Altogether, 608 and 457 terms in MF had score above zero among deregulated genes in cell lines and tumors, and in BP the respective numbers were 1019 and 710. Terms shown in italics were common among most enriched terms in cell lines and tumors. ^aExamples of common genes are presented according to their respective GO term in tumors. Abbreviations: BP, biological process; GO, gene ontology; GOC, gene ontology categorizer; MF, molecular function.

different sizes implying that this region harbors multiple target genes. We were also able to show that in the case of the 11q13 amplicon, DNA–RNA level correlation, exemplified by *FADD*, could be observed also at the protein level. More thorough functional analyses of 11q13 amplicon-derived transcripts are needed to gain complete knowledge of the target genes. In addition to 11q13, we identified many interesting, previously not well-characterized amplified regions in HNSCC, such as 8p11–p12, 9p24, 12q15 and 17q25.1–q25.3 harboring genes with significant overexpression. At 12q15, *DYRK2* with highest copy number in our data, was first recognized to be amplified and overexpressed in esophageal and lung adenocarcinomas (Miller *et al.*, 2003) but to our knowledge has not been reported in HNSCC. At 9p24, multiple genes were deregulated including *AK3L1* (*AK3*), *RNAC*, *MDS030* and

PDCD1L1 (*CD274*), which were also significantly overexpressed across the samples. Strome *et al.* (2003) demonstrated that freshly isolated HNSCC tissues expressed *PDCD1L1* protein.

We were able to narrow the number of target genes identified by statistical analysis to 40 by combining the data from both the tumors and cell lines. The GO analysis revealed that many highly represented terms were shared between tumors and cell lines. Therefore, this could implicate that important processes in LSCC are identified, even though different pathways might be activated. Similarly, pathways, to which different deregulated genes mapped, could be important because the activation of similar cellular processes may be achieved by multiple complex regulatory networks.

In some cases, the same gene showed either significant over- or underexpression accompanied with respective

copy number alteration implicating multiple mechanisms of genetic alterations involved in LSCC. For example, *BIRC2* gene (11q22) featured underexpression and deletion in some samples, but cell line UT-SCC-42A presented a clear amplicon at 11q22 including simultaneous overexpression of *BIRC2* as well as *BIRC3* and *PORIMIN* (Supplementary Figure 6). Additionally, several other genes, such as *MMP1*, 3, 7, 10, 12, 13, 20 and 27 were amplified at 11q22 but not overexpressed in this cell line. Imoto *et al.* (2001) reported amplification of this region in esophageal SCC as well as overexpression of *BIRC2*, and recently, Baldwin *et al.* (2005) identified region in oral cancer with no integration to expression data. Owing to our data allowed a direct integration of genomewide expression and CGH data, we could distinguish those genes that are deregulated through copy number alteration and are not just bystanders in an altered region. Our focus in this study was on those genes whose expression was increased together with copy number, since in many cancers gene amplification is an important mechanism to regulate gene expression. A recent review of HNSCC expression microarray publications reported only 32 genes with common alterations in transcriptional overexpression across multiple studies, including *COL4A1* identified also in our study as amplified and overexpressed (Choi and Chen, 2005). As authors conclude, only a few of these expression studies investigate possible mechanisms of altered gene expression observed by microarray analyses. Our study sheds light on this problem by combining copy number and expression microarray data. The statistical method applied in present approach is ideal for finding genes whose expression is driven by copy number changes and therefore, genes regulated by other mechanism than amplification, were excluded from the analyses.

The effect of copy number reduction to gene expression was not as clear, especially in tumors, where only 8% of genes showed underexpression when a copy number was less than 0.7. This might be due to the normal cell content of the tumor samples that will particularly influence on low copy number gains and losses that are most challenging to detect from heterogeneous tumor material. The samples, which lacked clear genomic changes, demonstrated poor association between RNA and DNA levels. Furthermore, on cDNA array platform amplifications are readily detected, whereas distinction of single copy losses is much more challenging. However, one of the genes showing association with deletion and underexpression in our data was a well-known tumor-suppressor gene *CDKN2A* at 9p21. This deletion is often homozygous but in UT-SCC-8 cell line, the raw ratio value on cDNA microarray was 0.45. We further investigated this deletion with whole-genome 44k oligo CGH arrays (Agilent Technologies, data not shown) confirming the homozygous deletion of *CDKN2A* gene in UT-SCC-8 cell line (raw ratio value 0.043). This example illustrates the potential to identify tumor-suppressor genes by integrating the information from CGH and gene expression microarray platforms.

We utilized improved custom-developed bioinformatics tools to analyse our data. To determine the copy number alterations in LSCC cell lines and primary tumors, we applied CGH-Plotter (Autio *et al.*, 2003). To be able to optimize the parameters for CGH analysis, CGH-Plotter requires user input and knowledge about typical alterations in the studied material. However, the automation for recognizing changes in CGH data makes analysis faster and more reproducible. We developed a new interpolation option in CGH-Plotter that enables integration of data derived from two different platforms lacking extensive overlap. Applying interpolation option, we retrieved CGH information of over 50% more clones than based on solely unique identifiers. To verify our approach, we analysed few cases by whole-genome 44k oligo CGH arrays and identified similar copy number changes (data not shown). Furthermore, the ECN-tool with new user-friendly graphical user interface provided an easy way to analyse and visualize expression data with CGH data derived directly from the CGH-Plotter analysis software.

Our results illustrate that microarray-based methods provide high-resolution information about cancer genomes. They facilitate an easy integration of CGH and expression data focusing our interest to those genes, which have altered expression owing to copy number changes. Integrated data can be applied for example to pinpoint potential therapeutic target genes, whose role in tumorigenesis can be followed further with functional assays. The data created in this genomewide screen is publicly available, offering a valuable resource for additional data mining to other investigators in HNSCC research field.

Materials and methods

LSCC cell lines and tumor samples

LSCC cell lines UT-SCC-8, UT-SCC-11, UT-SCC-19A, UT-SCC-19B, UT-SCC-29, UT-SCC-34, UT-SCC-38, UT-SCC-42A, UT-SCC-49 and UT-SCC-75 were kindly provided by Department of Otorhinolaryngology at Turku University Central Hospital (Turku, Finland). Cell lines UT-SCC-19A and UT-SCC-19B originated from the same patient. UT-SCC-11 and UT-SCC-19B were derived after radiation treatment. Cell culture conditions were as described previously (Grénman *et al.*, 1992; Lansford *et al.*, 1999). Total cellular RNA was extracted using TRIzol[®] reagent (Invitrogen, Carlsbad, CA, USA) followed by purification of the RNA using Qiagen's RNeasy columns (Valencia, CA, USA) according to the manufacturer's instructions. Genomic DNA was isolated either using Gentra's PureGene kit (Minneapolis, MN, USA) or Promega's Wizard Genomic DNA purification kit (Madison, WI, USA).

Tumor samples derived from larynx were obtained from Helsinki and Turku University Central Hospitals (Finland). Only the specimens, where at least 50% of cell population comprised tumor cells were taken to the analysis. RNA and DNA from tumor samples were extracted using Qiagen's RNA/DNA kit (Valencia, CA, USA). The use of sample material was approved by the Research Ethics Board at the Department of Otorhinolaryngology, HUCH and the Joint Ethical Committee of TUCH and Turku University.

Clinical information of the 20 samples used in the study is presented in Table 1. All the cell lines were derived from male patients and median age of the patients was 57. Of the primary tumors, 90% were obtained from male and 10% from female patients, and median age was 66 years at the time of surgery. Many of the patients had a positive history for both smoking and alcohol use. The clinical parameters in Table 1 are provided here as background information and were not used to draw clinical correlations owing to clinical heterogeneity and small size of the sample material.

Copy number profiling of LSCC on cDNA microarrays

CGH profiling on cDNA arrays was performed as described earlier (Pollack *et al.*, 1999) with slight modifications (Monni *et al.*, 2001; Hyman *et al.*, 2002) on Human 1 cDNA microarray slides (Agilent Technologies, Palo Alto, CA, USA). First, the concentration and purity of the DNA samples were measured using spectrophotometer ($A_{260}/A_{280} > 1.7$). Next, genomic DNA was digested overnight using *AthI* and *RsaI* restriction enzymes (Life Technologies Inc., Rockville, MD, USA) producing 200–600 bp DNA fragments. Sex-matched DNA, extracted with Qiagen Blood and Cell Culture Maxi Kit from white blood cells of healthy individuals, was used as a reference for CGH microarray hybridizations. Digested DNA samples were purified by phenol/chloroform extraction. From both test and reference samples, 4–6 μg of digested DNA was labeled with Cy5-dUTP and Cy3-dUTP (Amersham Biosciences, Piscataway, NJ, USA) in a random priming reaction using RadPrime DNA Labeling system (Invitrogen, Carlsbad, CA, USA). Hybridization conditions and washes were as described earlier (Pollack *et al.*, 1999).

Gene expression profiling of LSCC on oligonucleotide microarrays

Expression analysis was performed on Agilent's Human 1A oligo array according to the manufacturer's instructions. Before hybridization, the yield and integrity of total RNA was evaluated by spectrophotometer and 2100 Bioanalyzer (Agilent Technologies). Briefly, 20 μg of test RNA and 20 μg of reference RNA, composed of a pool from 10 different cancer cell lines (non-HNSCC), were labeled with Cy5-dCTP and Cy3-dCTP (Amersham Biosciences), and hybridizations were performed according to the manufacturer's recommendations (Agilent Technologies). Both raw expression microarray data as well as copy number microarray data are available at http://www.helsinki.fi/biochipcenter/Jarvinen_Supplement/.

Preprocessing of microarray data

A laser confocal scanner (Agilent Technologies) was used to measure signal intensities for both CGH and expression microarray hybridizations. Feature Extraction software (version A.6.1.1, Agilent Technologies) with the manufacturer's recommended settings were applied in microarray image analysis. However, for oligoarrays, we modified program's PolyOutlierFlagger settings slightly using the same values than in cDNA arrays (0.05; 0.09). Features, which did not pass quality control procedures in Feature Extraction software, were flagged automatically and their values were removed from further analysis. From nonunique probes, a median value was calculated for data analysis. Expression oligoarray data were median centered across the genes in 10 cell lines, and similarly independently across 10 clinical specimens when performing integration between CGH and expression data.

Each gene on both oligo and cDNA microarray was assigned to a UniGene cluster (Build 163; <http://ncbi.nih.gov/>). In addition to annotations provided by Agilent, we updated

annotations using RESOURCERER (<http://www.tigr.org/tigr-scripts/magic/r1.pl>; version 8) and Source databases (source.stanford.edu). A custom-made database was created including the genomic sequence alignment information for all available mRNA sequences according to the assembly by the University of California Santa Cruz's (UCSC) Genome Browser database (April 2003 freeze) (<http://genome.ucsc.edu/>), as well as the UniGene information. Genomic base pair localizations of the clones were retrieved by assigning each clone to its UniGene cluster and then relating the data to the genomic alignment of the mRNA sequence according to the largest alignment size from the corresponding cluster. In all subsequent analysis, we mapped individual clones into genome using start base pairs of the genomic alignments.

The base pair information was available for on average 11 200 (11 213 and 11 199 depending on batch) clones on cDNA arrays, providing an average spacing of 270 kb throughout the human genome. Similarly, information was available from 15 496 sequences on oligo microarrays providing a theoretical genomewide resolution of 200 kb. After removing duplicate clones, the number of unique clones were 7663 (7670 and 7656) on cDNA array and 13 643 on oligo array providing theoretical resolutions of 400 and 225 kb.

Analysis of copy number microarray data

The CGH copy number data were ordered according to the location of the clones along chromosomes starting from 1p and ending to Yq. A custom-developed software, CGH-Plotter (Autio *et al.*, 2003), was utilized to identify copy number changes on a genomewide scale by cDNA microarrays. We made modifications to the original version of CGH-Plotter to be able to determine the altered regions in the genome more accurately than with our previous version of the software. Both versions of the software can be downloaded at <http://www.cs.tut.fi/~bsmg/download.html>. We improved the original data smoothing options by allowing the user to define a genomic distance around each individual clone instead of using fixed number of clones for smoothing. Moreover, the method now acknowledges the known gaps in the genome that can be defined by the user. Here, we took into account telomeric, centromeric, and heterochromatin regions, as well as other gaps spanning more than 150 kb.

For smoothing of the data, we applied a moving median of the ratios with a 750 kb window (375 kb from both sides if possible). In the CGH-Plotter, the cutoff value in dynamic programming for gain or amplification was defined > 1.24 and for loss or deletion < 0.76 . These threshold values for altered copy number were determined based on the normal variation in the control hybridization; 99.5% of the raw copy number ratios in a self-versus-self experiment were between the cutoff values.

As a result, CGH-Plotter provides a text file describing amplified and deleted regions. In the CGH-analysis of the individual samples, we only analysed further those regions, which included four or more adjacent clones that had smoothed logarithmic ratio > 1.24 (gain) or < 0.76 (loss). The exact breakpoints were defined from the smoothed data in a manner that the clone in the border of altered region was normal. Altered regions separated by one clone with normal copy number, were joined to represent one region.

Integration of the gene expression and copy number data

The modified CGH-Plotter includes an option to interpolate data points in a case of missing values or when integrating data between two different platforms. In this study, we applied this feature to match as many data points as possible between oligo

(expression) and cDNA (copy number) microarrays. Briefly, if there were no matching cDNA clone for a gene or transcript present on the oligo microarray, we used location-based linear interpolation algorithm with a 750 kb window (375 kb from both sides) to interpolate CGH values from the cDNA data taking gaps in the genome into account similarly as described earlier in the data smoothing. When implementing interpolation, we assumed that CGH values from a certain genomic region are linearly dependent on the adjacent values. Theoretically, our cDNA and oligo data sets had 7030 common unique genes. After preprocessing and using of 750 kb interpolation window, 11 761 genes had both expression and CGH values at least in one sample.

Interpolated CGH data smoothed by the CGH-Plotter, was processed further using ECN-tool (<http://www.cs.tut.fi/~bsmg/download.html>). ECN is a MATLAB toolbox for plotting gene copy number data that is annotated with the gene expression values of the analysed sample. This plot enables easy and rapid visual integration of the expression data with the genomic information and identification of genes that are altered both in copy number and expression in one sample. The copy number ratios are annotated with expression ratios using color coding – red for upregulated genes (upper 7th percentile of expression ratios in analysed sample) and green for downregulated genes (lower 7th percentile of expression ratios in analysed sample). Furthermore, interesting genes with their copy number and gene expression values can be selected and saved in a text file format.

The influence of gene copy number on gene expression level was further evaluated across the samples by statistical method as described previously (Hyman *et al.*, 2002; Hautaniemi *et al.*, 2004). Briefly, for each gene, the smoothed CGH ratios were represented by a vector that was labeled '1' for gain (ratio > 1.28) and '0' for no gain. In opposite analysis, we labeled '1' for loss (ratio < 0.75) and '0' for no loss. As this method is based on individual genes rather than regions, we applied strict cutoff values for aCGH data in order to ensure that only those genes with true copy number alterations were taken into the analysis. A weight was calculated for each gene as described by Hautaniemi *et al.* (2004). To assess the statistical significance of each weight, we performed 10 000 random permutations of the label vector and gained *P*-value for each weight. This process was performed independently for cell lines and primary tumors. After the permutation test, we considered only genes with *P*-value below 0.05 to have association with copy number and gene expression.

References

- Autio R, Hautaniemi S, Kauraniemi P, Yli-Harja O, Astola J, Wolf M *et al.* (2003). *Bioinformatics* **19**: 1714–1715.
- Baldwin C, Garnis C, Zhang L, Rosin MP, Lam WL. (2005). *Cancer Res* **65**: 7561–7567.
- Baselga J, Arteaga CL. (2005). *J Clin Oncol* **23**: 2445–2459.
- Baudis M, Cleary ML. (2001). *Bioinformatics* **17**: 1228–1229.
- Belbin TJ, Singh B, Barber I, Succi N, Wenig B, Smith R *et al.* (2002). *Cancer Res* **62**: 1184–1190.
- Choi P, Chen C. (2005). *Cancer* **104**: 1113–1128.
- Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D *et al.* (2004). *Cancer Cell* **5**: 489–500.
- Cohen EE, Rosen F, Stadler WM, Recant W, Stenson K, Huo D *et al.* (2003). *J Clin Oncol* **21**: 1980–1987.
- Cromer A, Carles A, Millon R, Ganguli G, Chalmel F, Lemaire F *et al.* (2004). *Oncogene* **23**: 2484–2498.
- Dong Y, Sui L, Watanabe Y, Yamaguchi F, Hatano N, Tokuda M. (2005). *Clin Cancer Res* **11**: 259–266.
- Garcia MJ, Pole JC, Chin SF, Teschendorff A, Naderi A, Ozdag H *et al.* (2005). *Oncogene* **24**: 5235–5245.
- Gene Ontology Consortium (2000). *Nat Genet* **25**: 25–29.
- Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, Pambuccian SE *et al.* (2004). *Cancer Res* **64**: 55–63.
- Gollin SM. (2001). *Head Neck* **23**: 238–253.
- Grénman R, Pekkola-Heino K, Joensuu H, Aitasalo K, Klemi P, Lakkala T. (1992). *Arch Otolaryngol Head Neck Surg* **118**: 542–547.
- Hautaniemi S, Ringnér M, Kauraniemi P, Autio R, Edgren H, Yli-Harja O *et al.* (2004). *J Franklin Inst* **341**: 77–88.
- Heidenblad M, Lindgren D, Veltman JA, Jonson T, Mahlamaki EH, Gorunova L *et al.* (2005). *Oncogene* **24**: 1794–1801.
- Hopkins AL, Groom CR. (2002). *Nat Rev Drug Discov* **1**: 727–730.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E *et al.* (2002). *Cancer Res* **62**: 6240–6245.

Immunohistochemistry

We performed immunohistochemistry on paraffin-embedded tissue blocks from eight LSCC tumors using rabbit monoclonal antibodies for CCND1 (Lab Vision, Fremont, CA, USA), and rabbit polyclonal antibodies for FADD (Lab Vision, Fremont, CA, USA). Biotin-streptavidin detection system kits (Rabbit IgG Vectastain Elite ABC Kits, Vector Laboratories, Burlingame, CA, USA, and HRP-DAB Anti-goat Cell and Tissue Staining Kit, R&D Systems, Minneapolis, MN, USA) were used according to the manufacturers' instructions. First, the sections were treated with 0.01 M citrate buffer pH 6.0 in microwave oven (5 + 5 min) followed by cooling for 20 min. After blocking the endogenous peroxidase, sections were incubated with primary antibodies (diluted in 1.5% normal serum in PBS) for 30–60 min at RT. Immunoreaction was visualized with diaminobenzidine (DAB) and the sections were counterstained with Mayer's hematoxylin. Omission of the primary antibody in staining procedure was performed for control purposes.

Analyzing GO and pathway distribution of deregulated genes

To analyse which ontology classes and pathways were represented among the genes deregulated in LSCC identified by statistical method described in previous section, we applied GOC (Joslyn *et al.*, 2004) and MetaCore software (GeneGo, St Joseph, MI, USA). Ensembl IDs for GO analysis were retrieved from Ensembl database (<http://www.ensembl.org>; database 31; NCBI 35). Altogether, 692 and 300 deregulated genes had Ensembl Gene IDs available. In MF GO branch, the number of genes in the whole space was 13 375 whereas in BP the number was 12 347.

Acknowledgements

We thank Tuula Airaksinen and Marita Potila for excellent technical assistance, Mari Hero for help with clinical samples, Henrik Edgren for helpful discussions and comments especially with GO issues, Sami Kilpinen for advice with MATLAB, and Sampsa Hautaniemi for critical reading of the manuscript. This work was supported in part by Helsinki Biomedical Graduate School, Sigrid Juselius Foundation, Biocentrum Helsinki, Helsinki University Research Funds and Helsinki University Central Hospital Research Funds.

- Imoto I, Yang ZQ, Pimkhaokham A, Tsuda H, Shimada Y, Imamura M *et al.* (2001). *Cancer Res* **61**: 6629–6634.
- Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafour A *et al.* (2005). *CA Cancer J Clin* **55**: 10–30.
- Joslyn CA, Mniszewski SM, Fulmer A, Heaton G. (2004). *Bioinformatics* **20**(Suppl 1): i169–i177.
- Lansford CD, Grenman R, Bier H, Somers KD, Kim SY, Whiteside TL *et al.* (1999). In: Masters J and Palsson B (eds). *Head and Neck Cancers. Human Cell Culture, vol. 2, Cancer Cell Lines Part 2*. Kluwer Academic Press: Dordrecht (Holland). pp 185–255.
- Lin M, Smith LT, Smiraglia DJ, Kazhiyur-Mannar R, Lang JC, Schuller DE *et al.* (2006). *Oncogene* **25**: 1424–1433.
- Mao EJ, Schwartz SM, Daling JR, Beckmann AM. (1998). *J Oral Pathol Med* **27**: 297–302.
- Mao L, Hong WK, Papadimitrakopoulou VA. (2004). *Cancer Cell* **5**: 311–316.
- Miller CT, Aggarwal S, Lin TK, Dagenais SL, Contreras JJ, Orringer MB *et al.* (2003). *Cancer Res* **63**: 4136–4143.
- Monni O, Bärlund M, Mousses S, Kononen J, Sauter G, Heiskanen M *et al.* (2001). *Proc Natl Acad Sci USA* **98**: 5711–5716.
- Nawata S, Nakamura K, Hirakawa H, Sueoka K, Emoto T, Murakami A *et al.* (2003). *Electrophoresis* **24**: 2277–2282.
- Newton K, Harris AW, Strasser A. (2000). *EMBO J* **19**: 931–941.
- Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF *et al.* (1999). *Nat Genet* **23**: 41–46.
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE *et al.* (2002). *Proc Natl Acad Sci USA* **99**: 12963–12968.
- Redon R, Hussenet T, Bour G, Caulee K, Jost B, Muller D *et al.* (2002). *Cancer Res* **62**: 6211–6217.
- Reed AL, Califano J, Cairns P, Westra WH, Jones RM, Koch W *et al.* (1996). *Cancer Res* **56**: 3630–3633.
- Serra-Pagès C, Kedersha NL, Fazikas L, Medley Q, Debant A, Streuli M. (1995). *EMBO J* **14**: 2827–2838.
- Sinclair CS, Rowley M, Naderi A, Couch FJ. (2003). *Breast Cancer Res Treat* **78**: 313–322.
- Soulieres D, Senzer NN, Vokes EE, Hidalgo M, Agarwala SS, Siu LL. (2004). *J Clin Oncol* **22**: 77–85.
- Strojan P, Oblak I, Svetic B, Smid L, Kos J. (2004). *Br J Cancer* **90**: 1961–1968.
- Strome SE, Dong H, Tamura H, Voss SG, Flies DB, Tamada K *et al.* (2003). *Cancer Res* **63**: 6501–6505.
- Wolf M, Mousses S, Hautaniemi S, Karhu R, Huusko P, Allinen M *et al.* (2004). *Neoplasia* **6**: 240–247.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>).