Neoplasia • Vol. 6, No. 2, March/April 2004, pp. 143–149 143

www.neoplasia.com

Clustering of Molecular Alterations in Gastroesophageal Carcinomas[1]

Natalie Koon, Alexander Zaika*, Christopher A. Moskaluk†, Henry F. Frierson†, Sakari Knuutila‡, Steven M. Powell* and Wa'el El-Rifai**

*Digestive Health Center of Excellence and †Department of Pathology, University of Virginia Health System, Charlottesville, VA 22908-0708, USA; ‡Department of Medical Genetics, University of Helsinki, Helsinki, Finland

Abstract

Gene expression levels are regulated at many levels. Integration of genome-wide analyses for the study of DNA and RNA provides a unique tool to detect genetic alterations in the cancer genome. In this study, we generated and integrated DNA amplification data from comparative genomic hybridization (CGH) and serial analyses of gene expression (SAGE) in order to obtain a molecular profile of gastroesophageal junction (GEJ) carcinomas. DNA amplifications mapped to specific chromosomal regions and were frequently seen at 1q, 4q, 5q, 6p, 7p, 8q, 17q, and 20q. Using SAGE, we obtained over 156,432 tags from GEJ adenocarcinomas and normal gastric mucosa. These tags were assigned to UniGene clusters. Chromosomal positions for overexpressed genes were obtained to produce a GEJ carcinoma transcriptome map. A total of 123 genes was significantly overexpressed (more than fivefold; $P < .01$) in one or more SAGE libraries. This gene overexpression map was integrated and compared to the chromosomal CGH ideogram. Several chromosomal arms that had frequent DNA amplifications showed frequent gene expression alterations such as chromosomes 1 (15 genes), 2 (9 genes), 6 (6 genes), 11 (6 genes), 12 (8 genes), and 17 (13 genes). Despite the relatively large DNA amplification regions, overexpressed genes frequently mapped and clustered to small chromosomal regions at early-replicating (Giemsa light) bands such as 1q21.3 (nine genes), 6p21.3 (five genes), and 17q21 (eight genes). These results provide a comprehensive tool to search for DNA amplifications and overexpressed genes in GEJ carcinoma. The observed phenomenon of the presence of large amplification areas, yet clustering of overexpressed genes to relatively small loci, may suggest a high organization of chromatin and cancer-related genes in the nucleus.

Neoplasia (2004) **6**, 143–149

Keywords: Gastroesophageal junction cancer, comparative genomic hybridization, serial analysis of gene expression, gene clustering, transcriptome analyses.

Introduction

Gastroesophageal junction (GEJ) carcinomas have the most rapidly rising incidence of all visceral malignancies in the United States and Western world [1,2]. The majority of GEJ carcinomas are sporadic and exhibit various levels of DNA ploidy and chromosomal instability [3]. Comprehensive DNA copy number analyses using comparative genomic hybridization (CGH) have been widely used to characterize the DNA alteration in several cancer types. CGH can demonstrate recurrent DNA copy number changes and map them to chromosomal locations [4]. The development of serial analyses of gene expression (SAGE) technology has enabled genome-wide unlimited comprehensive profile of gene expression in a given cell population, representing the entire transcriptome [5,6]. This method has been valuable in studies of several tumor types including adenocarcinomas of the colon [7,8], prostate [9], pancreas [10], ovary [11], and breast [12].

Analyses of the human transcriptome map have shown clustering of highly expressed genes in chromosomal domains [13]. Chromosomal arms and bands are known to occupy specific locations within the nucleus known as chromosome territories (CTs). The positioning of a gene(s) can influence its access to the machinery responsible for specific nuclear functions such as transcriptional level and splicing [14]. In this study, we have globally explored the genome of GEJ carcinomas at the DNA and RNA levels, and mapped the DNA and gene expression changes to chromosomal positions, thereby generating a comprehensive genetic map of this deadly disease.

Materials and Methods

CGH

CGH was performed on 18 xenografted carcinomas that were generated from fresh tissues from surgically resected carcinomas of the lower esophagus ($n = 3$) or GEJ ($n = 15$).

Address all correspondence to: Wa'el El-Rifai, MD, PhD. Digestive Health Center of Excellence, University of Virginia Health System, PO Box 800708, Charlottesville, VA 22908-0708, USA. E-mail: wme8n@virginia.edu

[1]This work was supported by a grant award from the National Cancer Institute (1 R01 CA106176 to W.E.R.) and by the Cancer Center and Research and Development funds at the University of Virginia. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute or the University of Virginia.

Received 7 October 2003; Revised 11 November 2003; Accepted 19 November 2003.

Xenografting was performed as previously described [15]. We and others have shown that xenografted tumor tissues provide a pure source to study gene amplification and expression similar to primary tumor samples [16,17]. Mice were examined for tumor growth, and neoplasms were harvested and frozen upon reaching approximately 1 cm in diameter. High molecular weight genomic DNA was prepared from these frozen xenografted tumors by standard organic extraction methods. Histologic confirmation of the xenografted tumors was performed on cryostat-sectioned slides stained with hematoxylin and eosin (H&E). The CGH experiments were performed using a mixture of fluorochromes conjugated to dCTP and dUTP nucleotides for nick translation. Hybridizations, washings, and ISIS digital image analysis (Metasystems GmbH, Altlussheim, Germany) were performed as described elsewhere [18]. All the CGH results were confirmed using a 99% confidence interval. In each CGH experiment, a negative control (peripheral blood DNA from a healthy donor) and a positive control were included. Based on our earlier reports and the control results, we used 1.17 for DNA amplifications and 1.50 for high-level amplifications (HLAs).

SAGE

High-quality total RNA (500 μg) was extracted using RNeasy kit (QIAGEN GmbH, Hilden, Germany) from two dissected GEJ adenocarcinomas and a pool of four normal gastric epithelia biopsy samples that came from four patients who were referred for endoscopy for dyspepsia or for screening prior to gastric bypass surgery for obesity. The two tumors that were used in SAGE had closely similar pathological characters. GSM757 is a GEJ adenocarcinoma ($T_4N_0M_0$, moderately differentiated). GSM2385 is a GEJ adenocarcinoma ($T_4N_0M_0$, poorly differentiated). All normal samples had histologically normal mucosa confirmed on review of H&E–stained sections. Importantly, histopathology examination confirmed that none of the normal samples had any areas of inflammation or necrosis. The tumors selected for SAGE analysis were estimated to consist of more than 80% tumor cells. All samples were collected after obtaining informed consent in accordance with the Human Investigation Committee regulations at the University of Virginia.

SAGE libraries were constructed using *Nla*III as the anchoring enzyme and *Bsm*FI as the tagging enzyme, as described in SAGE protocol version 1.0e, June 23, 2000, which includes a few modifications of standard protocol [5]. A detailed protocol and schematic of the method is available (http://www.sagenet.org/sage_protocol.htm). Two thousand clones were sequenced for each case by the Cancer Genome Anatomy Project (CGAP). We used eSAGE 1.2a software to extract SAGE tags, remove duplicate ditags, tabulate tag contents, and link SAGE tags in the database to UniGene clusters using the recently reported ehm-TagMapping method [19,20]. The resulting tag libraries tags were compared to UniGene cluster and to the SAGE tag "reliable" mapping database (ftp://ftp.ncbi.nih.gov/pub/sage/map/Hs/Nla3/), and statistical analyses were performed using the eSAGE software. We have only included tags that had reliable gene hits. Tags with multiple gene hits were considered nonspecific and excluded from further analyses. Significant changes in levels of expression ($P \leq .01$) were determined.

Quantitative Real-Time Polymerase Chain Reaction (PCR)

For quantitative real-time PCR, 20 primary GEJ cancers and 13 normal gastric epithelial samples were collected. All tumors and normal gastric mucosal epithelial tissues were verified by our histopathologist (H.F. and C.M.). The collected tumors ranged from well-differentiated (WD) to poorly differentiated (PD) stages I–IIIa, and there was a mix of intestinal and diffuse-type tumors. The mRNA was isolated using RNeasy kit (QIAGEN GmbH). Single-stranded cDNA was synthesized using Advantage™ RT-for-PCR Kit (Clontech, Palo Alto, CA). Quantitative PCR was performed using iCycler (Bio-Rad, Hercules, CA) and threshold cycle number was determined using iCycler software

Table 1. DNA Copy Number Gains and Amplifications Using CGH on 18 GEJ and Esophageal Adenocarcinomas.

Case Number	Gains and HLAs (in bold)
1	12p13, 17q, 20q12-q13.1
2	1p31.2-pter (**1p34.2-pter**), 1q , 3p14-pter, 3q13.2-q24, 5q31-qter, **7p**, Xp, 8p, **8q22-qter**, 10p, 11q (**11q23.3-qter**), 16p, 17q12-q21, 18, **19q**, **20q**
3	1p32-pter, **5p**, 6p, 7p, 10p, 15q21-qter, 16p, 17q, 20q (**20q12-13.1**), 22
4	1q, 2pter-q14.3, 3p22-pter, 3q21-qter (**3q24-q26.3**), **4p**, 5p, 6q22-qter, 7q21.2-q22.1, **8p** (**8p21-pter**), **8q21.1-qter**, **11q13**, 12pter-q15, **12q23-qter**, 14q13-qter, 15q22-qter, 17q, 20 (**20q13**), 22
5	1q, 2q31-q33, 7p, 8q (**8q24**), 10q, 12, **15q24-qter**, 17q22, 19q, 20
6	1q, 3q21-qter, 5p, 6p11-p21, 7pter-q22, 8q21.2-qter, 10q24-qter, 17q, 18, **20q**, 21
7	1q21-qter, 2p, 3 (**3q25-q26.1**, **3q27-q28**), **5p12-p13.1**, 6p, **7p**, 8q, 11p14-qter, **12p12**, **12q14-q15, 12q24-q25**, 14q11-q21, 17q, 19q, 20 (**20q**)
8	1q21-qter, 2q, 3p24-pter, 7, 8q, 10pter-q22, 10q25-qter, 12p12, 12q14-q15, 12q24.2-qter, 14q11-q21, 17q, 18, 19q, 20q
9	1q21-qter, **2q31**, **2q33-q34**, 5p, 6pter-q24 (**6p**), **7p**, **8q22-qter**, 10, **11q13.4-q14**, 12, **13**, 15q21-qter, 16q12-qter, **17q12**, **20**, **22**
10	1q32-qter, 3q13.3-q23, 3q27-qter, 7q32-qter, 9q, 17q, 20q
11	2q (**2q32-q34**), 3q, 6p, **7q11-q31**, 10q, 12, 13q13-q21, 17, 18p, 20
12	3p21-pter, 7p14-q31 (**7q21-q31**), 8, **14q11-q24.1**, 17q, 20, 22
13	3q13.3-q23, **6p21.2-p21.3**, 7p, 8q24, 12p12-pter, 14q21-qter, **19q**, **20q**
14	4q11-q13, **6q22-q24**, 7p21-pter, 8q23-qter, 11q11-q13, 17q, 19q, 20
15	4q11-q21, **7q21**, **8p22-p23**, **12q14-q15**, **15q25-q26**, 17q21.2-qter, 19q, 20q
16	**6p21**, 7p (**7p13-pter**), 8q21.3-qter (**8q23-qter**)
17	7pter-q21 (**7p11-p13**), 8q (**8q24**), 10pter-q22, 12q21-qter, 17q (**17q11-q21.2**), **19q**, **20q**, Xp
18	8q22-qter, 9q, 10q21-q22, **11q13-q14**, 12p, 17q22-qter, 20

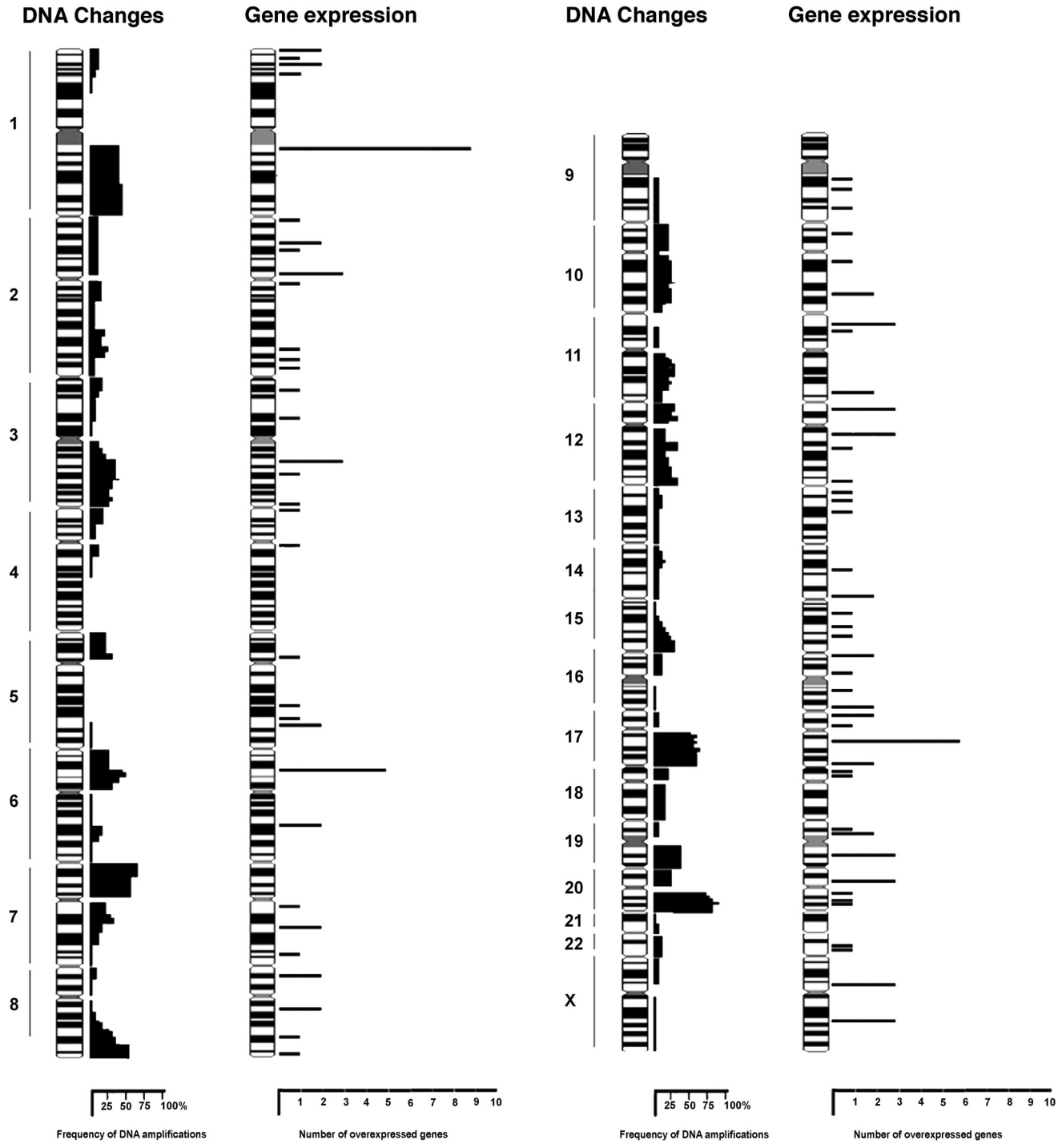

Figure 1. *Chromosomal ideogram showing DNA amplifications and gene overexpression alterations in GEJ carcinomas. Chromosomes are arranged vertically next to each other. Chromosome numbers are shown on the left-hand side. DNA amplifications are shown at the right-hand side of the chromosome. The scale at the bottom indicates the frequency (%) of the change. Gene expression alterations are also shown. Overexpression changes are shown at the right-hand side of the chromosome. The length of the horizontal bar correlates with the number of altered genes in a given chromosomal location. The scale is shown at the bottom. The highest gene clusters are seen at 1q21.3, 6p21.3, and 17q21.*

version 3.0. Reactions were performed in triplicate and threshold cycle numbers were averaged. Gene-specific primers for 11 genes (*LGALS3BP*, *PPP1R1B*, *HSPA5*, *TACSTD1*, *ANXA1*, *TOP2A*, *S100A6*, *S100A7*, *S100A8*, *S100A9*, and *S100A10*) were designed. These genes were chosen to cover different chromosomal locations and to be representative of different levels of gene overexpression in SAGE data. The primers used for RT-PCR were obtained from GeneLink (Hawthorne, NY), and their sequences are available on request. The results were normalized to β-amyloid precursor protein (APP), which had minimal variation in all normal and neoplastic GEJ samples that we tested. Fold overexpression was calculated according to the formula $2^{(R_t-E_t)}/2^{(R_n-E_n)}$, as earlier described [21],

Table 2. Gene Overexpression in GEJ Adenocarcinomas Using Serial Analyses of Gene Expression.*

TAG	UniGene Symbol	UniGene ID	Locus	Tag Number			Expression Ratio and *P* values			
				GSM784	GSM2385	GSM757	GSM2385 *vs* GSM784	P_{val} GSM2385	GSM757 *vs* GSM784	P_{val} GSM757
GGTTTGGCTT	UQCRH	285761	1p33	2	30	34	5.90	<0.01	6.50	<0.01
CGCCGACGAT	G1P3	287721	1p35	2	57	62	11.20	<0.001	11.90	<0.001
TTTCCTCTCA	SFN	184510	1p35	0	27	27	10.60	<0.001	10.30	<0.001
TTTGCACCTT	HMGN2	181163	1p36.1	0	9	17	3.50	0.073	6.50	<0.01
CTCTAAGAAG	C1QA	9641	1p36.3	0	28	22	11.00	<0.001	8.40	<0.01
AATCTGCGCC	G1P2	833	1p36.3	0	53	18	20.80	<0.001	6.90	<0.01
GGCTGGGGGC	COAS3	352407	1q21.3	7	108	78	6.10	<0.001	4.30	<0.001
GATCTCTTGG	S100A2	413843	1q21.3	0	2	46	−1.30	0.741	17.60	<0.001
CAGGCCCCAC	S100A11	417004	1q21.3	0	27	32	10.60	<0.001	12.20	<0.001
CCCCCTGGAT	S100A6	275243	1q21.3	2	83	57	16.30	<0.001	10.90	<0.001
GAGCAGCGCC	S100A7	112408	1q21.3	0	0	75	N/A	N/A	28.70	<0.001
TACCTGCAGA	S100A8	416073	1q21.3	0	0	284	N/A	N/A	108.60	<0.001
GTGGCCACGG	S100A9	112405	1q21.3	0	0	334	N/A	N/A	127.70	<0.001
GGCTTCTAAC	SPR-2A	505352	1q21.3	0	0	27	N/A	N/A	10.30	<0.001
AGCAGATCAG	S100A10	143873	1q21.3	1	48	95	18.80	<0.001	36.30	<0.001
GGCTGGTCTG	MGC4677	446688	2p11.2	0	18	14	7.10	<0.01	5.40	0.016
GAAACCCCAG	IGKC	306357	2p11.2	0	2	33	−1.30	0.741	12.60	<0.001
GGGGAAATCG	THYB10	446574	2p11.2	6	88	104	5.80	<0.001	6.60	<0.001
AGTTTGTTAG	TACSTD1	692	2p21	0	29	56	11.40	<0.001	21.40	<0.001
TTGTTGTTGA	CALM2	425808	2p21	0	19	6	7.50	<0.01	2.30	0.207
TCATCTTTAT	RPS7	444012	2p25.3	0	16	0	6.30	<0.01	N/A	N/A
TTCCTGGTAG	U5-200 KD	246112	2q11.2	1	22	11	8.60	<0.01	4.20	0.177
TAAATAATTT	HSPE1	1197	2q33.1	1	43	14	16.90	<0.001	5.40	0.080
ATCTTGTTAC	FN1	418138	2q35	0	34	12	13.30	<0.001	4.60	0.030
TCACAGTGCC	FLNB	81008	3p14.3	1	26	2	10.20	<0.01	−1.30	1.385
GTGTTAACCA	RPL15	74267	3p24.1	2	13	47	2.60	0.258	9.00	<0.001
CCCTCCCGAA	MUC13	5940	3q21	0	5	16	2.00	0.274	6.10	<0.01
ATCCTTGCTG	CSTA	412999	3q21	1	3	87	1.20	1.132	33.30	<0.001
GGGACGAGTG	TM4SF1	351316	3q21	1	26	32	10.20	<0.01	12.20	<0.001
ACCTTTACTG	TFRC	185726	3q29	0	17	5	6.70	<0.01	1.90	0.286
CCTGGTCCCA	SH3BP2	167679	4p16.3	0	2	33	−1.30	0.741	12.60	<0.001
CATATCATTA	IGFBP7	435795	4q12	1	41	50	16.10	<0.001	19.10	<0.001
ACTAATCGTT	RPL37	80545	5p13.1	0	18	6	7.10	<0.01	2.30	0.207
TTCACTGTGA	TA-WDRP	175596	5q22.1	4	96	108	9.40	<0.001	10.30	<0.001
GTGACAACAC	VDAC1	404814	5q31.1	2	18	31	3.50	0.080	5.90	<0.01
ATGTGAAGAG	SPARC	111779	5q31.3	1	65	75	25.50	<0.001	28.70	<0.001
AATTTGCAAC	H2AFY	75258	5q31.3	0	16	9	6.30	<0.01	3.40	0.078
CAGAGATGAA	HSPA1A	75452	6p21.3	1	7	43	2.70	0.461	16.40	<0.001
ACCCTTTAAC	HLA-E	381008	6p21.3	0	14	19	5.50	0.013	7.30	<0.01
AGCTTCTACC	HLA-A	181244	6p21.3	0	0	24	N/A	N/A	9.20	<0.001
GGGCATCTCT	HLA-DRA	409805	6p21.3	3	61	87	8.00	<0.001	11.10	<0.001
GTACTGTGGC	CLIC1	414565	6p21.3	0	2	18	−1.30	0.741	6.90	<0.01
AAGGCAATTT	TSPYL	278479	6q22.1	1	21	3	8.20	<0.01	1.10	1.153
CCCAAGCTAG	HSPB1	76067	7q11.2	0	13	94	5.10	0.019	36.00	<0.001
TGCACAATAT	MUC3B	129782	7q22	0	22	0	8.60	<0.001	N/A	N/A
GTTCCACAGA	COL1A2	232115	7q22	0	29	4	11.40	<0.001	1.50	0.396
CTGCCAAGTT	ZYX	75873	7q34	0	11	19	4.30	0.038	7.30	<0.01
TGGGTGAGCC	CTSB	135226	8p22	3	38	42	5.00	<0.01	5.40	<0.001
GAACGCCTAA	DPYSL2	173381	8p22	0	25	8	9.80	<0.001	3.10	0.108
GCTAATAATG	SULF1	409602	8q13.1	0	18	5	7.10	<0.01	1.90	0.286
ATTATTTTTC	RPL7	421257	8q13.2	2	117	38	23.00	<0.001	7.30	<0.001
TAAGTGGAAT	YWHAZ	386834	8q23.1	1	28	7	11.00	<0.01	2.70	0.482
CACTTCAAGG	LY6E	77667	8q24.3	0	12	29	4.70	0.027	11.10	<0.001
AGAAAGATGT	ANXA1	287558	9q21.1	0	56	138	22.00	<0.001	52.80	<0.001
AGCTGTATTC	CKS2	83758	9q22.2	0	17	11	6.70	<0.01	4.20	0.041
TTTTCTGAAA	TXN	395309	9q31	1	29	27	11.40	<0.001	10.30	0.002
TGCATCTGGT	HSPA5	310769	9q33.3	3	60	29	7.80	<0.001	3.70	0.018
TCCAAATCGA	VIM	435800	10p13	0	3	29	1.20	0.532	11.10	<0.001
AAAATACTAG	DKK1	40499	10q21.1	0	17	0	6.70	<0.01	N/A	N/A
TGATAATTCA	MGC14697	171625	10q24.3	3	38	10	5.00	<0.01	1.30	0.860
TTTGGTTTTC	RAB22A	21431	20q13.2	0	90	6	35.30	<0.001	2.30	0.207
ATATGTATAT	CD44	306278	11p13	0	16	8	6.30	<0.01	3.10	0.108
TCTTGTGCAT	LDHA	2795	11p15.1	1	51	19	20.00	<0.001	7.30	0.020
TAATAAATGC	TTS-2.2	118463	11p15.5	0	31	7	12.20	<0.001	2.70	0.150
GGATTTGGCC	RPLP2	437594	11p15.5	25	316	164	5.00	<0.01	2.50	<0.01
GCTGGTGCCT	THY1	134643	11q23.3	0	2	16	−1.30	0.741	6.10	<0.01
ACAGGCTACG	TAGLN	433401	11q23.3	1	0	28	−2.50	0.405	10.70	<0.01
GTCTCCTAAT	RAI3	194691	12p13.1	0	18	10	7.10	<0.01	3.80	0.050

continued on next page

Table 2. Continued

TAG	UniGene Symbol	UniGene ID	Locus	Tag Number			Expression Ratio and *P* values			
				GSM784	GSM2385	GSM757	GSM2385 *vs* GSM784	P_{val} GSM2385	GSM757 *vs* GSM784	P_{val} GSM757
TGGTTTGAGC	NDUFA9	75227	12p13.3	0	5	22	2.00	0.274	8.40	<0.001
AAGATTGGTG	CD9	387579	12p13.3	0	10	53	3.90	0.050	20.30	<0.001
GAAGCACAAG	KRT6A	334309	12q13.1	0	0	15	N/A	N/A	5.70	0.011
GGCAGAGAAG	KRT4	371139	12q13.1	0	0	29	N/A	N/A	11.10	<0.001
CTGTTGATTG	HNRPA1	356721	12q13.1	6	123	57	8.00	<0.001	3.60	<0.001
ATGTAAAAAA	LYZ	234734	12q15	5	75	378	5.90	<0.001	28.90	<0.001
ACTCCAAAAA	RNP24	75914	12q24.3	5	116	6	9.10	<0.001	−2.20	1.831
TGAAAGTGTG	HSP105B	36927	13q12.2	0	28	11	11.00	<0.001	4.20	0.041
TAATTTTTGC	GW112	273321	13q14.3	0	0	211	N/A	N/A	80.70	<0.001
CATCTGTACT	PCDH20	391781	13q21.2	0	0	19	N/A	N/A	7.30	<0.001
TTCACTGTGA	LGALS3	411701	14q22	4	96	108	9.40	<0.001	10.30	<0.001
TACTAGTCCT	HSPCA	446579	14q32.3	4	121	32	11.90	<0.001	12.20	<0.001
GAAATAAAGC	IGHG3	413826	14q32.3	12	302	389	9.90	<0.001	12.40	<0.001
CAAACTAACC	IGHM	153261	14q32.3	2	2	38	−2.50	1.720	7.30	<0.001
CAGGAGGAGT	GRP58	308709	15q15.3	0	1	16	−2.50	1.032	6.10	<0.01
CTTCCAGCTA	ANXA2	437110	15q22.2	2	27	29	5.30	<0.01	5.50	<0.01
TACTTGTGTG	SDFR1	389371	15q24.1	1	23	5	9.00	<0.01	1.90	0.762
GCGACCGTCA	ALDOA	273415	16p11.2	1	0	52	−2.50	0.405	19.90	<0.001
CCCCCTGCAG	MSLN	408488	16p13.3	0	105	5	41.20	<0.001	1.90	0.286
ACCGCCGTGG	CYBA	68877	16q24	0	0	46	N/A	N/A	17.60	<0.001
CCCAGAGCTC	HSD17B2	155109	16q24	0	16	3	6.30	<0.01	1.10	0.548
AACTAATACT	MGC40157	270232	17p11.2	6	108	14	7.10	<0.001	−1.10	1.287
GAAACCCCAG	HSA011916	84359	17p13	0	2	33	−1.30	0.741	12.60	<0.001
ATAGACATAA	C1QBP	78614	17p13	1	24	8	9.40	<0.01	3.10	0.378
GATCAATCAG	CCL18	16530	17q21??	0	16	3	6.30	<0.01	1.10	0.548
CTTCCTTGCC	KRT17	2785	17q21.2	0	2	53	−1.30	0.741	20.30	<0.001
AAAGCGGGGC	KRT13	433871	17q21.2	0	0	73	N/A	N/A	27.90	<0.001
CTGTTCCGGC	PPP1R1B	286192	17q21.2	0	0	10	N/A	N/A	4.00	0.050
CTCAGCAATG	TOP2A	156346	17q21.2	0	4	5	1.60	0.700	2.00	0.050
GTGTGGGGGG	JUP	2340	17q21.2	1	32	28	12.60	<0.001	10.70	<0.001
TTCGGTTGGT	COL1A1	172928	17q21.3	1	66	41	25.90	<0.001	15.70	<0.001
TCTCCAGGAA	CGI-69	237924	17q21.3	0	21	7	8.20	0.001	2.70	0.150
ATGCTCCCTG	LGALS3BP	79339	17q25.3	0	0	18	N/A	N/A	6.90	<0.01
TCTCTGATGC	TIMP2	6441	17q25.3	1	40	23	15.70	<0.001	8.80	<0.01
CAACTTAGTT	MLC-B	233936	18p11.3	2	19	29	3.70	0.062	5.50	<0.01
CCTCCTATTA	RIOK3	209061	18q11.2	0	2	22	−1.30	0.741	8.40	<0.01
ACCCCCCCGC	JUND	2780	19p13.1	0	1	19	−2.50	1.032	7.30	<0.01
AGAGGGTGGG	DNAJB1	82646	19p13.1	1	30	36	11.80	<0.001	13.80	<0.001
CCTCCACCTA	PRDX2	432121	19p13.2	2	28	15	5.50	<0.01	2.90	0.178
CGAGGGGCCA	ACTN4	443619	19q13.2	1	1	39	−2.50	1.613	14.90	<0.001
CGACCCCACG	APOE	169401	19q13.2	0	22	51	8.60	<0.001	19.50	<0.001
TGGCCCCAGG	APOC1	268571	19q13.2	0	34	56	13.30	<0.001	21.40	<0.001
GTACACACCC	CST1	123114	20p11.21	0	1	19	−2.50	1.032	7.30	<0.01
GTACACACAC	CST4	56319	20p11.21	0	36	3	14.10	<0.001	1.10	0.548
TGTTCCACTC	ENTPD6	438431	20p11.21	0	26	4	10.20	<0.001	1.50	0.396
AATGTGAGTC	C20orf129	70704	20q11.23	0	20	17	7.80	<0.01	6.50	<0.01
TTGAATCCCC	PI3	112341	20q13.1	0	0	62	N/A	N/A	23.70	<0.001
GCCCCCAATA	LGALS1	407909	22q13.1	2	38	26	7.50	<0.001	5.00	0.012
GAATTTTATA	BZRP	202	22q13.2	1	22	9	8.60	<0.01	3.40	0.295
GAGAGTGTCT	TIMP1	446641	Xp11.2	0	2	20	−1.30	0.741	7.60	<0.01
ATTATCCAGG	RBM3	301404	Xp11.2	0	18	0	7.10	<0.01	N/A	N/A
CCCCCACCTA	PLP2	77422	Xp11.2	0	23	5	9.00	<0.001	1.90	0.286
TATGTGTGCT	SYTL4	376981	Xq22.1	0	18	0	7.10	<0.01	N/A	N/A
GTTAACGTCC	RPL36A	432485	Xq22.1	2	34	17	6.70	<0.01	3.30	0.112
ATAGAGGCAA	MRGX	411358	Xq22.2	0	19	10	7.50	<0.01	3.80	0.050

Analyses were performed using eSAGE 1.2a software [19,20].
GSM784 = normal gastric mucosa; GSM2385 = gastroesophageal adenocarcinoma; GSM757 = gastroesophageal adenocarcinoma; P_{val} = *P* value.*

where R_t is the threshold cycle number for the reference gene observed in the tumor, E_t is the threshold cycle number for the experimental gene observed in the tumor, R_n is the threshold cycle number for the reference gene observed in the normal sample, and R_t is the threshold cycle number for the reference gene observed in the tumor sample. R_n and E_n values were an average for the 13 normals that were analyzed.

Results and Discussion

CGH analyses demonstrated complex DNA changes in all GEJ carcinomas (Table 1; Figure 1). A total of 164 DNA

amplifications was detected, with 54 HLAs in different chromosomal areas (Table 1). These amplifications were frequently (>45%) seen in chromosome arms such as 1q, 6p, 7p, 8q, and 17q. Our results are in agreement with a similar CGH study on GEJ carcinomas [22].

Global analyses of gene expression using SAGE libraries produced more than 156,432 expressed tags. Comparison to UniGene cluster, release January 2003, identified 26,633 unique SAGE tags. The most attractive feature of SAGE is its ability to evaluate the expression pattern of thousands of genes in a quantitative manner without prior sequence information [23–26]. The genome-wide transcriptome analyses for significant ($P \leq .01$) high expression changes ($\geq$5-fold) revealed that 123 genes are overexpressed in GEJ carcinomas as compared to normal mucosa (Table 2). The overexpressed genes ($P \leq .01$) included several genes that are cancer-related such as S100A proteins (A2, 6, 7, 8, 9, and 10), heat shock proteins (HSPE1, HSPA1A, HSPA5, and HSPCA), protease inhibitors (SKALP, TIMP1, and TIMP2), and proliferation markers (TOP2A, TGFA, RPS7, and IGFBP7). In addition, there are several keratins that are specific for epithelial cells such as keratin 4, 6A, 13, and 17 (Table 2). Sequence data from our SAGE libraries are publicly available (http://www.ncbi.nlm.nih.gov/SAGE/) with Geo Accession Omnibus nos. GSM757 and GSM2385 for GEJ cancer and GSM784 for normal stomach. Our quantitative real-time reverse transcription (RT) PCR for 11 overexpressed genes confirmed the SAGE data (Figure 2). However, we were not able to assess correlations with histopathology data or clinical outcomes because of the relatively small sample number in the pilot study.

We mapped gene expression alterations to chromosomal positions using the University of California Santa Cruz's November freeze 2002 assembly of the human genome sequence (http://genome.ucsc.edu) (Table 2). We Compared the DNA amplification regions to the gene expression mapping. The chromosome mapping of altered genes revealed clustering of several genes to small chromosomal domains, suggesting a high order of organization of the cancer genome. Some chromosomal arms that had frequent DNA changes had also frequent gene expression alterations such as chromosomes 1 (15 genes), 2 (9 genes), 6 (6 genes), 11 (6 genes), 12 (8 genes), and 17 (13 genes) (Table 2; Figure 1). Overexpressed genes clustered at specific chromosomal positions such as 1q21.3 (9 genes), 6p21.3 (5 genes), and 17q21 (8 genes). The gene expression profile indicated that although the amplification regions are often large, the expressed genes are clustered and mapped to small chromosomal regions such as 1q21.3 and 17q21.2. This observation indicates that the majority of genes located in areas involved in chromosomal amplifications remain highly regulated and only few critical genes may be deregulated and overexpressed. Although the DNA amplification is one mechanism responsible for the expression changes, other cellular mechanisms of gene regulation are often involved (Table 2; Figure 1).

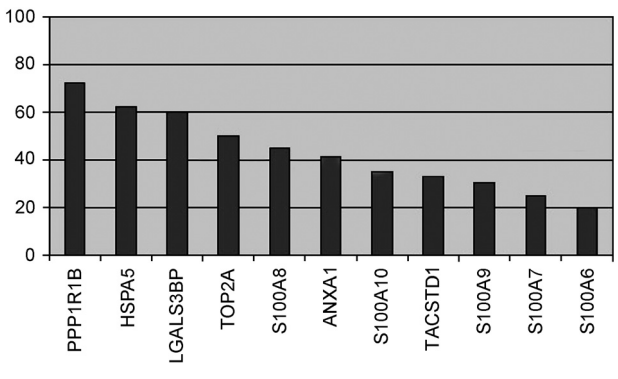

Figure 2. *Frequency histogram of gene overexpression in UGC (Upper Gastrointestinal Carcinoma). Quantitative real-time RT-PCR analyses of 11 representative genes detected in SAGE analyses (LGALS3BP, PPP1R1B, HSPA5, TACSTD1, ANXA1, TOP2A, S100A6, S100A7, S100A8, S100A9, and S100A10). The gene expression in 20 primary GEJ cancers was compared to 13 normal gastric epithelial samples. Gene overexpression was considered at a ratio $\geq$ 5.0. The Overexpression fold was calculated as described earlier [21]. The vertical bars demonstrate the frequency of overexpression of the corresponding gene in the GEJ cancer samples that we studied.*

Despite the genome-wide chromosomal instability in GEJ carcinomas, DNA gains and amplifications mapped to specific regions in the chromosomes such as 1q, 6q, and 17q. When we reviewed the DNA changes that were reported in 73 tumor types from 283 reports [4], we observed that DNA gains/amplifications map to chromosomal regions different from losses in most human cancer types. Thus, the DNA alterations are not randomly distributed but have a rather unique distribution over the chromosomal domains.

Our analyses of the transcriptome in GEJ cancer demonstrated clusters of overexpressed genes in a number of early-replicating chromatin (Giemsa light bands) chromosomal domains such as 1q21.3, 6p21.3, and 17q21.2 (Figure 1). A recent comprehensive study of the human transcriptome map demonstrated a similar clustering of highly expressed genes in chromosomal domains [13]. There are growing evidences that chromosomes occupy discrete CTs in the cell nucleus. The compartment for gene-dense, early-replicating chromatin (Giemsa light bands) is separated from the compartments for mid- to late-replicating chromatin (Giemsa dark bands) [14,27]. A novel theory for explaining gene expression has been recently explored where the transcriptional status of genes correlates with gene positioning in CTs where dynamic repositioning of genes with respect to centromeric heterochromatin has a role in gene silencing and activation [14,27]. Therefore, it is possible that the effect of DNA changes on gene expression alterations may not be limited to their respective gene copy numbers but also to the overall impact of the massive DNA amplifications on the chromatin repositioning in the nucleus. The existence of other regulatory mechanisms such as upstream gene regulation should not be overlooked.

Our results provide a comprehensive tool to search for DNA changes and genes that are overexpressed in GEJ carcinoma. The presence of large amplification areas, yet clustering of overexpressed genes to relatively small loci, may suggest a high organization of chromatin and cancer-related genes in the nucleus. The impact of massive DNA

changes that we detected in GEJ carcinomas on the nuclear organization of the chromatin and the repositioning of genes in CT requires further investigation. Moreover, the transcriptome data provide us, as well as others, the opportunity to develop functional and cell biology assays for particular genes of interest that may serve as prognostic or therapeutic targets. This is expected to add to the overall understanding of the biology of this genetically complex and deadly cancer.

References

[1] Blot WJ, Devesa SS, Kneller RW, and Fraumeni JF Jr (1991). Rising incidence of adenocarcinoma of the esophagus and gastric cardia. *JAMA* **265,** 1287–1289.

[2] Locke GRR, Talley NJ, Carpenter HA, Harmsen WS, Zinsmeister AR, and Melton LJR (1995). Changes in the site- and histology-specific incidence of gastric cancer during a 50-year period. *Gastroenterology* **109,** 1750–1756.

[3] El-Rifai W and Powell S (2002). Molecular and biologic basis of upper gastrointestinal malignancy gastric carcinoma. *Surg Oncol Clin North Am* **11,** 273–291.

[4] Knuutila S, Autio K, and Aalto Y (2000). Online access to CGH data of DNA sequence copy number changes. *Am J Pathol* **157,** 689.

[5] Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW (1995). Serial analysis of gene expression. *Science* **270,** 484–487.

[6] Velculescu VE, Vogelstein B, and Kinzler KW (2000). Analysing uncharted transcriptomes with SAGE. *Trends Genet* **16,** 423–425.

[7] St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, and Kinzler KW (2000). Genes expressed in human tumor endothelium. *Science* **289,** 1197–1202.

[8] Parle-McDermott A, McWilliam P, Tighe O, Dunican D, and Croke DT (2000). Serial analysis of gene expression identifies putative metastasis-associated transcripts in colon tumour cell lines. *Br J Cancer* **83,** 725–728.

[9] Culp LA, Holleran JL, and Miller CJ (2001). Tracking prostate carcinoma micrometastasis to multiple organs using histochemical marker genes and novel cell systems. *Histol Histopathol* **16,** 945–953.

[10] Argani P, Rosty C, Reiter RE, Wilentz RE, Murugesan SR, Leach SD, Ryu B, Skinner HG, Goggins M, Jaffee EM, Yeo CJ, Cameron JL, Kern SE, and Hruban SE (2001). Discovery of new markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma. *Cancer Res* **61,** 4320–4324.

[11] Hough CD, Sherman-Baust CA, Pizer ES, Montz FJ, Im DD, Rosenshein NB, Cho KR, Riggins GJ, and Morin PJ (2000). Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. *Cancer Res* **60,** 6281–6287.

[12] Seth P, Krop I, Porter D, and Polyak K (2002). Novel estrogen and tamoxifen induced genes identified by SAGE (serial analysis of gene expression). *Oncogene* **21,** 836–843.

[13] Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus MC, van Asperen R, Boon K, Voute PA, Heisterkamp S, van Kampen A, and Versteeg R (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291,** 1289–1292.

[14] Cremer T and Cremer C (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2,** 292–301.

[15] Hahn SA, Seymour AB, Hoque AT, Schutte M, da Costa LT, Redston MS, Caldas MS, Weinstein CL, Fischer A, and Yeo CJ (1995). Allelotype of pancreatic adenocarcinoma using xenograft enrichment. *Cancer Res* **55,** 4670–4675.

[16] Yustein AS, Harper JC, Petroni GR, Cummings OW, Moskaluk CA, and Powell SM (1999). Allelotype of gastric adenocarcinoma. *Cancer Res* **59,** 1341–1437.

[17] Varis A, Wolf M, Monni O, Vakkari ML, Kokkola A, Moskaluk C, Frierson H Jr, Powell SM, Knuutila S, Kallioniemi A, and El-Rifai W (2002). Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res* **62,** 2625–2629.

[18] Miettinen MM, El-Rifai W, Sarlomo-Rikala M, Andersson LC, and Knuutila S (1997). Tumor size–related DNA copy number changes occur in solitary fibrous tumors but not in hemangiopericytomas. *Mod Pathol* **10,** 1194–1200.

[19] Margulies EH, Kardia SL, and Innis JW (2001). A comparative molecular analysis of developing mouse forelimbs and hindlimbs using serial analysis of gene expression (SAGE). *Genome Res* **11,** 1686–1698.

[20] Margulies EH and Innis JW (2000). eSAGE: managing and analysing data generated with serial analysis of gene expression (SAGE). *Bioinformatics* **16,** 650–651.

[21] El-Rifai W, Smith MF Jr, Li G, Beckler A, Carl VS, Montgomery E, Knuutila S, Moskaluk CA, Frierson HF Jr, and Powell SM (2002). Gastric cancers overexpress DARPP-32 and a novel isoform, t-DARPP. *Cancer Res* **62,** 4061–4064.

[22] van Dekken H, Geelen E, Dinjens WN, Wijnhoven BP, Tilanus HW, Tanke HJ, and Rosenberg C (1999). Comparative genomic hybridization of cancer of the gastroesophageal junction: deletion of 14Q31-32.1 discriminates between esophageal (Barrett's) and gastric cardia adenocarcinomas. *Cancer Res* **59,** 748–752.

[23] Lee S, Baek M, Yang H, Bang YJ, Kim WH, Ha JH, Kim DK, and Jeoung DI (2002). Identification of genes differentially expressed between gastric cancers and normal gastric mucosa with cDNA microarrays. *Cancer Lett* **184,** 197–206.

[24] Barrett MT, Yeung KY, Ruzzo WL, Hsu L, Blount PL, Sullivan R, Zarbl H, Delrow J, Rabinovitch PS, and Reid BJ (2002). Transcriptional analyses of Barrett's metaplasia and normal upper GI mucosae. *Neoplasia* **4,** 121–128.

[25] Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M, Kodama T, and Aburatani H (2002). Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res* **62,** 233–240.

[26] El-Rifai W, Frierson HJ, Harper J, Powell S, and Knuutila S (2001). Expression profiling of gastric adenocarcinoma using cDNA array. *Int J Cancer* **92,** 828–832.

[27] Dundr M and Misteli T (2001). Functional architecture in the cell nucleus. *Biochem J* **356,** 297–310.