Integrated analysis of expression and genome alteration reveals putative amplified target genes in esophageal cancer

TAKASHI SUGIMOTO[1,2,5], MAKOTO ARAI[3], HIDEAKI SHIMADA[4], AKIRA HATA[1] and NAOHIKO SEKI[2]

Departments of [1]Public Health, [2]Functional Genomics, [3]Medicine and Clinical Oncology, and [4]Frontier Surgery, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670; [5]DNA Chip Research Inc., 1-1-43 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Received March 30, 2007; Accepted May 14, 2007

Abstract. Microarray and comparative genomic hybridization (CGH) studies have provided a wide range of information about esophageal cancer, but the correlations between gene expression and copy number alteration are largely unknown. To identify putative amplification target genes in esophageal cancer, a survey of parallel DNA copy number and gene expression in 10 esophageal squamous cell carcinoma (ESCC) cell lines was performed using classical CGH and oligonucleotide microarrays. The gene expression and copy number data were subsequently integrated using signal-to-noise ratio analysis. The results revealed a set of 97 genes with elevated expression levels that were attributable to increased copy number. The set included genes previously reported as overexpressed in cancer as well as several novel genes associated with copy number elevation. These genes are involved in essential cellular processes (e.g., regulation of transcription, signal transduction, cell proliferation, the cell cycle and cell differentiation) that can also have an impact on cancer development. Thus, the integration of DNA and RNA profiles provides a highly productive entry point for the discovery of genes involved in the development and progression of esophageal cancer.

Introduction

The initiation and progression of human cancer is associated with the accumulation of alterations in the function of important regulatory genes. Many different factors, including changes in genome copy number, can perturb appropriate gene function. The consensus that certain genomic aberrations may involve genes that are important for tumor development (1,2) is particularly evident in cases involving changes in gene dosage; oncogene activities may be enhanced by amplification while tumor suppressor genes may be suppressed by a physical deletion (3). However, since a copy number change does not necessarily induce actual alterations in gene expression, functional consequences of recurrent abnormalities do not always appear (4). The issue is further complicated by the observation that many aberrations map to large chromosomal regions that contain multiple genes (1) that are not directly associated with tumor pathogenesis.

Microarray and comparative genomic hybridization (CGH) have been used to document solid tumor features. Gene expression analyses using microarrays have revealed novel candidate genes for cancer initiation, progression and malignancy in esophageal cancer (5-9). In CGH analysis, the pattern of aberration, which comprises the numbers and types of aberrations and the regions that are recurrently altered, is characteristic of each tumor type. In esophageal cancer, copy number gains have been observed at 3q26.3-27, 5p15, 8q24, 11q13, 14q32, 20q13.3 and Xq27-28. while recurrent losses have been observed at 9p13 (10,11). The frequent amplifications and their effects on expression levels have been monitored in recent studies with CGH microarrays. For example, in work done on breast cancer cell lines, 40% of highly amplified genes were overexpressed (12), whereas in a similar study of primary breast tumors, 62% of highly amplified genes were overexpressed (13). A study on head and neck squamous cell carcinoma concluded that large chromosomal regions are transcriptionally affected, although many genes appeared to be unrelated to malignant progression (14). Such studies suggest that genomic alterations can directly influence global expression patterns, and that the alterations may be selected because they alter the expression of multiple genes that coordinately promote tumor progression (1). However, a similar work on colorectal tumors produced a different result. An analysis of the gene expression data suggested that only a small minority of amplified genes are overexpressed in colorectal tumors (4). Putative target genes in pancreatic cancer were extracted when direct integration of CGH and microarray data, with bioinformatics analysis, was used to specifically identify those gene expression change events associated with copy number alterations (15,16). As for esophageal cancer, the relationship between changes in DNA content and gene expression remains to be discovered. Therefore, identification of the affected genes in these loci,

Correspondence to: Dr Naohiko Seki, Department of Functional Genomics, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670, Japan
E-mail: naoseki@faculty.chiba-u.jp

Key words: esophageal cancer, microarray, CGH, signal-to-noise ratio analysis

elucidation of their functions and association of these genes with cancer progression are required to fully understand esophageal tumorigenesis and progression.

In this study, we integrated CGH and expression profiling to reveal the associated copy-number-driven changes in gene expression and the effective entry points for cancer gene discovery in esophageal cancer. Analysis of 10 esophageal squamous cell carcinoma (ESCC) cell lines revealed a set of 97 genes whose expression levels were correlated with copy number increase. We hypothesize that these genes are likely to have a central role in the pathogenesis of esophageal cancer.

Materials and methods

Cell lines. Ten of the esophageal squamous cell carcinomas (TE1, TE2, TE3, TE10, TE13, TTn, YES1, YES2, YES4 and YES6) were generous gifts from Dr H. Shimada (Department of Frontier Surgery, Graduate School of Medicine, Chiba University, Japan) (17). All of the cell lines were maintained in RPMI-1640 (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT). These cells were cultured in 5% CO_2 at 37˚C.

CGH analysis

DNA isolation. Chromosomal DNA was isolated from the 10 cell lines using FlexiGene (Qiagen, Hilgen, Germany), according to the manufacturer's protocol. The purity and molecular weight of the DNA were estimated using agarose gels.

Labeling, hybridization and image analysis. CGH was performed as previously described (18). Briefly, the DNA from each cell lines was directly labeled with SpectrumGreen-dUTP (Vysis Inc., Downers Grove, IL) using a CGH nick translation kit (Vysis). Normal sex-matched reference DNA was labeled with SpectrumRed-dUTP (Vysis). Labeled cell line and reference DNAs, together with 10 μg of Cot-1 DNA, were denatured and hybridized to metaphase spreads that were prepared using standard protocol. The slides were washed and counterstained with DAPI. On the basis of these findings, the cut-off values were set at 1.3 and 0.7 with a 95% confidence limit.

Microarray analysis

Total-RNA isolation. Total-RNA was extracted by using TRIzol reagent (Invitrogen) following the manufacturer's protocol. RNA integrity was checked with the RNA 6000 Nano Assay kit and the 2100 Bioanalyzer (Agilent, Palo Alto, CA).

Labeling, hybridization and image analysis. The AceGene Human Oligo Chip 30K (DNA Chip Research Inc. and Hitachi Software Engineering, Yokohama, Japan) oligonucleotide microarray, which contained a total of 30,336 spots corresponding to 29,640 independent genes (the gene list is available at http://bio.hitachi-sk.co.jp/acegene/), was used for expression profiling. One microgram of total-RNA was used in each RNA amplification. Amino allyl-labeled antisense RNA was prepared with the Amino Allyl MessageAmp aRNA kit (Ambion, Austin, TX). Each aRNA from the 10 samples was labeled with Cyanine 5, whereas aRNA from Universal Reference Total-RNA (Clontech, Mountain View, CA) was labeled with Cyanine 3 as a control. Five micrograms of labeled aRNA was usually sufficient for one hybridization experiment. Probe purification, hybridization, and washings were performed according to the manufacturer's instructions. The arrays were scanned using the Packard GSI Lumonics ScanArray 4000 (Perkin-Elmer, Boston, MA). The data were analyzed by DNASIS array software (Hitachi Software Engineering, Yokohama, Japan), which converted the signal intensity of each spot into text format.

Data analysis. Log_2-ratios of the median subtracted background intensity levels were analyzed. Data from each microarray were normalized by global normalization.

Generation of the gene list. One-sample t-tests were used to identify genes that were different at the $p<0.01$ level with at least a 2-fold change in expression level.

Statistical analysis. The influence of genome aberrations on gene expression levels was evaluated as previously described (12,15,16). Briefly, microarray log-ratios in each cell line were z-transformed. The CGH data were represented by a vector that was labeled '1' for gain ratio >1.3 and '0' for no gain. The relation between CGH and microarray data was integrated by signal-to-noise ratio statistics (12,15,16). For the analysis, the regions of gain detected in five or more cell lines were used. A weight ω was calculated for each gene: $\omega = (m_1-m_0)/(\delta_1-\delta_0)$; where m_1, δ_1 and m_0, δ_0 denote the means and standard deviations of the expression levels in amplified and non-amplified cell lines, respectively. To assess the statistical significance of each weight, 10,000 random permutations of the label vector were performed. The probability that a gene had an equal or greater weight than the original weight by random permutation was denoted by α. A low α (<0.05) indicates a strong association between gene expression and chromosomal gain.

Gene functions. Gene functions were analyzed using PathwayAssist software (Ariadne Genomics, Rockville, MD). Functional characteristics of genes were annotated by searching for connections of genes with common regulators and finding the shortest paths between nodes.

Quantitative PCR (QPCR)

Genomic DNA. Primers were designed to amplify products of 100-200 bp within target and control sequences. Primers for control sequences in each cell line were designed within a region of euploid copy number as shown by CGH analysis. QPCR was performed by monitoring the increase in fluorescence of SYBR green dye (Qiagen) with an ABI7700 (Applied Biosystems, Foster City, CA). The relative genome aberration was calculated by the comparative C_t method. The threshold value of correlation coefficient with microarray data was designated 0.8 or more.

Total-RNA. cDNA of the 10 cell lines was synthesized using QuantiTect Reverse Transcription kit (Qiagen). The measurement of gene expression was performed as described above.

Results

Genomic profiling of ESCC cell lines by chromosomal CGH. To produce a comprehensive survey of genomic aberrations

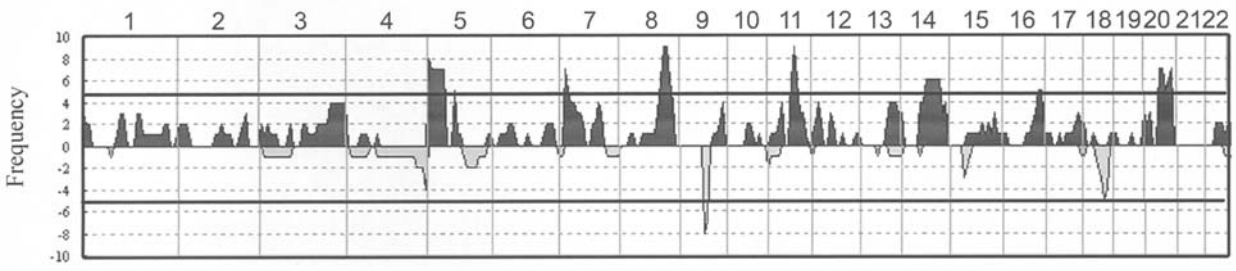

Figure 1. Gain/loss frequencies in 10 ESCC cell lines. Upper curve indicates the frequency of gains, whereas the lower curve shows the frequency of losses. Plots are shown with respect to chromosome order. Frequent over-representations (50% or more) were seen at 8q23-24, 11q13, 5p15.3, 5p15.2-12, 7p22, 20q11.2-12, 20q13.3, 8q22, 8q24.2, 11q12, 11q14, 14q13-24, 20q13.2, 5q11.2, 7p21, 16q23-24 and 20q13.1. Under-representations (50% or more) were found at 9q13, 9q21 and 18q22. Gains and losses were defined by $\log_2$-ratios of >1.3 and <0.7, respectively.

Table I. Highly alternated loci in ESCC.

Chromosome arm	Cytogenetic band	Frequency
Gain region		
5p	5p15.3	8/10
	5p15.2-12	7/10
5q	5q11.2	5/10
7p	7p22	7/10
	7p21	5/10
8q	8q22	6/10
	8q23-24	9/10
	8q24.2	6/10
11q	11q12	6/10
	11q13	9/10
	11q14	6/10
14q	14q13-24	6/10
16q	16q23-24	5/10
20q	20q11.2-12	7/10
	20q13.1	5/10
	20q13.2	6/10
	20q13.3	7/10
Loss region		
9q	9q13	8/10
	9q21	7/10
18q	18q22	5/10

Gains and losses were defined by $\log_2$-ratios of >1.3 and <0.7, respectively.

in ESCC, 10 cell lines were analyzed using classical CGH for patterns of chromosomal gains and/or losses. Many of these genome alterations were observed in all cell lines (Fig. 1). A region with over-representation in five or more cell lines was designated a 'gain region', and an under-represented region as a 'loss region'. Gain and loss regions are summarized in Table I. Gain regions were found at 8q23-24 (9/10 cell lines), 11q13 (9/10), 5p15.3 (8/10), 5p15.2-12 (7/10), 7p22 (7/10), 20q11.2-12 (7/10), 20q13.3 (7/10), 8q22 (6/10), 8q24.2 (6/10), 11q12 (6/10), 11q14 (6/10), 14q13-24 (6/10), 20q13.2 (6/10), 5q11.2 (5/10), 7p21 (5/10), 16q23-24 (5/10) and 20q13.1(5/10). Loss regions were found at 9q13 (8/10), 9q21 (7/10) and 18q22 (5/10).

Integration of genomic copy number and expression profile data. Expression levels in the 10 ESCC cell lines were analyzed using an oligonucleotide DNA microarray. The result of one-sample t-test analysis ($p<0.01$) led to the extraction of 767 genes, of which 631 were significantly over-expressed and 136 were underexpressed (data not shown).

Gene expression levels were then correlated with increased or decreased copy number by integrating gene expression and CGH data in each ESCC cell line using signal-to-noise ratio analysis. Only the gain regions on chromosomes 1-22 were used because resolution of DNA copy number losses by classical CGH is relatively unreliable. This analysis revealed significant correlations in 97 genes between gene expression and DNA copy number (Table II, $p<0.05$). The average expression log-ratio of the 97 genes was 0.548, which was higher than the average of all of the genes (0.096) ($p<0.001$, Student's t-test).

In Table III, the 97 genes were sorted by ω value. Top ranked genes were located at 11q12-14, indicating that genes located at 11q may be significant. Thus, to verify our approach, the status of 24 genes located at 11q12-14 was confirmed using QPCR. Analysis of the Pearson's correlation coefficients between microarray and QPCR data showed that the correlations of 21 genes (88%) were significant (Table IV). Gene copy number at 11q12-14 was also validated by genomic QPCR, and 16 genes (67%) were consistent with the CGH data.

The 97 genes and their loci are summarized in Table V. The genes and the loci were collected for each chromosome. Statistically significant genes were located with high probability at 20q11.2-13.3 (29/97 genes), 11q12-14 (24/97), and 14q13-24 (20/97). Conversely, genes with scarce transcripts were located at 16q23-24 and 7p22-21.

Obtaining the functional characteristics of the genes. The possible roles of the 97 genes in the pathogenesis of esophageal cancer were investigated using text-mining to annotate functional characteristics. A large fraction (49%) of the 47 genes was involved in key cellular processes including regulation of transcription, signal transduction, cell proliferation,

Table II. List of overexpressed genes with statistically significant correlation ($p<0.05$) between gene copy number and expression ratios.

Gene ID	Gene symbol	Map location	Description	ω	P-value
51585	PCF11	11q13	Pre-mRNA cleavage complex II protein Pcf11	5.78	0.005
5870	RAB6A	11q13.3	RAB6A, member RAS oncogene family	3.34	0.000
10938	EHD1	11q13	EH-domain containing 1	3.14	0.000
4221	MEN1	11q13	Multiple endocrine neoplasia I	2.74	0.000
78999	LRFN4	11q13.2	Leucine rich repeat and fibronectin type III domain containing 4	2.59	0.015
80227	WDR71	11q13.4	WD repeat domain 71	2.45	0.000
10825	NEU3	11q13.5	Sialidase 3 (membrane sialidase)	2.35	0.000
144097	LOC144097	11q13.1	Hypothetical protein BC007540	2.22	0.000
79139	DERL1	8q24.13	Der1-like domain family, member 1	2.11	0.000
84933	FLJ14825	8q24.13	Hypothetical protein FLJ14825	2.06	0.012
83940	TATDN1	8q24.13	TatD DNase domain containing 1	1.93	0.001
63930	C20orf51	20q13.33	Chromosome 20 open reading frame 51	1.86	0.000
10809	STARD10	11q13	START domain containing 10	1.81	0.000
116092	DNTTIP1	20q13.12	Deoxynucleotidyltransferase, terminal, interacting protein 1	1.71	0.000
9965	FGF19	11q13.1	Fibroblast growth factor 19	1.69	0.000
64979	MRPL36	5p15.3	Mitochondrial ribosomal protein L36	1.64	0.007
51526	C20orf111	20q13.11	Chromosome 20 open reading frame 111	1.52	0.003
8815	BANF1	11q13.1	Barrier to autointegration factor 1	1.51	0.008
58157	NGB	14q24	Neuroglobin	1.48	0.004
51111	SUV420H1	11q13.2	Suppressor of variegation 4-20 homolog 1 (*Drosophila*)	1.44	0.000
8500	PPFIA1	11q13.3	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), α 1	1.44	0.005
25879	WDSOF1	8q22.3	WD repeats and SOF1 domain containing	1.43	0.005
10992	SF3B2	11q13.1	Splicing factor 3b, subunit 2, 145 kDa	1.41	0.012
220074	LRRC51	11q13.4	Leucine rich repeat containing 51	1.40	0.004
11244	ZHX1	8q24.13	Zinc fingers and homeoboxes 1	1.35	0.000
5138	PDE2A	11q13.4	Phosphodiesterase 2A, cGMP-stimulated	1.31	0.010
871	SERPINH1	11q13.5	Serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	1.30	0.004
90199	WFDC8	20q13.12	WAP four-disulfide core domain 8	1.27	0.002
83658	DNCL2A	20q11.21	Dynein, cytoplasmic, light polypeptide 2A	1.25	0.025
55299	BXDC2	5p13.2	Brix domain containing 2	1.24	0.016
10923	PC4	5p13.3	Activated RNA polymerase II transcription cofactor 4	1.24	0.025
4726	NDUFS6	5p15.33	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13 kDa (NADH-coenzyme Q reductase)	1.18	0.027
51637	C14orf166	14q22.1	Chromosome 14 open reading frame 166	1.17	0.016
1477	CSTF1	20q13.31	Cleavage stimulation factor, 3' pre-RNA, subunit 1, 50 kDa	1.17	0.014
10001	MED6	14q24.2	Mediator of RNA polymerase II transcription, subunit 6 homolog (yeast)	1.15	0.009
6389	SDHA	5p15	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	1.15	0.015
1434	CSE1L	20q13	CSE1 chromosome segregation 1-like (yeast)	1.15	0.000
63916	ELMO2	20q13	Engulfment and cell motility 2 (ced-12 homolog, *C. elegans*)	1.15	0.005
51035	LOC51035	11q12.3	ORF	1.15	0.018
6227	RPS21	20q13.3	Ribosomal protein S21	1.14	0.015
54888	NSUN2	5p15.31	NOL1/NOP2/Sun domain family, member 2	1.14	0.005
6617	SNAPC1	14q22	Small nuclear RNA activating complex, polypeptide 1, 43 kDa	1.12	0.009
57143	ADCK1	14q24.3	aarF domain containing kinase 1	1.11	0.005
6749	SSRP1	11q12	Structure specific recognition protein 1	1.10	0.021
2778	GNAS	20q13.3	GNAS complex locus	1.10	0.047
10972	TMP21	14q24.3	Transmembrane trafficking protein	1.09	0.008
149954	C20orf186	20q11.21	Chromosome 20 open reading frame 186	1.08	0.018

Table II. Continued.

Gene ID	Gene symbol	Map location	Description	ω	P-value
112858	TP53RK	20q13.2	TP53 regulating kinase	1.08	0.008
51001	CGI-12	8q22.1	CGI-12 protein	1.08	0.005
5687	PSMA6	14q13	Proteasome (prosome, macropain) subunit, α type, 6	1.07	0.013
55905	ZNF313	20q13.13	Zinc finger protein 313	1.05	0.007
26205	GMEB2	20q13.33	Glucocorticoid modulatory element binding protein 2	1.04	0.017
140823	C20orf52	20q11.22	Chromosome 20 open reading frame 52	1.02	0.032
5203	PFDN4	20q13.2	Prefoldin 4	1.02	0.003
3895	KTN1	14q22.1	Kinectin 1 (kinesin receptor)	1.01	0.004
9791	PTDSS1	8q22	Phosphatidylserine synthase 1	0.99	0.008
51241	C14orf112	14q24.2	Chromosome 14 open reading frame 112	0.99	0.017
140685	BTBD4	20q13.33	BTB (POZ) domain containing 4	0.99	0.000
83853	ROPN1L	5p15.2	Ropporin 1-like	0.98	0.041
84181	CHD6	20q12	Chromodomain helicase DNA binding protein 6	0.95	0.009
8480	RAE1	20q13.32	RAE1 RNA export 1 homolog (S. pombe)	0.95	0.045
55251	C20orf36	20q13.33	Chromosome 20 open reading frame 36	0.93	0.015
57037	ANKMY2	7p21	Ankyrin repeat and MYND domain containing 2	0.92	0.024
64405	CDH22	20q13.1	Cadherin-like 22	0.92	0.012
10299	MARTH6	5p15.2	Membrane-associated ring finger (C3HC4) 6	0.91	0.009
1072	CFL1	11q13	Cofilin 1 (non-muscle)	0.91	0.021
55656	FLJ20530	8q22.1	Hypothetical protein FLJ20530	0.90	0.019
4054	LTBP3	11q12	Latent transforming growth factor ß binding protein 3	0.88	0.015
6590	SLPI	20q12	Secretory leukocyte protease inhibitor (antileukoproteinase)	0.87	0.031
26272	FBXO4	5p12	F-box protein 4	0.87	0.016
6156	RPL30	8q22	Ribosomal protein L30	0.86	0.024
10605	PAIP1	5p12	Poly(A) binding protein interacting protein 1	0.85	0.047
5494	PPM1A	14q23.1	Protein phosphatase 1A (formerly 2C), magnesium-dependent, α isoform	0.84	0.034
8813	DPM1	20q13.13	Dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit	0.84	0.031
55195	C14orf105	14q22.3	Chromosome 14 open reading frame 105	0.83	0.021
11065	UBE2C	20q13.12	Ubiquitin-conjugating enzyme E2C	0.81	0.023
91612	CHURC1	14q23.3	Churchill domain containing 1	0.80	0.029
5706	PSMC6	14q22.1	Proteasome (prosome, macropain) 26S subunit, ATPase, 6	0.80	0.049
54843	SYTL2	11q14	Synaptotagmin-like 2	0.80	0.030
2287	FKBP3	14q21.3	FK506 binding protein 3, 25 kDa	0.79	0.042
51528	C14orf100	14q23.1	Chromosome 14 open reading frame 100	0.76	0.037
54979	HRASLS2	11q12.3	HRAS-like suppressor 2	0.75	0.037
63935	C20orf67	20q13.12	Chromosome 20 open reading frame 67	0.74	0.037
58475	MS4A7	11q12	Membrane-spanning 4-domains, subfamily A, member 7	0.73	0.039
27304	MOCS3	20q13.13	Molybdenum cofactor synthesis 3	0.72	0.025
22990	PCNX	14q24.2	Pecanex homolog (*Drosophila*)	0.72	0.044
4053	LTBP2	14q24	Latent transforming growth factor ß binding protein 2	0.71	0.039
4147	MATN2	8q22	Matrilin 2	0.71	0.033
7764	ZNF217	20q13.2	Zinc finger protein 217	0.70	0.038
10490	VTI1B	14q24.1	Vesicle transport through interaction with t-SNAREs homolog 1B (yeast)	0.68	0.043
25962	DKFZP434I11	8q22.1	DKFZP434I116 protein	0.68	0.035
6729	SRP54	14q13.2	Signal recognition particle 54 kDa	0.66	0.039
2115	ETV1	7p21.3	ets variant gene 1	0.66	0.050
8611	PPAP2A	5q11	Phosphatidic acid phosphatase type 2A	0.65	0.044
6674	SPAG1	8q22.2	Sperm associated antigen 1	0.62	0.032
64841	GNPNAT1	14q22.1	Glucosamine-phosphate N-acetyltransferase 1	0.61	0.039
23613	PRKCBP1	20q13.12	Protein kinase C binding protein 1	0.58	0.036

Table III. A subset of which shows statistically significant genes grouped by each chromosome.

Cytogenetic band	Size (Mb)	No. of transcripts[a]	No. of transcripts on AceGene	
			Total[b]	Significant[c]
5p15-q11.2	60.56	316	136	11
7p22-21	18.79	154	83	2
8q22-24.2	43.26	214	199	12
11q12-14	36.74	632	349	24
14q13-24	45.27	441	238	20
16q23-24	13.94	159	44	0
20q11.2-13.3	31.73	437	312	29

[a]Number of transcripts are based on NCBI build 36. [b]Total, transcripts are represented on AceGene human 30K array. [c]Significant, a subset of which show statistical significance ($p<0.05$) for copy number correlation.

Table IV. Verification of expression and DNA copy number data by QPCR.

Gene symbol	Correlation coefficients	
	Microarray - QPCR	CGH - genomic QPCR
SSRP1	0.88[a]	0.85[a]
MS4A7	0.82[a]	0.84[a]
LOC51035	0.87[a]	0.80[a]
HRASLS2	1.00[a]	0.41
LOC144097	0.98[a]	0.81[a]
MEN1	0.98[a]	0.82[a]
EHD1	0.99[a]	0.43
LTBP3	0.77	0.62
CFL1	0.91[a]	0.87[a]
BANF1	0.99[a]	0.86[a]
SF3B2	0.80[a]	0.90[a]
LRFN4	0.77	1.00[a]
SUV420H1	0.83[a]	0.85[a]
FGF19	0.95[a]	-0.24
PPFIA1	0.95[a]	0.91[a]
LRRC51	-0.23	-0.56
PDE2A	0.95[a]	0.89[a]
STARD10	0.99[a]	0.85[a]
RAB6A	1.00[a]	0.93[a]
WDR71	0.92[a]	0.58
NEU3	0.95[a]	0.86[a]
SERPINH1	0.97[a]	0.86[a]
PCF11	0.99[a]	0.36
SYTL2	0.26	0.56

[a]Correlation coefficient is above 0.8.

cell cycle, and cell differentiation. The remaining 50 (51%) genes were novel and represented hypothetical proteins or known genes with no functional annotation.

Discussion

Esophageal cancer is a highly aggressive neoplasm. On a global basis, esophageal cancer is the sixth leading cause of cancer deaths. Gastric and esophageal cancers together accounted for nearly 1.3 million new cases and 980,000 deaths worldwide in 2,000, more than lung, breast or colorectal cancer (19). New methods for early detection, a better understanding of the biological mechanisms underlying cancer progression, and novel cancer-targeted treatments are urgently needed to reduce the mortality from this lethal disease.

The availability of global expression and genome aberration platforms, such as microarrays and CGH, has greatly facilitated the identification of novel tumor markers in many cancer types. The main focus of this study was to identify genes whose overexpression could be correlated with an increase in DNA copy number. Chromosomal alterations were first determined by performing CGH. Subsequent expression analysis using 30,336-spot microarrays permitted direct correlation between copy number and expression level on a gene-by-gene basis.

In this study, CGH analysis revealed altered loci (i.e., gain and loss regions) in the ESCC cell lines. The gain regions were located at 8q23-24, 11q13, 5p15.3, 5p15.2-12, 7p22, 20q11.2-12, 20q13.3, 8q22, 8q24.2, 11q12, 11q14, 14q13-24, 20q13.2, 5q11.2, 7p21, 16q23-24 and 20q13.1; the loss regions were found at 9q13, 9q21 and 18q22. The regions reported here match the region described in the published literature (10,11,20). These regions are also characteristic of other squamous cell carcinomas (head and neck, cervical, anal and oral carcinoma) and columnar epithelium (10).

Expression analysis of the 10 ESCC cell lines using oligonucleotide DNA microarrays extracted 767 genes with significantly altered expression profiles. Of these genes, 631 were overexpressed and 136 were under-expressed. Some of these genes have also exhibited altered expression profiles in esophageal cancer (e.g., *CDC2*, *ID1*, keratin, laminin, vimentin, insulin-like growth factor binding protein and S100 calcium binding protein) (5,7-9,21). Other studies have shown that it is possible to use gene expression microarrray data not only to identify characteristic genes that are associated with cancer development (7,22,23), but also to detect cytogenetic changes (24).

CGH and expression profiling data were integrated to evaluate copy-number-driven changes in gene expression. The integration of microarray data with CGH permits more relevant interpretations of the expression data by highlighting the dependence of gene expression on gene dosage. To facilitate the identification of possible amplification target genes, we applied a statistical approach, signal to noise ratio analysis, to investigate the effects of gene copy number on gene expression levels across all 10 cell lines (15,16). Only the gain regions were used for this analysis because resolution of DNA copy number losses by classical CGH is relatively unreliable.

Several previous studies have reported that chromosome 11 abnormalities are frequently found in esophageal cancer

Table V. Functional characteristics of the statistical significant genes.

Node	Count	Gene symbol
Regulation of transcription	12	BTBD4, CHURC1, ETV1, GMEB2, MED6, MEN1, PC4, PRKCBP1, SNAPC1, SSRP1, ZHX1, ZNF217
Signal transduction	9	CFL1, GNAS, LTBP2, MS4A7, PDE2A, PPAP2A, PPFIA1, RAB6A, UBE2C
Protein transport	6	CSE1L, DERL1, RAB6A, SYTL2, TMP21, VTI1B
Cell proliferation	4	CSE1L, PPAP2A, UBE2C, VTI1B
Apoptosis	3	CFL1, CSE1L, ELMO2
mRNA cleavage	3	CSTF1, PCF11, SF3B2
Phosphorylation	3	PPAP2A, PPM1A, TP53RK
Protein biosynthesis	3	MRPL36, RPL30, RPS21
Protein folding	3	FKBP3, PFDN4, SERPINH1
Cell adhesion	2	CDH22, PPFIA1
Cell cycle	2	LTBP2, UBE2C
Cell differentiation	2	RPS21, ZNF313
Electron transport	2	NDUFS6, SDHA
Metabolism	2	NEU3, PPAP2A
Microtubule-based movement	2	DNCL2A, KTN1
Protein secretion	2	GNAS, LTBP2
Signaling pathway	2	LTBP2, PPAP2A
Ubiquitin-dependent protein catabolism	2	PSMA6, PSMC6

(10), suggesting that oncogenes or genes which are involved in this locus are related to the initiation and/or progression of this cancer (10,25-27). QPCR analysis of 24 genes at 11q12-14 was used to verify microarray and CGH data. The Pearson's correlation coefficients between the microarray and QPCR data were high, with the correlations of 21 genes above 0.8 (Table IV). Validation of the copy number of these genes by genomic QPCR produced 16 genes that corresponded to the CGH data. These results strongly support the reliability and rationale of our approach. The confirmation rates between CGH and genomic QPCR might be low due to the possibly low resolution of CGH data.

Although gene copy number has been shown to be a major determinant of gene expression level, there are genes whose elevated expression levels are not correlated with increases in gene copy number (12). It was, therefore, expected that this statistical approach would identify an overlapping, but clearly separate, set of genes than those reported in studies where only expression levels were evaluated. This observation might be attributed to the multiple mechanisms responsible for normal and abnormal control of gene expression, including those related to mutation, promoter methylation and micro-RNA expression.

This systematic analysis identified 97 genes, including novel genes as well as previously described oncogenes, such as *RAB6A*, *MEN1*, *CSE1L*, *DNCL2A*, *GNAS*, *NEU3*, *NSUN2*, *PC4*, *PFDN4*, *PPFIA1*, *RPS21*, *SERRPINH1*, *SRF54*, *STARD10*, *UBE2C* and *ZNF217*. RAB6A is a known member of the RAS oncogene family; CSE1L plays a role in the mitotic spindle checkpoint, which assures genomic stability during cell division (28); and PC4 mediates enhancement of p53 DNA binding as a transcriptional coactivator (29). However, most of the genes identified by the statistical approach in this study are novel for esophageal cancer. These genes are expected to be related to the pathogenesis of this disease. An attempt to elucidate some of their functions using pathway analysis has not yielded significant results (data not shown).

Although this study was restricted to the gain regions, it is well accepted that DNA copy number aberrations owing to deletions and low-level gains are also of great importance in cancer pathology. Therefore, the analysis of copy number aberration is used at the gain regions, even though quantitative data for a single case might also be important in deciphering tumor-maintaining cancer gene networks.

In conclusion, the use of CGH, bioinformatics tools, and integration of expression profiles, led to the identification of genes with overexpression profiles that are associated with an alteration in DNA copy number in esophageal cancer. The high degree of genomic complexity, the recurrent nature of these lesions, and preliminary functional characterization of resident genes support the view that a large number of important oncogenes and tumor suppressor genes remain to be identified, opening potential therapeutic and diagnostic opportunities for this disease.

Acknowledgements

We thank honorary Professor K. Matsubara, President, DNA Chip Research Institute, Yokohama, Japan and Dr S. Kawasaki,

Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan for helpful advice and discussions.

References

1. Albertson DG, Collins C, McCormick F and Gray JW: Chromosome aberrations in solid tumors. Nat Genet 34: 369-376, 2003.
2. Vogelstein B and Kinzler KW: Cancer genes and the pathways they control. Nat Med 10: 789-799, 2004.
3. Alitalo K, Schwab M, Lin CC, Varmus HE and Bishop JM: Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. Proc Natl Acad Sci USA 80: 1707-1711, 1983.
4. Platzer P, Upender MB, Wilson K, *et al*: Silence of chromosomal amplifications in colon cancer. Cancer Res 62: 1134-1138, 2002.
5. Hu YC, Lam KY, Law S, Wong J and Srivastava G: Identification of differentially expressed genes in esophageal squamous cell carcinoma (ESCC) by cDNA expression array: overexpression of Fra-1, Neogenin, Id-1, and CDC25B genes in ESCC. Clin Cancer Res 7: 2213-2221, 2001.
6. Ishibashi Y, Hanyu N, Nakada K, *et al*: Profiling gene expression ratios of paired cancerous and normal tissue predicts relapse of esophageal squamous cell carcinoma. Cancer Res 63: 5159-5164, 2003.
7. Lu J, Liu Z, Xiong M, *et al*: Gene expression profile changes in initiation and progression of squamous cell carcinoma of esophagus. Int J Cancer 91: 288-294, 2001.
8. Su H, Hu N, Shih J, *et al*: Gene expression analysis of esophageal squamous cell carcinoma reveals consistent molecular profiles related to a family history of upper gastrointestinal cancer. Cancer Res 63: 3872-3876, 2003.
9. Tamoto E, Tada M, Murakawa K, *et al*: Gene-expression profile changes correlated with tumor progression and lymph node metastasis in esophageal cancer. Clin Cancer Res 10: 3629-3638, 2004.
10. Myllykangas S, Himberg J, Bohling T, Nagy B, Hollmen J and Knuutila S: DNA copy number amplification profiling of human neoplasms. Oncogene 25: 7324-7332, 2006.
11. Shinomiya T, Mori T, Ariyama Y, *et al*: Comparative genomic hybridization of squamous cell carcinoma of the esophagus: the possible involvement of the DPI gene in the 13q34 amplicon. Genes Chromosomes Cancer 24: 337-344, 1999.
12. Hyman E, Kauraniemi P, Hautaniemi S, *et al*: Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 62: 6240-6245, 2002.
13. Pollack JR, Sorlie T, Perou CM, *et al*: Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci USA 99: 12963-12968, 2002.
14. Masayesva BG, Ha P, Garrett-Mayer E, *et al*: Gene expression alterations over large chromosomal regions in cancers include multiple genes unrelated to malignant progression. Proc Natl Acad Sci USA 101: 8715-8720, 2004.
15. Aguirre AJ, Brennan C, Bailey G, *et al*: High-resolution characterization of the pancreatic adenocarcinoma genome. Proc Natl Acad Sci USA 101: 9067-9072, 2004.
16. Mahlamaki EH, Kauraniemi P, Monni O, Wolf M, Hautaniemi S and Kallioniemi A: High-resolution genomic and expression profiling reveals 105 putative amplification target genes in pancreatic cancer. Neoplasia 6: 432-439, 2004.
17. Nishihira T, Hashimoto Y, Katayama M, Mori S and Kuroki T: Molecular and cellular features of esophageal cancer cells. J Cancer Res Clin Oncol 119: 441-449, 1993.
18. Hashimoto Y, Oga A, Okami K, Imate Y, Yamashita Y and Sasaki K: Relationship between cytogenetic aberrations by CGH coupled with tissue microdissection and DNA ploidy by laser scanning cytometry in head and neck squamous cell carcinoma. Cytometry 40: 161-166, 2000.
19. Tew WP, Kelsen DP and Ilson DH: Targeted therapies for esophageal cancer. Oncologist 10: 590-601, 2005.
20. Tada K, Oka M, Hayashi H, Tangoku A, Oga A and Sasaki K: Cytogenetic analysis of esophageal squamous cell carcinoma cell lines by comparative genomic hybridization: relationship of cytogenetic aberrations to *in vitro* cell growth. Cancer Genet Cytogenet 117: 108-112, 2000.
21. Kan T, Shimada Y, Sato F, *et al*: Gene expression profiling in human esophageal cancers using cDNA microarray. Biochem Biophys Res Commun 286: 792-801, 2001.
22. Nishida K, Mine S, Utsunomiya T, *et al*: Global analysis of altered gene expressions during the process of esophageal squamous cell carcinogenesis in the rat: a study combined with a laser microdissection and a cDNA microarray. Cancer Res 65: 401-409, 2005.
23. Yamabuki T, Daigo Y, Kato T, *et al*: Genome-wide gene expression profile analysis of esophageal squamous cell carcinomas. Int J Oncol 28: 1375-1384, 2006.
24. Crawley JJ and Furge KA: Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. Genome Biol 3: RESEARCH0075, 2002.
25. Arai H, Ueno T, Tangoku A, *et al*: Detection of amplified oncogenes by genome DNA microarrays in human primary esophageal squamous cell carcinoma: comparison with conventional comparative genomic hybridization analysis. Cancer Genet Cytogenet 146: 16-21, 2003.
26. Ishizuka T, Tanabe C, Sakamoto H, *et al*: Gene amplification profiling of esophageal squamous cell carcinomas by DNA array CGH. Biochem Biophys Res Commun 296: 152-155, 2002.
27. Shiomi H, Sugihara H, Kamitani S, *et al*: Cytogenetic heterogeneity and progression of esophageal squamous cell carcinoma. Cancer Genet Cytogenet 147: 50-61, 2003.
28. Wellmann A, Flemming P, Behrens P, *et al*: High expression of the proliferation and apoptosis associated CSE1L/CAS gene in hepatitis and liver neoplasms: correlation with tumor progression. Int J Mol Med 7: 489-494, 2001.
29. Banerjee S, Kumar BR and Kundu TK: General transcriptional coactivator PC4 activates p53 function. Mol Cell Biol 24: 2052-2062, 2004.